金融时间序列分析备课笔记

李东风

2024-12-23

目录

说明

北京大学数学科学学院金融数学系本科生《金融时间序列分析》授课备课资料。课程采用何书元教授《应用时间序列分析》（北京大学 2003 年）作为主要教材（第 2 版已出版，但教材暂用第 1 版）。

课程内容

- 研究沿时间记录的数据。
- 主要针对序列相关性。
- 主要使用二阶矩、矩阵运算。
- 也用少部分的基础泛函分析、复变函数。
- 有广泛应用但本课程主要是打基础。

时间序列分析的应用领域:

- 经济与金融，比如宏观经济数据序列、证券价格、指数等。
- 工程，比如机械振动、无线电信号、音频信号。
- 科学研究，比如全球气温、人体生理节律、化学反应时间序列等。
- 社会，研究逐年、逐月记录的数据，如交通流量、幸福程度变化等。

参考：

- [HeSY:TSA2023] 主要教材。何书元. 2023. 应用时间序列分析第二版. 北京大学出版社.
- (何书元, 2003) 主要教材的第一版.
- (Brockwell & Davis, 1991) Brockwell, P. J., and R. A. Davis. 1991. Time Series: Theory and Methods. Springer-Verlag. 2nd Ed. 理论基

础很详细，很适合学过时变与泛函的同学详细研读。（有中译本: 田铮，《时间序列的理论与方法》，高等教育出版社，第二版，2001）

- (Shumway & Stoffer, 2019) Shumway, R. H. and Stoffer, D. S. 2019. Time Series: a Data Analysis Approach Using R Taylor and Francis Group, LLC. 比较浅显易懂，有很多 R 数据建模实例。图书馆有电子版。
- (谢衷洁, 1990) 谢衷洁. 1990. 时间序列分析. 北京大学出版社. 很好很全面的理论基础书籍。
- (Shumway & Stoffer, 2017) Shumway, R. H. and Stoffer, D. S. 2017 Time Series Analysis and Its Applications: with R Examples . 4th. edition. Springer International Publishing. 研究生水平的教材，内容比较全面，有 R 编程实例。图书馆有电子版。
- (Paul S.P. Cowpertwait, 2009) Paul S. P. Cowpertwait, Andrew V. Metcalfe. 2009. Introductory Time Series with R. Springer.
- (Cryer & Chan, 2008) Cryer, Jonathan D., and Kung-Sik Chan. 2008. Time Series Analysis with Applications in R. 2nd Ed. Springer.
- (Tsay, 2013) Tsay, Ruey S. 2013. 金融数据分析导论：基于 R 语言. 机械工业出版社. 很好的应用时间序列入门教材。
- (Tsay, 2010) Tsay, Ruey S. 2010. Analysis of Financial Time Series. 3rd Ed. John Wiley & Sons, Inc.
- (Tsay, 2014) Tsay, Ruey S. 2014. Multivariate Time Series Analysis with R and Financial Applications. John Wiley & Sons, Inc.
- (Lutkepohl & Kratzig, 2004) Lutkepohl, Helmut, and Markus Kratzig. 2004. Applied Time Series Econometrics. Cambridge University Press.
- (Christoffersen, 2003) Christoffersen, Peter F. 2003. Elements of Financial Risk Management. Elsevier Science (USA).
- (吴喜之 & 刘苗, 2018) 吴喜之 and 刘苗. 2018. 应用时间序列分析——R 软件陪同. 第二版. 机械工业出版社.
- 王燕，应用时间序列分析，中国人民大学出版社，第二版，2008。
- 肖枝洪，郭明月 (2009) 时间序列分析与 SAS 应用，武汉大学出版社
- 李正辉, 李庭辉 (1010) 时间序列分析实验，中国统计出版社

Part I

时间序列概念与性质

Chapter 1

时间序列概念

时间序列定义: 按时间顺序排列的随机变量序列。

观测样本: 时间序列各随机变量的观测样本。一定是有限多个。

一次实现 (一条轨道): 时间序列的一组实际观测。

时间序列分析的任务: 数据建模，解释、控制或预报。

记号: $\{X_t\}$, $\{x_t\}$, $X(t)$, $x(t)$.

例：北京地区洪涝灾害数据

北京地区从 1949 年到 1964 年的受灾面积与成灾面积（单位：万亩）数据。

```
flood.data <- matrix(c(
1949 ,  1 ,  331.12 ,   243.96 ,
1950 ,  2 , 380.44 ,   293.90 ,
1951 ,  3  , 59.63   , 59.63,
1952 ,  4  , 37.89   , 18.09,
1953 ,  5  , 103.66  ,72.92,
1954 ,  6  , 316.67  , 244.57,
1955 ,  7  , 208.72  , 155.77,
1956 ,  8  , 288.79  , 255.22,
1957 ,  9  , 25.00   , 0.50,
```

```
1958 , 10  , 84.72   , 48.59,
1959 , 11  , 260.89   ,202.96,
1960,  12  , 27.18   ,15.02,
1961,  13  , 20.74   ,17.09,
1962 , 14  , 52.99   ,14.66,
1963 , 15  , 99.25   , 45.61,
1964 , 16  , 55.36   ,41.90),
    byrow=T, ncol=4,
    dimnames=list(1949:1964,
      c("year", "t", "area1", "area2")))
flood <- ts(flood.data, start=1949, frequency=1)
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 -
## v ggplot2 3.3.6          v purrr   1.0.1
## v tibble  3.2.1          v dplyr   1.1.2
## v tidyr   1.3.0          v stringr 1.5.0.9000
## v readr   2.1.2          v forcats 0.5.2

## Warning: package 'tibble' was built under R version 4.2.3

## Warning: package 'tidyr' was built under R version 4.2.3

## Warning: package 'purrr' was built under R version 4.2.3

## Warning: package 'dplyr' was built under R version 4.2.3

## -- Conflicts ------------------------------------------ tidyverse_conflicts() -
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
ggplot(data = as_tibble(flood.data)) +
  geom_line(mapping = aes(
    x = year, y = area1, color = " 受灾面积")) +
  geom_line(mapping = aes(
    x = year, y = area2, color = " 成灾面积")) +
  scale_color_manual(
```

```
    name = NULL,
    breaks = c(" 受灾面积", " 成灾面积"),
    values = c(
      " 受灾面积" = "orangered2",
      " 成灾面积" = "turquoise")) +
  labs(y = NULL) +
  theme(
    legend.position = c(0.98, 0.98),
    legend.justification = c(1,1)  )
```

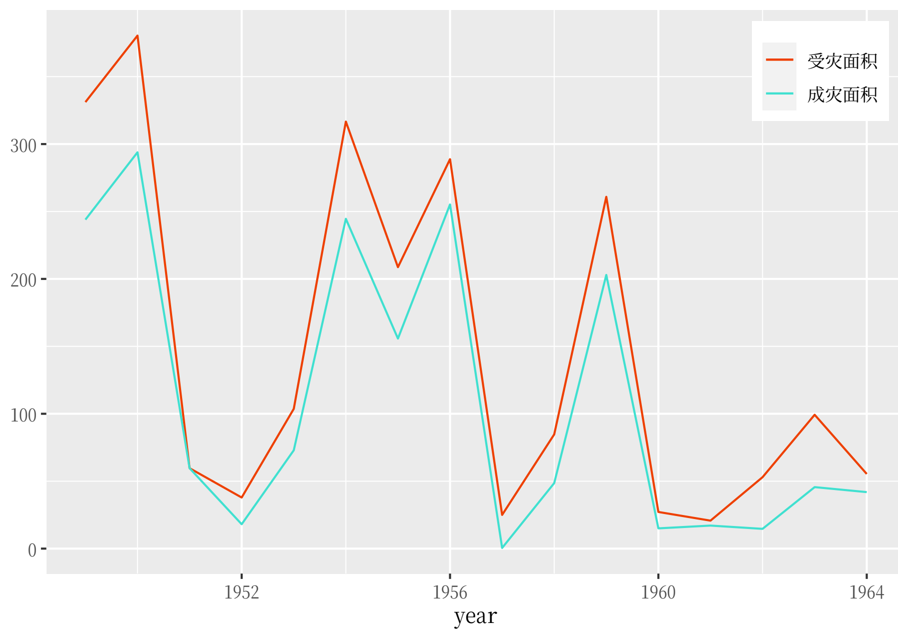


图 1.1: 北京地区洪灾数据

1.1 时间序列分解

$$X_t = T_t + S_t + R_t,\ t = 1, 2, \ldots$$

- 趋势项

- 季节项
- 随机项
- 有时还有随机周期项（Cycle）

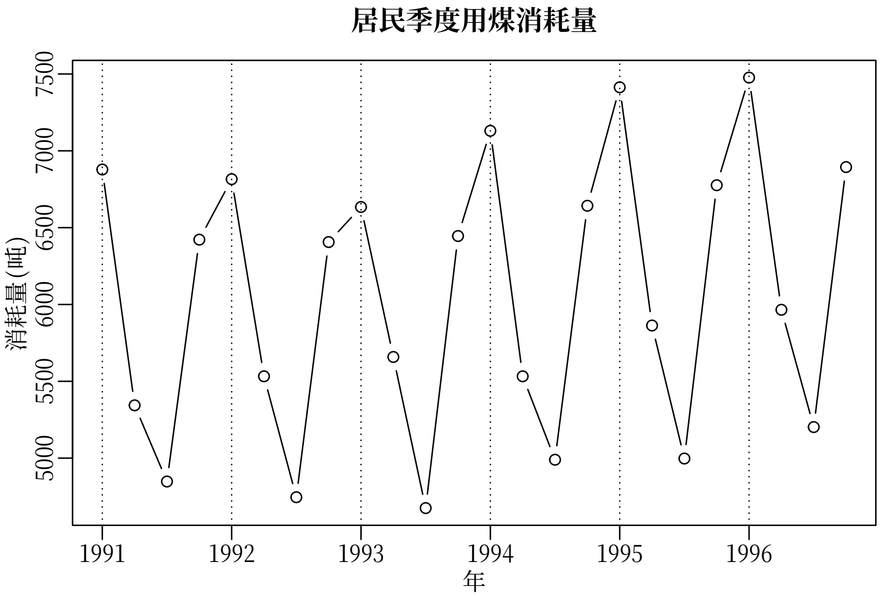
季节项模型有：* 固定的周期季节项: $S(t+s)=S(t), \forall t$. 只需要 $S_1, \dots, S_s$ 且可设 $\sum_{j=1}^{s} S_j = 0$

关于随机项，可设 $ER_t = 0, \forall t$。

1.1.1 例子：居民用煤消耗季度值

```
coal.consumption <-
  ts(c(
     6878.4 ,   5343.7 ,   4847.9 ,  6421.9 ,
     6815.4  ,  5532.6 ,   4745.6 ,  6406.2  ,
     6634.4  ,  5658.5 ,   4674.8 ,  6445.5  ,
     7130.2  ,  5532.6 ,   4989.6 ,  6642.3  ,
     7413.5  ,  5863.1 ,   4997.4 ,  6776.1  ,
     7476.5  ,  5965.5 ,   5202.1 ,  6894.1 ),
     frequency=4, start=c(1991,1))
```

```
demo.coal.data <- function(){
  opar <- par(mar=c(3,3,3,1), mgp=c(1.5,0.5,0))
  on.exit(par(opar))
  plot(coal.consumption, lty=1, type='b',
       main='居民季度用煤消耗量',
       xlab='年', ylab='消耗量 (吨)')
  abline(v=1991:1996, lty=3)
}
demo.coal.data()
```



1.1.1.1 分解方法

- 估计趋势 $\{\hat{T}_t\}$ 后，$X_t - \hat{T}_t$ 主要由季节项和随机项组成，季节项可以用 $X_t - \hat{T}_t$ 每个季节平均得到；
- 趋势估计可用:
- 每年平均；
- 线性回归拟合直线；
- 二次曲线回归；
- 滑动平均估计；
- 得到 $\hat{T}_t$ 和 $\hat{S}_t$ 后可以再从 $X_t - \hat{S}_t$ 估计新的 $\hat{\hat{T}}_t$。
- 季节项估计还可以直接列入模型用哑变量表示 (注意共线问题)。
- 用 $\hat{T}_t + \hat{S}_t$ 拟合或预报 y_t。
- 还可以建立趋势、季节项、随机项的状态空间模型。

1.1.1.2 用每年的平均值作为趋势项估计，季度平均作为季节项

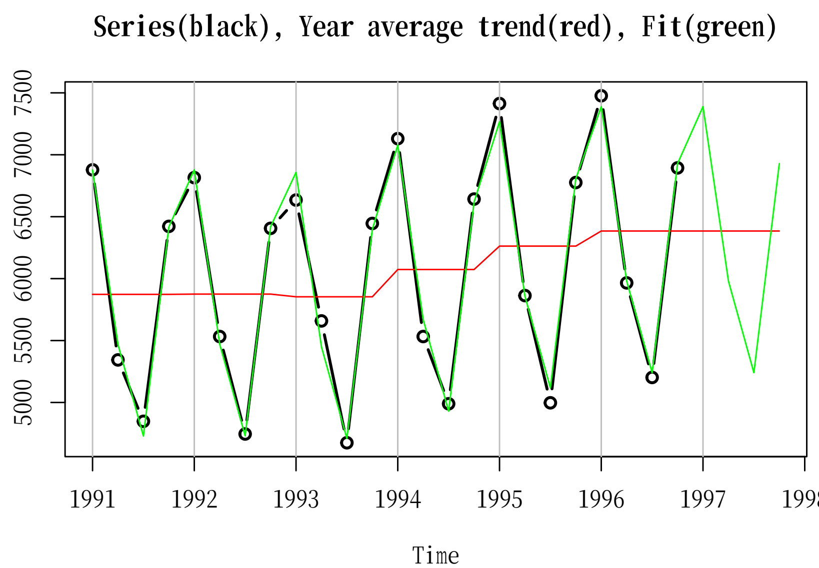
这样的趋势是阶梯函数，每年的 4 个季度的趋势相等。去趋势后同季度平均作为季节项。

```
y <- coal.consumption
ymore <- ts(c(y, rep(NA,4)), start=start(y), frequency=4)
ymat <- matrix(c(y), byrow=TRUE, nrow=6, ncol=4)

cols <- rainbow(20)
ic <- 1

## 用同季度的值平均得到四个季节项
get.season <- function(yd){ # input: Detrended series
  ymat <- matrix(c(yd), byrow=TRUE, ncol=4)
  cy <- as.vector(cycle(yd))

  ## season
  seas0 <- tapply(yd, cy, mean, na.rm=TRUE)

  seas0
}

## 画去除趋势后的序列、季节项和随机项
plot.season <- function(yd, seas0){ # input: Detrended series
  ## season
  seas <- rep(seas0, 6)
  seas <- ts(seas, start=c(1991,1), frequency=4)

  ## Error
  r <- as.vector(yd) - seas
  r <- ts(r, frequency=4, start=c(1991,1))

  plot(yd, type='b', lwd=2,
       main="Detrended Series(black), Seasonal(red) and Random(cyan)",
```

```
      xlim=c(1991,1998), ylab="")
  abline(v=1991:1996, col="gray")
  abline(h=0, col="gray")
  lines(seas, type="l", col="red")
  lines(r, type="l", col="cyan")
}
```

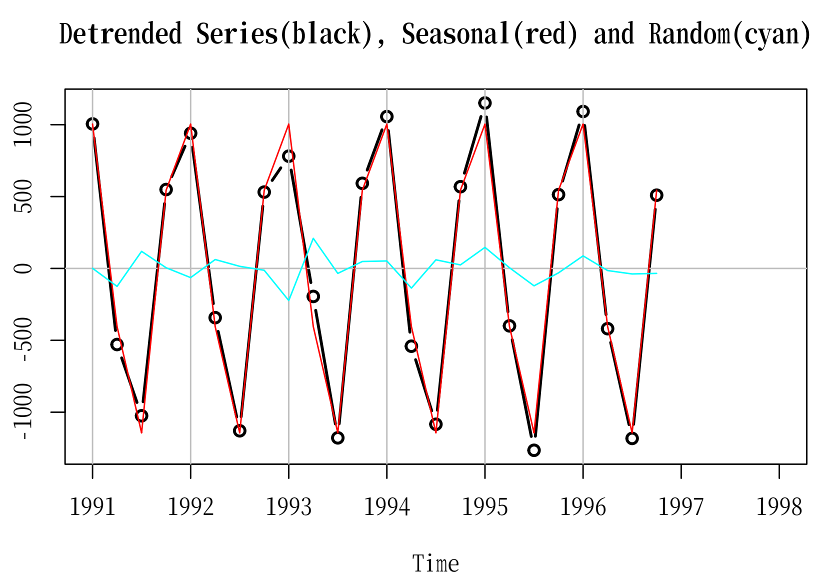
下图为原始序列、趋势、拟合（包括趋势与季节项）：

```
tr0 <- rowMeans(ymat, na.rm=TRUE)
tr <- rep(tr0, each=4)
tr <- ts(tr, frequency=4, start=c(1991,1))
y.detrended <- y - tr

seas0 <- get.season(y.detrended)

tr.more <- ts(c(tr, rep(tr[length(tr)],4)),
              start=start(y), frequency=4)
seas.more <- ts(rep(seas0, 7),
                start=start(y), frequency=4)
y.pred <- tr.more + seas.more

plot(ymore,
     main="Series(black), Year average trend(red), Fit(green)",
     lwd=2,
     type="b", col="black",
     ylab="")
abline(v=1991:1997, col="gray")
lines(tr.more, col="red", type="l")
lines(y.pred, col="green", type="l")
```



下图为去掉了趋势后的序列、季节项、取掉了趋势与季节项后的随机项：

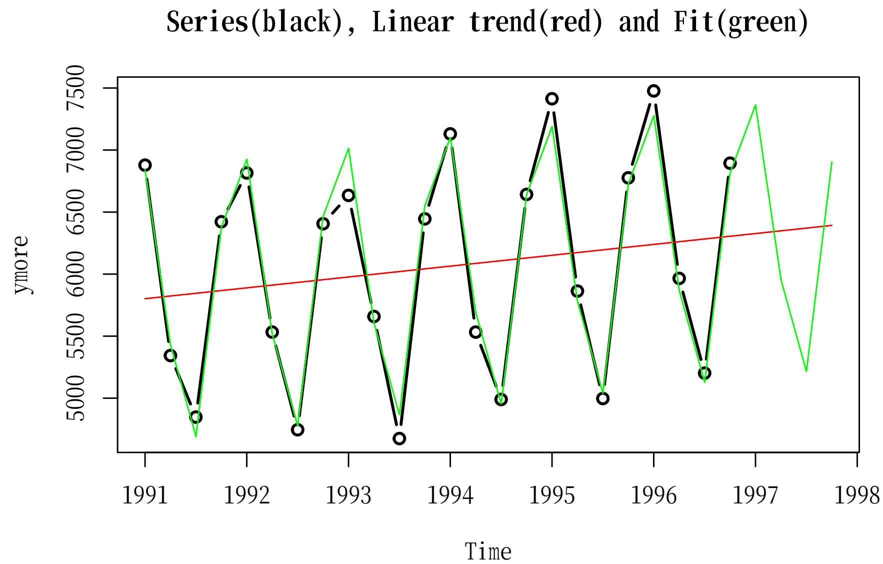
```
plot.season(y.detrended, seas0)
```



1.1.1.3 用时间的线性回归作为趋势项估计，季度平均作为季节项

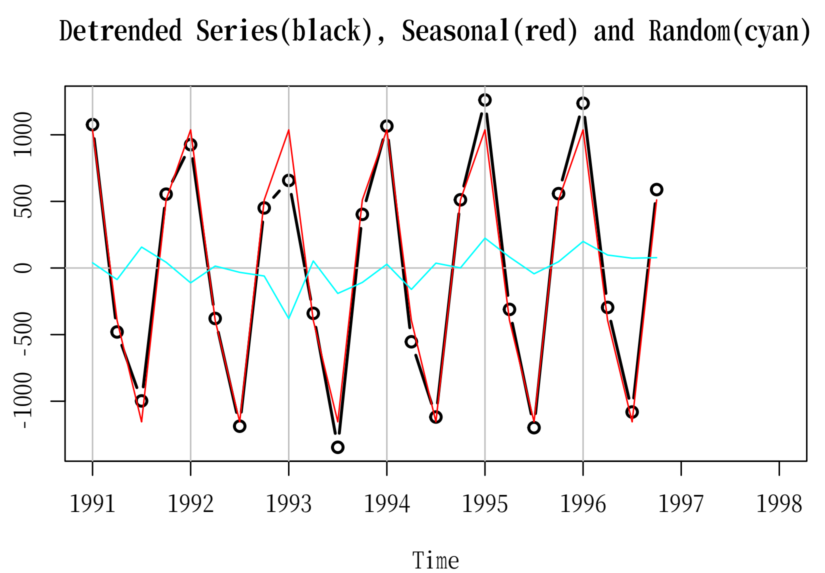
可以设趋势为 $T_t = a + bt$，用一元线性回归程序估计趋势并预测。下图绘制了原始数据、估计的趋势、拟合值（包括趋势与季节项）：

```
yy <- as.vector(y)
tt <- seq(along=y)
lmr <- lm(yy ~ tt)
tr.more <- ts(predict(lmr, newdata=list(tt=seq(length(y)+4))),
              frequency=4, start=c(1991,1))

## season
y.detrended <- y - tr.more[1:length(y)]
seas0 <- get.season(y.detrended)
seas.more <- ts(rep(seas0, 7),
                start=start(y), frequency=4)
y.pred <- tr.more + seas.more

plot(ymore, main="Series(black), Linear trend(red) and Fit(green)",
     lwd=2,
     type="b", col="black")
lines(tr.more, col="red", type="l")
lines(y.pred, col="green", type="l")
```



下图为去掉了趋势后的序列、季节项、取掉了趋势与季节项后的随机项：

```
plot.season(y.detrended, seas0)
```

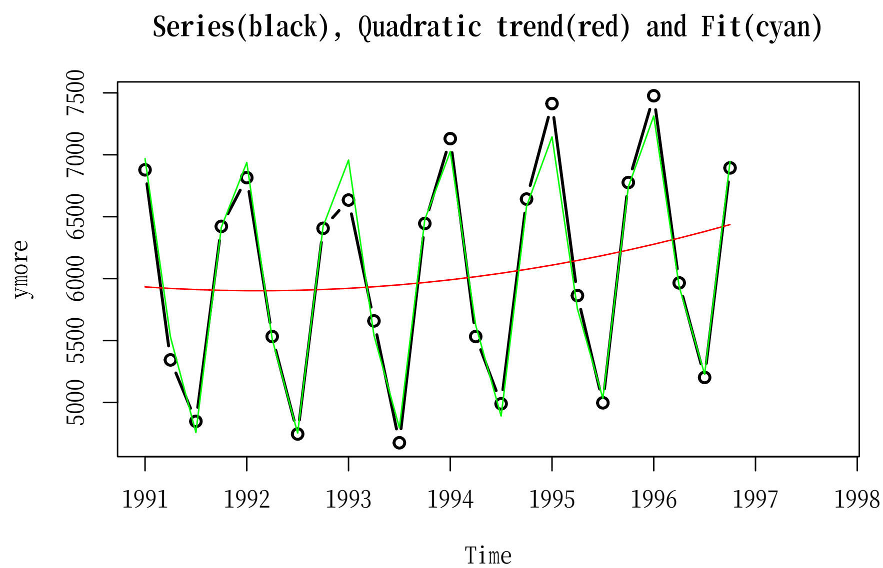


1.1.1.4 二次曲线趋势

可以用 $T_t = a + bt + ct^2$ 作为趋势模型，用多元线性回归程序估计模型并预测。

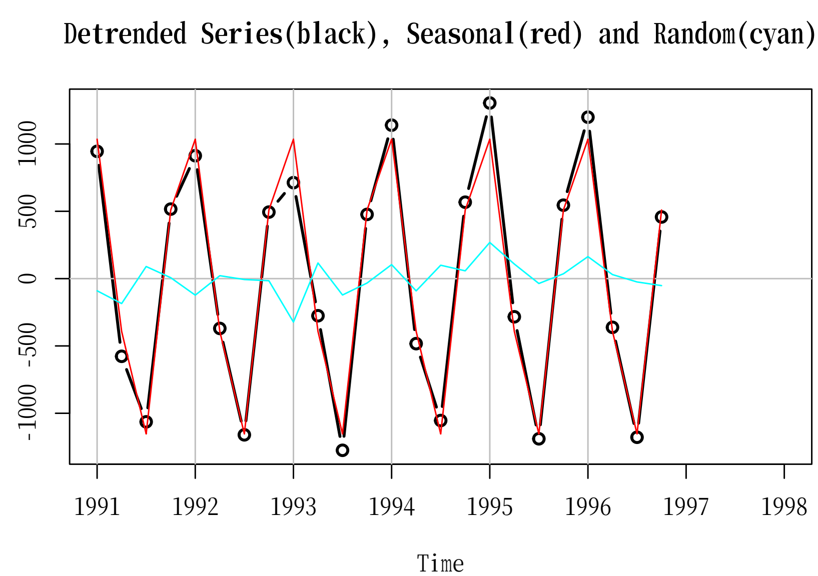
下图绘制了原始数据、估计的趋势、拟合值（包括趋势与季节项）：

```
yy <- as.vector(y)
tt <- seq(along = y)
lmr <- lm(yy ~ tt + I(tt^2))
tr.more <- ts(predict(lmr, new.data=list(tt=seq(length(y)+4))),
              frequency=4, start=c(1991,1))

## season
y.detrended <- y - tr.more[1:length(y)]
seas0 <- get.season(y.detrended)
seas.more <- ts(rep(seas0, 7),
                start=start(y), frequency=4)
y.pred <- tr.more + seas.more

plot(ymore, main="Series(black), Quadratic trend(red) and Fit(cyan)",
     lwd=2,
     type="b", col="black")
lines(tr.more, col="red", type="l")
lines(y.pred, col="green", type="l")
```



下图为去掉了趋势后的序列、季节项、取掉了趋势与季节项后的随机项：

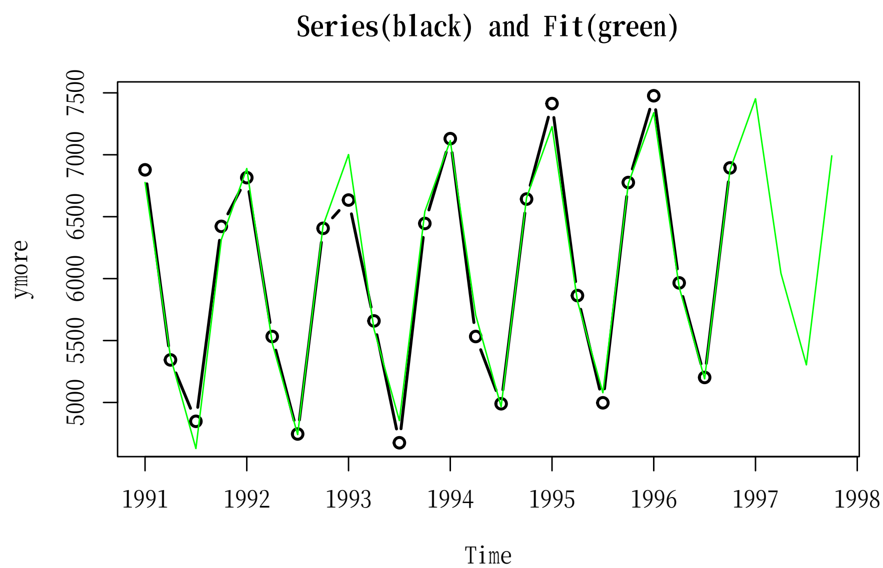
```
plot.season(y.detrended, seas0)
```



1.1.1.5 用线性回归同时估计趋势与季节项

可以设趋势加季节项为 $T_t+S_t = bt+c_1 I_1+c_2 I_2+c_3 I_3+c_4 I_4$，其中 I_1, I_2, I_3, I_4 是各个季度的示性函数。用线性回归程序估计趋势和季节项并预测。下图绘制了原始数据、估计的趋势、拟合值（包括趋势与季节项）：

```
yy <- as.vector(y)
tt <- seq(along = y)
ss <- factor(as.vector(cycle(y)), levels=1:4)
lmr <- lm(yy ~ -1 + tt + ss)
y.pred <- ts(
  predict(lmr,
          newdata=data.frame(
            tt=seq(length(y)+4),
          ss=factor(rep(1:4, 7), levels=1:4))),
  frequency=4, start=c(1991,1))

plot(ymore, main="Series(black) and Fit(green)",
     lwd=2,
     type="b", col="black")
lines(y.pred, col="green", type="l")
```



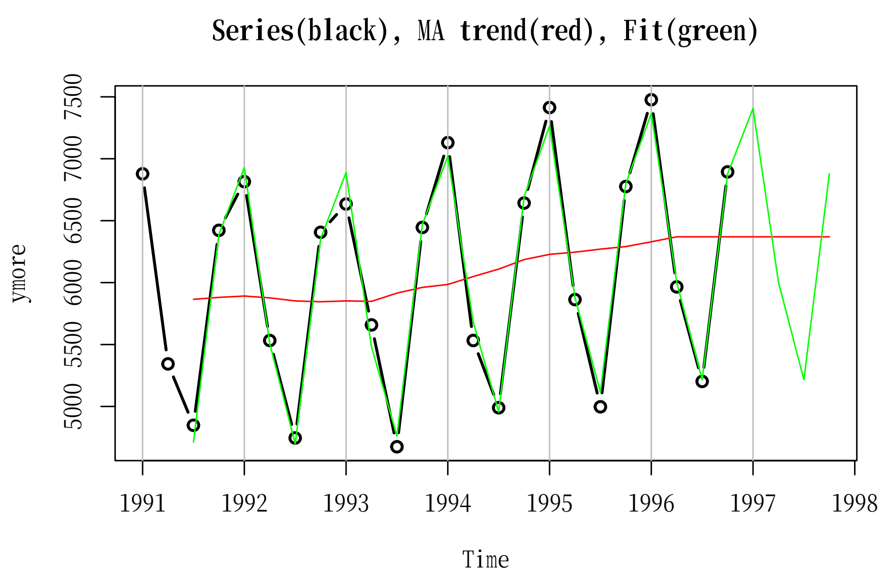
1.1.1.6 加权移动平均趋势

均线法是常用的估计趋势的方法。这里用非等权的 5 点滑动平均方法估计趋势，权重为 1/8, 1/4, 1/4, 1/4, 1/8。

下图绘制了原始数据、估计的趋势、拟合值（包括趋势与季节项）：

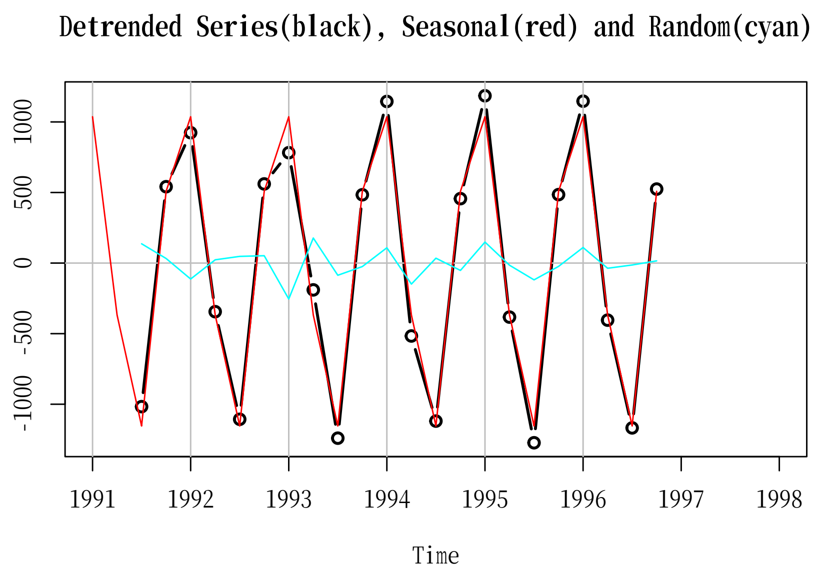
```
yy <- c(y)
tr <- ts(stats::filter(yy, c(1/8, 1/4, 1/4, 1/4, 1/8),
  method="convolution"),
  frequency=4, start=c(1991,1))
ind2 <- max(which(!is.na(tr)))
tr.more <- ts(c(tr, rep(NA,4)),
              start=start(y), frequency=4)
tr.more[(ind2+1):length(tr.more)] <- tr.more[ind2]
y.detrended <- y - tr.more[1:length(y)]
seas0 <- get.season(y.detrended)
seas.more <- ts(rep(seas0, 7),
                start=start(y), frequency=4)
```

```
y.pred <- tr.more + seas.more

plot(ymore, main="Series(black), MA trend(red), Fit(green)",
     lwd=2,
     type="b", col="black")
abline(v=1991:1997, col="gray")
lines(tr.more, col="red", type="l")
lines(y.pred, col="green", type="l")
```



下图为去掉了趋势后的序列、季节项、取掉了趋势与季节项后的随机项：

```
plot.season(y.detrended, seas0)
```

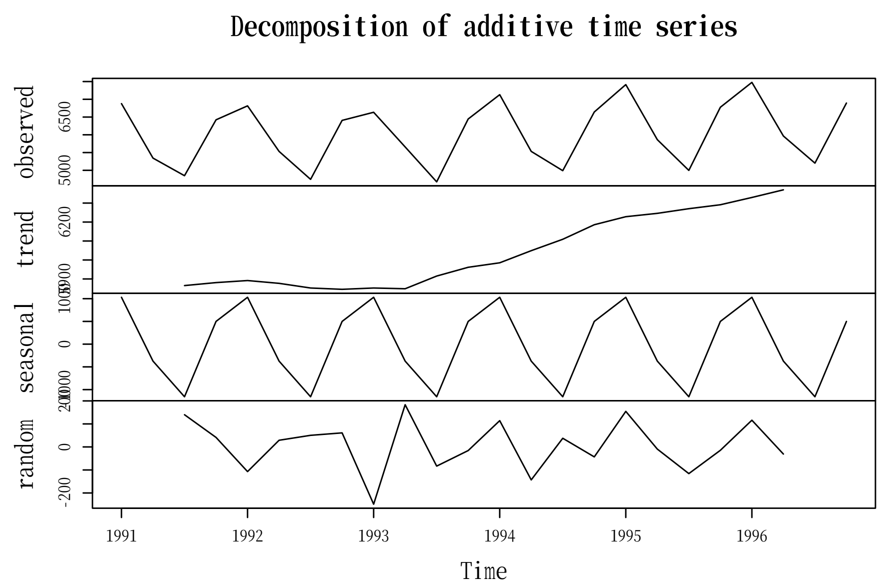


1.1.2 decompose() 函数

R 的 stats 包的 `decompose()` 函数输入一个时间序列，将其分解为趋势项、季节项和随机项。去趋势的方法是中心对称滑动平均。可以用 `type="additive"` 或 `type="multiplicative"` 指定各项之间是相加还是相乘。

比如，城市季度用煤量序列的分解：

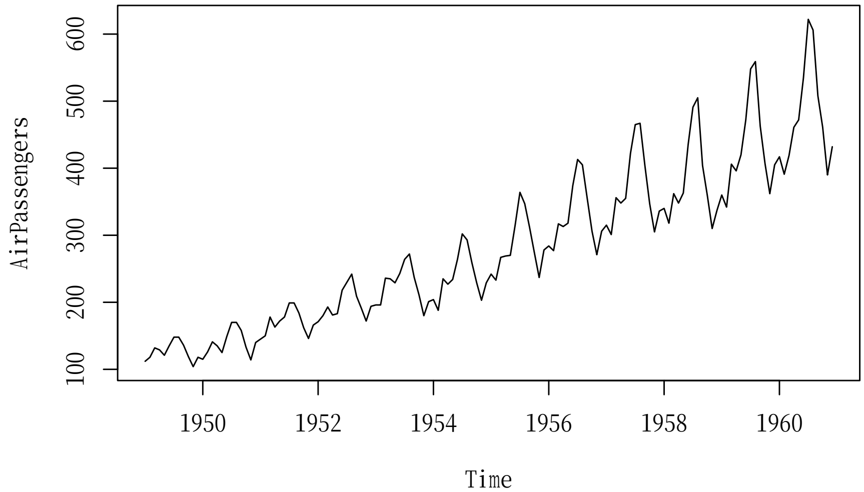
```
res1 <- decompose(coal.consumption)
plot(res1)
```



分别是原始序列、趋势估计、季节项估计和随机项估计。函数输出结果为一个列表，各项为 `x`, `seasonal`, `trend`, `random` 等。

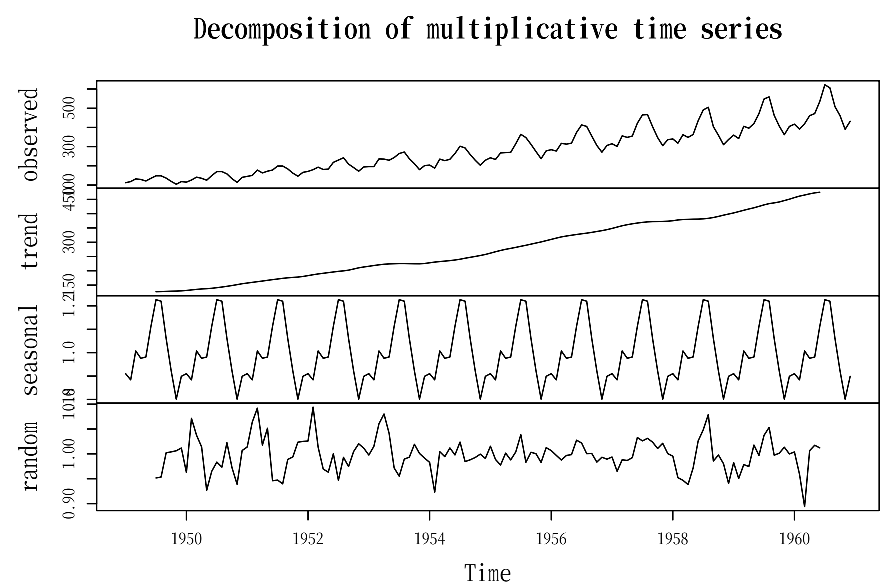
又比如，考虑著名的美国泛美航空公司 1949-1960 的国际航班订票数的月度数据（单位：千人），12 年 144 个月的数据。序列图：

```
plot(AirPassengers)
```



分解：

```r
plot(decompose(AirPassengers, type="multiplicative"))
```



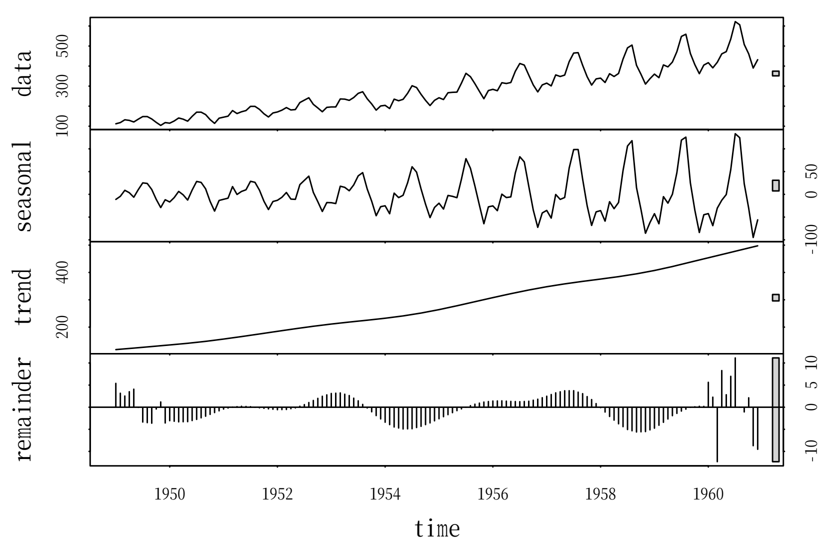
1.1.3 stl() 函数

滑动平均是平滑 (smoothing) 的一种，中心对称的滑动平均会丢失序列开始和结束部分的若干点。R 的 stats 包提供了函数 `stl()`，该函数基于 loess 局部加权回归估计季节项，可以减少异常点的影响，属于稳健回归。用同月份（季度）的数值估计平滑的季节变动，减去季节项后再用平滑方法估计趋势。参见 [R.B.Cleveland1990]。

滑动平均和 loess 局部加权回归都不能表示成公式模型。平滑的另一说法是“滤波”。

如

```
plot(stl(AirPassengers, s.window=3))
```

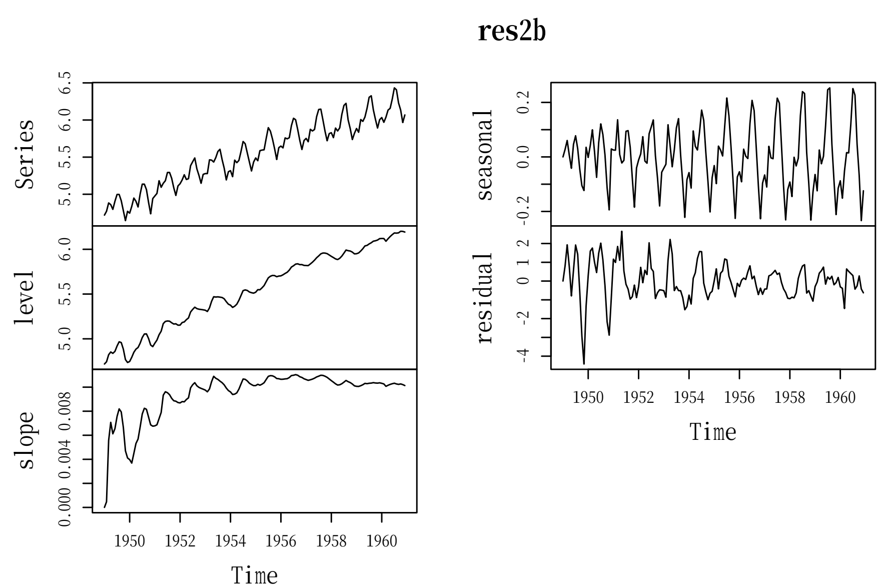


1.1.4 StructTS() 函数

R 的 stats 包的 `StructTS()` 函数用状态空间模型表示时间序列分解，用最大似然方法估计各个成分。结果列表中的 `fitted` 有三列 `level`, `slope`, `sea`。

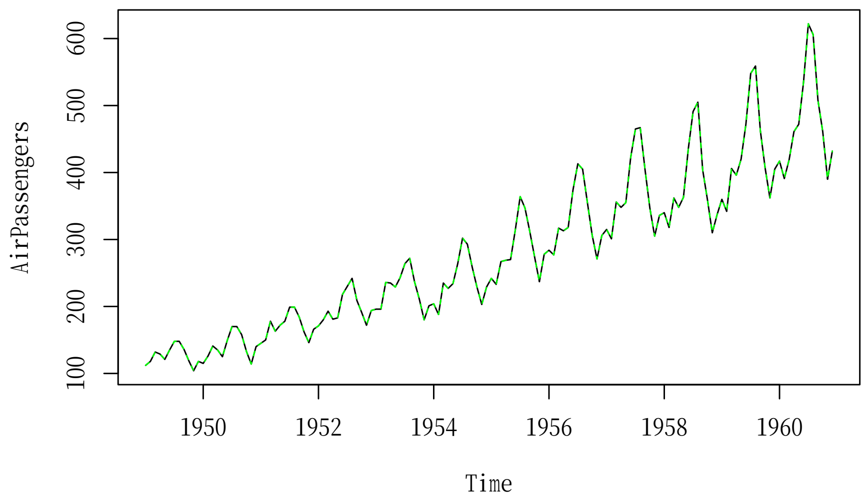
如：

```
res2 <- StructTS(log(AirPassengers), type="BSM")
res2b <- cbind(log(AirPassengers), fitted(res2), resid(res2))
colnames(res2b) <- c(
  "Series", "level", "slope", "seasonal", "residual"
)
plot(res2b)
```



序列 (实线) 与拟合值 (绿色虚线):

```
plot(AirPassengers)
lines(exp(fitted(res2)[,"level"] + fitted(res2)[,"sea"]),
      lty=2, col="green")
```



1.2 平稳序列

随机过程的概念: 指标集 $\mathcal{T}$。$\{X_t : t \in \mathcal{T}\}$ 称为随机过程。

时间序列: $\mathcal{T}$ 为全体整数或正整数时，随机过程称为随机序列；把整数下标看成时间则称随机序列为时间序列。

连续时随机过程、连续时时间序列: $\mathcal{T}$ 为全体实数或全体非负实数时称随机过程为连续时随机过程。下标看成时间时称为连续时的时间序列。

离散采样: 连续时的时间序列记录下来就变成了离散时间。

平稳性: 去除趋势项和季节项后的随机部分经常具有平稳性。

序列分解中趋势和季节部分可以用非随机函数描述，但也可以用随机模型。随机项通常是平稳的。表现：水平没有明显变化；方差没有明显变化；相关性结构不随时间变化。

独立序列不能预报；平稳序列可以用历史值预报。

记号:

- $\mathbb{Z}$—所有整数的集合;。

- $\mathbb{N}_+$—所有正整数的集合；。
- $\mathbb{N}$—表示 $\mathbb{Z}$ 或 $\mathbb{N}_+$。

1.2.1 平稳序列及其自协方差函数

定义 1.1 (平稳序列)**.** 如果时间序列 $\{X_t\} = \{X_t : t \in \mathbb{N}\}$ 满足

(1) 对任何 $t \in \mathbb{N}$, $EX_t^2 < \infty$;
(2) 对任何 $t \in \mathbb{N}$, $EX_t = \mu$;
(3) 对任何 $t, s \in \mathbb{N}$, $E[(X_t - \mu)(X_s - \mu)] = \gamma_{t-s}$, 则称 $\{X_t\}$ 是**平稳时间序列**, 简称为**平稳序列**或**平稳列**. 称实数列 $\{\gamma_t\}$ 为 $\{X_t\}$ 的**自协方差函数**.

性质：

- 期望、方差与 t 无关。
- 时间平移不影响两时刻的相关系数。
- 又称平稳序列为**二阶矩平稳序列**, 还称为**宽平稳列**或**弱平稳列**。

自协方差函数性质:

(1) 对称性: $\gamma_k = \gamma_{-k}, \forall k \in \mathbb{Z}$.

(2) 非负定性:

$$\Gamma_n = (\gamma_{k-j})_{k,j=1}^n = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ .. & .. & \cdots & .. \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix} \text{非负定}(\forall n \in \mathbb{N}_+)$$

(3) 有界性：$|\gamma_k| \le \gamma_0$, $\forall k \in \mathbb{Z}$。

任何满足上述三个性质的实数列都被称为**非负定序列**. 所以平稳序列的自协方差函数是非负定序列. 可以证明, 每个非负定序列都可以是一个平稳序列的自协方差函数.

Γ_n 的元素通项：

$$\Gamma_n = \left(\gamma_{|i-j|}\right)_{i=1,2,\dots,nj=1,2,\dots,n}$$

记

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

则 $\Gamma_n = \mathrm{Var}(X)$。

关于随机向量 X 与矩阵 A, B，有

$$\begin{aligned} E(A+BX) =& A+BE(X) \\ \mathrm{Var}(A+BX) =& B\mathrm{Var}(X)B^T \end{aligned}$$

且 X 的协方差阵 $\mathrm{Var}(X)$ 总是非负定的。

非负定性及随机变量的线性相关:

因为 Γ_n 是协方差阵，所以非负定。事实上，设 $\alpha = (a_1, \dots, a_n)^T$, 则

$$\begin{aligned} &\mathrm{Var}(\alpha^T X) \\ =&\alpha^T \Gamma_n \alpha \\ =&\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) \geq 0. \end{aligned}$$

Γ_n 退化 (不满秩) 当且仅当存在 $\alpha \neq 0$ 使得

$$\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) = 0$$

这时称随机变量 $X_1, \dots, X_n$ 是**线性相关**的。即 $X_1, \dots, X_n$ 的非零线性组合是退化随机变量。如果 $X_1, \dots, X_n$ 线性相关，则 $m \geq n$ 时 $X_1, \dots, X_m$ 线性相关。

Schwarz 不等式:

$$|E(XY)| \leq \sqrt{EX^2 EY^2}$$

$$|\mathrm{Cov}(X,Y)| \leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

推论:

$$E|X| \le \sqrt{E|X|^2}$$

$$|\gamma_t| = |\mathrm{Cov}(X_1, X_{t+1})| \le \gamma_0$$

例 1.1 (平稳序列的线性变换)**.** 设 $\{X_t\}$ 为平稳序列，期望 μ，自协方差函数 $\gamma(t)$。

令 $Y_t = a + bX_t, t \in \mathbb{Z}$。则 $EY_t = a + b\mu$, $\mathrm{Cov}(Y_s, Y_{s+t}) = b^2\mathrm{Cov}(X_t, X_{t+s}) = b^2\gamma(t)$. 可见 $\{Y_t\}$ 平稳。

若取

$$Y_t = \frac{X_t - \mu}{\sqrt{\gamma_0}}$$

则 $EY_t = 0$, $\mathrm{Var}(Y_t) = 1$，称 $\{Y_t\}$ 为 $\{X_t\}$ 的标准化序列。

定义 1.2 (自相关系数)**.** 设平稳序列 $\{X_t\}$ 的标准化序列是 $\{Y_t\}$. $\{Y_t\}$ 的自协方差函数

$$\rho_k = \gamma_k/\gamma_0, \quad k \in \mathbb{Z},$$

称为平稳序列 $\{X_t\}$ 的**自相关系数** (ACF, auto-correlation function)，或自相关函数.

事实上，$\rho_k = \mathrm{corr}(X_t, X_{t+k})$, 其中 $\mathrm{corr}(\cdot,\cdot)$ 表示两个随机变量的相关系数。自相关系数 $\{\rho_t\}$ 是满足 $\rho_0 = 1$ 的自协方差函数, 从而也是非负定序列.

例 1.2 (调和平稳序列)**.** 设 a, b 是常数, 随机变量 U 在 $(-\pi, \pi)$ 内均匀分布, 则

$$X_t = b\cos(at + U), \quad t \in \mathbb{Z}$$

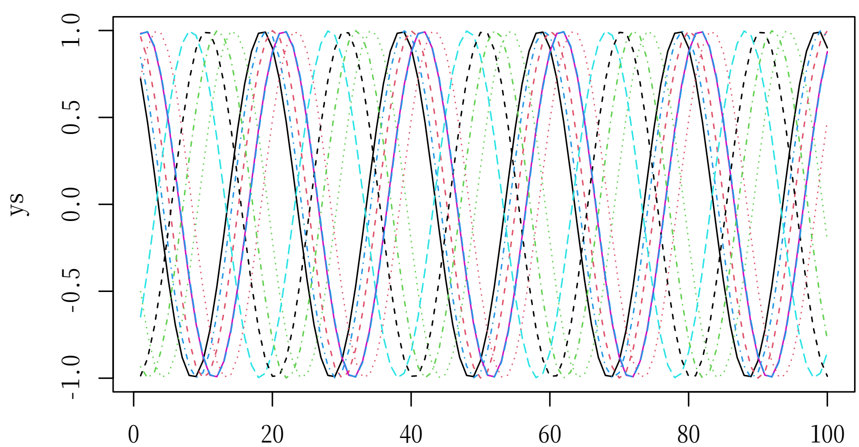
是平稳序列，称为调和平稳列.

$$\begin{aligned}
EX_t &= \frac{1}{2\pi}\int_{-\pi}^{\pi} b\cos(at+u)\,du = 0, \\
E(X_tX_s) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} b^2\cos(at+u)\cos(as+u)\,du \\
&= \frac{1}{2}b^2\cos((t-s)a),
\end{aligned}$$

这个平稳序列的观测样本和自协方差函数 $\gamma_k = 0.5b^2\cos(ak)$ 都是以 a 为角频率, 以 $2\pi/a$ 为周期的函数.

这个例子告诉我们, 平稳序列也可以有很强的周期性. $\{X_t\}$ 的一次实现是一个周期函数，不表现出随机性。

```
sim_harm <- function(a=2*pi*0.05, b=1, n=100){
  tt = seq(n)
  us <- runif(10, -pi, pi)
  ys = outer(tt, us, function(t, u) b*cos(a*t + u))
  matplot(ys, type="l")
}
set.seed(102)
sim_harm()
```



1.2.2 白噪声

定义 1.3 (白噪声)**.** 设 $\{\varepsilon_t\}$ 是一个平稳序列. 如果对任何 $s, t \in \mathbb{N}$,

$$\begin{aligned} E\varepsilon_t =& \mu, \\ \mathrm{Cov}(\varepsilon_t, \varepsilon_s) =& \sigma^2 \delta_{t-s} = \begin{cases} \sigma^2, & t = s, \\ 0, & t \neq s, \end{cases} \end{aligned}$$

则称 $\{\varepsilon_t\}$ 是一个**白噪声**, 记做 $\mathrm{WN}(\mu, \sigma^2)$.

设 $\{\varepsilon_t\}$ 是白噪声, 当 $\{\varepsilon_t\}$ 是独立序列时, 称 $\{\varepsilon_t\}$ 是**独立白噪声**。

当 $\mu = 0$ 时, 称 $\{\varepsilon_t\}$ 为**零均值白噪声**。白噪声的另一种定义要求零均值，本书中用到的白噪声一般都是零均值的。

当 $\mu = 0, \sigma^2 = 1$ 时, 称 $\{\varepsilon_t\}$ 为**标准白噪声**。

当 ε_t 服从正态分布时, 称 $\{\varepsilon_t\}$ 是**正态白噪声**或**高斯白噪声**。正态白噪声总是独立白噪声。

利用 Kronecker 函数 δ_t, 白噪声满足条件 $\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = \sigma^2 \delta_{t-s}$.

例 1.3 (Poisson 白噪声)**.** 这是服从 Poisson 独立白噪声的例子。

如果连续时的随机过程 $\{N(t) : t \in [0, \infty)\}$ 满足 (1) $N(0) = 0$, 且对任何 $s \geq 0, t > 0$ 和非负整数 k,

$$P(N(t+s) - N(s) = k) = \frac{(\lambda t)^k}{k!} \exp[-\lambda t], (\lambda > 0,)$$

(2) $\{N(t)\}$ 有独立增量性: 对任何 $n > 1$ 和 $0 \leq t_0 < t_1 < \cdots < t_n$, 随机变量 $N(t_j) - N(t_{j-1}), \quad j = 1, 2, \cdots, n$ 相互独立,

则称 $\{N(t)\}$ 是一个强度 λ 的 **Poisson 过程**. $EN(t) = \lambda t$, $\mathrm{Var}(N(t)) = \lambda t$.

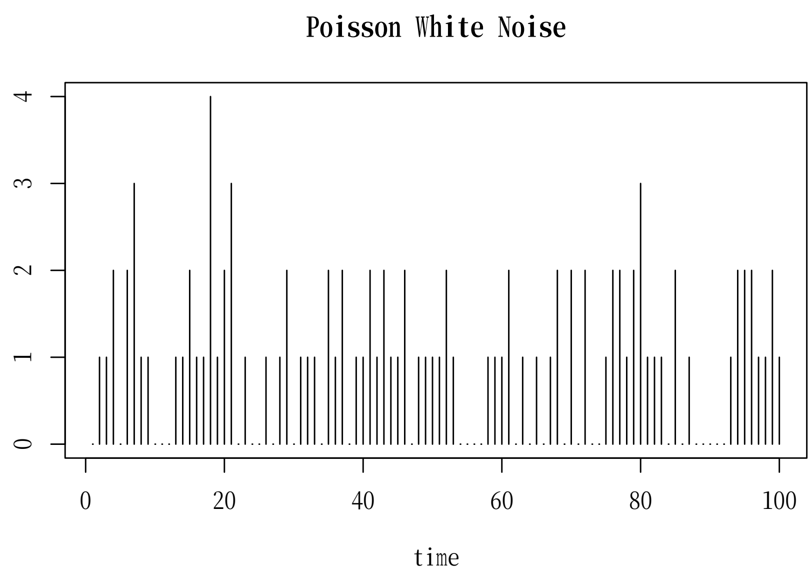
令

$$\varepsilon_n = N(n+1) - N(n) - \lambda, \quad n = 1, 2, \dots,$$

则 $E\varepsilon_n = 0$, $\mathrm{Var}(\varepsilon_n) = \lambda$, $\{\varepsilon_n\}$ 是独立白噪声, 称为 **Poisson 白噪声**.

模拟的 Poisson 白噪声：

```
set.seed(1); x <- rpois(100, 1)
plot(x, type="h", xlab="time", ylab="", main="Poisson White Noise")
```



例 1.4 (布朗运动)**.** 这是一种典型的随机过程。

如果连续时的随机过程 $\{B(t) : t \in [0, \infty)\}$ 满足

(1) $B(0) = 0$, 且对任何 $s \geq 0, t > 0$, $B(t+s) - B(s)$ 服从正态分布 $N(0, t)$;
(2) $\{B(t)\}$ 有独立增量性,

则称 $\{B(t)\}$ 是一个**标准布朗运动**.

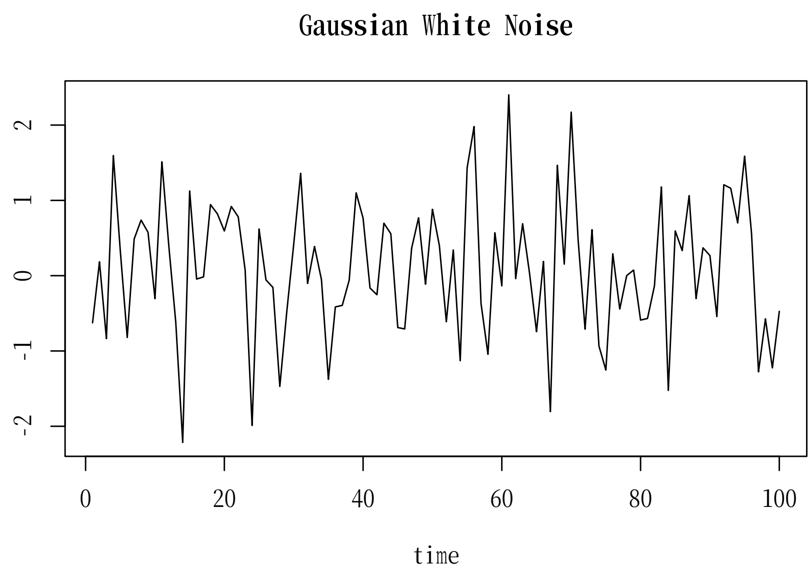
定义

$$\varepsilon_n = B(n+1) - B(n), \quad n = 1, 2, \dots,$$

则 $\{\varepsilon_n\}$ 是一个标准正态白噪声.

模拟的标准正态白噪声：

```
set.seed(1); x <- rnorm(100)
plot(x, type="l", xlab="time", ylab="", main="Gaussian White Noise")
```



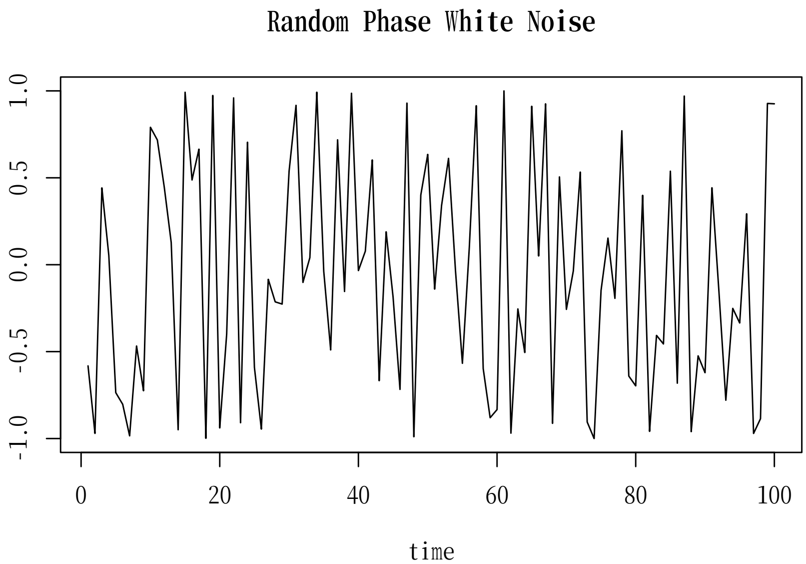
例 1.5 (随机相位)**.** 设 $U_1, U_2, \cdots$ iid $\mathrm{U}(0, 2\pi)$. 令

$$X_t = b\cos(at + U_t), \quad t \in \mathbb{Z},$$

其中 a, b 为常数。则 X_t 为独立序列；$EX_t = 0$, $\mathrm{Var}(X_t) = 0.5b^2$，即 X_t 是独立白噪声。

模拟的随机相位白噪声：

```
set.seed(1)
x <- cos(2*pi/12*(1:100) + runif(100, 0, 2*pi))
plot(x, type="l", xlab="time", ylab="",
     main="Random Phase White Noise")
```



1.2.3 正交平稳序列

两个平稳列 $\{X_t\}$ 和 $\{Y_t\}$ 称为**平稳相关**，如果 $\mathrm{Cov}(X_t, Y_s)$ 仅依赖于 $t-s$。正交和不相关的平稳列是平稳相关的平稳列的最简单情况。

随机变量 X, Y 不相关：$\mathrm{Cov}(X,Y)=0$；正交：$E(XY)=0$.

对零均值随机变量正交与不相关等价。对平稳列 $\{X_t\}$ 和 $\{Y_t\}$，

- 两个序列不相关: $\mathrm{Cov}(X_t, Y_s)=0$, $\forall t, s$.
- 两个序列正交：$E(X_t Y_s)=0$, $\forall t, s$.

定理 1.1. 对平稳列 $\{X_t\}$, $\{Y_t\}$, 设自协方差函数分别为 $\gamma_X(t)$, $\gamma_Y(t)$, 期望分别为 μ_X, μ_Y。令

$$Z_t = X_t + Y_t, t \in \mathbb{Z}$$

(1) 如果 $\{X_t\}$ 和 $\{Y_t\}$ 正交, 则 $\{Z_t\}$ 是平稳序列, 有自协方差函数

$$\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k) - 2\mu_X\mu_Y, \quad k = 0, 1, 2, ...$$

(2) 如果 $\{X_t\}$ 和 $\{Y_t\}$ 不相关，则 $\{Z_t\}$ 是平稳序列，有自协方差函数

$$\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k),\ k = 0, 1, 2, \dots .$$

(3) 当 $\mu_X = \mu_Y = 0$ 且两序列正交时第 *2* 条的结论成立。

1.3 附录：补充知识

定理 1.2. 设 $\{\gamma_k, k \in \mathbb{Z}\}$ 是非负定列，必存在平稳列 $\{X_t\}$ 使得 $\{X_t\}$ 的自协方差函数为 $\{\gamma_k\}$。

证明

由 §4.3定理4.4可得此结论。也可参见 (Brockwell & Davis, 1987)§1.5 定理1.5.1。

注

为了验证某个实数列 $\{\gamma_k\}$ 是否非负定列，直接按照定义检查有时比较困难，这时，如果能找到某个平稳列以 $\{\gamma_k\}$ 为自协方差函数，就证明了它是非负定列。或者，根据 §6.2.1的谱函数、谱密度与自协方差函数的关系，如果存在 $[-\pi, \pi]$ 上单调不减右连续函数 $G(\lambda)$ 使得 $\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda} dG(\lambda)$，或者存在非负可积函数 $f(\lambda)$ 使得 $\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda)\, d\lambda$，则可证明 $\{\gamma_k\}$ 是非负定列。

Chapter 2

线性平稳序列和线性滤波

2.1 有限运动平均

线性平稳序列是白噪声的线性组合得到的序列。最简单的线性平稳序列是有限运动平均。

设 $\{\varepsilon_t\} = \{\varepsilon_t : t \in \mathbb{Z}\}$ 是 $\mathrm{WN}(0, \sigma^2)$. 对于非负整数 q 和常数 $a_0, a_1, \dots, a_q (a_0 \neq 0, a_q \neq 0)$, 我们称

$$X_t = \sum_{j=0}^{q} a_j \varepsilon_{t-j} = a_0 \varepsilon_t + a_1 \varepsilon_{t-1} + \cdots + a_q \varepsilon_{t-q}, \quad t \in \mathbb{Z}$$

是白噪声 $\{\varepsilon_t\}$ 的 (有限) 运动平均或滑动平均, 简称为 MA (Moving Average).

MA 的平稳性:

$$\begin{aligned} EX_t =& 0, \\ E(X_{t+k} X_t) =& \begin{cases} \sigma^2 \sum_{j=0}^{q-k} a_j a_{j+k}, & 0 \leq k \leq q, \\ 0, & k > q \end{cases} \end{aligned}$$

可见 $\{X_t\}$ 平稳。$\gamma_k = 0$, $\forall k > q$, 称这样的序列为 q 相关的。

2.2 线性平稳序列

2.2.1 期望与极限交换次序

随机变量有限平均到随机变量无穷级数的推广需要概率论的极限理论。

定理 2.1 (单调收敛定理)**.** 如果非负随机变量序列 $\{\xi_n\}$ 单调不减: $0 \le \xi_1 \le \xi_2 \le \cdots$, 则当 $\xi_n \to \xi$ *a.s.* 时, 有 $E\xi = \lim_{n\to\infty} E\xi_n$.

这里的随机变量是广义随机变量，允许取 $+\infty$ 值。

对于任何时间序列 $\{Y_t\}$, 利用单调收敛定理得到

$$E\left[\sum_{t=-\infty}^{\infty} |Y_t|\right] = \lim_{n\to\infty} E\left[\sum_{t=-n}^{n} |Y_t|\right] = \lim_{n\to\infty} \sum_{t=-n}^{n} E|Y_t| = \sum_{t=-\infty}^{\infty} E|Y_t|.$$

定理 2.2 (控制收敛定理)**.** 如果随机变量序列 $\{\xi_n\}$ 满足 $|\xi_n| \le \xi_0$ *a.s.* 和 $E|\xi_0| < \infty$, 则当 $\xi_n \to \xi$, *a.s.* 时, $E|\xi| < \infty$ 并且 $E\xi_n \to E\xi$.

在定理2.1和定理2.2条件下均有，

$$\lim_{n\to\infty} E\xi_n = E \lim_{n\to\infty} \xi_n.$$

即期望与极限可以交换次序。

推论 2.1. 如果随机变量序列 $\{Y_n\}$ 满足 $\sum_{i=-\infty}^{\infty} E|Y_i| \le \infty$, 则 $\sum_{i=-\infty}^{\infty} Y_i$ *a.s.* 收敛, $E|\sum_{i=-\infty}^{\infty} Y_i| < \infty$ 且

$$E\sum_{n=-\infty}^{\infty} Y_n = \sum_{n=-\infty}^{\infty} EY_n$$

证明令

$$\eta = \sum_{i=-\infty}^{\infty} |Y_i|$$

则 η 是随机变量，由单调收敛定理知 $E\eta < \infty$, $\eta < \infty$, a.s.。

令

$$\xi_n = \sum_{i=-n}^{n} Y_i$$

则

$$|\xi_n| \le \eta,$$

记事件 $A = \{\eta < \infty\}$，则 $P(A) = 1$，对 $\omega \in A$, 有 $\sum_{i=-\infty}^{\infty} |Y_i(\omega)| < \infty$ 所以 $\sum_{i=-\infty}^{\infty} Y_i(\omega)$ 收敛，$\lim_{n\to\infty} \xi_n(\omega)$ 收敛，记为 $\xi(\omega)$；对 $\omega \notin A$, 定义 $\xi(\omega) = 0$，则 ξ 是随机变量，且

$$\lim_{n\to\infty} \xi_n = \xi, \text{ a.s.}$$

记

$$\xi = \sum_{i=-\infty}^{\infty} \xi_i.$$

由控制收敛定理

$$\lim_{n\to\infty} E\xi_n = E\xi,$$

而

$$\lim_{n\to\infty} E\xi_n = \lim_{n\to\infty} E \sum_{i=-n}^{n} Y_i = \lim_{n\to\infty} \sum_{i=-n}^{n} EY_i = \sum_{i=-\infty}^{\infty} EY_i,$$

于是有

$$E \sum_{i=-\infty}^{\infty} \xi_i = \sum_{i=-\infty}^{\infty} E\xi_i.$$

如果实数列 $\{a_j\}$ 满足

$$\sum_{j=-\infty}^{\infty} |a_j| < \infty,$$

则称 $\{a_j\}$ 是**绝对可和**的. 记 $\{a_j\} \in l_1$.

注意: $\{a_j\} \in l_1$ 则 $\{a_j\} \in l_2$ (即 $\sum_j a_j^2 < \infty$). 反之不一定成立。

2.2.2 线性序列定义

对于绝对可和的实数列 $\{a_j\}$, 定义零均值白噪声 $\{\varepsilon_t\}$ 的无穷滑动和如下

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}.$$

则 $\{X_t\}$ 是平稳序列。$EX_t = 0$,

$$\gamma_k = \sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+k}, k \in \mathbb{Z}. \tag{2.1}$$

令 $c_k = \sum_{j=-\infty}^{\infty} a_j a_{j+k}$, $k \in \mathbb{Z}$。

2.2.3 线性序列的 a.s. 收敛性

作为无穷和 $\{X_t\}$ 有没有定义？由 Schwarz 不等式，

$$E|\varepsilon_t| = E(|\varepsilon_t| \cdot 1) \le \sqrt{E\varepsilon_t^2 \cdot 1} = \sigma,$$

所以

$$\sum_{j=-\infty}^{\infty} E|a_j\varepsilon_{t-j}| = \sum_{j=-\infty}^{\infty} |a_j| E|\varepsilon_{t-j}| \le \sigma \sum_{j=-\infty}^{\infty} |a_j| < \infty,$$

由推论2.1可知 $\sum_{j=-\infty}^{\infty} a_j\varepsilon_{t-j}$ a.s. 收敛。

2.2.4 线性序列的 L1 收敛性

级数的余项的 L_1 模

$$E\left|\sum_{|j|>N} a_j\varepsilon_{t-j}\right| \le \sum_{|j|>N} |a_j| E|\varepsilon_{t-j}| \le \sigma \sum_{|j|>N} |a_j| \to 0$$

即线性平稳列在 L_1 意义下收敛。

2.2.5 线性序列的平稳性

由推论2.1可知

$$E\sum_{j=-\infty}^{\infty} a_j\varepsilon_{t-j} = \sum_{j=-\infty}^{\infty} a_j E\varepsilon_{t-j} = 0$$

由 Schwarz 不等式知 $E|\epsilon_{t-j}\epsilon_{t+k-l}| \le \sigma^2$, 于是

$$\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} E|a_j a_l \epsilon_{t-j}\epsilon_{t+k-l}| \le \sigma^2 \sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} |a_j|\,|a_l| = \sigma^2 \left(\sum_{j=-\infty}^{\infty} |a_j|\right)^2 < \infty,$$

由推论2.1，

$$\begin{aligned} EX_tX_{t+k} &= E\sum_{j=-\infty}^{\infty} a_j\varepsilon_{t-j}\sum_{l=-\infty}^{\infty} a_l\varepsilon_{t+k-l} \\ &= \sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} a_ja_lE(\varepsilon_{t-j}\varepsilon_{t+k-l}) \\ &= \sigma^2\sum_{j=-\infty}^{\infty} a_ja_{j+k} \end{aligned}$$

即 $\{X_t\}$ 是平稳序列, $EX_t = 0$，$\gamma_k = \sigma^2\sum_{j=-\infty}^{\infty} a_ja_{j+k}$.

2.2.6 线性序列的 L2 收敛性

设 $\{a_j\} \in l_2$，即 $\sum_j a_j^2 < \infty$，则

$$X_t = \sum_{j=-\infty}^{\infty} a_j\varepsilon_{t-j}\ (L^2)$$

也是平稳序列。期望为零，自协方差函数同上。

X_t 定义的无穷级数是 L^2 收敛的。证明需要应用 Hilbert 空间性质。见5.1.5。

注意 $\{a_j\} \in l_1 \Longrightarrow \{a_j\} \in l_2$。

2.2.7 线性序列的自协方差函数收敛性

当 $\{a_j\} \in l_1$ 时

$$\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$$

事实上，

$$\begin{aligned} &\sum_{k=-\infty}^{\infty} |\gamma_k| \leq \sigma^2\sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} |a_ja_{j+k}| \\ =&\ \sigma^2\sum_{j=-\infty}^{\infty} |a_j|\sum_{k=-\infty}^{\infty} |a_{j+k}| = \sigma^2(\sum_{j=-\infty}^{\infty} |a_j|)^2 < \infty \end{aligned}$$

定理 2.3. 当 $\{a_j\} \in l_2$ 时，自协方差函数 $\lim_{k\to\infty}\gamma_k = 0$.

证明: 利用 Cauchy 不等式 $|\sum a_j b_j| \leq \left(\sum a_j^2 \sum b_j^2\right)^{1/2}$ 得到

$$\begin{aligned}
|\gamma_k| =& \sigma^2 \left|\sum_{j=-\infty}^{\infty} a_j a_{j+k}\right| \\
\leq& \sigma^2 \sum_{|j|\leq k/2} |a_j a_{j+k}| + \sigma^2 \sum_{|j|>k/2} |a_j a_{j+k}| \\
\leq& \sigma^2 \left[\sum_{j=-\infty}^{\infty} a_j^2 \sum_{|j|\leq k/2} a_{j+k}^2\right]^{1/2} + \sigma^2 \left[\sum_{|j|>k/2} a_j^2 \sum_{j=-\infty}^{\infty} a_j^2\right]^{1/2} \\
\leq& 2\sigma^2 \left[\sum_{j=-\infty}^{\infty} a_j^2\right]^{1/2} \left[\sum_{|j|\geq k/2} a_j^2\right]^{1/2} \to 0 \quad (k \to \infty)
\end{aligned}$$

线性序列的应用：

线性序列描述了自协方差函数衰减到零的时间序列。只要样本自协方差函数衰减到零就可以用线性序列来描述。

单边线性序列：

$$X_t = \sum_{j=\mathbf{0}}^{\infty} a_j \varepsilon_{t-j},\ t \in \mathbb{Z}$$

称为单边运动平均 (MA)，或单边无穷滑动和。这样的 X_t 有因果性：X_t 只受 $s \leq t$ 的 ε_s 影响而不受 t 时刻以后的 ε_s 影响。

$$\gamma_k = \begin{cases} \sum_{j=0}^{\infty} a_j a_{j+k}, & k \geq 0 \\ \gamma_{-k} & k < 0 \end{cases}$$

2.3 时间序列的线性滤波

对序列 $\{X_t\}$ 进行滑动求和：

$$Y_t = \sum_{j=-\infty}^{\infty} h_j X_{t-j}, \quad t \in \mathbb{Z}$$

称为对 $\{X_t\}$ 进行**线性滤波**。其中绝对可和的 $\{h_j\}$ 称为一个**保时线性滤波器**。如果输入信号 $\{X_t\}$ 是平稳列则输出 $\{Y_t\}$ 也是平稳列。$\{Y_t\}$ 的收敛性可以用推论2.1得出。

关于线性滤波的性质，在7.1中还有进一步讨论。

当 $\{X_t\}$ 平稳时，由推论2.1，

$$\mu_Y = EY_t = \sum_{j=-\infty}^{\infty} h_j EX_{t-j} = \mu_X \sum_{j=-\infty}^{\infty} h_j$$

$$\begin{aligned}\gamma_Y(n) &= \mathrm{Cov}(Y_{n+1}, Y_1)\\ &= \sum_{j,k=-\infty}^{\infty} h_j h_k E[(X_{n+1-j}-\mu)(X_{1-k}-\mu)]\\ &= \sum_{j,k=-\infty}^{\infty} h_j h_k \gamma_{n+k-j}\end{aligned}$$

例 2.1 (矩形窗滤波器). 取

$$h_j = \begin{cases} \frac{1}{2M+1}, & |j| \le M \\ 0, & |j| > M. \end{cases}$$

则

$$Y_t = \frac{1}{2M+1} \sum_{j=-M}^{M} x_{t-j}$$

是 X_t 的滑动平均，这种滤波称为**矩形窗滤波器**。可以平滑 $\{X_t\}$，抑制高频信号。高频信号表现是粗糙和复杂的曲线，低频信号表现为缓慢和光滑的变化。

例 2.2 (余弦波信号的滤波). 设

$$X_t = S_t + \varepsilon_t = b\cos(\omega t + U) + \varepsilon_t, \quad t \in \mathbb{Z}$$

其中 $U \sim \mathrm{U}(0, 2\pi)$, $\{\varepsilon_t\}$ 零均值平稳, U 与 $\{\varepsilon_t\}$ 独立。信号 $\{S_t\}$ 方差 $b^2/2$，噪声 $\{\varepsilon_t\}$ 方差 σ^2, 信号与方差之比称为**信噪比**，等于 $b^2/(2\sigma^2)$。用矩形窗滤波。

$$
\begin{aligned}
Y_t =& \frac{1}{2M+1} \sum_{j=-M}^{M} X_{t-j} \\
=& \frac{b \sin[\omega(M+0.5)]}{(2M+1)\sin(\omega/2)} \cos(\omega t + U) + \eta_t \\
\eta_t =& \frac{1}{2M+1} \sum_{j=-M}^{M} \varepsilon_{t-j}
\end{aligned}
$$

除了 η_t 项之外，结果与原始信号相比只有幅度有了成比例变化。上式的化简可以用复数的极坐标表示来推导有关三角函数求和的公式。$\text{Var}(\eta_t) = \frac{\sigma^2}{2M+1}$.

新的信噪比为

$$
\frac{b^2}{2\sigma^2} \frac{\sin^2[\omega(M+0.5)]}{(2M+1)\sin^2(\omega/2)}.
$$

特别当 $\omega(M+0.5) = \pi/2$ 时信噪比为

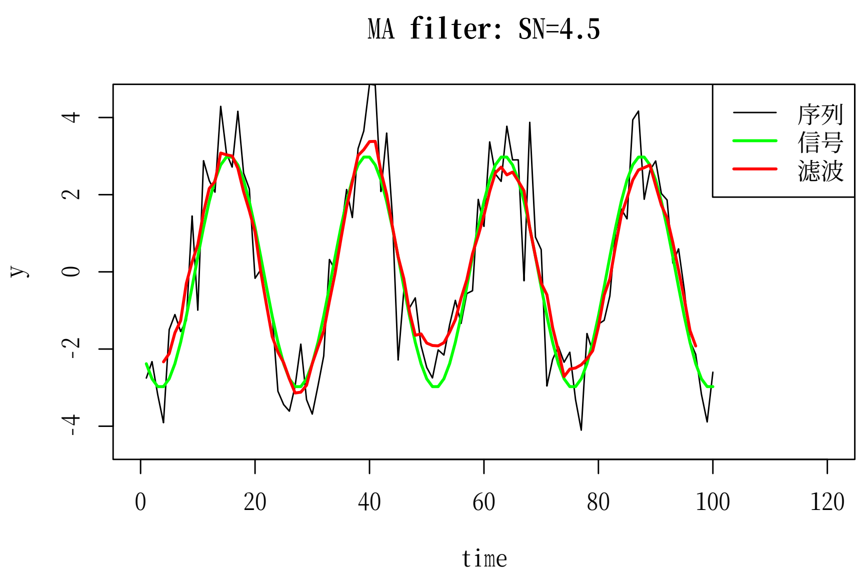
$$
\frac{b^2 \omega}{2\pi\sigma^2 \sin^2(\omega/2)} > \frac{2b^2}{\pi\omega\sigma^2} = \frac{b^2}{2\sigma^2} \cdot \frac{4}{\pi\omega}
$$

信噪比至少增大到 $4/(\pi\omega)$ 倍，ω 越小信噪比提高越多。ω 越小，M 应越大。

演示：

```r
demo.mafilt <- function(b=3, M=3){
  n <- 100
  ##om <- pi/7
  om <- pi/12
  sigma <- 1.0
  eps <- rnorm(n, 0, sigma)
  sn0 <- b^2 / (2*sigma^2)
  tt <- seq(n)
  u <- runif(1)
  signal <- b * cos(om*tt + 2*pi*u)
  y <- signal + eps
  filt <- rep(1/(2*M+1), 2*M+1)
  yf <- filter(y, filt, method="convolution")
  rg <- range(c(y, yf))
  sn <- round(b^2 / (2*sigma^2), 3)
```

```
  plot(tt, y, main=paste("MA filter: SN=", sn, sep=""),
       type="l", xlab='time', ylab='y',
       xlim=c(0,120),
       ylim=c(-b*1.5,b*1.5))
  lines(tt, signal, col="green", lwd=2)
  lines(tt, yf, col="red", lwd=2)
  legend("topright", lty=c(1,1,1), lwd=c(1,2,2),
         col=c("black", "green", "red"),
         legend=c(" 序列", " 信号", " 滤波"))
}
demo.mafilt()
```



2.4 附录：补充证明

2.4.1 不独立的白噪声示例

考虑如下的模型。设 $\{\varepsilon_t\}$ 为 iid 标准正态分布随机变量列，$\alpha_0 > 0, 0 < \alpha_1 < 1$。令

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2$$

令 $\mathscr{F}_t = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$，则 $\varepsilon_t \in \mathscr{F}_t$, ε_t 与 $\mathscr{F}_{t-1}$ 独立。应可证明 $\sigma_t \in \mathscr{F}_{t-1}$, $a_t \in \mathscr{F}_t$, $\mathscr{F}_t = \sigma\{a_t, a_{t-1}, \dots\}$ (严格证明?)

于是

$$\begin{aligned} Ea_t =& E[E(\sigma_t \varepsilon_t | \mathscr{F}_{t-1})] \\ =& E[\sigma_t E(\varepsilon_t | \mathscr{F}_{t-1})] \\ =& E[\sigma_t E(\varepsilon_t)] = 0 \end{aligned}$$

对 $k = 1, 2, \dots$,

$$\begin{aligned} E(a_t a_{t-k}) =& E[E(\sigma_t \sigma_{t-k} \varepsilon_t \varepsilon_{t-k} | \mathscr{F}_{t-1})] \\ =& E[\sigma_t \sigma_{t-k} \varepsilon_{t-k} E(\varepsilon_t | \mathscr{F}_{t-1})] \\ =& E[\sigma_t \sigma_{t-k} \varepsilon_{t-k} E(\varepsilon_t)] = 0 \end{aligned}$$

所以 $\{a_t\}$ 是不相关列，而

$$\begin{aligned} \text{Var}(a_t | a_{t-1}, a_{t-2}, \dots) =& E(a_t^2 | a_{t-1}, a_{t-2}, \dots) \\ =& E(\sigma_t^2 \varepsilon_t^2 | a_{t-1}, a_{t-2}, \dots) \\ =& \sigma_t^2 E(\varepsilon_t^2 | a_{t-1}, a_{t-2}, \dots) \\ =& \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 \end{aligned}$$

可见 a_t 与 a_{t-1} 不独立。

如果模型中的 $\{a_t\}$ 平稳，则 $Ea_t^2 = Ea_{t-1}^2$，有

$$\begin{aligned} E(a_t^2) =& E[E(a_t^2 | \mathscr{F}_{t-1})] \\ =& E[E(\sigma_t^2 \varepsilon_t^2 | \mathscr{F}_{t-1})] \\ =& E[\sigma_t^2 E(\varepsilon_t^2 | \mathscr{F}_{t-1})] \\ =& E[\sigma_t^2 E(\varepsilon_t^2)] \\ =& E[\sigma_t^2] = \alpha_0 + \alpha_1 E a_{t-1}^2 \end{aligned}$$

解得

$$\mathrm{Var}(a_t) = E(a_t^2) = \frac{\alpha_0}{1-\alpha_1}$$

但是 a_t 平稳的证明不显然。

2.4.2 三角级数求和的简化推导

$$\begin{aligned}
&\sum_{j=-M}^{M} b\cos(\omega(t-j)+U) \\
=&\Re\left\{\sum_{j=-M}^{M} b\exp\{i[\omega(t-j)+U]\}\right\} \\
=&\Re\left\{be^{i(\omega t+U)}\sum_{j=-M}^{M} e^{-i\omega j}\right\}
\end{aligned}$$

其中

$$\begin{aligned}
&\sum_{j=-M}^{M} e^{-i\omega j} = \frac{e^{i\omega M} - e^{-i\omega(M+1)}}{1-e^{-i\omega}} \\
=&\frac{\left(e^{i\omega M} - e^{-i\omega(M+1)}\right)(1-e^{i\omega})}{|1-e^{-i\omega}|^2} \\
=&\frac{e^{i\omega M} - e^{-i\omega(M+1)} - e^{i\omega(M+1)} + e^{-i\omega M}}{(1+\cos\omega)^2+\sin^2\omega} \\
=&\frac{2\cos(M\omega) - 2\cos[(M+1)\omega]}{2+2\cos\omega} \\
=&\frac{-4\cos[(2M+1)\omega/2]\sin(-\frac{1}{2}\omega)}{e\sin^2\frac{\omega}{2}} \\
=&\frac{\sin[(M+\frac{1}{2})\omega]}{\sin\frac{\omega}{2}}
\end{aligned}$$

所以

$$\sum_{j=-M}^{M} b\cos(\omega(t-j)+U) = b\cos(\omega t+U)\frac{\sin[(M+\frac{1}{2})\omega]}{\sin\frac{\omega}{2}}$$

Chapter 3

正态时间序列和随机变量的收敛性

3.1 随机向量的数学期望和方差

矩阵随机变量 $M = (M_{i,j})_{m\times n}$：即矩阵每个元素都是一个随机变量。

矩阵随机变量的期望为每个元素取期望：

$$E(M) = (EM_{i,j})_{m\times n} = (\mu_{ij})_{m\times n}.$$

若 A, B, C 是常值矩阵，$C + AMB$ 有意义，则

$$E(C + AMB) = C + A \cdot E(M) \cdot B$$

随机向量 $X = (X_1, X_2, \dots, X_n)^T$. 则协方差阵为

$$\Sigma_X = (\sigma_{ij})_{n\times n} = \text{Var}(X) = E[(X-\mu)(X-\mu)^T]$$

其中 $\sigma_{ii} = \text{Var}(X_i)$，$\sigma_{ij} = \text{Cov}(X_i, X_j)$。

Σ_X 对称非负定 (半正定)。

$$\Sigma_X = E(XX^T) - E(X)E(X)^T.$$

若

$$Y = a + BX, \tag{3.1}$$

则有

$$EY = a + BEX, \qquad \mathrm{Var}(Y) = B\Sigma_X B^T. \tag{3.2}$$

由(3.2), 对 $Y = \alpha^T X$, 有

$$0 \le \mathrm{Var}(Y) = \mathrm{Var}(\alpha^T X) = \alpha^T \mathrm{Var}(X)\alpha.$$

由此可以证明随机向量协方差阵非负定（半正定）。

设随机向量 $X = (X_1, X_2, \dots, X_n)^T$，$Y = (Y_1, Y_2, \dots, Y_m)^T$，则两个随机向量的协方差阵为

$$\mathrm{Cov}(X, Y) = E\big[(X - EX)(Y - EY)^T\big].$$

这是一个 $n \times m$ 矩阵，其 (i, j) 元素为 $\mathrm{Cov}(X_i, Y_j)$。有恒等式

$$\mathrm{Cov}(X, Y) = E(XY^T) - (EX)(EY)^T.$$

设 μ, ν 为非随机的向量，A, B 为非随机的矩阵，则

$$\mathrm{Cov}(\mu + AX, \nu + BY) = A\mathrm{Cov}(X, Y)B^T.$$

3.2 多元正态分布

定义 3.1 (多元正态分布)**.** 称随机向量 $Y = (Y_1, Y_2, \cdots, Y_m)^T$ 服从 m 元 (或多元，或多维，或 m 维) 正态分布, 如果存在 m 维常数列向量 μ, $m \times n$ 常数矩阵 B 和 iid 的标准正态随机变量 $X_1, X_2, \dots, X_n$ 使得

$$Y = \mu + BX.$$

这时

$$\begin{aligned} EY =& \mu, \\ \Sigma =& \mathrm{Var}(Y) = BB^T. \end{aligned}$$

每个 X_j 的特征函数为

$$E\left(e^{itX_j}\right) = e^{-t^2/2}$$

随机向量 $X=(X_1,X_2,\dots,X_n)^T$ 的特征函数为

$$\begin{aligned}\phi_X(t) =&Ee^{it^TX}=E\prod_{j=1}^n e^{it_jX_j}\\ =&\prod_{j=1}^n Ee^{it_jX_j}=\prod_{j=1}^n e^{-t_j^2/2}=e^{-t^Tt/2}\end{aligned}$$

其中 $t=(t_1,t_2,\dots,t_n)^T$。

于是, Y 的特征函数为

$$\begin{aligned}\phi_Y(t) =&Ee^{it^TY}\\ =&Ee^{i(t^T\mu+t^TBX)}\\ =&e^{it^T\mu}Ee^{i(t^TB)X}\\ =&e^{it^T\mu}e^{-(t^TB)(t^TB)^T/2}\\ &(\text{注意}Ee^{is^TX}=e^{-s^Ts/2},\text{令}s^T=(t^TB))\\ =&\exp\left[it^T\mu-\frac{1}{2}t^TBB^Tt\right]\\ =&\exp\left[it^T\mu-\frac{1}{2}t^T\Sigma t\right].\end{aligned}$$

这是多维正态分布的等价定义。

定义 3.2 (多元正态分布等价定义)**.** 称随机向量 $Y=(Y_1,Y_2,\cdots,Y_m)^T$ 服从 m 元正态分布, 若其特征函数为

$$\phi_Y(t) =Ee^{it^TY}=\exp\left[it^T\mu-\frac{1}{2}t^T\Sigma t\right]. \tag{3.3}$$

其中 μ 为 m 为常数列向量，Σ 为 $m\times m$ 非负定阵。

多维正态分布记为 $Y\sim\mathrm{N}(\mu,\Sigma)$，其中 $\mu=E(Y)$，$\Sigma=\mathrm{Var}(Y)$。Y 的分布完全由 μ,Σ 决定，与定义3.1中矩阵 B 的选择无关。

当 $\Sigma>0$(正定) 时，Y 有密度

$$p(y)=(2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right\}$$

若 $|\Sigma|=0$，则 Y 的分量由两部分 Y_1 和 Y_2 组成，$\mathrm{Var}(Y_1)>0$，Y_2 为 Y_1 的线性组合。(可递推证明)

定理 3.1. $Y = (Y_1, Y_2, \dots, Y_n)^T \sim N(\mu, \Sigma)$ 的充分必要条件是: 对任何 $a = (a_1, a_2, \cdots, a_n)^T \in \mathbb{R}^n$,

$$W = a^T Y \ \sim N(a^T \mu, a^T \Sigma a). \tag{3.4}$$

定理(3.4)说明多维正态分布的任意线性组合是一元正态分布。但是，这里的一元正态分布是推广的 $\mathrm{N}(\mu, \sigma^2)$，允许 $\sigma^2 = 0$。

证明:

必要性: 由(3.3)得 W 的特征函数为

$$\begin{aligned}\phi(t) =& E \exp(itW) \\ =& E \exp[ita^T Y] \\ =& E \exp[i(ta^T) Y] \\ =& \exp\left[ita^T \mu - \frac{1}{2} t^2 a^T \Sigma a\right]\end{aligned} \tag{3.5}$$

这是一元正态分布的特征函数，所以 $W \sim \mathrm{N}(a^T \mu, a^T \Sigma a)$。

充分性:

若(3.4)成立，则(3.5)成立，取 $t = 1$，对任意 a 有

$$E \exp(ia^T Y) = \exp\left(ia^T \mu - \frac{1}{2} a^T \Sigma a\right).$$

即 Y 的特征函数为(3.3), 于是 Y 服从多维正态分布。

3.3 正态平稳序列

定义 3.3. 对于时间序列 $\{X_t\}$, 如果对任何 $n \geq 1$ 和 $t_1, t_2, \cdots, t_n \in \mathbb{Z}$, 有 $(X(t_1), X(t_2), \dots, X(t_n))$ 服从多元正态分布, 则称 $\{X_t\}$ 是**正态时间序列**. 特别当 $\{X_t\}$ 还是平稳序列时, 又称为**正态平稳列**.

$\{X_t : t \in \mathbb{N}_+\}$ 是正态时间序列 $\Longleftrightarrow$ 对任何正整数 m, $(X_1, X_2, \dots, X_m)$ 服从 m 维正态分布。

$\{X_t : t \in \mathbb{Z}\}$ 是正态时间序列 $\iff$ 对任何正整数 m, $(X_{-m}, X_{-m+1}, \dots, X_m)$ 服从 $2m+1$ 维正态分布.

正态分布对线性运算的封闭性为其理论研究提供了便利。另外，正态分布和线性模型之间有一种内在的联系。

3.4 概率极限

设 $\xi_n \sim F_n(x), \xi \sim F(x)$。如果在 F 的每个连续点 x 有 $F_n(x) \to F(x)$, 则称 ξ_n **依分布收敛**到 ξ, 记做 $\xi_n \xrightarrow{d} \xi$。

如果对任取 $\epsilon > 0$ 有 $P(|\xi_n - \xi| \geq \epsilon) \to 0$, 则称 ξ_n **依概率收敛**到 ξ, 或称 ξ_n 相合于 ξ, 或 ξ_n 弱收敛到 ξ, 记做 $\xi_n \xrightarrow{\text{P}} \xi$。

如果 $E|\xi_n - \xi| \to 0$, 则称 ξ_n L^1 收敛到 ξ (很少用)。

如果 $E|\xi_n - \xi|^2 \to 0$, 则称 ξ_n L^2 **收敛**到 ξ, 或称 ξ_n **均方收敛**到 ξ, 记做 $\xi_n \to \xi$ (L^2)。

对 $p > 0$，如果 $E|\xi_n|^p$ 和 $E|\xi|^p$ 都有限，且 $E|\xi_n - \xi|^p \to 0$，则称称 ξ_n L^p **收敛**到 ξ。因为 $0 < p < q$ 时 $E|X|^p \leq 1 + E|X|^q$，所以 L^q 收敛推出 L^p 收敛。

如果

$$P(\lim_{n\to\infty} \xi_n = \xi) = 1$$

则称 ξ_n **a.s. 收敛**到 ξ。

定理 3.2. *L^2 收敛 $\Rightarrow$ L^1 收敛 $\Rightarrow$ 依概率收敛 $\Rightarrow$ 依分布收敛。*

证明略。

定理 3.3. *$a.s.$ 收敛 $\Rightarrow$ 依概率收敛 $\Rightarrow$ 依分布收敛。*

证明略。

定理 3.4. *若 ξ_n 依概率收敛到 ξ，则存在子序列 $\{n_k\}$ 使得 ξ_{n_k} $a.s.$ 收敛到 ξ。*

证明略。

定理 3.5. 依概率极限如果存在，就 *a.s.* 唯一。

证明

设随机变量序列 $\{\xi_n\}$ 依概率收敛到 ξ 和 η。则 $\forall \epsilon > 0$，

$$\begin{aligned}\lim_{n\to\infty} P(|\xi_n - \xi| > \epsilon) =& 0,\\ \lim_{n\to\infty} P(|\xi_n - \eta| > \epsilon) =& 0.\end{aligned}$$

于是

$$\begin{aligned}&P(|\xi - \eta| > 2\epsilon)\\ =&P(|(\xi_n - \eta) - (\xi_n - \xi)| > 2\epsilon)\\ \leq&P(|\xi_n - \eta| + |\xi_n - \xi| > 2\epsilon)\\ \leq&P(|\xi_n - \eta| > \epsilon \text{ 或} |\xi_n - \xi| > \epsilon)\\ \leq&P(|\xi_n - \eta| > \epsilon) + P(|\xi_n - \xi| > \epsilon)\\ \to&0 \ (n \to \infty).\end{aligned}$$

于是

$$\begin{aligned}P(\xi \neq \eta) =&P(|\xi - \eta| > 0)\\ =&P\left(\bigcup_{n=1}^{\infty}\left\{|\xi - \eta| > \frac{1}{n}\right\}\right)\\ \leq&\sum_{n=1}^{\infty} P\left(|\xi - \eta| > \frac{1}{n}\right)\\ =&0.\end{aligned}$$

即 $\xi = \eta$, a.s.，证毕。

推论 3.1. 以概率 *1* 收敛极限、L^2 极限、L^1 极限、依概率极限如果存在，则 *a.s.* 唯一；如果这些极限中的几个同时存在，则极限也 *a.s.* 相等。

定理 3.6. ξ_n 依分布收敛到 ξ，当且仅当对任意 $\mathbb{R}$ 上的一元有界实值连续函数 $f(\cdot)$ 都有

$$Ef(\xi_n) \to Ef(\xi), \ n \to \infty.$$

由此，也称依分布收敛为**弱收敛**。证明略。

定理 3.7. ξ_n 依分布收敛到 ξ，当且仅当对任意 $\forall t \in \mathbb{R}$

$$Ee^{it\xi_n} \to Ee^{it\xi},\ n \to \infty.$$

即依分布收敛等价于特征函数收敛。证明略。

定理 3.8. ξ_n 依分布收敛到常数 c，当且仅当 ξ_n 依概率收敛到常数 c。

证明略。

定理 3.9. 随机向量 ξ_n *a.s.* (或者 L^p、依概率) 收敛到随机向量 ξ，当且仅当对应的分量 *a.s.* (或者 L^p、依概率) 收敛关系成立。

证明略。

定理 3.10. 如果正态序列 $\xi_n \sim N(\mu_n, \sigma_n^2), n \in \mathbb{N}$ 依分布收敛到随机变量 ξ, 则极限

$$\lim \mu_n = \mu,\ \lim \sigma_n^2 = \sigma^2$$

存在，且 $\xi \sim N(\mu, \sigma^2)$.

证明参见王梓坤《随机过程论》P.18。

定理 3.11. $\{\varepsilon_t\}$ 是正态 $WN(0, \sigma^2)$ 序列，实数列 $\{a_j\}$ 绝对可和，则线性序列

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$$

是零均值正态平稳列，自协方差函数为

$$\gamma_k = \sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+k},\ k \in \mathbb{Z}. \tag{3.6}$$

当 $\{a_j\} \in l_2$ 时结论仍成立。

证明:

由 §2.2知 $\{X_t\}$ 是零均值平稳列，自协方差函数为(3.6)。

只要证明 $\{X_t\}$ 是正态序列,只要证明 $\forall m \in \mathbb{N}_+$, $X = (X_{-m}, \dots, X_0, \dots, X_m)^T$ 服从多元正态分布。要使用定理3.1（多元正态与一元正态关系）和定理3.10（一元正态分布的依分布极限仍为正态分布）。

记 $\Sigma = (\gamma_{|i-j|})_{i,j=-m,\dots,m}$，来证明 $X \sim \mathrm{N}(0, \Sigma)$。记

$$\eta_t(n) = \sum_{j=-n}^{n} a_j \varepsilon_{t-j},\ t = -m, \dots, m$$

这是 X_t 的部分和。由控制收敛定理可知

$$E|\eta_t(n) - X_t| \leq \sum_{|j|>n} |a_j| \sigma \to 0 (n \to \infty)$$

对 $\forall b = (b_{-m}, \dots, b_0, \dots, b_m)^T \in \mathbb{R}^{2m+1}$, 记

$$\begin{aligned} Y =& b^T X = \sum_{t=-m}^{m} b_t X_t \\ \eta(n) =& \sum_{t=-m}^{m} b_t \eta_t(n) \end{aligned}$$

则当 $n \to \infty$ 时

$$E|\eta(n) - Y| \leq \sum_{t=-m}^{m} |b_t| \cdot E|\eta_t(n) - X_t| \to 0$$

即 $\eta(n) \xrightarrow{L_1} Y$, 于是 $\eta(n) \xrightarrow{d} Y$，由定理3.1知 $\eta(n)$ 服从正态分布，由定理3.10知 Y 服从正态分布，易见 $EY = 0$,

$$\mathrm{Var}(Y) = \mathrm{Var}(\sum_{t=-m}^{m} b_t X_t) = b^T \Sigma b$$

即有 $Y \sim \mathrm{N}(0, b^T \Sigma b)$，从而由定理3.1可知 X 服从多元正态分布，从而 $\{X_t\}$ 为正态序列。

3.4.1 一些反例

例 3.1. a.s. 收敛推出依概率收敛，但是反之不然。给出反例。

设 $\Omega=[0,1]$,$P(\cdot)$ 为 $[0,1]$ 上的勒贝格测度。令 $f_{mk}=I_{[\frac{k-1}{m},\frac{k}{m}]}$, $k=1,2,\dots,m$, $m=1,2,\dots$。将 $\{f_{mk}\}$ 排序为 f_{11}, f_{21}, f_{22}, f_{31}, f_{32}, f_{33}, ...，记这个随机变量序列为 $X_n(\omega)(\omega\in[0,1])$。则对任意 $\epsilon\in(0,1)$，

$$P(|X_n(\omega)|>\epsilon)=P(X_n(\omega)=1)\to 0\ (n\to\infty).$$

即 X_n 依概率收敛到 0。但是对任意 $\omega\in[0,1]$，总有无数个 n 使得 $X_n(\omega)=1$，从而 X_n 不 a.s. 收敛到 0。

例 3.2. L_1 收敛和 L_2 收敛都推出依概率收敛，但是反之不然。给出反例。

设 $\Omega=[0,1]$，$P(\cdot)$ 为 $[0,1]$ 上的勒贝格测度。令 $X_n(\omega)=n^2 I_{[0,\frac{1}{n}]}(\omega)$, 则对任意 $\epsilon\in(0,1)$ 有

$$P(|X_n-0|>\epsilon)=P(X_n\neq 0)=\frac{1}{n}\to 0,\ n\to\infty$$

即 X_n 依概率收敛到 0。但是，

$$E|X_n-0|=EX_n=\int_0^1 X_n(\omega)\,d\omega=n^2\cdot\frac{1}{n}=n\to\infty\ (n\to\infty)$$

所以 X_n 不 L_1 收敛到 0, 也不 L_2 收敛到 0。

例 3.3. 依概率收敛推出依分布收敛，但是反之不然。给出反例。

设 $X, X_1, X_2, \dots$ 独立同 N(0,1) 分布。则 X_n 的分布函数 $F_n(x)$ 与 X 的分布函数 $F(x)$ 处处相等，当然有 $\lim_{n\to\infty}F_n(x)=F(x)$，对任意 $x\in(-\infty,\infty)$ 成立，即 X_n 依分布收敛到 X。但是对任意 $\epsilon>0$，因为 $X_n-X\sim \mathrm{N}(0,2)$，所以

$$P(|X_n-X|>\epsilon)=2(1-\Phi(\epsilon/\sqrt{2}))$$

为正常数，因此 X_n 不能依概率收敛到 X。上式中 $\Phi(\cdot)$ 表示标准正态分布函数。

3.5 补充

3.5.1 联合密度

性质：若 Y 服从多元正态分布 $N(\mu, \Sigma)$ 且 Σ 正定，则 Y 有联合密度

$$p(y) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right\}$$

证明:

当 Σ 为对称正定阵时，由线性代数知识可知 Σ 有特征值分解 $\Sigma = U\Lambda U^T$，其中 $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$，$\lambda_j > 0, j = 1, 2, \dots, n$，$U$ 为正交阵 $U^TU = I_n$。令 $\Lambda^{-1/2} = \mathrm{diag}(\lambda_1^{-1/2}, \dots, \lambda_n^{-1/2})$，$\Sigma^{-1/2} = U\Lambda^{-1/2}U^T$，令 $X = \Sigma^{-1/2}(Y-\mu)$，则 $Y = \mu + \Sigma^{1/2}X$, X 的特征函数为

$$\begin{aligned}
\phi(t) =& E\exp\left\{it^TX\right\} \\
=& E\exp\left\{it^T\Sigma^{-1/2}Y\right\} \cdot \exp\left\{-it^T\Sigma^{-1/2}\mu\right\} \\
=& \exp\left\{it^T\Sigma^{-1/2}\mu - \frac{1}{2}t^T\Sigma^{-1/2}\Sigma\Sigma^{-1/2}t\right\} \cdot \exp\left\{-it^T\Sigma^{-1/2}\mu\right\} \\
=& \exp\left\{-\frac{1}{2}t^Tt\right\}
\end{aligned}$$

这说明 X 为 n 维标准正态分布随机向量，于是 X 的联合密度函数为 $p_X(x) = (2\pi)^{-n/2}\exp\{-\frac{1}{2}x^Tx\}$。从 X 到 Y 的变换的逆变换为 $X = \Sigma^{-1/2}(Y-\mu)$，这是 $\mathbb{R}^n$ 上的一一变换，逆变换的 Jacobi 行列式为 $|\Sigma^{-1/2}| = |\Sigma|^{-1/2}$。由随机向量的变换的密度公式可得 Y 的密度为

$$\begin{aligned}
p_Y(y) =& p_X(\Sigma^{-1/2}(y-\mu)) \cdot |\Sigma|^{-1/2} \\
=& (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right\}
\end{aligned}$$

性质：对 $\mu \in \mathbb{R}^n$ 和 n 阶对称非负定阵 Σ，设 Σ 的秩为 $m \le n$，则存在列满秩矩阵 $B_{n\times m}$ 和 m 元的标准多元正态分布随机向量 X 使得 $Y = \mu + BX$ 服从多元正态分布 $\mathrm{N}(\mu, \Sigma)$ 分布。

证明：

由线性代数知识，Σ 有如下的特征值分解：

$$\Sigma = U\mathrm{diag}(\lambda_1, \dots, \lambda_m, 0, \dots, 0)U^T,$$

其中 $\lambda_1 \geq \cdots \geq \lambda_m > 0$ 是 Σ 的正特征值，U 为 n 阶正交阵，记 $U = (U_1\ U_2)$，其中 U_1 是 U 的前 m 列组成的矩阵，记 $\Lambda_1 = \mathrm{diag}(\lambda_1, \dots, \lambda_m)$, $\Lambda_1^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \dots, \lambda_m^{1/2})$, 则

$$\Sigma = (U_1\ U_2) \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix} = U_1 \Lambda_1 U_1^T$$

令 $B = U_1 \Lambda_1^{1/2}$，则 $BB^T = \Sigma$，于是若 X 服从 m 元的标准多元正态分布，则 $Y = \mu + BX$ 服从多元正态分布 $\mathrm{N}(\mu, BB^T)$ 即 $\mathrm{N}(\mu, \Sigma)$。

3.5.2 二元正态分布

二元正态分布的协方差阵为

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

行列式 $|\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2)$, $|\Sigma| = 0$ 当且仅当 $\rho = \pm 1$。$|\rho| < 1$ 时有联合密度

$$\begin{aligned} p_Y(y) =& \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 \right.\right. \\ & \left.\left. -2\rho \left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) \right] \right\} \end{aligned}$$

3.5.3 正态条件分布

设

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}(\mu, \Sigma),\ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

则 $X_2 = x_2$ 条件下 X_1 的条件分布为

$$\mathrm{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

条件方差不依赖于 x_2 的值。

令

$$\begin{aligned} X_{1\cdot 2} =& X_1 - E(X_1|X_2) = X_1 - \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \\ \Sigma_{11\cdot 2} =& \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

则 $(X_2, X_{1\cdot 2})$ 独立，$X_{1\cdot 2} \sim \mathrm{N}(0, \Sigma_{11\cdot 2})$。

3.5.4 多元正态分布等价定义证明

如果随机向量 Z 有特征函数

$$\phi(t) = \exp(it^T\mu - \frac{1}{2}t^T\Sigma t),$$

其中 Σ 是 n 阶非负定矩阵，则存在分量独立同标准正态分布的随机向量 ε 和常数矩阵 B 使得 $Z = \mu + B\varepsilon$。

证明：令

$$Y = Z - \mu$$

则 Y 的特征函数为

$$\phi_Y(t) = E\exp\left[it^TY\right] = \exp\left[-\frac{1}{2}t^T\Sigma t\right]$$

设 $\mathrm{rank}(\Sigma) = m \le n$, 做特征值分解

$$\begin{aligned}
&\Sigma = P^T\Lambda P,\ P^TP = I_n, \\
&\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_m, 0, \dots, 0)
\end{aligned}$$

($\text{其中 } \lambda_j > 0, j = 1, 2, \dots, m$)。令

$$\begin{aligned}
A =& \mathrm{diag}(\lambda_1^{-\frac{1}{2}}, \lambda_2^{-\frac{1}{2}}, \dots, \lambda_m^{-\frac{1}{2}}, 1, \dots, 1) \\
W =& APY
\end{aligned}$$

则

$$Y = P^TA^{-1}W \triangleq DW,$$

其中

$$\mathrm{Var}(W) = \mathrm{Var}(APY) = AP\Sigma P^TA = A\Lambda A = \mathrm{diag}(1, 1, \dots, 1, 0, \dots, 0)$$

所以

$$W = \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix}$$

其中 ε 为 m 维。记

$$G = (I_m \quad 0)_{m\times n}$$

则 $\varepsilon = GW = GAPY$, ε 的特征函数为

$$\begin{aligned}
\phi_\varepsilon(t) =& E\exp\left[it^TGAPY\right] = \phi_Y(P^TAG^Tt) \\
=& \exp\left[-\frac{1}{2}t^TGAP\Sigma P^TAG^Tt\right] = \exp\left[-\frac{1}{2}t^Tt\right]
\end{aligned}$$

即 $\varepsilon \sim \mathrm{N}(0, I_m)$。则

$$\begin{aligned} Y =& DW = (D_1\ D_2) \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix} \\ =& D_1 \varepsilon \triangleq B\varepsilon \\ Z =& \mu + B\varepsilon. \end{aligned}$$

证毕。

Chapter 4

严平稳序列及其遍历性

4.1 严平稳

随机向量同分布是指其联合分布函数相同。

时间序列 $\{X_t\}$ 与 $\{Y_t\}$ 同分布，当且仅当 $\forall n \in \mathbb{N}_+$ 和 $t_1, \dots, t_n \in \mathbb{Z}$, $(X(t_1), \dots, X(t_n))^T$ 与 $(Y(t_1), \dots, Y(t_n))^T$ 同分布。

定义 4.1 (严平稳列)**.** 时间序列 $\{X_t\}$ 对 $\forall n \in \mathbb{N}_+$ 和 $k \in \mathbb{Z}$ 都有

$$(X_1, X_2, \dots, X_n)^T \text{ 和} (X_{1+k}, X_{2+k}, \dots, X_{n+k})^T \text{ 同分布.}$$

即分布平移不变，称 $\{X_t\}$ 为**严平稳**时间序列。

若 $\{X_t\}$ 严平稳，对任多元函数 $\phi(x_1, x_2, \dots, x_m)$ 令

$$\{Y_t = \phi(X_{t+1}, \dots, X_{t+m}),\ t \in \mathbb{Z}\}$$

则 $\{Y_t\}$ 仍是严平稳列。

严平稳与宽平稳关系：

- 二阶矩有限的严平稳为宽平稳。
- 宽平稳一般不是严平稳。
- 正态平稳列既是宽平稳也是严平稳。

- 平稳序列 = 宽平稳序列 = 弱平稳序列。
- 严平稳序列 = 强平稳序列。

4.2 遍历性

时间序列一般只有一条轨道。要用时间序列 $\{X_t\}$ 的一次实现 $x_1, x_2, ...x_T$ 推断 $\{X_t(\omega), t \in \mathbb{N}, \omega \in \Omega\}$ 的统计性质. 遍历性可以保证从一条轨道可以推断整体的统计性质。

如果严平稳序列是遍历的, 从它的一次实现 $x_1, x_2, \dots, x_T$ 就可以推断出这个严平稳序列的所有有限维分布:

$$\begin{aligned}&F(x_1, x_2, ..., x_m)\\=&P(X_1 \leq x_1, X_2 \leq x_2, ..., X_m \leq x_m), \quad m \in \mathbb{N}.\end{aligned}$$

有遍历性的严平稳序列被称作**严平稳遍历序列**.

严平稳遍历的严格定义依赖于用测度论叙述的保测变换、不变集、不变随机变量概念，详见王梓坤《随机过程通论》第 197–204 页（北京师范大学出版社，1996）。

定理 4.1 (遍历定理)**.** *如果 $\{X_t\}$ 是严平稳遍历序列, 则有如下的结果:*

1. *强大数律: 如果 $E|X_1| < \infty$ 则*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} X_t = EX_1, \ a.s.$$

2. *对任何多元函数 $\phi(x_1, x_2, \cdots, x_m)$,*

$$Y_t = \phi(X_{t+1}, X_{t+2}, \cdots, X_{t+m})$$

是严平稳遍历序列.

定理 4.2. *如果 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$, 实数列 $\{a_j\}$ 平方可和, 则线性平稳序列*

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z},$$

是严平稳遍历的。

这说明在独立同分布白噪声条件下线性平稳列满足严平稳遍历条件。所以，在许多教材中线性平稳列都假定独立同分布白噪声条件。

例 4.1. 对严平稳序列 $\{X_t\}$, 设一条轨道为 $X_1, \dots, X_T$，给出有限维分布的强相合估计。

解答：定义严平稳序列

$$\begin{aligned} Y_t =& I[X(t+t_1) \le y_1, X(t+t_2) \le y_2, \cdots, X(t+t_m) \le y_m], \\ & t \in \mathbb{Z}. \end{aligned}$$

这里 $I[A]$ 是事件 A 的示性函数. 因为 $\{X_t\}$ 是遍历的, 由定理4.1的第 2 条知道 $\{Y_t\}$ 也是遍历的, 并且有界. 利用定理4.1的第 1 条 (强大数律) 得到

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} Y_t = EY_1 \\ =& P(X(t_1) \le y_1, X(t_2) \le y_2, \cdots, X(t_m) \le y_m), \text{ a.s.} \end{aligned}$$

这个例子说明, 在几乎必然的意义下, 严平稳遍历序列 $\{X_t\}$ 的每一次观测在观测长度趋于无穷时都可以决定 $\{X_t\}$ 的有限维分布.

例 4.2. 对严平稳序列 $\{X_t\}$, 设其二阶矩有限，设一条轨道为 $X_1, \dots, X_T$，给出相关函数 $E(X_t X_{t+k})$ 的强相合估计。

解答：这时 $Y_t = X_t X_{t+k}$ 也是严平稳遍历序列。于是

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Y_t \to E(Y_1) = E(X_t X_{t+k}), \text{ a.s.}$$

由此可知

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n) \to \gamma_k, \text{ a.s.}$$

4.3 附录：随机过程知识

随机过程的分布由其所有有限维分布决定，有限维分布族对次序交换保持一致，对取边缘分布保持一致，称为 Kolmogorov 相容性条件。

定理 4.3 (存在性定理)**.** *给定足标集和满足 Kolmogorov 相容性条件的有限维分布函数族，必存在相应分布族的随机过程。*

构造不唯一。见王梓坤《随机过程论》。

定理 4.4 (正态过程存在性定理)**.** *设 T 为足标集，a_t 为实值函数，$\sigma_{s,t}$ 为二元实值函数，对称，非负定，则必存在正态过程 $\{\xi_t, t \in T\}$ 使其均值函数为 a_t，自协方差函数为 $\sigma_{s,t}$。*

见 (谢衷洁, 1990)P5。

复值正态分布: 实部和虚部为联合正态分布。

Chapter 5

Hilbert 空间中的平稳序列

5.1 Hilbert 空间

本章中设 $\{X_t\}$ 是平稳序列。X_t 可以用线性组合 $\sum_{j=1}^k a_j x_{t-j}$ 预测。应该推广到用无穷的线性组合 $\sum_{j=1}^\infty a_j x_{t-j}$ 预测。这就需要研究这样的无穷线性组合的性质。

在线性代数的欧式空间理论中，一个向量 y 到一个子空间 S 的最短距离，是 y 与 y 在 S 上的投影向量的距离。将预测误差按均方误差度量，也可以得到类似结论，但是需要考虑无穷维空间的问题。

所有二阶矩有限的随机变量组成的一个无穷维函数空间 L^2，所有 $\{X_t\}$ 及其有限和无穷线性组合也构成一个无穷维函数空间，是 L^2 的子空间。下面从有限线性组合讲起。

5.1.1 平稳列导出的线性空间

设 $\{X_t\}$ 是平稳序列。令

$$L^2(X) = \left\{ \sum_{j=1}^k a_j X(t_j) \middle| a_j \in \mathbb{R}, t_j \in \mathbb{Z}, 1 \le j \le k, k \in \mathbb{N}_+ \right\}$$

$\forall X, Y, Z \in L^2(X), a, b \in \mathbb{R}$ 有

1. $X + Y = Y + X \in L^2(X)$, $(X + Y) + Z = X + (Y + Z)$;

2. $0 \in L^2(X)$, $X + 0 = X$, $X + (-X) = 0 \in L^2(X)$;
3. $a(X+Y) = aX + aY \in L^2(X)$, $(a+b)X = aX + bX$, $a(bX) = (ab)X$.

即 $L^2(X)$ 是一个线性空间。

5.1.2 将线性空间推广为 Hilbert 空间

令 $L^2 = \{Y : EY^2 < \infty\}$，易见 L^2 是线性空间，$L^2(X)$ 是 L^2 空间的子空间。

在 L^2 中定义内积 $\langle X, Y\rangle = E(XY)$，则

$$\langle X, Y\rangle = \langle Y, X\rangle$$
$$\langle aX + bY, Z\rangle = a\langle X, Z\rangle + b\langle Y, Z\rangle$$

$\langle X, X\rangle \geq 0$, 并且 $\langle X, X\rangle = 0$ 当且仅当 $X = 0$, a.s., 所以 L^2 又是内积空间. L^2 中的 0 元素，定义为 a.s. 等于 0 的随机变量。

内积有 Schwarz 不等式 (习题 1.6.2)

$$|\langle X, Y\rangle| \leq [\langle X, X\rangle\langle Y, Y\rangle]^{1/2}.$$

我们还需要对无穷线性组合的封闭性。需要极限的概念。

由内积定义模 (范数)

$$\|X\| = (\langle X, X\rangle)^{1/2}$$

Schwarz 不等式可以写成

$$|\langle X, Y\rangle| \leq \|X\| \cdot \|Y\|.$$

由 Schwarz 不等式可以证明模满足三角不等式

$$\|X + Y\| \leq \|X\| + \|Y\|.$$

由模定义距离

$$\|X - Y\| = (\langle X - Y, X - Y\rangle)^{1/2}.$$

则 $\|X - Y\| = \|Y - X\| \geq 0$, 且 $\|X - Y\| = 0$ 当且仅当 $X = Y$, a.s.. 距离满足三角不等式:

$$\|X - Y\| \leq \|X - Z\| + \|Z - Y\|.$$

对有限维线性空间，定义了内积已经足够，因为有限维内积空间与 $\mathbb{R}^n$ 同构。对无穷维空间，要考虑极限问题。

定义 5.1. 对 $\xi_n \in L^2$, $\xi_0 \in L^2$:

(1) 如果 $\lim_{n\to\infty}\|\xi_n-\xi_0\|=0$, 则称 ξ_n 在 L^2 中 (或均方) 收敛到 ξ_0, 记做 $\xi_n \overset{\text{m.s.}}{\to} \xi_0$ 或 $\xi_n \overset{L^2}{\to} \xi_0$.

(2) 如果当 $n,\ m\to\infty$ 时, $\|\xi_n-\xi_m\|\to 0$, 则称 $\{\xi_n\}$ 是 L^2 中的基本列或 Cauchy 列.

定理 5.1. *如果 $\{\xi_n\}$ 是 L^2 中的基本列, 则 (在 a.s. 的意义下) 有惟一的 $\xi \in L^2$ 使得 $\xi_n \overset{m.s.}{\to} \xi$.*

证明略。(同学们自学)

定义 5.2 (完备的内积空间)**.** 每个基本列都有极限在空间内的内积空间称为**完备**的内积空间，又称 **Hilbert 空间**。

由定理5.1，L^2 是 Hilbert 空间。

用 $\bar{L}^2(X)$ 表示 L^2 中包含 $L^2(X)$ 的最小闭子空间，则 $\bar{L}^2(X)$ 是 Hilbert 空间，称为由平稳序列 $\{X_t\}$ 生成的 Hilbert 空间。

5.1.3 内积的连续性

定理 5.2 (内积的连续性)**.** *在内积空间中, 如果 $\|\xi_n-\xi\|\to 0$, $\|\eta_n-\eta\|\to 0$ 则有*

1. $\|\xi_n\|\to\|\xi\|$,
2. $\langle\xi_n,\eta_n\rangle\to\langle\xi,\eta\rangle$.

注意：由第 2 条当然有

$$\langle\xi_n,\eta\rangle\to\langle\xi,\eta\rangle$$

证明:

$$\begin{aligned}
(1)\quad & \Big| \|\xi_n\| - \|\xi\| \Big| \\
\le & \|\xi_n - \xi\| \to 0 \ (n \to \infty) \\
(2)\quad & \Big| \langle \xi_n, \eta_n \rangle - \langle \xi, \eta \rangle \Big| \\
= & \Big| \langle \xi_n, \eta_n - \eta \rangle + \langle \xi_n - \xi, \eta \rangle \Big| \\
\le & \Big| \langle \xi_n, \eta_n - \eta \rangle \Big| + \Big| \langle \xi_n - \xi, \eta \rangle \Big| \\
\le & \|\xi_n\| \cdot \|\eta_n - \eta\| + \|\xi_n - \xi\| \cdot \|\eta\| \\
\to & 0. \ (n \to \infty)
\end{aligned}$$

例 5.1 (n 维欧式空间)**.** $\mathbb{R}^n$ 是线性空间，定义内积 $\langle a, b \rangle = a^T b$ 则为内积空间。$\mathbb{R}^n$ 是完备的内积空间。$|a| = \sqrt{a^T a}$ 为欧氏模。此空间有限维：由 n 个元素（n 维向量）组成基。

5.1.4 平稳列的有限维欧式空间

设 $\{X_t\}$ 是零均值平稳列，$X = (X_1, \dots, X_n)$。令 $L_n = \text{sp}\{X_1, \dots, X_n\} = \{a^T X : a \in \mathbb{R}^n\}$, 则 L_n 是 Hilbert 空间，称为由 X 生成的 Hilbert 空间。

L_n 是线性空间和内积空间易验证，下面证明其完备性。

先设 $\{X_t\}$ 是标准白噪声 WN$(0, 1)$。对任何线性组合 $\xi_k = a_k^T X$, 只要

$$\|\xi_k - \xi_m\|^2 = \|a_k^T X - a_m^T X\|^2 = (a_k - a_m)^T (a_k - a_m) \to 0,$$

由例5.1知道有 $a \in \mathbb{R}^n$ 使得

$$|a_k - a| \to 0$$

当 $k \to \infty$, 取 $\xi = a^T X$ 时,

$$\|\xi_k - \xi\|^2 = (a_k - a)^T (a_k - a) \to 0.$$

可见 $\{X_t\}$ 是标准白噪声 WN$(0, 1)$ 时 L_n 是完备的.

对一般的零均值平稳序列, 可以设协方差阵 $\Gamma = E(XX^T)$ 的秩是 m, $m \le n$. 对 Γ 做特征值分解得 (注意这是随机向量的常用手法)

$$\begin{aligned}
\Gamma = & P^T \Lambda P \\
\Lambda = & \text{diag}\{\lambda_1, \dots, \lambda_m, 0, \dots, 0\}
\end{aligned}$$

令

$$\begin{aligned} A =& \text{diag}\{\lambda_1^{-1/2}, \dots, \lambda_m^{-1/2}, 1, \dots, 1\} \triangleq \Lambda^{-1/2} \\ Y =& APX, \quad X = P^T A^{-1} Y \end{aligned}$$

则

$$\begin{aligned} \text{Var}(Y) =& AP\text{Var}(X)P^T A = APP^T \Lambda PP^T A \\ =& \text{diag}\{1, \dots, 1, 0, \dots, 0\} \end{aligned}$$

因此 $Y_1, \dots, Y_m$ 是某零均值白噪声列的某段。X 的线性组合即 $Y_1, \dots, Y_m$ 的线性组合。因此 L_n 是完备的。

5.1.5 L^2 意义下的线性序列

考虑 L^2 中的零均值白噪声列 $\{\varepsilon_t\}$, 设 $\text{Var}(\varepsilon_t) = \sigma^2$. 设 $\{a_j\} \in l_2$. 令

$$\xi_n(t) = \sum_{j=-n}^{n} a_j \varepsilon_{t-j}, \ t \in \mathbb{Z}$$

则 $\xi_n(t) \in L^2$。对 $m < n$，当 $m \to \infty$ 时

$$\begin{aligned} \|\xi_n - \xi_m\|^2 =& \| \sum_{m<|j|\le n} a_j \varepsilon_{t-j} \|^2 \\ =& \sigma^2 \sum_{m<|j|\le n} a_j^2 \le \sigma^2 \sum_{|j|>m} a_j^2 \to 0 \end{aligned}$$

由 L^2 完备性知存在 $X_t \in L^2$ 使得 $\xi_n(t) \stackrel{\text{m.s.}}{\to} X_t$。

记 $\xi_n(t)$ 在 L^2 中的极限 X_t 为

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \ t \in \mathbb{Z}$$

来证明 $\{X_t\}$ 平稳。

由 L^2 中内积连续性得

$$EX_t = \lim_{n\to\infty} \langle \xi_n(t), 1 \rangle = \lim_{n} E\xi_n(t) = 0$$

以及

$$\begin{aligned}
EX_tX_{t+k} =&\lim_n \langle \xi_n(t), \xi_n(t+k)\rangle \\
=&\lim_n \langle \sum_{j=-n}^{n} a_j\varepsilon_{t-j}, \sum_{l=-n}^{n} a_l\varepsilon_{t+k-l}\rangle \\
=&\lim_n \sum_{j=-n}^{n}\sum_{l=-n}^{n} a_ja_l\delta_{k-l+j}\sigma^2 \\
=&\sigma^2 \sum_{j=-\infty}^{\infty} a_ja_{j+k}
\end{aligned}$$

这就证明了 $\{X_t\}$ 为平稳列。

5.2 复值时间序列

设 X, Y 为随机变量，$Z = X + iY$ 称为复值随机变量。$EZ = EX + iEY$.

$$\mathrm{Cov}(Z_1, Z_2) = E\left((Z_1 - EZ_1)(Z_2 - EZ_2)^*\right)$$

(Z^* 表示 Z 的共轭转置)

对于两个复随机向量 Z_1, Z_2，

$$\mathrm{Cov}(Z_1, Z_2) = E\left((Z_1 - EZ_1)(Z_2 - EZ_2)^*\right)$$

$E|Z|^2 = EX^2 + EY^2 < \infty$ 时称 Z 是二阶矩有限的复随机变量。所有二阶矩有限复随机变量的集合 H 在定义内积 $\langle X, Y\rangle = E(XY^*)$ 后构成 Hilbert 空间。

复值随机变量的序列 $\{Z_n\}$ 称为**复值时间序列**. 若 $EZ_n = \mu$, $\mathrm{Cov}(Z_n, Z_m) = \gamma_{n-m}$，$\forall n, m \in \mathbb{Z}$, 则称 $\{Z_n\}$ 是**复值平稳序列**。

$$\gamma_{-k} = \gamma_k^*.$$

若复值零均值平稳列 $\{\varepsilon_t\}$ 满足

$$\mathrm{Cov}(\varepsilon_n, \varepsilon_m) = \sigma^2\delta_{n-m}, \quad \forall n, m \in \mathbb{Z}$$

则称 $\{\varepsilon_t\}$ 为**复值零均值白噪声**。

例 5.2. 设 $Y \sim \mathrm{U}[-\pi,\pi]$。定义

$$\varepsilon_n = e^{inY},\ n \in \mathbb{Z}.$$

则

$$\begin{aligned} E\varepsilon_n =& \delta_n \\ E(\varepsilon_n \bar{\varepsilon}_m) =& \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)y}\, dy \\ =& \delta_{n-m} \end{aligned}$$

即 $\{\varepsilon_t\}$ 为复值白噪声 (需要将 ε_0 定义为 0)。

对于平方可和的实数列 $\{a_j\}$, 定义

$$Z_n = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{n-j} = \sum_{j=-\infty}^{\infty} a_j e^{i(n-j)Y}, \quad n \in \mathbb{Z}.$$

由内积的连续性得到

$$\begin{aligned} EZ_n =& \sum_{j=-\infty}^{\infty} a_j E\varepsilon_{n-j} = a_n \\ E(Z_n \bar{Z}_m) =& \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k E(\varepsilon_{n-j} \varepsilon_{m-k}) \\ =& \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k \delta_{n-m-j+k} \\ =& \sum_{k=-\infty}^{\infty} a_k a_{k+n-m} \end{aligned}$$

因为均值非常数，所以 $\{Z_n\}$ 不是复值平稳列。

另一方面，

$$\begin{aligned} & E(Z_n \bar{Z}_m) \\ =& E\left[\sum_{j=-\infty}^{\infty} a_j e^{i(n-j)Y} \sum_{k=-\infty}^{\infty} a_k e^{-i(m-k)Y} \right] \\ =& \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\sum_{j=-\infty}^{\infty} a_j e^{i(n-j)y} \right) \left(\sum_{k=-\infty}^{\infty} a_k e^{-i(m-k)y} \right)\, dy \\ =& \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{j=-\infty}^{\infty} a_j e^{-ijy} \right|^2 e^{i(n-m)y}\, dy \end{aligned}$$

即

$$\sum_{j=-\infty}^{\infty} a_j a_{j+k} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{j=-\infty}^{\infty} a_j e^{-ijy}\right|^2 e^{iky}\,dy.$$

定义

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|\sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda}\right|^2 = \frac{\sigma^2}{2\pi}\left|\sum_{j=-\infty}^{\infty} a_j e^{ij\lambda}\right|^2$$

就得到公式

$$\sigma^2\sum_{j=-\infty}^{\infty} a_j a_{j+k} = \int_{-\pi}^{\pi} f(y)e^{iky}\,dy \tag{5.1}$$

5.3 附录：补充知识

5.3.1 绝对可和系数的线性序列与 L2 的线性序列

系数绝对可和的线性平稳列是系数平方可和的线性平稳列的特例。系数绝对可和的线性平稳列的极限是 a.s. 极限，设 $\sum_j |a_j| < \infty$,

$$X_t = \sum_{j=-\infty}^{\infty} a_j\varepsilon_{t-j},\ \text{a.s.}$$

而

$$Y_t = \sum_{j=-\infty}^{\infty} a_j\varepsilon_{t-j},\ (L^2)$$

记

$$\xi_n = \sum_{j=-n}^{n} a_j\varepsilon_{t-j},$$

则

$$\xi_n \to X_t,\ \text{a.s.},\quad \xi_n \to Y_t,\ (L^2)$$

这时

$$\xi_n \overset{\text{Pr}}{\to} X_t,\quad \xi_n \overset{\text{Pr}}{\to} Y_t,$$

由测度论可知同一序列如果有两个依概率的极限则这两个极限 a.s. 相等。

5.3.2 随机过程的均方连续性

定义 5.3 (均方连续).

$$\lim_{h\to 0}\|\xi_{t_0+h}-\xi_{t_0}\|=0$$

称 $\{\xi_t\}$ 在 $t=t_0$ **均方连续**，如果对所有 $t_0\in\mathbb{R}$ 都成立则称称 $\{\xi_t\}$ 在 $\mathbb{R}$ 上均方连续。

均方连续不一定轨道连续。若 $\{\xi_t\}$ 是平稳过程 (连续时)，则 $\{\xi_t\}$ 在 $\mathbb{R}$ 上均方连续 $\iff$ $\{\xi_t\}$ 在 $t=0$ 均方连续 $\iff$ $\gamma(\tau)$ 在 $\mathbb{R}$ 连续 $\iff$ $\gamma(\tau)$ 在 $\tau=0$ 连续。

Chapter 6

时间序列的谱

遍历的时间序列可以从一次实现的时间分布进行统计分析，称为**时域分析**。平稳时间序列的二阶性质也可以从其频率分解来研究，称为**频域分析**。频谱的典型代表是声音。

6.1 声音频谱演示

利用Audacity软件演示声音波形和频谱。

1. 在 Audacity 中显示并播放笛子曲片段 music-demo1.wav。界面见图6.1。

2. 其中有一个 C5 音主频率为 523Hz，相应的片段存入了 music-demo1b.wav, 在 Audacity 中打开此片段并聆听，见图6.2。放大局部见图6.3。谱密度图见图6.4。其中最大值点就是乐器当前振动的频率，音阶为 C5(523Hz 左右)。还有一些其他的峰，在基础频率的倍数的位置。谱估计的纵轴用了对数尺度。

3. 我们人为用正弦波生成了频率为 523Hz 的声音文件，存入了 music-demo1c.wav。在 Audacity 中聆听、看波形、频谱图。放大的波形图见图6.5，这是一个严格的正弦波。频谱图见6.6。一个频率为 f_0 的正弦波声音，其振动波形的方程为

$$y_t = \sin(2\pi f_0 t), t > 0$$

t 为以秒为单位的连续时间。

连续波形必须按一定的采样频率变成离散时间的时间序列，比如每秒 8000 点，称为采样频率 8kHz, 这只要按每秒钟按上述方程生成 8000 个 y_t 值即可。常用的采样频率如 CD 音质，为 44.1kHz。可以记录单声道或者双声道，每个采样点的值一般用若干二进制位表示，比如 CD 音质使用 16bit 记录一个采样点的值，波形取值在 $\pm 2^{15}$ 之间。

保存声音的数据文件可以进行压缩，比如 MP3 格式就是一种压缩的有失真的压缩声音文件格式。用比特率表示保存每秒所需要的二进制位数，比如 MP3 常用的标准音质为 128kbps，即每秒钟 16K 字节，或每分钟 0.9M 字节。

4. 在 R 中读入 music-demo2.wav，这是笛子演奏片段，有三个音阶 `C5--D#5--C5`。在 R 中画 C5(523Hz)0.05 秒片段时间序列图、谱密度估计图、动态谱峰图。见 §6.4.4。
5. 在 R 中人为生成 523Hz 正弦波，画时间序列图、谱密度估计和动态谱峰图。见 §6.4.4。保存为 music-demo2b.wav。可在 Audacity 中查看和聆听。
6. 在 R 中直接生成的两个单音 C4 和 C5 叠加的复音的序列图和谱密度估计、动态谱峰图。保存为 music-demo2c.wav。可在 Audacity 中查看和聆听。

音调和频率的对照表：

音调	低音 (3)	中音 (4)	高音 (5)
C	131	262	523
D	147	294	587
E	165	330	659
F	175	349	698
G	196	392	784
A	220	440	880
B	247	494	988

频率 262 是中音 C 或 C4，是钢琴键盘的最中间 8 度音的 C 音，这一组是小字 1 组；频率 440 是中音 A 或 A440，频率 440 是国际标准音高。

在十二音阶中，相差八度音的两个音调（比如 C4, 262Hz 与 C5, 523Hz）的频率相差一倍；每个半音的频率为上一个半音频率的 $\sqrt[12]{2}$ 倍。

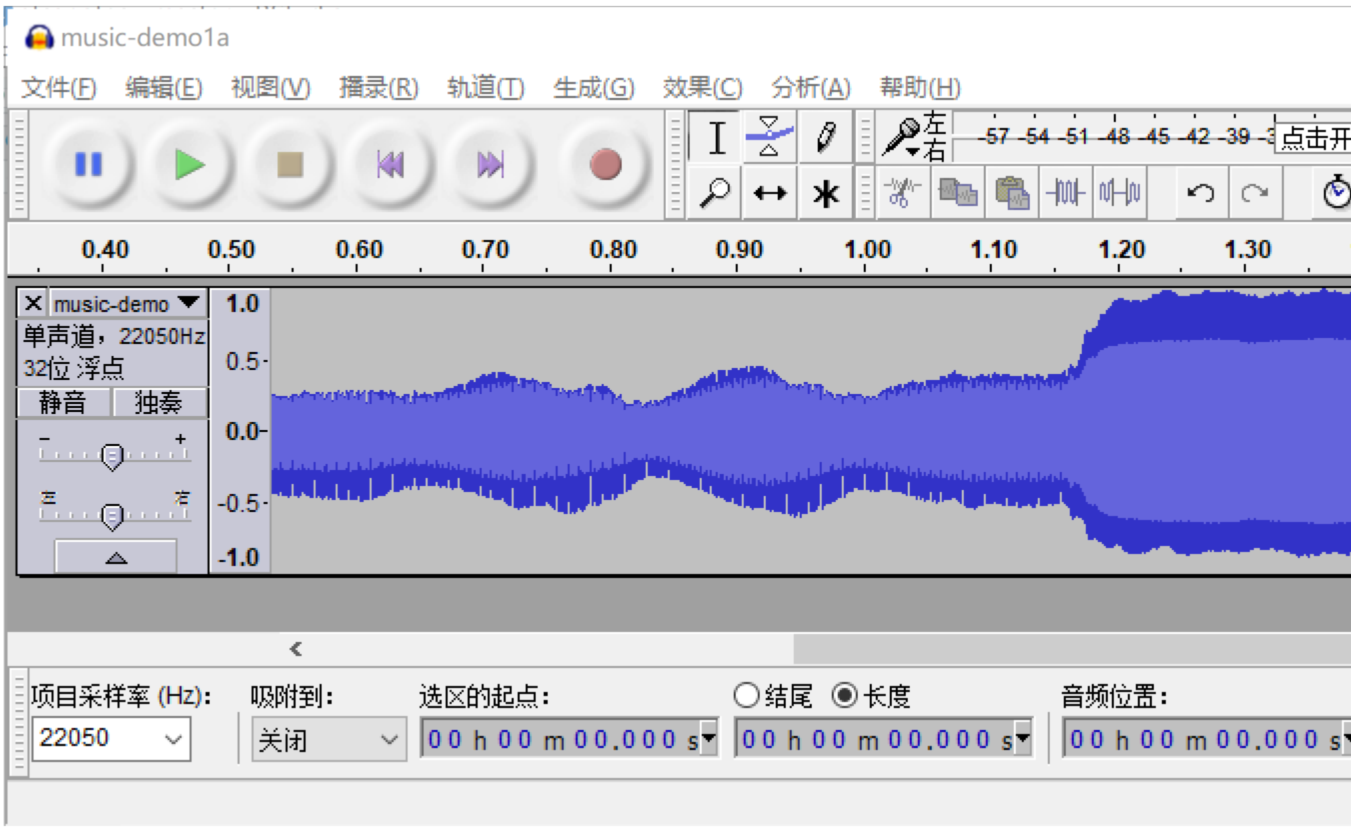


图 6.1: Audacity 运行界面

6.2 平稳序列的谱

6.2.1 平稳序列的谱函数

设平稳序列 $\{X_t\}$ 有自协方差函数 $\{\gamma_k\}$. 如果有 $[-\pi, \pi]$ 上的单调不减右连续的函数 $F(\lambda)$ 使得

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda}\, dF(\lambda), \quad F(-\pi) = 0, \quad k \in \mathbb{Z}, \tag{6.1}$$

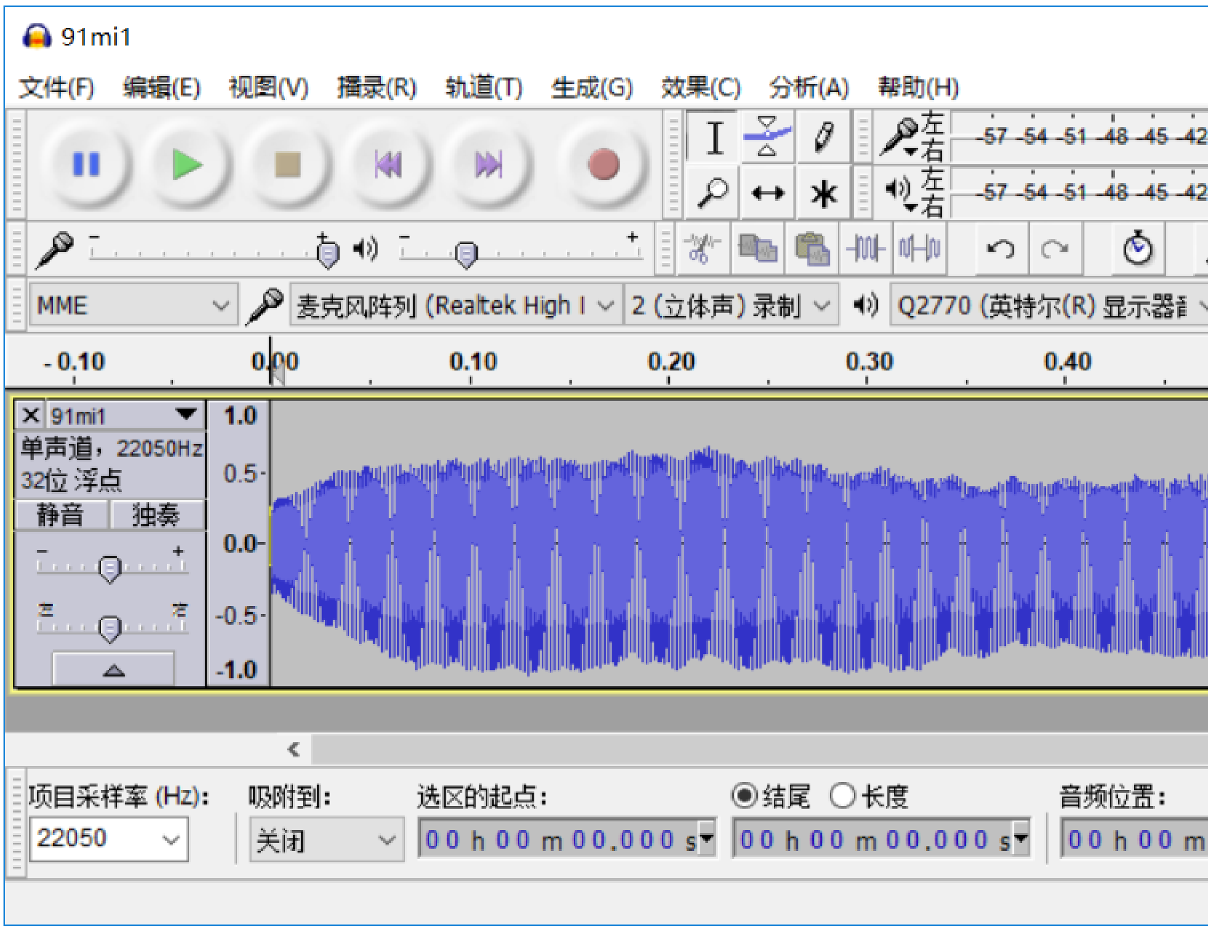


图 6.2: Audacity 中笛子音 C5

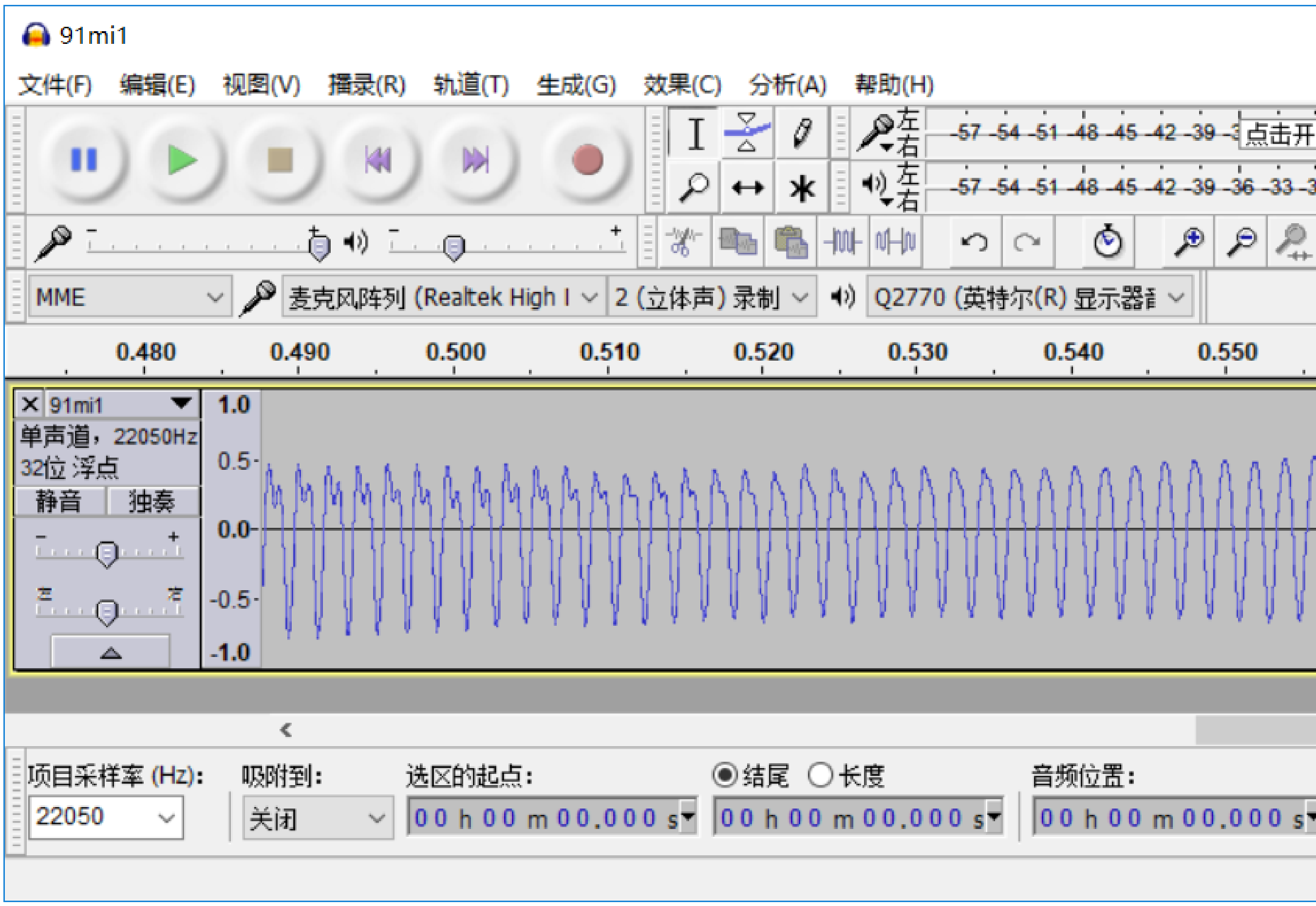


图 6.3: Audacity 中笛子音 C5 画面放大

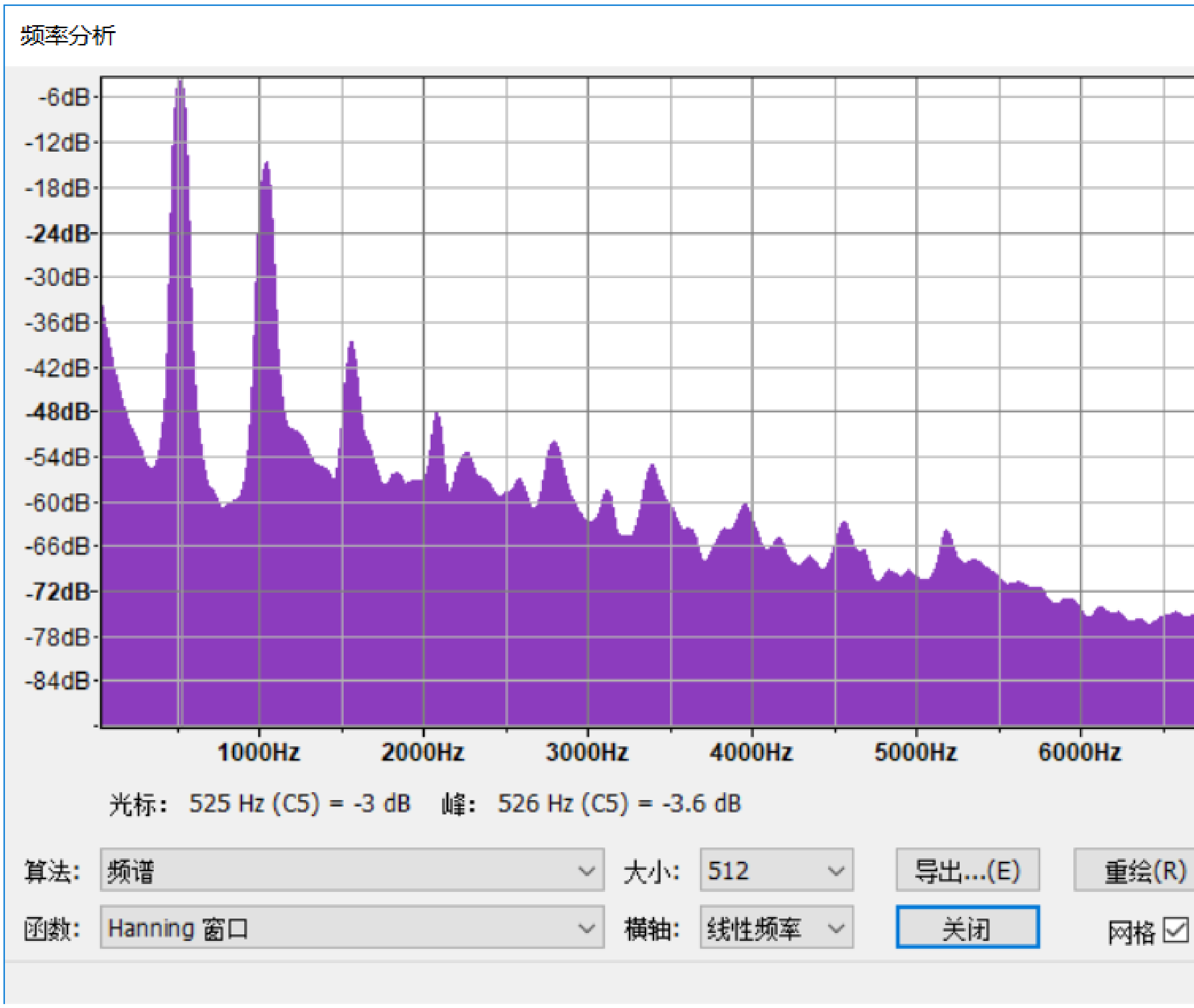


图 6.4: Audacity 中笛子音 C5 片段的谱密度

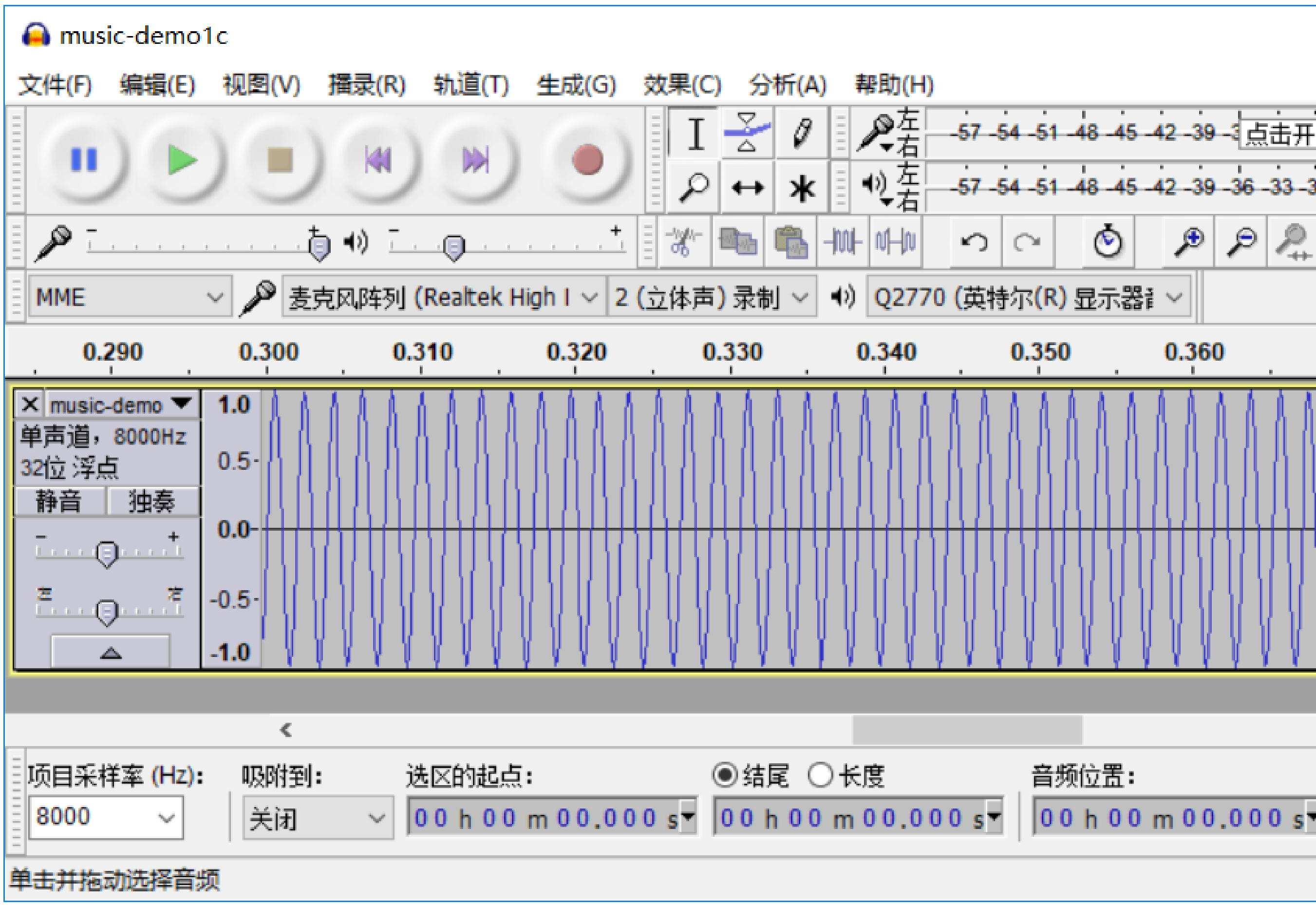


图 6.5: Audacity 中单音 C5 片段放大

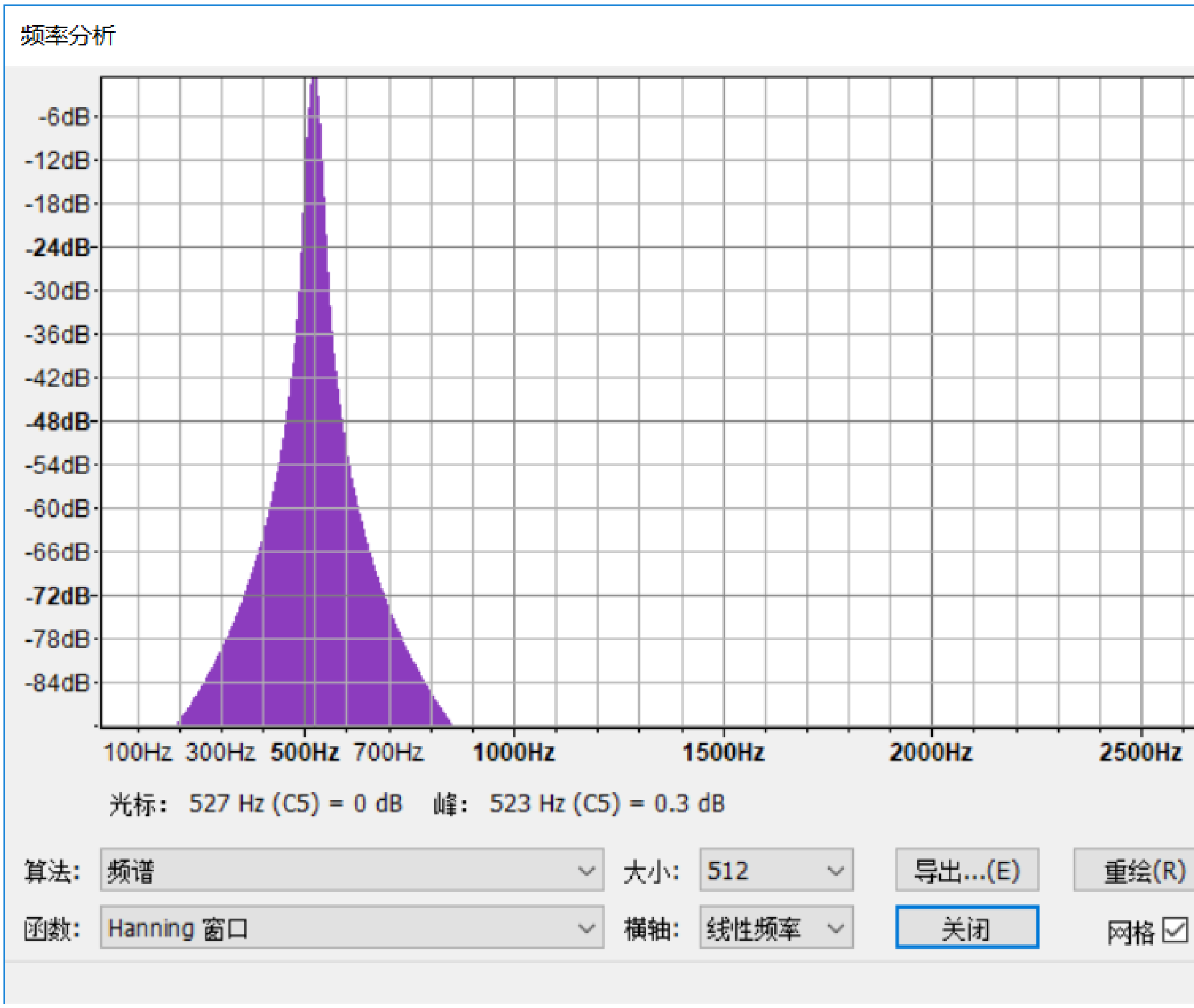


图 6.6: Audacity 中单音 C5 谱密度

则称 $F(\lambda)$ 是 $\{X_t\}$ 或 $\{\gamma_k\}$ 的**谱分布函数**, 简称为**谱函数**; 如果有 $[-\pi, \pi]$ 上的非负函数 $f(\lambda)$ 使得

$$\gamma_k = \int_{-\pi}^{\pi} f(\lambda) e^{ik\lambda} \, d\lambda, \quad k \in \mathbb{Z}, \tag{6.2}$$

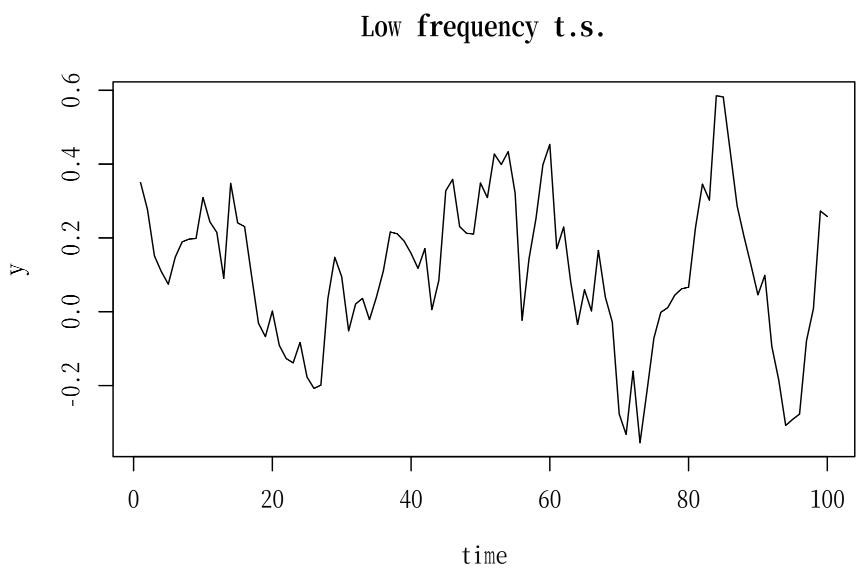
则称 $f(\lambda)$ 是 $\{X_t\}$ 或 $\{\gamma_k\}$ 的谱密度函数或功率谱密度, 简称为**谱密度**或功率谱.

谱反映了平稳序列的相关结构。谱密度是将原始序列看成许多个不同频率的余弦波的叠加时，不同频率的振幅平方大小，谱密度越高的地方，对应的频率成分的振幅越大。

例 6.1. 演示：高频时间序列和低频时间序列的序列图和密度.

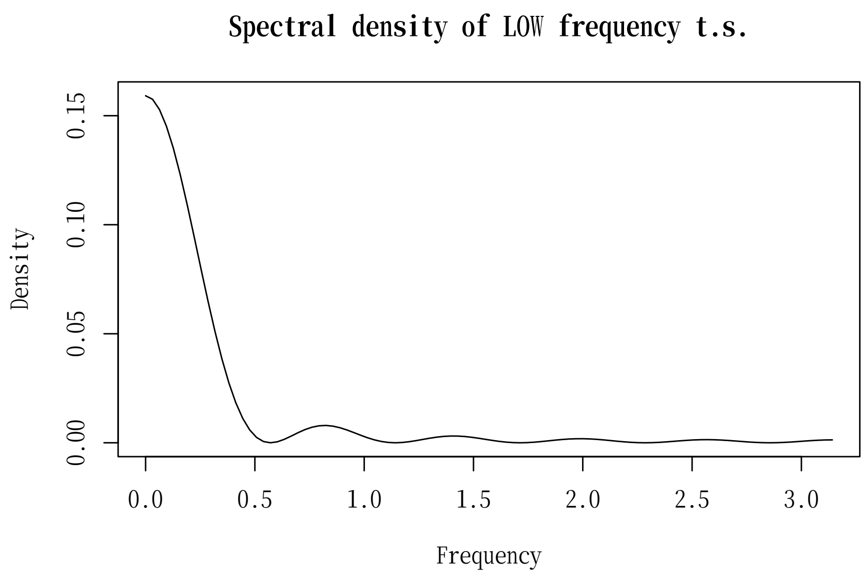
演示：低频时间序列图形

```r
set.seed(1)
m <- 10
n <- 100+m+1
fil <- rep(1/(m+1), m+1)
eps <- rnorm(n)
x <- filter(eps, fil, method="convolution", sides=1)
eps <- eps[-seq(m+1)]
x <- x[-seq(m+1)]
ts.plot(x, col="black", main="Low frequency t.s.",
        xlab="time", ylab="y")
```



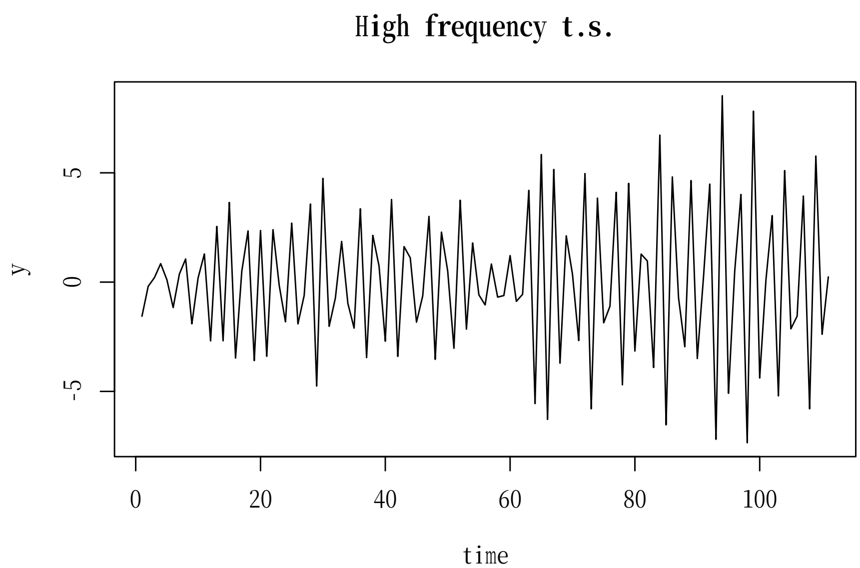
低频时间序列的谱密度样例（横坐标为 $[0, \pi]$ 内的角频率）：

```
fr <- seq(0, pi, length=100)
sp <- numeric(length(fr))
for(j in seq(along=fr)){
  sp[j] <- abs(sum(complex(mod=1, arg=-(0:m)*fr[j])))^2
}
sp <- sp/(2*pi*(m+1)^2)
plot(fr, sp, type='l', main="Spectral density of LOW frequency t.s.",
     xlab="Frequency", ylab="Density")
```



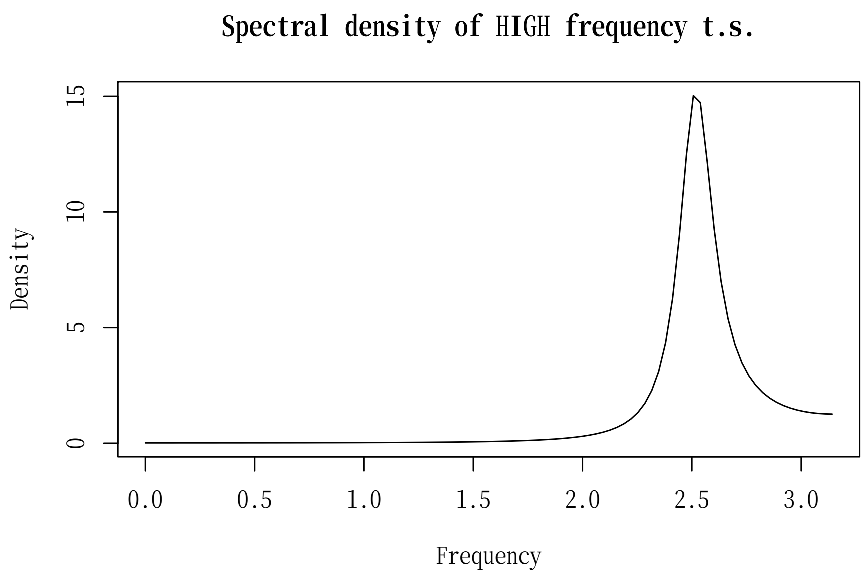
高频时间序列的模拟数据：

```
r0 <- complex(mod=1.1, arg=4/5*pi)
poly1 <- c(1, -2*Re(r0), Mod(r0)^2)
poly2 <- poly1 / poly1[3]
poly2 <- rev(poly2)
eps <- rnorm(n+100)
y <- numeric(length(eps))
for(j in 3:length(y)){
  y[j] = -poly2[2]*y[j-1] - poly2[3]*y[j-2] + eps[j]
}
y <- y[-seq(100)]
ts.plot(y, main="High frequency t.s.",
        xlab="time", ylab="y")
```



高频时间序列的谱密度样例：

```
sp2 <- numeric(length(fr))
for(j in seq(along=fr)){
  sp2[j] <- abs(sum(poly2 * complex(mod=1, arg=(0:2)*fr[j])))^2
}
sp2 <- 1/(2*pi) / sp2
plot(fr, sp2, type='l', main="Spectral density of HIGH frequency t.s.",
     xlab="Frequency", ylab="Density")
```



定理 6.1 (Herglotz 定理). 平稳序列的谱函数是惟一存在的.

(参见 (谢衷洁, 1990)P21 定理 1.6)

若 $F(\lambda)$ 和 $f(\lambda)$ 同时存在则

$$F(\lambda) = \int_{-\pi}^{\lambda} f(s)\, ds \tag{6.3}$$

若 $F(\lambda)$ 绝对连续则其几乎处处导数 $F'(\lambda)$ 为谱密度。若 $F(\lambda)$ 是连续函数，除去有限点外导函数存在且连续则

$$f(\lambda) = \begin{cases} F'(\lambda) & \text{当 } F'(\lambda) \text{ 存在} \\ 0 & \text{当 } F'(\lambda) \text{ 不存在} \end{cases}$$

是谱密度。

6.2.2 白噪声列的谱密度

设 $\{\varepsilon_t\}$ 为白噪声列 $\mathrm{WN}(\mu, \sigma^2)$。自协方差函数为

$$\gamma_k = \sigma^2 \delta_k, \quad k \in \mathbb{Z}$$

显然，令

$$f(\lambda) = \frac{\sigma^2}{2\pi}, \quad \lambda \in [-\pi, \pi]$$

有

$$\gamma_k = \int_{-\pi}^{\pi} f(\lambda) e^{ik\lambda} d\lambda, \quad k \in \mathbb{Z}$$

即白噪声列有常数谱密度。

反之，有常数谱密度的平稳列一定是白噪声列。

6.2.3 线性平稳列的谱密度

定理 6.2. 如果 $\{\varepsilon_t\}$ 是 $WN(0, \sigma^2)$, 实数列 $\{a_j\}$ 平方可和, 则线性平稳序列

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z},$$

有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} a_j e^{ij\lambda} \right|^2. \tag{6.4}$$

证明:

§5.1.5已经证明 $\{X_t\}$ 的自协方差函数为

$$\gamma_k = \sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+k}$$

由 §5.2的式(5.1)，可得到

$$\sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+k} = \int_{-\pi}^{\pi} f(y) e^{iky} dy, \quad k \in \mathbb{Z}$$

其中

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} a_j e^{ij\lambda} \right|^2$$

于是

$$\gamma_k = \int_{-\pi}^{\pi} f(y) e^{iky} dy, \quad k \in \mathbb{Z}$$

即 $f(\lambda)$ 是 $\{X_t\}$ 的谱密度。

实际上，存在谱密度也是线性序列的充分条件：

定理 6.3. *若平稳列 $\{X_t\}$ 有谱密度，则存在白噪声列 $\{\varepsilon_t\}$ 和平方可和的数列 $\{a_j\}$ 使得*

$$x_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j},\ t \in \mathbb{Z}.$$

参见 (谢衷洁, 1990) P.49 定理 1.14。

6.2.4 两正交序列的谱

定理 6.4. *设 $\{X_t\}$ 和 $\{Y_t\}$ 是相互正交的零均值平稳序列, c 是常数, 定义*

$$Z_t = X_t + Y_t + c, \quad t \in \mathbb{Z}.$$

1. *如果 $\{X_t\}$ 和 $\{Y_t\}$ 分别有谱函数 $F_X(\lambda)$ 和 $F_Y(\lambda)$, 则平稳序列 $\{Z_t\}$ 有谱函数 $F_Z(\lambda) = F_X(\lambda) + F_Y(\lambda)$.*

2. *如果 $\{X_t\}$ 和 $\{Y_t\}$ 分别有谱密度 $f_X(\lambda)$ 和 $f_Y(\lambda)$, 则 $\{Z_t\}$ 有谱密度 $f_Z(\lambda) = f_X(\lambda) + f_Y(\lambda)$.*

证明: 主要由

$$\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k)$$

及谱函数和谱密度定义可得。

6.2.5 线性滤波与谱

设平稳序列 $\{X_t\}$ 有谱函数 $F_X(\lambda)$ 和自协方差函数 $\{\gamma_k\}$. $H = \{h_j\}$ 是一个绝对可和的保时线性滤波器 (参见 §2.3). 当输入过程是 $\{X_t\}$ 时, 输出过程是

$$Y_t = \sum_{j=-\infty}^{\infty} h_j X_{t-j}, \quad t \in \mathbb{Z}.a.s. \tag{6.5}$$

输出的自协方差为

$$\gamma_Y(k) = \sum_{l,j=-\infty}^{\infty} h_l h_j \gamma(k+l-j). \tag{6.6}$$

求和意义为实数级数绝对收敛。

由控制收敛定理

$$
\begin{aligned}
\gamma_Y(k) &= \sum_{l,j=-\infty}^{\infty} h_l h_j \int_{-\pi}^{\pi} \exp(i(k+l-j)\lambda)\ dF_X(\lambda) \\
&= \int_{-\pi}^{\pi} \sum_{l,j=-\infty}^{\infty} h_l h_j \exp(i(l-j)\lambda) e^{ik\lambda}\ dF_X(\lambda) \\
&= \int_{-\pi}^{\pi} \left| \sum_{j=-\infty}^{\infty} h_j \exp(-ij\lambda) \right|^2 e^{ik\lambda}\ dF_X(\lambda) \\
&= \int_{-\pi}^{\pi} |H(e^{-i\lambda})|^2 e^{ik\lambda}\ dF_X(\lambda),
\end{aligned} \tag{6.7}
$$

其中

$$
H(z) = \sum_{j=-\infty}^{\infty} h_j z^j,\ |z| \leq 1. \tag{6.8}
$$

线性滤波输出 $\{Y_t\}$ 的谱函数为

$$
F_Y(\lambda) = \int_{-\pi}^{\lambda} |H(e^{-is})|^2\ dF_X(s). \tag{6.9}
$$

当 $\{X_t\}$ 有谱密度 $f_X(\lambda)$ 时

$$
F_Y(\lambda) = \int_{-\pi}^{\lambda} |H(e^{-is})|^2 f_X(s)\ ds. \tag{6.10}
$$

即 $\{Y_t\}$ 谱密度为

$$
f_Y(\lambda) = |H(e^{-i\lambda})|^2 f_X(\lambda) \tag{6.11}
$$

结论归纳成如下定理.

定理 6.5. 设 $\{X_t\}$ 是平稳序列, $H = \{h_j\}$ 是绝对可和的保时线性滤波器, $\{Y_t\}$ 为滤波输出

$$
Y_t = \sum_{j=-\infty}^{\infty} h_j X_{t-j}, \quad t \in \mathbb{Z}.a.s. \tag{6.12}
$$

$H(z)$ 是滤波器 H 的特征多项式

$$
H(z) = \sum_{j=-\infty}^{\infty} h_j z^j,\ |z| \leq 1. \tag{6.13}
$$

1. 如果 $\{X_t\}$ 有谱函数 $F_X(\lambda)$, 则 $\{Y_t\}$ 有谱函数

$$F_Y(\lambda) = \int_{-\pi}^{\lambda} |H(e^{-is})|^2 \, dF_X(s). \tag{6.14}$$

2. 如果 $\{X_t\}$ 有谱密度 $f_X(\lambda)$, 则 $\{Y_t\}$ 有谱密度

$$f_Y(\lambda) = |H(e^{-i\lambda})|^2 f_X(\lambda) \tag{6.15}$$

由于(6.12)，序列 $\{h_j\}$ 看成 $j \in \mathbb{Z}$ 的函数，称为保时线性滤波器的**脉冲响应函数**。由于(6.15), $H(e^{-i\lambda})|^2$ 称为**频率响应函数**。

脉冲响应的含义是，如果输入 $\{X_t\}$ 中仅 $X_0 = 1$，其它都等于 0，则 $Y_j = h_j$，即输入在 $t = 0$ 时的一个单位的扰动，由于线性滤波的影响，输出在 $t = j$ 时受到的影响是 h_j。

频率响应的含义是，原来在 λ_0 频率附近的震荡能量，经过滤波后被放大到了原来的 $|H(e^{-i\lambda_0})|^2$ 倍。

6.2.6 采样定理

设 $\{X(t), t \in \mathbb{R}\}$ 是连续时的实值随机过程，若

$$\begin{aligned} &EX(t) \equiv \mu, \quad \forall t \in \mathbb{R} \\ &\text{Cov}(X(t), X(s)) = \gamma(t-s), \quad t, s \in \mathbb{R} \end{aligned}$$

则称 $\{X(t)\}$ 为**连续时平稳过程**。$\{\gamma(\tau), \tau \in \mathbb{R}\}$ 称为 $\{X(t)\}$ 的自协方差函数。

对连续时平稳过程 $\{X(t), t \in \mathbb{R}\}$，设自协方差函数为 $\{\gamma(\tau)\}$，若有 $(-\infty, \infty)$ 上的单调不减右连续的函数 $F(\lambda)$ 使得

$$\gamma(\tau) \ = \int_{-\infty}^{\infty} e^{i\tau\lambda} \, dF(\lambda), \quad F(-\infty) = 0, \quad \tau \in \mathbb{R},$$

则称 $F(\lambda)$ 是 $\{X_t\}$ 或 $\{\gamma(\tau)\}$ 的谱函数, 在一定理论条件下也有类似于 Herglotz 定理的结论。

如果有 $(-\infty, \infty)$ 上的非负函数 $f(\lambda)$ 使得

$$\gamma(\tau) \ = \int_{-\infty}^{\infty} f(\lambda) e^{i\tau\lambda} \, d\lambda, \quad \tau \in \mathbb{R},$$

则称 $f(\lambda)$ 是 $\{X_t\}$ 或 $\{\gamma(\tau)\}$ 的谱密度函数或功率谱密度, 简称为**谱密度**或功率谱.

连续时的平稳过程只能以一定的时间间隔记录，称为**采样**。如何保证不损失信息？

定理 6.6 (采样定理)**.** 设 $\{X(t), t \in \mathbb{R}\}$ 为连续时平稳过程，有谱函数 $F(\lambda)$ 且谱函数满足

$$\int_{|\lambda| \geq 2\pi f_0} dF(\lambda) = 0$$

则

$$X_t = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2f_0}\right) \frac{\sin(2\pi f_0 t - n\pi)}{2\pi f_0 t - n\pi} \quad (L^2), \quad t \in \mathbb{R}$$

这时记 $Y_n = X\left(\frac{n}{2f_0}\right), n \in \mathbb{Z}$, 则平稳时间序列 $\{Y_n\}$ 是 $X(t)$ 以 $\Delta = \frac{1}{2f_0}$ 等间隔取样的结果。从采样 $\{Y_n, n \in \mathbb{Z}\}$ 可以恢复原信号 $\{X(t), t \in \mathbb{R}\}$。

参见：(谢衷洁, 1990) P50 定理 1.15。

f_0 称为上界频率，采样间隔不大于 $\Delta = \frac{1}{2f_0}$ 采样恢复原始信号。

如果采样间隔大于 Δ，信号的高频部分有丢失和混杂入低频的问题。

6.2.6.1 声音信号采样

数字音频信号以一定频率采样，并把信号强度数字化为 8 位、16 位等。采样频率越高、数字化位数越多则信号保真度越高。CD 音质采样频率为 44.1kHz(即采样间隔为 $\Delta = 1/44100$ 秒, 上界频率为 22.05kHz), 信号强度用每声道 16 位记录。这样，立体声音频每秒钟记录的数据比特数为 $44100 \times 16 \times 2 = 1411200$ (称为码率 1411.2kbps)。每分钟约 80MB 数据。

MP3 编码使用比较低的采样速率和数字化位数，并利用听觉心理学进行了压缩，码率一般在 $32 \sim 320$kbps。

6.3 离散谱序列及其周期性

6.3.1 简单的离散谱序列

设随机变量 ξ,η。$E\xi=0, E\eta=0$; $E(\xi\eta)=0$, $E\xi^2=E\eta^2=\sigma^2$。常数 $\lambda_0\in(0,\pi]$。令

$$Z_t=\xi\cos(\lambda_0 t)+\eta\sin(\lambda_0 t), t\in\mathbb{N}_+ \tag{6.16}$$

写成极坐标表示:

$$A=\sqrt{\xi^2+\eta^2},\quad \cos\theta=\frac{\xi}{A},\ \sin\theta=\frac{\eta}{A} \tag{6.17}$$

$$Z_t=A\cos(t\lambda_0-\theta) \tag{6.18}$$

其实现是一个相位为 $-\theta$ 的角频率为 λ_0 的余弦函数的离散采样，表现并不随机，随机性表现在多个实现样本。

如果 ξ, η 服从独立的标准正态分布，A 服从 Rayleigh 分布，θ 服从 $\mathrm{U}(-\pi,\pi)$，A 和 θ 独立。

易见 $EZ_t=0$。自协方差函数为

$$\begin{aligned}
\gamma_{t-s}=&E(Z_tZ_s)\\
=&\Big[E(\xi^2)\cos(t\lambda_0)\cos(s\lambda_0)+E(\eta^2)\sin(t\lambda_0)\sin(s\lambda_0)\\
&+E(\xi\eta)\cos(t\lambda_0)\sin(s\lambda_0)+E(\xi\eta)\cos(s\lambda_0)\sin(t\lambda_0)\Big]\\
=&\sigma^2\left[\cos(t\lambda_0)\cos(s\lambda_0)+\sin(t\lambda_0)\sin(s\lambda_0)\right]+0+0\\
&(\text{注意}\xi,\eta\text{正交})\\
=&\sigma^2\cos((t-s)\lambda_0)
\end{aligned}$$

所以(6.18)定义的 $\{Z_t\}$ 平稳，有自协方差函数 $\{\gamma_k\}$。

现

$$\gamma_k=\sigma^2\cos(k\lambda_0),\ k\in\mathbb{Z}$$

若 $\lambda_0\neq\pi$, 谱函数为

$$F(\lambda)=\sigma^2[0.5I_{[-\lambda_0,\pi]}(\lambda)+0.5I_{[\lambda_0,\pi]}(\lambda)]$$

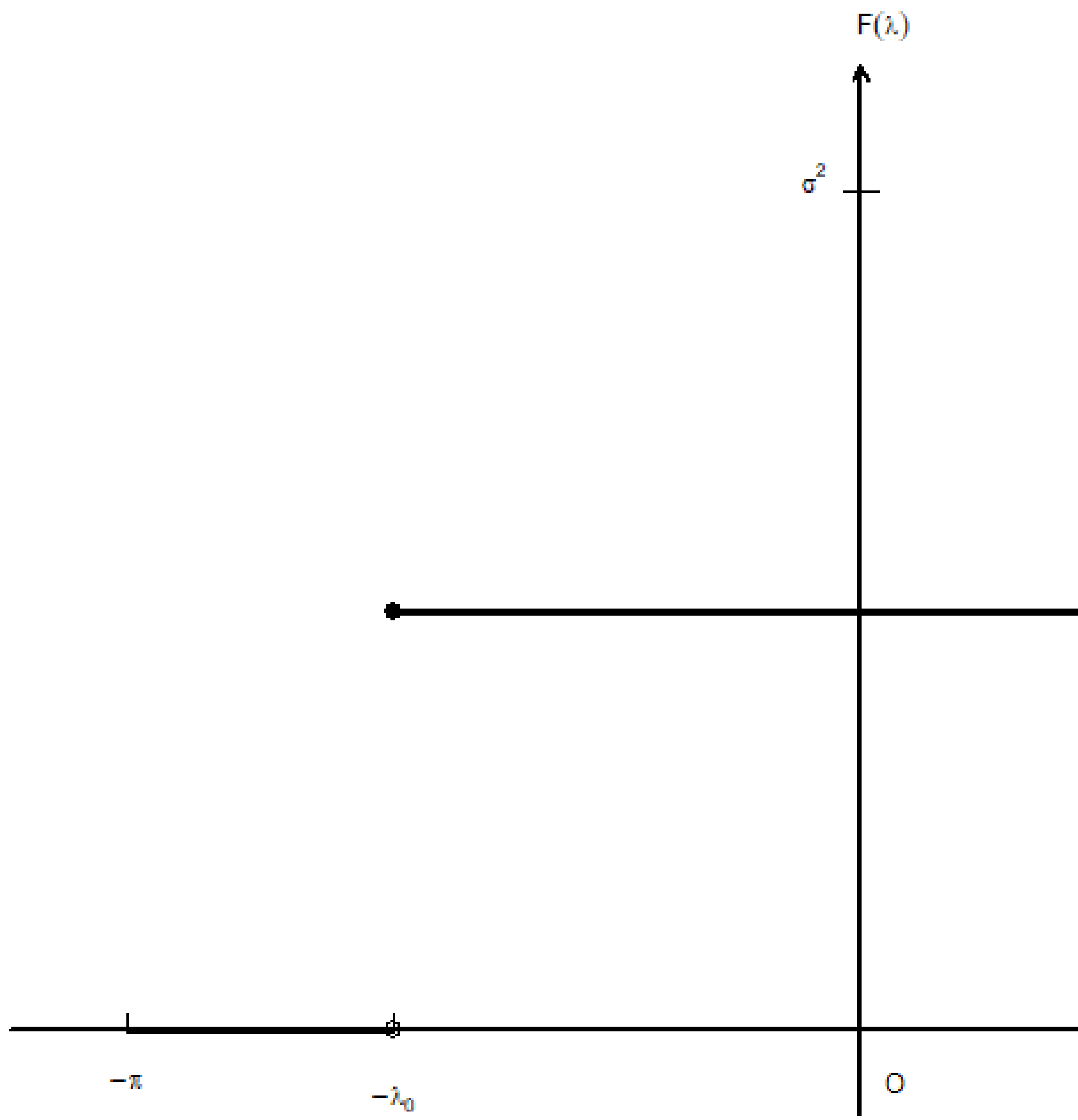


图 6.7: 简单离散谱序列的谱函数

这样的阶梯函数形式的谱函数表示在 $[-\pi,\pi]$ 上的一个测度，此测度仅在 $-\lambda_0$ 和 λ_0 两个点上有质量 $\sigma^2/2$。由积分与测度的关系可知

$$\int_{-\pi}^{\pi} e^{ik\lambda} dF(\lambda) = \frac{\sigma^2}{2}[e^{-ik\lambda_0} + e^{ik\lambda_0}] = \gamma_k$$

若 $\lambda_0 = \pi$, 则 $\gamma_k = \sigma^2 \cos(k\pi)$,

$$F(\lambda) = \sigma^2 I_{\{\pi\}}(\lambda),\ \lambda \in [-\pi,\pi]$$

如果将 γ_k 看成是 $k \in [0,\infty)$ 的函数，这是以 $\frac{2\pi}{\lambda_0}$ 为周期的一个余弦函数，当 k 等于周期的倍数时达到最大值，也就是说，$\rho(X_t, X_{t+k})$ 当 k 为轨道的周期的倍数时最大，而轨道的周期恰好是谱函数的跳跃点或者谱密度的峰值点。注意谱函数和谱密度是序列中不同频率振动的强度的度量，这个例子告诉我们，如果从谱函数和谱密度中看出某个频率占优，则相应的周期上的相关性也最强。γ_k 与 $f(\lambda)$ 是傅立叶系数与傅立叶级数函数之间的关系，从时间序列来看，γ_k 与 ρ_k 反映了相关性大小，谱密度反映不同频率的振动能量大小，相关性强的滞后值 k 会对应于谱密度 $\frac{2\pi}{k}$ 附近的峰值。

6.3.2 离散谱

阶梯函数的谱函数称为**离散谱函数**，相应的平稳序列称为**离散谱序列**。离散谱函数对应 $[-\pi,\pi]$ 上的一个只在离散点上取值的测度。

离散谱函数没有对应的谱密度，但是可以逼近。令

$$\begin{aligned} f_n(\lambda) =& n\sigma^2 I_{[-\lambda_0,-\lambda_0+1/(2n)]\cup[\lambda_0-1/(2n),\lambda_0]}(\lambda) \\ F_n(\lambda) =& \int_{-\pi}^{\lambda} f_n(u)du \end{aligned}$$

则 f_n 是一个仅在两个长度为 $\frac{1}{2n}$ 的小区间上非零的分段函数。$F_n(\lambda)$ 为连续的折线函数，仅在上述两个小区间上为线性增函数，在其它位置为水平线。有极限

$$F_n(\lambda) \to F(\lambda),\ \lambda \neq \pm\lambda_0,\ n \to \infty.$$

$f_n(\lambda)$ 是某平稳列的谱密度。当 $n \to \infty$ 时

$$
\begin{aligned}
g_k(n) \triangleq & \int_{-\pi}^{\pi} e^{ik\lambda} f_n(\lambda) d\lambda \\
= & n\sigma^2 \left(\int_{-\lambda_0}^{-\lambda_0+\frac{1}{2n}} e^{ik\lambda} d\lambda + \int_{\lambda_0-\frac{1}{2n}}^{\lambda_0} e^{ik\lambda} d\lambda \right) \\
= & 2n\sigma^2 \int_{\lambda_0-\frac{1}{2n}}^{\lambda_0} \cos(k\lambda) d\lambda \\
& \quad (\sin \text{ 在关于原点对称的区间积分为零}) \\
= & 2n\sigma^2 \cos(k(\lambda_0 + \alpha_n)) \cdot \frac{1}{2n} \quad (\text{积分中值定理}) \\
= & \sigma^2 \cos(k\lambda_0 + k\alpha_n) \\
\to & \sigma^2 \cos(k\lambda_0) = \gamma_k, \ n \to \infty
\end{aligned}
$$

其中数列 $\{\alpha_n\}$ 满足 $\alpha_n \in (-\frac{1}{2n}, 0), n \in \mathbb{N}_+$。

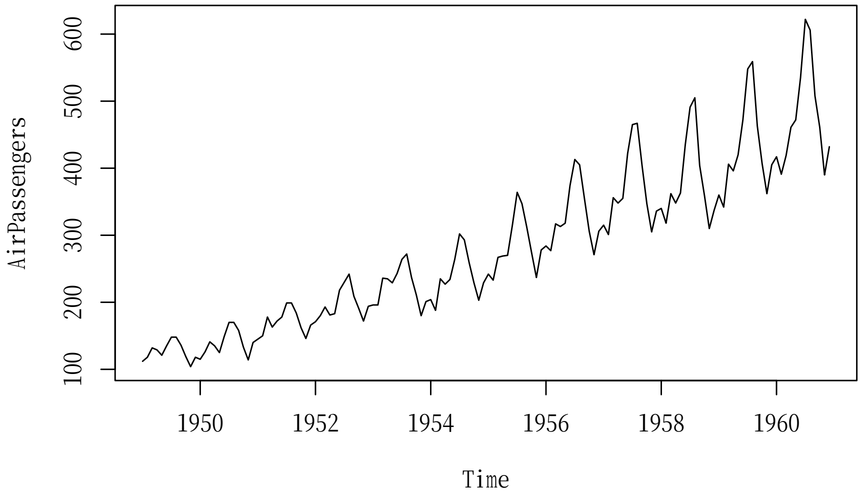
当 n 很大时，$f_n(\lambda)$ 对应的平稳列和 $\{Z_t\}$ 的表现已经很接近，其轨道表现为近似周期函数形式。

推广来看，如果某平稳列的谱密度在某处有很高的峰，则此序列的轨道在峰对应的频率 (角频率) 处应该表现出周期性。比如，月度数据的谱密度在角频率 $\frac{2\pi}{12}$ 处表现出高的峰值。设 ω 为角频率，f 为频率，T 为周期，则

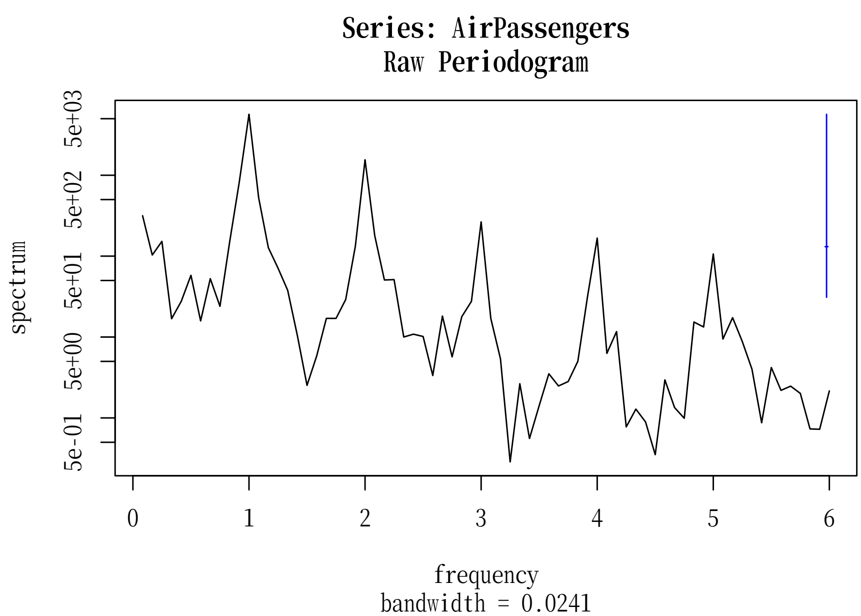
$$
\omega = 2\pi f, \quad T = \frac{1}{f}
$$

例 6.2. 国际航空订票数据的随机项的谱密度在周期 12 处有显著峰值。考察数据的周期性变化是谱密度估计的重要应用之一。

```
plot(AirPassengers)
```




```r
spec.pgram(AirPassengers)
```



这里的横坐标频率单位是以 $1/T$ 为单位，这里 $T = 12$。

6.3.3 离散谱序列

实际中的离散谱序列经常会有多个频率成分。设有 $2p$ 个随机变量 $\xi_k, \eta_k, k = 1, \dots, p$, 所有 $2p$ 个两两正交，满足

$$E\xi_j = E\eta_j = 0, \quad E\xi_j^2 = E\eta_j^2 = \sigma_j^2 \tag{6.19}$$

设 $\lambda_j \in (0, \pi], j = 1, \dots, p$，定义

$$\begin{aligned} Z_t &= \sum_{j=1}^{p} [\xi_j \cos(\lambda_j t) + \eta_j \sin(\lambda_j t)] \\ &= \sum_{j=1}^{p} A_j \cos(\lambda_j t - \theta_j), \quad t \in \mathbb{N}_+ \end{aligned} \tag{6.20}$$

其轨道表现为有 p 个频率成分的非随机函数。

$\{Z_t\}$ 为零均值平稳列,

$$\gamma_k = \sum_{j=1}^{p} \sigma_j^2 \cos(k\lambda_j), \quad k \in \mathbb{Z}. \tag{6.21}$$

设所有 $\lambda_j \neq \pi$, 这时谱函数为

$$F(\lambda) = \sum_{j=1}^{p} \frac{\sigma_j^2}{2} \left[I_{[-\lambda_j, \pi]}(\lambda) + I_{[\lambda_j, \pi]}(\lambda) \right], \ \lambda \in [-\pi, \pi]$$

此谱函数表现为在 $\pm\lambda_j$ 处有跳跃 $\frac{\sigma_j^2}{2}$ 的阶梯函数，表明谱的能量集中在这 $2p$ 个频率上。

§6.1讲谱时的单音例子就是离散谱的例子。

离散谱序列可以由可列个简单的离散谱序列叠加而成。设 $\xi_j, \eta_k, (j, k = 1, 2, \dots)$ 两两正交，满足

$$E\xi_j = E\eta_j = 0, \quad E\xi_j^2 = E\eta_j^2 = \sigma_j^2 \tag{6.22}$$

且

$$\sum_{j=1}^{\infty} \sigma_j^2 < \infty \tag{6.23}$$

设 $\lambda_j \in (0, \pi], j = 1, 2, ...$。

定义

$$Z_t = \sum_{j=1}^{\infty} [\xi_j \cos(\lambda_j t) + \eta_j \sin(\lambda_j t)] \quad t \in \mathbb{N}_+ \tag{6.24}$$

改写为

$$Z_t = \sum_{j=1}^{\infty} \left[\sigma_j \cos(t\lambda_j)(\xi_j/\sigma_j) + \sigma_j \sin(t\lambda_j)(\eta_j/\sigma_j)\right], \tag{6.25}$$

其中 $\{\xi_j/\sigma_j\}$ 和 $\{\eta_j/\sigma_j\}$ 可以合并为一个 WN(0,1)，级数中组合系数平方可和：

$$\sum_{j=1}^{\infty} \left[(\sigma_j \cos(t\lambda_j))^2 + (\sigma_j \sin(t\lambda_j))^2\right] = \sum_{j=1}^{\infty} \sigma_j^2 < \infty$$

所以级数(6.24)的右端均方收敛。

由 L^2 中内积的连续性, 对 $\forall t, s \in \mathbb{Z}$ 有

$$\begin{aligned}
E(Z_t) =& E(Z_t \cdot 1) \\
=& \sum_{j=1}^{\infty} E[\xi_j \cos(\lambda_j t) + \eta_j \sin(\lambda_j t)] \\
=& 0 \\
E(Z_t Z_s) =& \lim_{n\to\infty} \Big\{ \sum_{j=1}^{n} [\xi_j \cos(\lambda_j t) + \eta_j \sin(\lambda_j t)] \\
& \cdot \sum_{j=1}^{n} [\xi_j \cos(\lambda_j s) + \eta_j \sin(\lambda_j s)] \Big\} \\
=& \lim_{n\to\infty} \sum_{j=1}^{n} \sigma_j^2 \cos[(t-s)\lambda_j] \\
=& \sum_{j=1}^{\infty} \sigma_j^2 \cos[(t-s)\lambda_j]
\end{aligned}$$

所以，由可列个简单离散谱序列叠加得到的序列(6.24)是零均值平稳列。它由可列个正弦波和余弦波叠加构成。有自协方差函数

$$\gamma_k = \sum_{j=1}^{\infty} \sigma_j^2 \cos(k\lambda_j),\ k \in \mathbb{Z} \tag{6.26}$$

谱函数为

$$F(\lambda) = \sum_{j=1}^{\infty} \frac{\sigma_j^2}{2} \left[I_{[-\lambda_j,\pi]}(\lambda) + I_{[\lambda_j,\pi]}(\lambda) \right], \ \lambda \in [-\pi, \pi] \tag{6.27}$$

是一个有可列个跳跃点的阶梯函数，在 $\pm\lambda_j$ 有跳跃 $\sigma_j^2/2$，对应于 $[-\pi, \pi]$ 的一个离散测度。如果某个 $\lambda_j = \pi$，它对 $F(\lambda)$ 的贡献应该写成, $\sigma_j I_{\{\pi\}}(\lambda)$。

尽管离散谱序列是随机的，但它的每一次观测是确定的三角函数相加在整数点上的取值。实际工作中也经常把这样的模型看作非随机的，这时

$$Z_t = \sum_{j=1}^{p} A_j \cos(\lambda_j t + \theta_j), \ t \in \mathbb{Z}$$

其中 A_j, θ_j 非随机。称这样的模型为**调和模型** (harmonic model)。

6.4 附录：用 R 程序处理声音

6.4.1 tuneR 包

tuneR 包可以读写 WAV 格式的声音文件，并作一些简单操作。还可以读入 MP3 格式，读入 Midi 格式，分辨音符并输出为 lilipond 格式。

读入一个 WAV 文件，用如

```
library(tuneR)
mus <- readWave("data/music-demo2.wav")
```

还可以指定 `from`, `to`, `units="seconds"` 来读取 WAV 文件的一个片段。读入的左声道数据为 `mus@left`, 采样率为 `mus@samp.rate`。读入的声音数据没有自动标准化，取值可能是在 $\pm 2^{\text{bit}-1}$ 范围而不是 $[-1, 1]$ 范围。

`readMP3()` 函数可以读取 MP3 文件。

用 writeWave 将声音保存为 WAV 文件，如

```
y2 <- tuneR::sine(freq=523, samp.rate=8000, duration=1, xunit="time")
writeWave(y2, filename="data/music-demo2b.wav")
```

可以用 `stereo()` 函数将两个单声道组合成立体声，可以用 `mono()` 函数从立体声取出其中一个声道。

6.4.2 sound 包

R 软件的 sound 包可以读写 WAV 格式的声音文件并进行简单的操作。可以在 WAV 格式与矩阵格式之间进行转换。

读入一个 WAV 文件，用如

```
library(sound)
mus <- loadSample("data/music-demo2.wav")
```

这样读入一个声音对象。用 `rate(mus)` 求采样频率。

用 saveSample 把声音片段保存为 WAV 格式，如

```
saveSample(mus, filename="tmp.wav")
```

6.4.3 seewave 包

seewave 包可以分析、合成声音，进行时间、振幅、频率分析，估计声音之间的数量差别，生成新的声音用于回放。

读入的声音，可以是向量、矩阵、时间序列 (ts, mts)，或者 tuneR 包的 Wave 类型，audio 包的 audioSample 类型。

对于普通向量，矩阵，只要指定采样频率，就可以看成是声音。

对于 ts 或者 mts 类型的时间序列，其 frequency 对应于声音的采样频率。多元时间序列在 seewave 中只使用其中第一列。

`oscillo()` 绘制声音波形图。`spec()` 作谱密度估计图。`spectro()` 作动态谱峰图。

6.4.4 演示程序

如下的程序读入 music-demo2.wav 音乐文件，此文件包含了笛子演奏片段，有三个音符：

```
library(seewave)
library(tuneR)
mus <- readWave("data/music-demo2.wav")
```

```
## Warning in readChar(con, 4): 在non-UTF-8 MBCS语言环境里只能读取字节

## Warning in readChar(con, 4): 在non-UTF-8 MBCS语言环境里只能读取字节

## Warning in readChar(con, 4): 在non-UTF-8 MBCS语言环境里只能读取字节

## Warning in readChar(con, 4): 在non-UTF-8 MBCS语言环境里只能读取字节
```

```
f <- mus@samp.rate
cat(" 采样频率 =", f, "\n")
```

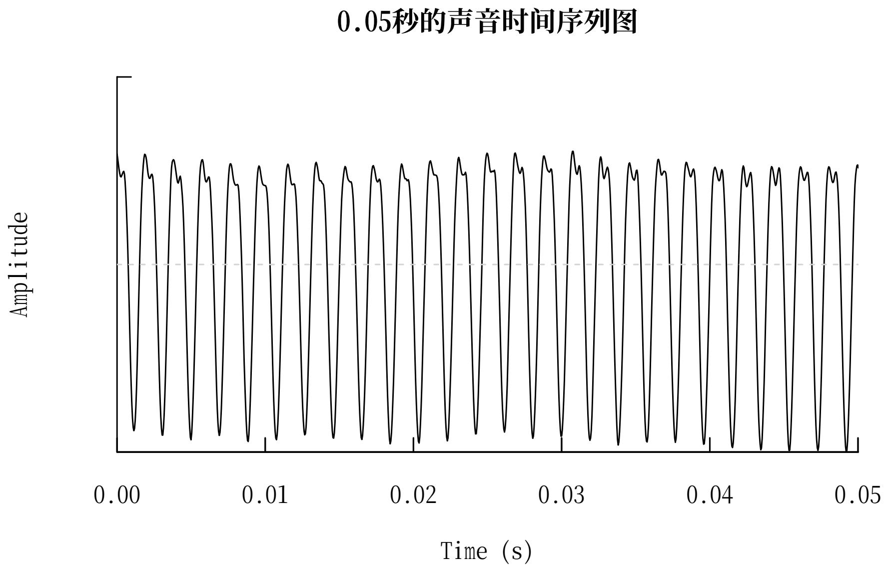
```
## 采样频率= 22050
```

用 seewave::cutw() 函数截取从 0.125 秒到 1.125 秒的音高为 C5 的片段，以及其中 0.05 秒的片段：

```
## mi1: 1 秒长的一段 C5 音阶。
mi1 <- cutw(mus, from=0.125, to=1.125,
            output="Wave")
## mi1a: 0.05 秒长的一段 C5 音阶。
mi1a <- cutw(mus, from=0.500, to=0.550,
             output="Wave")
```

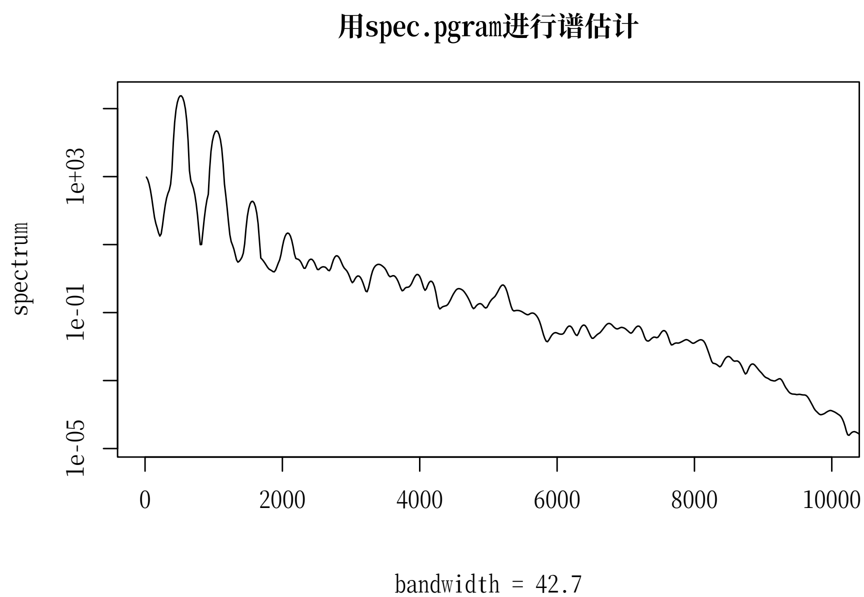
显示 0.05 秒的序列图，标题中的 f=22050 是采样频率，音阶是 C5(523Hz)：

```
oscillo(mi1a,
        title='0.05 秒的声音时间序列图')
```



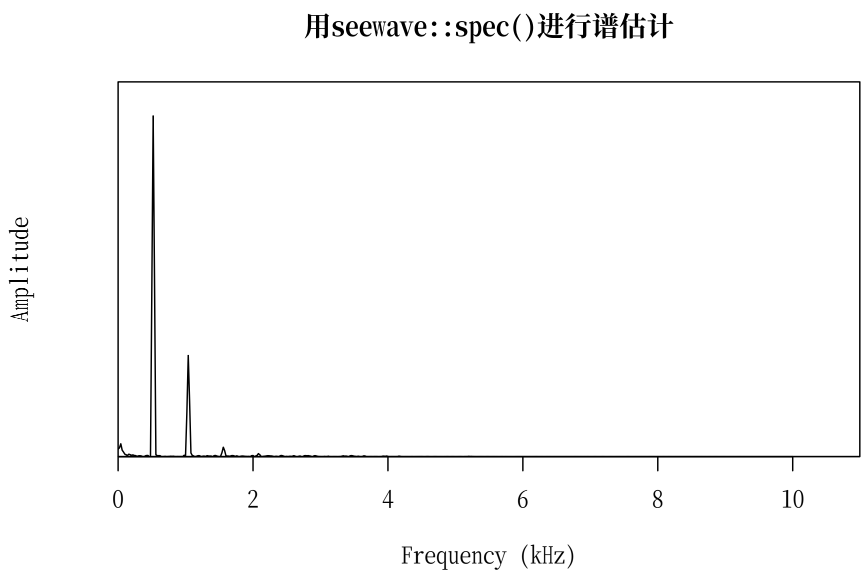
对 0.05 秒的 C5 音阶录音用时间序列分析的 spec.pgram() 函数做加窗谱估计，注意纵轴用了对数刻度:

```
spec.pgram(ts(mi1a@left, frequency=mi1a@samp.rate), spans=c(5,7),
           xlim=c(0,10000), main=" 用 spec.pgram 进行谱估计",
           xlab="")
```



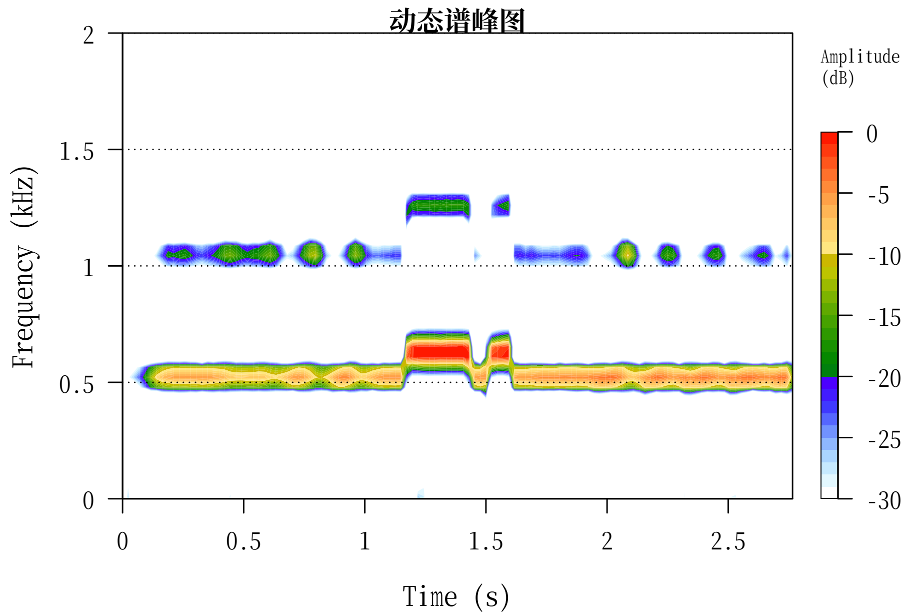
用 seewave 包的 `spec()` 函数显示 0.05 秒的 C5 音阶的频谱，纵轴没有用对数刻度：

```
seewave::spec(mi1a, main=" 用 seewave::spec() 进行谱估计")
```



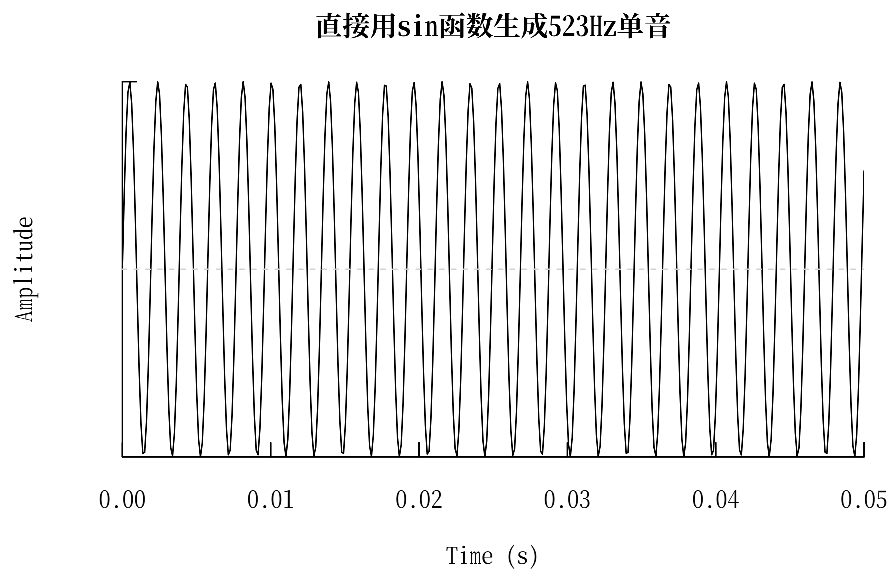
下面用 seewave 包的 `spectro()` 函数对包含三个音阶的笛子演奏片段做动态谱峰图, 动态显示随时间变化的主要谱峰位置。横坐标是时间，纵坐标是声音频率，颜色为在该时间、该频率上的振幅（注意同一时刻只有一个主要频率有振幅）能看出 `C5--D#5--C5` 的频率主峰变化。

```
seewave::spectro(mus, flim=c(0,2), main=" 动态谱峰图")
```



如下的程序用 R 的基本函数生成一个 523Hz 声音 (音调 C5) 的 1 秒的波形，作为一个 R 向量，并显示前 0.05 秒的波形图：

```
f8000 <- 8000 ## 采样频率 8kHz。
## timey: 为 [0,1] 中等间隔 8000 个点的时间。采样频率 8000Hz。
timey <- seq(0,1, length.out=f8000*1)
## y: 直接用 sin 函数生成的 523Hz 单音，向量长度为 8000 个点，组成 1 秒钟。
y <- sin(2*pi*523*timey)
## 用 oscillo 函数作波形图，仅取前 0.05 秒：
seewave::oscillo(
  cutw(y, f=f8000, from=0, to=0.05),
  f=f8000,
  title='直接用 sin 函数生成 523Hz 单音')
```

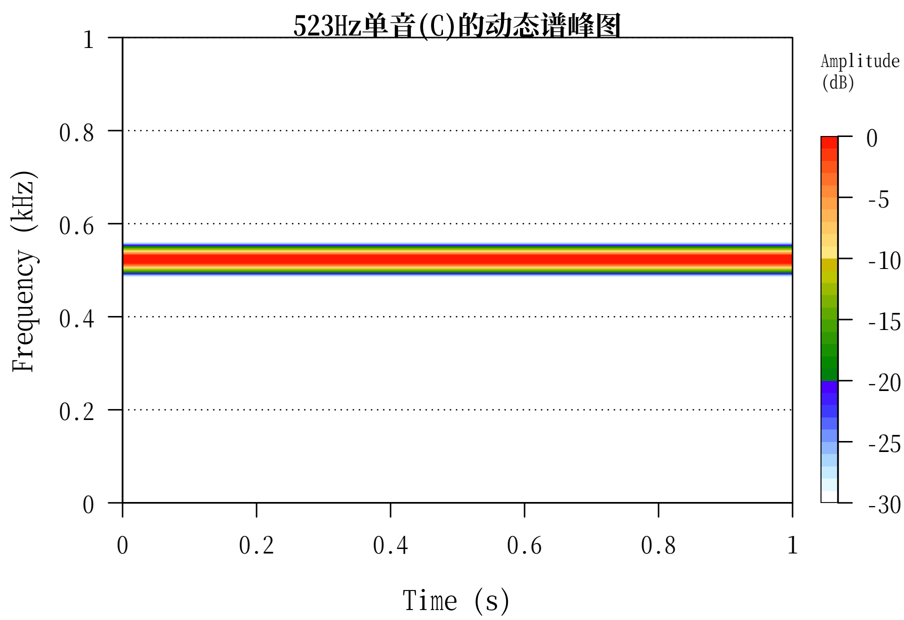


事实上，tuneR 包的 `sine()` 函数可以更容易地生成这种单音的文件，如

```
y2 <- tuneR::sine(freq=523, samp.rate=8000, duration=1, xunit="time")
```

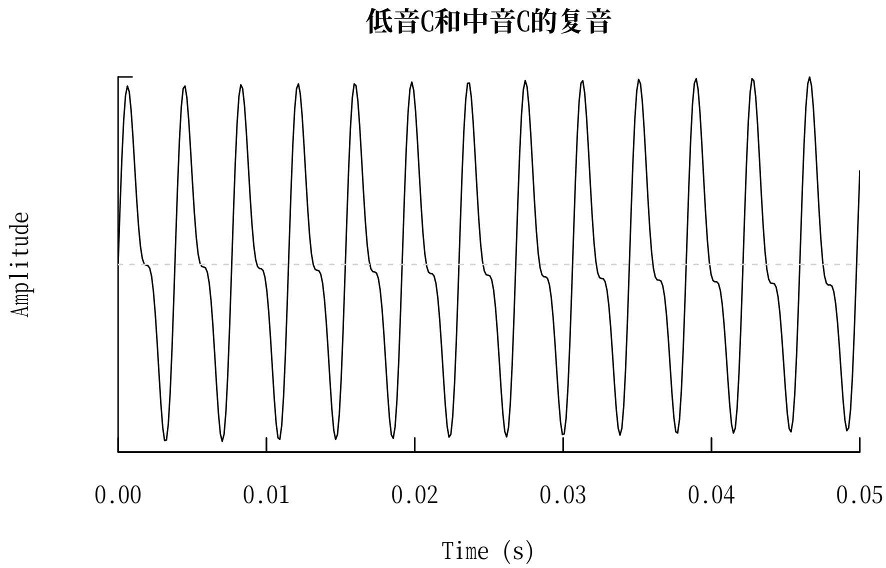
人造单音的动态频谱图:

```
seewave::spectro(y, f=f8000, flim=c(0,1.0), main="523Hz 单音 (C) 的动态谱峰图")
```



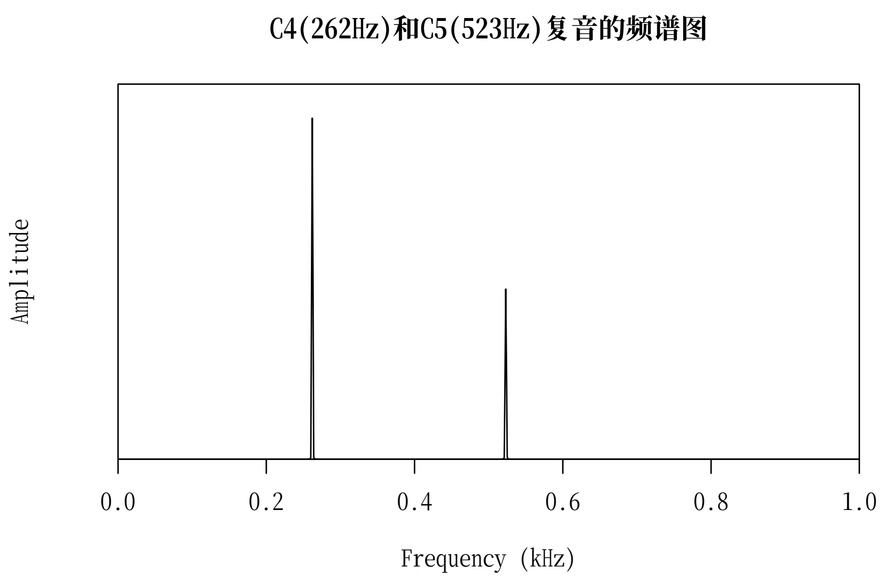
下面用基本 R 函数生成 C4+C5 复音，以 C4 为主（振幅比值为 2）。作前 0.05 秒的波形图：

```
f8000 <- 8000
y3 <- 0.5*sin(2*pi*262*seq(0,1, length.out=f8000*1)) + # C4 音阶
  0.25 * sin(2*pi*523*seq(0,1, length.out=f8000*1))
##writeWave(Wave(y3*2^15, samp.rate=f8000, bit=16),
##          filename="data/music-demo2c.wav")
oscillo(cutw(y3, f=f8000, from=0, to=0.05), f=f8000,
          title='低音 C 和中音 C 的复音')
```



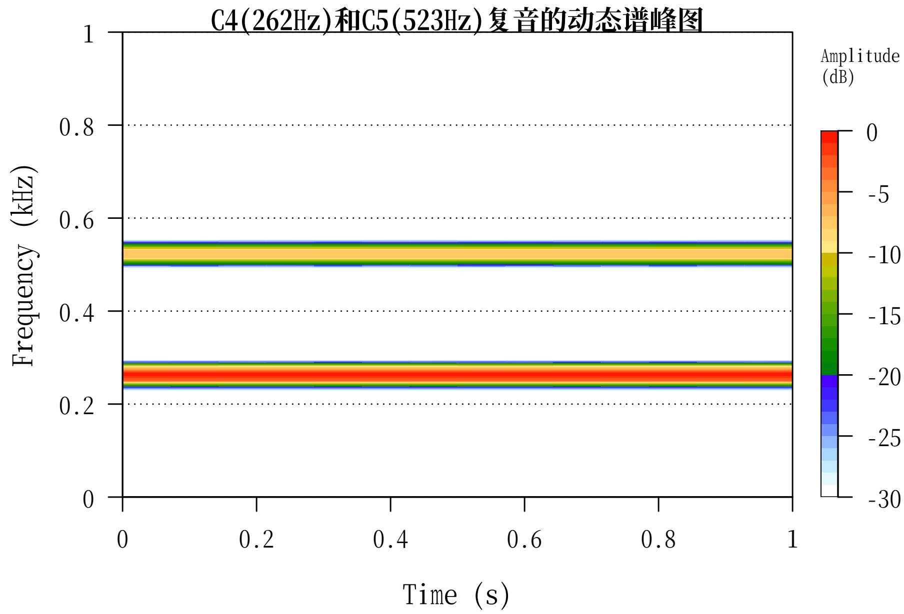
人造复音的频谱图:

```
seewave::spec(y3, f=f8000, main="C4(262Hz) 和 C5(523Hz) 复音的频谱图", flim=c(0, 1))
```



人造复音的动态频谱图:

```
seewave::spectro(y3, f=f8000, flim=c(0,1.0), main="C4(262Hz) 和 C5(523Hz) 复音的动
```



6.5 谱函数和谱密度补充

当平稳时间序列的自协方差函数绝对可和时，谱密度存在且与自协方差函数是傅立叶级数和傅立叶系数的关系。对于连续时平稳过程的自协方差函数绝对可积时，也有谱密度，且自协方差函数与谱密度为傅立叶变换关系。

设复值平稳列 $\{\xi_t, t \in \mathbb{Z}\}$ 的谱函数为 $F(\lambda)$，$\{\xi_t\}$ 张成的 Hilbert 空间 H_ξ 与

$$L^2(dF) = \{\varphi(\lambda) : \int_{-\pi}^{\pi} |\varphi(\lambda)|^2 dF(\lambda)\}$$

存在等距对应映射 $\mathcal{K}$ 把 $L^2(dF)$ 映射到 H_ξ 使得两空间同构，$\mathcal{K}$ 把 $e^{it\lambda}$ 映射为 ξ_t。

定理 6.7 (平稳序列谱表示). 对复值平稳时间序列 $\{\xi_t, t \in \mathbb{Z}\}$，存在 $[-\pi, \pi]$ 上的零均值连续时正交增量左 L^2 连续复值随机过程 $\{Z(\lambda), \lambda \in [-\pi, \pi]\}$，使得

$Z(-\pi)=0$,

$$E|Z(\lambda_2)-Z(\lambda_1)|^2 = F_\xi(\lambda_2)-F_\xi(\lambda_1),\ -\pi \le \lambda_1 < \lambda_2 \le \pi.$$

对任意 $\varphi \in L^2(dF_\xi)$，可以定义随机积分

$$\int_{-\pi}^{\pi} \varphi(\lambda) dZ(\lambda)$$

这时

$$\xi_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)$$

称为平稳时间序列的**谱表示**，含义是如下极限

$$\xi_t = \lim_{\max|\Delta\lambda_k|\to 0} \sum_k e^{it\lambda_k} Z^{(b)}(\Delta\lambda_k)$$

见 (谢衷洁, 1990)PP30–40。

定理 6.8. 有谱密度的平稳列必为系数平方可和的线性序列。

见 (谢衷洁, 1990)P49。

6.6 给定谱密度后的时间序列模拟生成

设某平稳列 $\{X_t\}$ 谱密度为 $f(\lambda)$，希望生成 $\{X_t\}$ 的模拟数据 $X_1, X_2, \dots, X_n$。

因为 $f(\lambda)$ 决定了自协方差函数 $\{\gamma_k\}$，可以先计算 $\gamma_0, \gamma_1, \dots, \gamma_{n-1}$，具体可以使用数值积分方法计算

$$\gamma_k = 2\int_0^{\pi} f(\lambda)\cos(k\lambda)\, d\lambda,\ k=0,1,\dots,n-1$$

得到随机向量 $X=(X_1,X_2,\dots,X_n)$ 的协方差阵 Γ_n。

当 n 不太大时，可以计算 Γ_n 的 Cholesky 分解 $\Gamma_n = BB^T$，B 为下三角阵且对角线元素为正值。生成独立同标准正态分布的 $Z_1, Z_2, \dots, Z_n$，令 $Z = (Z_1, Z_2, \dots, Z_n)^T$，令 $X = BZ$，则 $X_1, X_2, \dots, X_n$ 为自协方差函数为 $\{\gamma_k\}$，谱密度为 $f(\lambda)$ 的零均值正态平稳列的样本。

当 n 很大时，计算很多个 $\{\gamma_k\}$ 计算量很大也不必要，因为 $\gamma_k \to 0$；存储和计算 Cholesky 分解也可能需要许多存储空间与计算时间，而且由于计算误差可

能导致 Γ_n 不正定。这时可以用条件分布法，设定一个界限 p，认为条件分布 $X_k|X_{k-1},\dots,X_1$ 近似等于条件分布 $X_k|X_{k-1},\dots,X_{k-p}$，然后仅利用 Γ_p 产生条件期望和条件方差，逐步向前推进模拟。可以从 $\gamma_0,\dots,\gamma_p$ 求解出一个 $\text{AR}(p)$ 模型，模拟这个 $\text{AR}(p)$ 代替要模拟的序列。

作为例子，我们来模拟6.3.2中用阶梯函数谱密度逼近离散谱序列的问题。只有一个频率的离散谱序列为

$$X_t = \xi\cos(\lambda_0 t) + \eta\sin(\lambda_0 t)$$

其中 ξ,η 不相关，零均值，方差都等于 σ^2。自协方差函数为

$$\gamma_k = \sigma^2\cos(k\lambda_0)$$

$\{X_t\}$ 的谱函数为

$$F(\lambda) = \sigma^2[0.5I_{[-\lambda_0,\pi]}(\lambda) + 0.5I_{[\lambda_0,\pi]}(\lambda)]$$

是仅在 $-\lambda_0$ 和 λ_0 处有跳跃的阶梯函数，见图6.7。

取谱密度

$$\begin{aligned} f_n(\lambda) =& n\sigma^2 I_{[-\lambda_0,-\lambda_0+1/(2n)]\cup[\lambda_0-1/(2n),\lambda_0]}(\lambda) \\ F_n(\lambda) =& \int_{-\pi}^{\lambda} f_n(u)du \end{aligned}$$

则 f_n 是一个仅在两个长度为 $\frac{1}{2n}$ 的小区间上非零的分段函数。$F_n(\lambda)$ 为连续的折线函数，仅在上述两个小区间上为线性增函数，在其它位置为水平线。有极限

$$F_n(\lambda) \to F(\lambda),\ \lambda \neq \pm\lambda_0,\ n \to \infty.$$

来模拟生成谱密度为 $f_n(\lambda)$ 的平稳列。对应的协方差函数为

$$\begin{aligned} g_k(n) =& \int_{-\pi}^{\pi} f(\lambda)\cos k\lambda \\ =& \frac{2n\sigma^2}{k}\left[\sin(k\lambda_0) - \sin(k(\lambda_0 - \frac{1}{2n}))\right], k = 1,2,\dots, \\ g_0(n) =& \sigma^2 \end{aligned}$$

取 $\lambda_0 = \frac{2\pi}{4}$，对应于周期 4，取 $n = 10$。下面的函数中, `T` 是要模拟的序列长度，`T0` 是离散谱的周期，对应于 $\frac{2\pi}{\lambda_0}$，n 是用来近似的谱密度 $f_n(\lambda)$ 的逼近程度，n 越大，模拟的谱函数越接近于离散谱函数。

```
## 一般的给定自协方差函数的模拟程序
sim.fromgam <- function(gamfun){
  n <- length(gamfun)
  Gmat <- outer(1:n, 1:n, function(i, j) gamfun[abs(i-j) + 1])
  B <- chol(Gmat) # 结果是 n*n 上三角矩阵
  c(rnorm(n) %*% B)
}
sim.disc2dens <- function(T = 64, T0 = 4, n = 10, sigma = 1){
  ## 先计算\gamma_0, \dots, \gamma_{T-1}
  lam0 <- 2*pi / T0
  lam1 <- lam0 - 1/(2*n)
  kk <- 1:(T-1)
  gamfun <- (2*n*sigma^2) * c(1/(2*n), 1/kk * (
    sin(kk*lam0) - sin(kk*lam1) ) )
  sim.fromgam(gamfun)
}
```

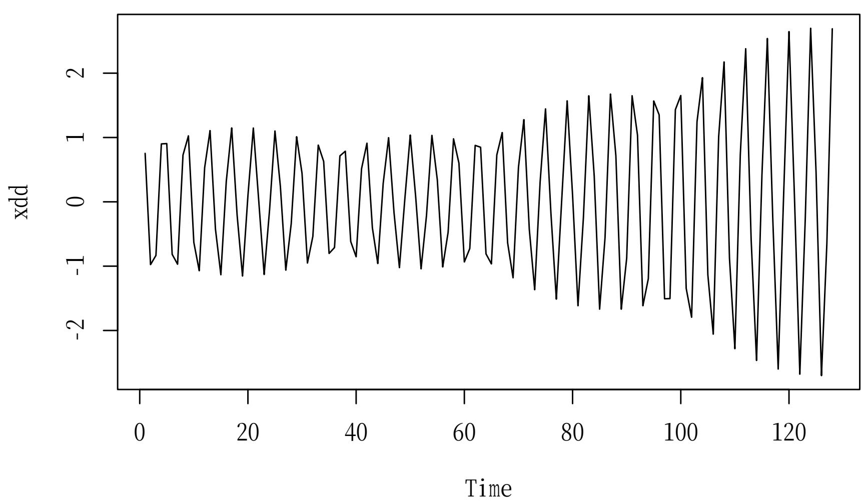
模拟一次：

```
xdd <- sim.disc2dens(T = 64, T0 = 4, n = 10)
## Error in chol.default(Gmat) : the leading minor of order 11 is not positive definite
```

经过试验发现，由于数值误差以及近似为离散谱的原因，即使对比较小的 n，Γ_T 矩阵也很快变得不满秩。由于计算误差的影响，上面结果中的 Γ_T 矩阵计算出的特征值从第 28 个开始就为负值了。所以，我们取一个较低阶的 AR(p) 作为近似。取 $p = 6$。

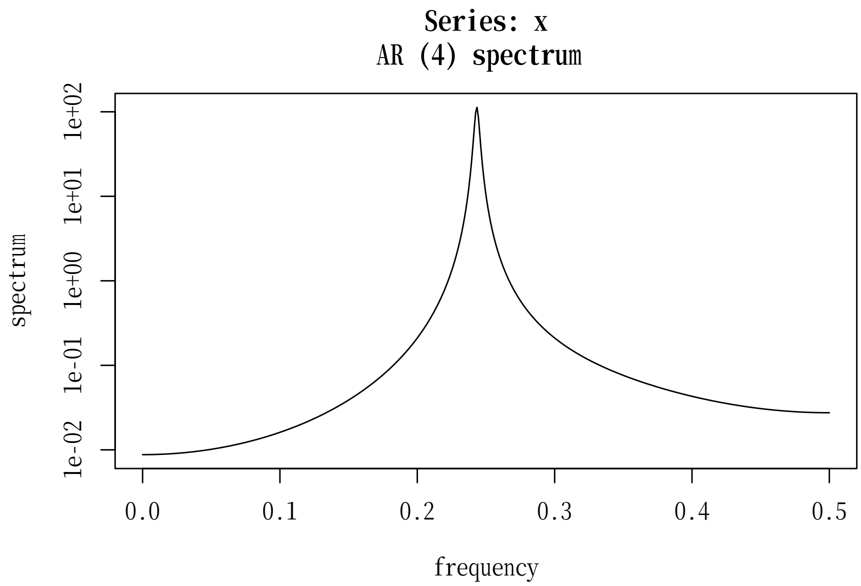
```
sim.disc2dens <- function(T = 128, T0 = 4, n = 10, p = 6, sigma = 1){
  ## 先计算\gamma_0, \dots, \gamma_{p}
  lam0 <- 2*pi / T0
  lam1 <- lam0 - 1/(2*n)
  kk <- 1:p
  gamfun <- (2*n*sigma^2) * c(1/(2*n), 1/kk * (
    sin(kk*lam0) - sin(kk*lam1) ) )
  ## 求解 Y-W 方程得到 a_1, \dots, a_p 和 AR 模型的白噪声方差
```

```
  Gmat <- outer(1:p, 1:p, function(i, j) gamfun[abs(i-j) + 1])
  avec <- solve(Gmat, gamfun[-1])
  s2ar <- gamfun[1] - sum(gamfun[-1] * avec)
  arima.sim(model = list(ar = avec), n = T) * sqrt(s2ar)
}
xdd <- sim.disc2dens(T = 128, T0 = 4, n = 10, p = 6)
ts.plot(xdd)
```



结果中可以看出明显的周期性，但是又和离散谱序列轨道哪种严格的周期函数不同。估计其谱密度：

```
spectrum(xdd, method = "ar")
```



估计的谱密度在 $\lambda_0 = \frac{2\pi}{4} = 0.25$ 处有唯一一个高峰。

Part II

ARMA 模型

Chapter 7

推移算子和常系数差分方程

7.1 推移算子

对任何时间序列 $\{X_t\}$ 和无穷级数

$$\psi(z) = \sum_{j=-\infty}^{\infty} b_j z^j$$

只要级数

$$\sum_{j=-\infty}^{\infty} b_j X_{t-j}$$

在某种意义下收敛 (例如 a.s. 收敛, 依概率收敛, 均方收敛), 就定义

$$\psi(\mathscr{B}) = \sum_{j=-\infty}^{\infty} b_j \mathscr{B}^j, \tag{7.1}$$

$$\psi(\mathscr{B}) X_t = \sum_{j=-\infty}^{\infty} b_j \mathscr{B}^j X_t = \sum_{j=-\infty}^{\infty} b_j X_{t-j}. \tag{7.2}$$

并且称 $\mathscr{B}$ 是时间 t 的向后推移算子或滞后算子, 简称为**推移算子**.

显然 $\mathscr{B} X_t = X_{t-1}$. 如果 $\{X_t\}$ 是平稳列，$\mathscr{B}$ 确实是 Hilbert 空间 $\bar{L}^2(X)$ 上的一个算子，也可以扩充到 L^2 空间上。这里我们只给出它的简单性质。

- (1) 对和 t 无关的随机变量或者常数 Y, 有 $\mathscr{B} Y = Y$.
- (2) 对常数 a，$\mathscr{B}^n (a X_t) = a \mathscr{B}^n X_t = a X_{t-n}$.

- (3) $\mathscr{B}^{n+m}X_t = \mathscr{B}^n(\mathscr{B}^m X_t) = X_{t-n-m}$.
- (4) 对多项式 $\psi(z) = \sum_{j=0}^{p} c_j z^j$, 有 $\psi(\mathscr{B})X_t = \sum_{j=0}^{p} c_j X_{t-j}$.
- (5) （交换律）对于多项式 $\psi(z) = \sum_{j=0}^{p} c_j z^j$ 和 $\phi(z) = \sum_{j=0}^{q} d_j z^j$ 的乘积 $A(z) = \psi(z)\phi(z)$, 有

$$A(\mathscr{B})X_t = \psi(\mathscr{B})[\phi(\mathscr{B})X_t] = \phi(\mathscr{B})[\psi(\mathscr{B})X_t].$$

- (6) 对于时间序列 $\{X_t\}$, $\{Y_t\}$, 多项式 $\psi(z) = \sum_{j=0}^{p} c_j z^j$, 和随机变量 U, V, W, 有

$$\psi(\mathscr{B})(UX_t + VY_t + W) = U\psi(\mathscr{B})X_t + V\psi(\mathscr{B})Y_t + W\psi(1).$$

性质证明:

(1) 对 $t \in \mathbb{Z}$, 定义 $X_t = Y$, $\forall t$，对 $j \neq 1$ 定义 $b_j = 0$。由(7.2)得

$$\mathscr{B}Y = \mathscr{B}X_t = \mathscr{B}X_{t-1} = Y$$

(2) 令 $Y_t = aX_t$，由(7.2)得到

$$\mathscr{B}^n(aX_t) = \mathscr{B}^n Y_t = Y_{t-n} = aX_{t-n}$$

(3)

$$\mathscr{B}^n[\mathscr{B}^m X_t] = \mathscr{B}^n X_{t-m} = X_{t-m-n} = \mathscr{B}^{n+m} X_t$$

(4) 对 $j < 0$ 和 $j > p$ 取 $b_j = 0$，由(7.2)即可得

$$\psi(\mathscr{B})X_t = \sum_{j=0}^{n} b_j X_{t-j}$$

(5) 对于多项式 $\psi(z) = \sum_{j=0}^{p} c_j z^j$ 和 $\phi(z) = \sum_{j=0}^{q} d_j z^j$ 的乘积 $A(z) = \psi(z)\phi(z)$, 记

$$Z_t = \phi(\mathscr{B})X_t = \sum_{j=0}^{q} d_j X_{t-j}$$

则

$$\begin{aligned}\psi(\mathscr{B})[\phi(\mathscr{B})X_t] =&\psi(\mathscr{B})Z_t = \sum_{k=0}^{p} c_k Z_{t-k}\\ =&\sum_{k=0}^{p} c_k \sum_{j=0}^{k} d_j X_{t-k-j}\\ =&\sum_{k=0}^{p}\sum_{j=0}^{k} c_k d_j X_{t-k-j}\\ =&A(\mathscr{B})X_t\end{aligned}$$

同理可证 $A(\mathscr{B})X_t = \psi(\mathscr{B})[\phi(\mathscr{B})X_t]$。

(6) 略。

如果 $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$，$\{X_t\}$ 为平稳列，则

$$\Psi(\mathscr{B})X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

是平稳列 $\{X_t\}$ 的线性滤波，$\{\psi_j\}$ 为一个保时线性滤波器，$\Psi(\mathscr{B})X_t$ 在 a.s. 和均方意义下收敛，见2.3和5.1.5。这时性质 (5)(交换律) 和性质 (6)(线性性质) 仍成立。

考虑无穷阶推移算子多项式的交换律。有如下定理：

定理 7.1 (线性滤波的交换率)**.** 设 $\{a_k\}$, $\{b_j\}$ 为两个绝对可和的实数列，则实数列

$$d_m = \sum_{j=-\infty}^{\infty} a_j b_{m-j}$$

绝对可和（$\{d_m\}$ 称为 $\{a_k\}$ 和 $\{b_j\}$ 的**离散卷积**），记

$$A(z) = \sum_k a_k z^k, \quad B(z) = \sum_j b_j z^j, \quad D(z) = \sum_m d_m z^m.$$

(1) 若 $\{y_t\}$ 为有界的数列：$|y_t| \le M, t \in \mathbb{Z}$，则

$$A(\mathscr{B})[B(\mathscr{B})y_t] = B(\mathscr{B})[A(\mathscr{B})y_t] = D(\mathscr{B})y_t$$

(2) 若 $\{X_t\}$ 为平稳列，则

$$A(\mathscr{B})[B(\mathscr{B})X_t] = B(\mathscr{B})[A(\mathscr{B})X_t] = D(\mathscr{B})X_t,\ a.s.\ 和 L^2.$$

证明 上述的 $\{a_k\}$, $\{b_j\}$ 绝对可和保证了 $\{d_m\}$ 也绝对可和：

$$\begin{aligned}\sum_{m=-\infty}^{\infty}|d_m| &\le \sum_{m=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}|a_j||b_{m-j}| \\ &= \sum_{j=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}|a_j||b_{m-j}| \\ &= \sum_{j=-\infty}^{\infty}|a_j|\cdot\sum_{m=-\infty}^{\infty}|b_{m-j}| \\ &= \sum_{j=-\infty}^{\infty}|a_j|\cdot\sum_{k=-\infty}^{\infty}|b_k| < \infty\end{aligned}$$

(1)

$$\begin{aligned}&\sum_k\sum_j|a_k||b_j||y_{t-k-j}| \\ \le &M(\sum_k|a_k|)(\sum_j|b_j|) < \infty\end{aligned} \tag{*1}$$

所以

$$\begin{aligned}A(\mathscr{B})B(\mathscr{B})y_t &= \sum_k a_k\sum_j b_j y_{t-k-j} \\ =&\sum_k\sum_j a_k b_j y_{t-k-j} \\ =&\sum_k\sum_m a_k b_{m-k} y_{t-m} \quad (\text{令 } m=k+j,\ j=m-k) \\ =&\sum_m\sum_k a_k b_{m-k} y_{t-m} \quad (\text{无穷级数次序交换用到 (*1) 式}) \\ =&\sum_m d_m y_{t-m}\end{aligned}$$

同样可证 $B(\mathscr{B})A(\mathscr{B})y_t = D(\mathscr{B})y_t$。

(2) 由推论2.1，因为

$$\sum_k\sum_j|a_k||b_j|E|X_{t-k-j}| \le \sqrt{\gamma_0}(\sum_k|a_k|)(\sum_j|b_j|) < \infty,$$

由单调收敛定理有

$$E\sum_k\sum_j|a_k||b_j||X_{t-k-j}| \le \sqrt{\gamma_0}(\sum_k|a_k|)(\sum_j|b_j|) < \infty,$$

所以 $A(\mathscr{B})[B(\mathscr{B})X_t]$, $B(\mathscr{B})[A(\mathscr{B})X_t]$, $D(\mathscr{B})X_t$ 都 a.s. 绝对收敛，所以两重级数可以交换次序，并且可以用换元法证明都等于 $D(\mathscr{B})X_t$。

类似可证明 $A(\mathscr{B})[B(\mathscr{B})X_t]$, $B(\mathscr{B})[A(\mathscr{B})X_t]$, $D(\mathscr{B})X_t$ 都 L^2 收敛。当同时 a.s. 收敛和 L^2 收敛时，极限 a.s. 相等。

注：若所有 $a_k = 0(k < 0)$, $b_j = 0(j < 0)$，则

$$d_m = \sum_{j=0}^{\infty} a_j b_{m-j}$$

这时 $A(z)$, $B(z)$，$D(z)$ 在 $|z| \leq 1$ 绝对一致收敛。

如果不满足这样的条件，若 $a_k, k < 0$ 和 $b_j, j < 0$ 满足一些条件，$A(z)$、$B(z)$、$D(z)$ 可以在包含单位圆的圆环内解析。

比如, 如果存在 $0 < \nu < 1$ 使得

$$a_k = o(\nu^{-k}), \quad k < 0$$

则取 $\nu < \nu_1$，对 $\nu_1 \leq |z| \leq 1$ 有

$$\begin{aligned}
&\sum_{k=-\infty}^{0} |a_k z^k| = \sum_{k=-\infty}^{0} |a_k| \cdot |z|^k \\
\leq &\sum_{k=-\infty}^{0} |a_k| \cdot \nu_1^k \leq \sum_{k=-\infty}^{0} c_1 \nu_{-k} \cdot \nu_1^k \\
= &c_1 \sum_{k=-\infty}^{0} c_1 \left(\frac{\nu}{\nu_1}\right)^{-k} < \infty
\end{aligned}$$

所以，$A(z)$, $B(z)$, $D(z)$ 在双边的时候也可以是有意义的。定理7.1的证明不需要 $A(z)$、$B(z)$、$D(z)$ 的级数收敛性。

考虑有限阶推移算子多项式的逆算子。有如下定理：

定理 7.2 (线性滤波的逆). 设实系数多项式 $A(z) = \sum_{j=0}^{p} \phi_j z^j$ 满足最小相位条件:

$$A(z) \neq 0, \quad \forall |z| \leq 1$$

则存在 $\delta > 0$ 使

$$A^{-1}(z) \triangleq \frac{1}{A(z)} = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq 1+\delta$$

其中 $\psi_j = o((1+\delta)^{-j})$, $(j \to \infty)$，从而 $\sum_{j=0}^{\infty} |\psi_j| < \infty$，且有

(1) 若 $\{y_t\}$ 为有界数列，则

$$y_t = A^{-1}(\mathscr{B})A(\mathscr{B})y_t = A(\mathscr{B})A^{-1}(\mathscr{B})y_t$$

(2) 若 $\{X_t\}$ 为平稳列，则

$$X_t = A^{-1}(\mathscr{B})A(\mathscr{B})X_t = A(\mathscr{B})A^{-1}(\mathscr{B})X_t, \quad a.s.$$

证明: 设多项式 $A(z)$ 的根为 $z_1, \dots, z_p$，取

$$1 < 1+\delta < \min_j |z_j|$$

则 $A(z)$ 在 $|z| \leq 1+\delta$ 无零点，$A^{-1}(z)$ 在 $|z| \leq 1+\delta$ 解析，可以展开为 Taylor 级数

$$A^{-1}(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq 1+\delta$$

由于 $\sum_{j=0}^{\infty} \psi_j (1+\delta)^j$ 收敛所以 $\psi_j (1+\delta)^j \to 0$ 当 $j \to \infty$，即 $\psi_j = o((1+\delta)^{-j})$ $(j \to \infty)$。由此知 $\{\psi_j\}$ 绝对可和。定义 $\phi_j = 0$, 对 $j < 0$ 或 $j > p$，定义 $\psi_j = 0$ 对 $j < 0$，注意到

$$A(z)A^{-1}(z) = 1$$

令 $d_0 = 1, d_m = 0$ 对 $m \neq 0$，由定理7.1可得所需结论。

上面的定理也可以推广到双边无穷阶滤波的情形：

定理 7.3 (双边线性滤波的逆)**.** 设实系数多项式 $A(z) = \sum_{j=-\infty}^{\infty} a_j z^j$ 满足*:*

$$A(z) \neq 0, \quad \forall \alpha < |z| < \beta \tag{7.3}$$

（其中 $0 < \alpha < 1 < \beta$）则存在 $\{\psi_j\}$ 使得 $\psi_j = o(\rho^{-|j|})$, $\rho > 1$, 且

$$A^{-1}(z) \triangleq \frac{1}{A(z)} = \sum_{j=-\infty}^{\infty} \psi_j z^j, \quad \alpha < |z| < \beta$$

收敛。有

(1) 若 $\{y_t\}$ 为有界数列，则

$$y_t = A^{-1}(\mathscr{B})A(\mathscr{B})y_t = A(\mathscr{B})A^{-1}(\mathscr{B})y_t$$

(2) 若 $\{X_t\}$ 为平稳列，则

$$X_t = A^{-1}(\mathscr{B})A(\mathscr{B})X_t = A(\mathscr{B})A^{-1}(\mathscr{B})X_t, \quad a.s.$$

注意条件(7.3)保证了系数 $\{a_j\}$ 绝对可和。证明类似。

7.2 常系数齐次线性差分方程

给定 p 个实数 $a_1, a_2, \cdots, a_p$, $a_p \neq 0$, 我们称

$$X_t - [a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p}] = 0, \quad t \in \mathbb{Z}, \tag{7.4}$$

为 p 阶**齐次常系数线性差分方程**, 简称为齐次差分方程. 满足(7.4)的实数列 (或复数列)$\{X_t\}$ 称为(7.4)的解. 满足(7.4)的实值 (或复值) 时间序列 $\{X_t\}$ 也称为(7.4)的解.

(7.4)的解 $\{X_t\}$ 可以由它的 p 个初值 $X_0, X_1, \cdots, X_{p-1}$ 逐步递推得到:

$$\begin{aligned} X_t &= [a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p}], \quad t \geq p, \\ X_{t-p} &= \frac{1}{a_p}[x_t - a_1 X_{t-1} - \cdots - a_{p-1} X_{t-p+1}], \ t - p < 0 \end{aligned}$$

若初值是随机变量则递推得到的 $\{X_t\}$ 是时间序列。

用推移算子把差分方程写成

$$A(\mathscr{B})X_t = 0, \ t \in \mathbb{Z}, \ 其中 \ A(z) = 1 - \sum_{j=1}^{p} a_j z^j. \tag{7.5}$$

diff-lagop0103) \end{align} $A(z)$ 称为差分方程 (1.2) 的特征多项式。

解有线性性质: $\{X_t\}$, $\{Y_t\}$ 是解则

$$\xi X_t + \eta Y_t$$

也是解。这样，差分方程只要有非零解，就有无穷多个解，下面找出方程的所有的线性独立的解。

7.2.1 差分方程基础解

根据代数基本定理，设实系数 p 次多项式 $A(z)$ 有 k 个互不相同的零点 $z_1, z_2, \cdots, z_k$, 其中 z_j 是 $r(j)$ 重零点，$\sum_{j=1}^{k} r(j) = p$。可以证明对每一 z_j 有

$$A(\mathscr{B}) t^l z_j^{-t} = 0,\ l = 0, 1, \dots, r(j) - 1 \tag{7.6}$$

证明:

$$A(z) = \prod_{j=1}^{k} (1 - z_j^{-1} z)^{r(j)}$$

$$A(\mathscr{B}) = \prod_{j=1}^{k} (1 - z_j^{-1} \mathscr{B})^{r(j)}$$

只要证明

$$(1 - z_j^{-1}\mathscr{B})^{l+1}(t^l z_j^{-t}) = 0,\ l = 0, 1, \dots, r(j) - 1 \tag{7.7}$$

用归纳法。$l = 0$ 时

$$(1 - z_j^{-1}\mathscr{B}) z_j^{-t} = z_j^{-t} - z_j^{-1} z_j^{-(t-1)} = 0$$

设对 $0 \le l \le m - 1$ 已经证明

$$(1 - z_j^{-1}\mathscr{B})^{l+1} t^l z_j^{-t} = 0$$

则对 $l = m$ 有

$$\begin{aligned}
&(1 - z_j^{-1}\mathscr{B})^{m+1} t^m z_j^{-t} \\
=&(1 - z_j^{-1}\mathscr{B})^{m}(1 - z_j^{-1}\mathscr{B}) \left(t^m z_j^{-t}\right) \\
=&(1 - z_j^{-1}\mathscr{B})^{m} \left(t^m z_j^{-t} - z_j^{-1}(t-1)^m z_j^{-(t-1)}\right) \\
=&(1 - z_j^{-1}\mathscr{B})^{m} \left(t^m - (t-1)^m\right) z_j^{-t} \\
=&(1 - z_j^{-1}\mathscr{B})^{m} \left(c_1 t^{m-1} + c_2 t^{m-2} + \cdots + c_m\right) z_j^{-t} \\
=&0
\end{aligned}$$

于是(7.7)成立，从而(7.6)成立。

这样的基础解一共有 p 个，组成了 p 个线性独立的解。把基础解线性组合可以得到齐次线性差分方程的通解。通解是方程的含有可取任意值的系数的解，而且方程的每个解都包含在通解中。

7.2.2 齐次线性差分方程的通解

定理 7.4. *设 $A(z)$ 有 k 个互不相同的零点 $z_1, z_2, \dots, z_k$, 其中 z_j 是 $r(j)$ 重零点. 则*

$$t^l z_j^{-t}, \quad l = 0, 1, \cdots, r(j) - 1, \quad j = 1, 2, \cdots, k \tag{7.8}$$

是(7.4)的 p 个解; 而且, (7.4)的任何解 $\{X_t\}$ 都可以写成这 p 个解的线性组合:

$$X_t = \sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t}, \quad t \in \mathbb{Z}, \tag{7.9}$$

*其中的随机变量 $U_{l,j}$ 可以由 $\{X_t\}$ 的初值 $X_0, X_1, \cdots, X_{p-1}$ 惟一决定. (7.9)称为齐次线性差分方程(7.4)的***通解***。*

此定理关于时间序列叙述，实际上对差分方程的复数或实数列解也是成立的。

证明见 (Brockwell & Davis, 1987)。

(7.9)中 z_j 可以是复数。记

$$U_{l,j} = V_{l,j} e^{i\theta_{l,j}}, \quad z_j = \rho_j e^{i\lambda_j},$$

则

$$\begin{aligned} \text{Re}\left(U_{l,j} t^l z_j^{-t}\right) =& \text{Re}\left(V_{l,j} t^l \rho_j^{-t} e^{-i(\lambda_j t - \theta_{l,j})}\right) \\ =& V_{l,j} t^l \rho_j^{-t} \cos(\lambda_j t - \theta_{l,j}), \end{aligned}$$

差分方程(7.4)的实值解可以表示为

$$\sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} V_{l,j} t^l \rho_j^{-t} \cos(\lambda_j t - \theta_j), \quad t \in \mathbb{Z}$$

$\{V_{l,j}, \theta_{l,j}\}$ 可以由初值 $X_0, X_1, \dots, X_{p-1}$ 唯一决定。

7.2.3 通解的收敛性

如果差分方程的特征多项式 $A(z)$ 的根都在单位圆外: $|z_j| > 1, \quad j = 1,2,\dots,k$, 或 $A(z) \neq 0, \quad \forall |z| \leq 1$, 取 $1 < \alpha < \min\{|z_j| : j = 1,\dots,k\}$，则

$$t^l |z_j|^{-t} = t^l (\alpha/|z_j|)^t \alpha^{-t} = o(\alpha^{-t})$$

于是方程的任意解 X_t 满足

$$|X_t| = o(\alpha^{-t}) \quad a.s.,\ t \to \infty \tag{7.10}$$

称 X_t 以负指数阶收敛到零。

如果特征多项式有单位根 $z_j = \exp(i\lambda_j)$，则方程有一个周期解

$$X_t = a\cos(\lambda_j t), \quad t \in \mathbb{Z}$$

如果单位圆内有根 $z_j = \rho_j \exp(i\lambda_j), \rho_j < 1$，则方程有一个爆炸解（发散解）

$$X_t = a\left(\frac{1}{\rho_j}\right)^t \cos(\lambda_j t), \quad t \in \mathbb{Z}.$$

7.3 非齐次线性差分方程

设 $\{Y_t\}$ 为实值时间序列。

$$A(\mathscr{B})X_t = Y_t, \quad t \in \mathbb{Z} \tag{7.11}$$

满足(7.11)的时间序列 $\{X_t\}$ 称为(7.11)的解。

如果有(7.11)的某个解 (称为特解)$\{X_t^{(0)}\}$，则通解可以写成

$$X_t = X_t^{(0)} + \sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t}, \quad t \in \mathbb{Z}. \tag{7.12}$$

7.4 附录：复变函数复习

讨论推移算子多项式用到了一些复变函数知识，这里进行复习。

若复变函数 $f(z)$ 在复平面中点 z_0 的某个邻域内的任一点都可导，称 $f(z)$ 在 z_0 **解析**。若 $f(z)$ 在区域 D 内每一点解析，称 $f(z)$ 在 D 内解析，或称 $f(z)$ 是 D 内一个**解析函数**。

区域D 指 D 是连通开集。D 连通，即 D 中任意两个点之间都存在一条连续曲线。复平面中的连续曲线，就是从 $[\alpha, \beta]$ 到复数域的一个连续映射，或平面上用参数方程表示的一条连续曲线。

若 $f(z), g(z)$ 在区域 D 解析，则其和、差、积也在 D 内解析。若 $g(z) \neq 0, z \in D$，则 $f(z)/g(z)$ 也在 D 内解析。所以，多项式的倒数在不含分母零点的区域内解析。

把复变函数写成两个二元实函数：

$$f(z) = f(x+iy) = u(x,y) + iv(x,y)$$

则 $f(z)$ 解析当且仅当 $u(x,y)$ 和 $v(x,y)$ 都可微且满足柯西-黎曼条件

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

这时

$$\begin{aligned} f'(z) =& \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \\ =& \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y} \end{aligned}$$

解析函数必无穷阶可微。

调和函数: 二元实函数 $\varphi(x,y)$ 在区域 D 内有二阶连续偏导且

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0.$$

解析函数的实部和虚部都是调和函数。

幂级数: 复变级数

$$\sum_{k=0}^{\infty} a_n z^n$$

称为幂级数。如果

$$R = \lim_{n\to\infty} \frac{|a_n|}{|a_{n+1}|}$$

存在，则

- 当 $|z| < R$ 时，幂级数绝对收敛；
- 当 $|z| > R$ 时，幂级数发散；
- 当 $|z| = R$ 时，幂级数可能收敛也可能发散。

R 称为幂级数的收敛半径，幂级数的收敛区域叫做收敛圆。

泰勒级数: 若函数 $f(z)$ 在 $|z-b|<R$ 内解析，则在此范围可以展开成如下泰勒级数

$$f(z)=\sum_{k=0}^{\infty}a_k(z-b)^k,\quad |z-b|<R$$

系数为

$$a_n=\frac{1}{2\pi i}\oint_L\frac{f(\zeta)d\zeta}{(\zeta-b)^{n+1}}=\frac{f^{(n)}(b)}{n!}$$

反之，这样的一个级数在 z_0 收敛则在 $|z-b|<|z_0-b|$ 绝对收敛，解析。

罗朗级数: 含有负幂的级数

$$\sum_{n=-\infty}^{\infty}a_n(z-b)^n$$

称为**罗朗级数**。其中 $n\geq 0$ 部分称作**解析部分**, $n<0$ 部分称作**主要部分**。解析部分的收敛范围是 $|z|<R_1$, 主要部分的收敛范围是 $|z|>R_2$，如果这两个区域相交，则罗朗级数在

$$R_2<|z|<R_1$$

内收敛。若函数 $f(z)$ 在环域 $R_2<|z-b|<R_1$ 内部为解析，则 $f(z)$ 在该环域内可展开为罗朗级数。

7.5 附录：用差分方程给出 Fibonacci 数列通解

Fibonacci 数列定义为：

$$F_1=F_2=1,\ F_n=F_{n-1}+F_{n-2},n=2,3,...$$

这是一个齐次线性差分方程。特征多项式

$$1-z-z^2=0,$$

两个特征根为

$$z_1=\frac{\sqrt{5}-1}{2},\ z_2=-\frac{\sqrt{5}+1}{2}.$$

可求出通解，注意到黄金分割比

$$\frac{\sqrt{5}-1}{2}=\left(\frac{1+\sqrt{5}}{2}\right)^{-1}\approx 0.618$$

有

$$
\begin{aligned}
F_n =& c_1(\frac{\sqrt{5}-1}{2})^{-n} + c_2(-\frac{\sqrt{5}+1}{2})^{-n} \\
=& c_1(\frac{\sqrt{5}+1}{2})^{n} + c_2(-\frac{\sqrt{5}-1}{2})^{n}
\end{aligned}
$$

将 $F_1 = F_2 = 1$ 代入，就得到通项公式

$$
F_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right]
$$

程序验证：

```
fib <- function(n){
  gr <- (sqrt(5) - 1)/2
  grinv <- 1/gr
  1/sqrt(5)*(grinv^n - (-gr)^n)
}
fib(1:10)
```

```
##  [1]  1  1  2  3  5  8 13 21 34 55
```

关于 Fibonacci 数列，有级数性质

$$
\sum_{j=1}^{2n} F_j = \sum_{k=1}^{n} (F_{2k-1} + F_{2k}) = F_{2n+2} - F_2.
$$

7.6 附录：推移算子补充

设 $\{X_t\}$ 为时间序列，$\{a_j, j \in \mathbb{Z}\}$ 是绝对可和实数列。若 $\sup_t E|X_t| < \infty$，则

$$
\Psi(\mathscr{B}) X_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}
$$

以概率 1 收敛。如果进一步地 $\sup_t EX_t^2 < \infty$ 则级数还均方收敛到同一极限。(见 (Brockwell & Davis, 1987) §3.1)。当 $\{X_t\}$ 平稳时级数必以概率 1 收敛和均方收敛到同一极限，结果也是平稳序列。

7.7 附录：常系数齐次线性微分方程

设 $a_0, a_1, \dots, a_p$ 为实常数，$a_p \neq 0$，微分方程

$$a_0 y + a_1 y' + \cdots + a_{p-1} y^{(p-1)} + a_p y^{(p)} = 0$$

称为 p 阶常系数齐次线性微分方程。这是与常系数齐次线性差分方程类似的方程。定义其伴随方程 (auxiliary equation) 为

$$a_0 + a_1 z + \cdots + a_{p-1} z^{p-1} + a_p z^p = 0.$$

设伴随方程有 k 个不同的复根 z_j，$j = 1, 2, \dots, k$，设 z_j 为 $r(j)$ 重根，$\sum_{j=1}^n r(j) = p$。则

$$y = x^s e^{z_j x},\ s = 0, 1, \dots, r(j) - 1;\ j = 1, 2, \dots, k$$

是微分方程的 p 个线性独立的解，取其实部，就构成微分方程的 p 个线性独立的实函数解，通解为这 p 个基础解的线性组合。

7.8 附录：R 中多项式求根

在 R 软件中，用 polyroot() 函数求多项式的所有复根，输入为多项式的升幂表示的系数作为 R 向量。如：

$$z^2 - 2z + 1 = 0$$

```
polyroot(c(1, -2, 1))
```

```
## [1] 1+0i 1-0i
```

又如：

$$z^3 + z = 0$$

```
polyroot(c(0, 1, 0, 1))
```

```
## [1] 0+0i 0+1i 0-1i
```

用 abs() 函数求复数模，如：

```
abs(polyroot(c(0, 1, 0, 1)))
```

```
## [1] 0 1 1
```

Chapter 8

自回归模型及其平稳性

8.1 特例：AR(1)

$$X_t = aX_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \{\varepsilon_t\} \sim \text{WN}(0, \sigma^2) \tag{8.1}$$

从初值 X_0 出发。a 越小，初值影响减小越快。$|a|$ 接近于 1 时，初值和前面的 ε_{t-j} 影响减小越慢，序列振荡。

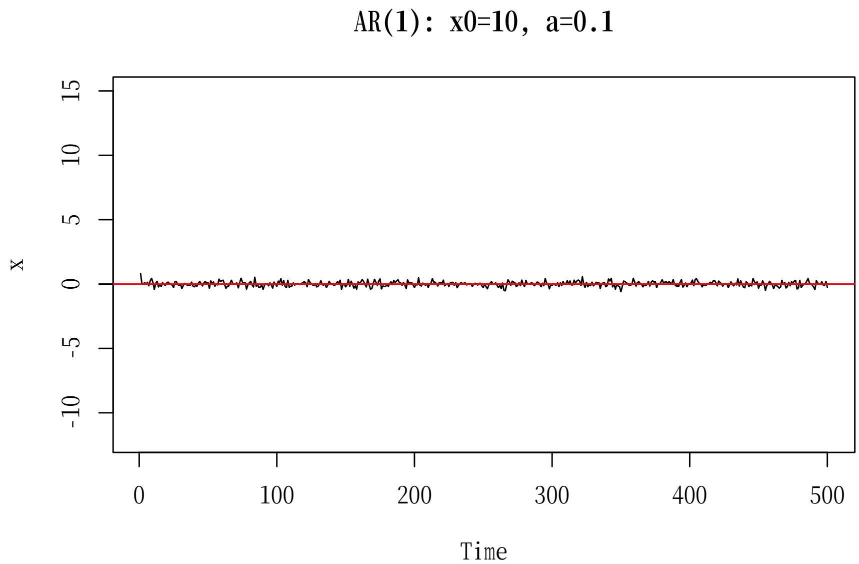
只要 $|a| < 1$，序列最终可以稳定下来。称系统(8.1)是**稳定的**。

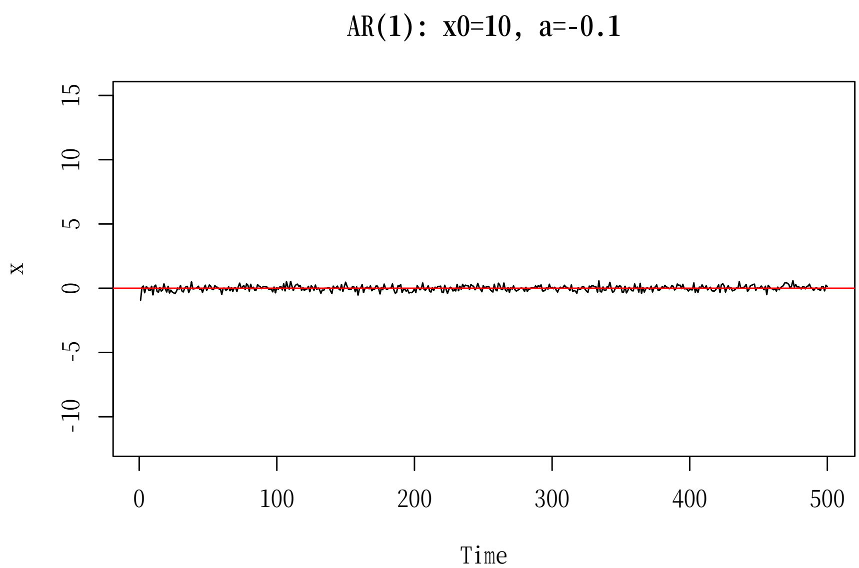
如果 $a = \pm 1$ 则序列振荡越来越大，呈爆炸型。

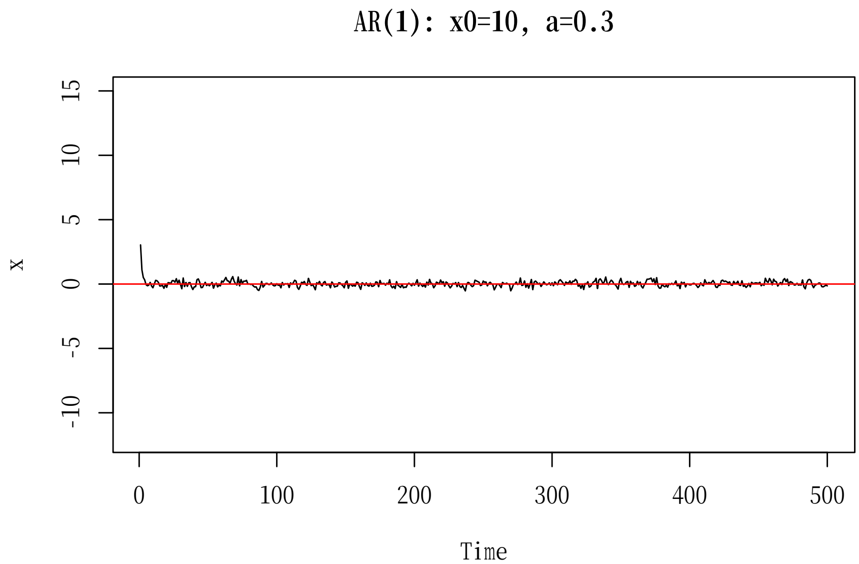
$|a| > 1$ 时序列也不能稳定。$|a| \geq 1$ 时称(8.1)是**非稳定的**.

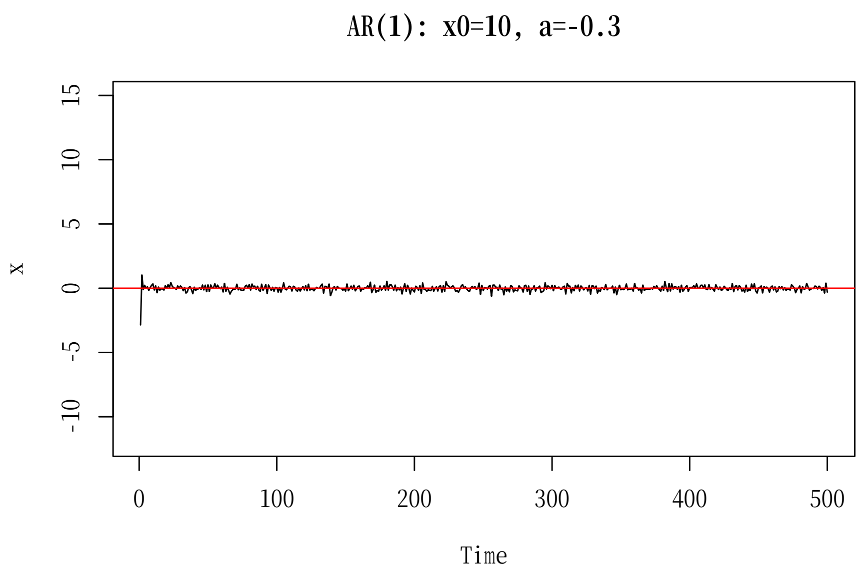
```
ar1.gen <- function(n, a, sigma=1.0,
                    plot.it=FALSE, n0=1000,
                    x0=numeric(length(a))){
  n2 <- n0 + n
  eps <- rnorm(n2, 0, sigma)
  x2 <- filter(eps, a, method="recursive", side=1, init=x0)
  x <- x2[(n0+1):n2]
  x <- ts(x)
```

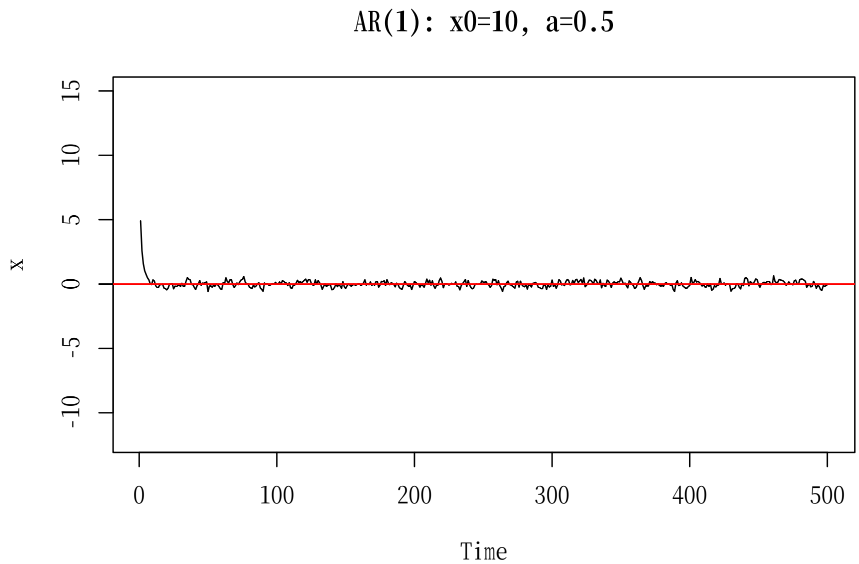
```r
  attr(x, "model") <- "AR(1)"
  attr(x, "a") <- a
  attr(x, "sigma") <- sigma
  if(plot.it) plot(x)
  x
}

demo.ar1 <- function(){
  as <- c(0.1, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99, 1.0, 1.001)
  x0 <- 10
  n <- 500
  for(a in as){
    for(a1 in c(a, -a)){
      x <- ar1.gen(n=n, a=a1, sigma=0.2,
                   n0=0, x0=x0)
      plot(x, main=paste("AR(1): x0=10, a=", a1, sep=""),
           ylim=c(-12, 15))
      abline(h=0, col="red")
    }
  }
}
set.seed(106)
demo.ar1()
```

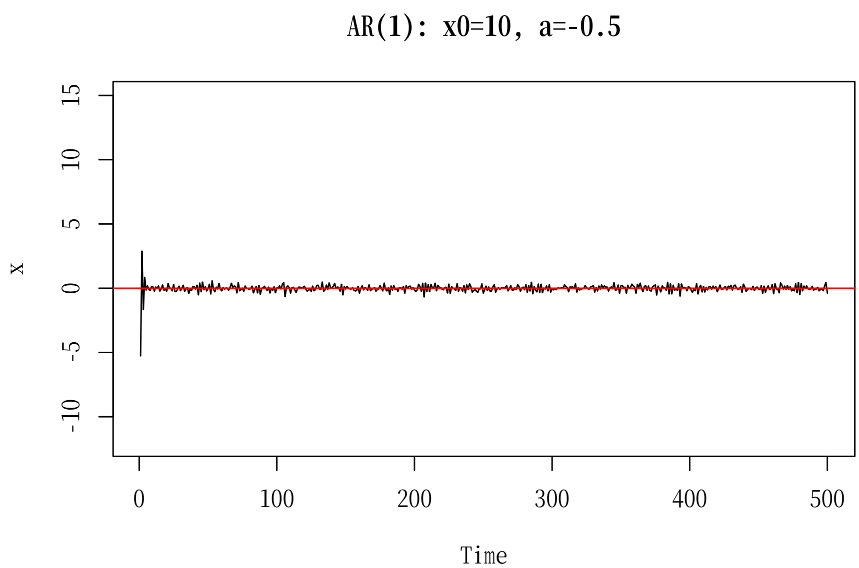

AR(1): x0=10, a=0.1
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

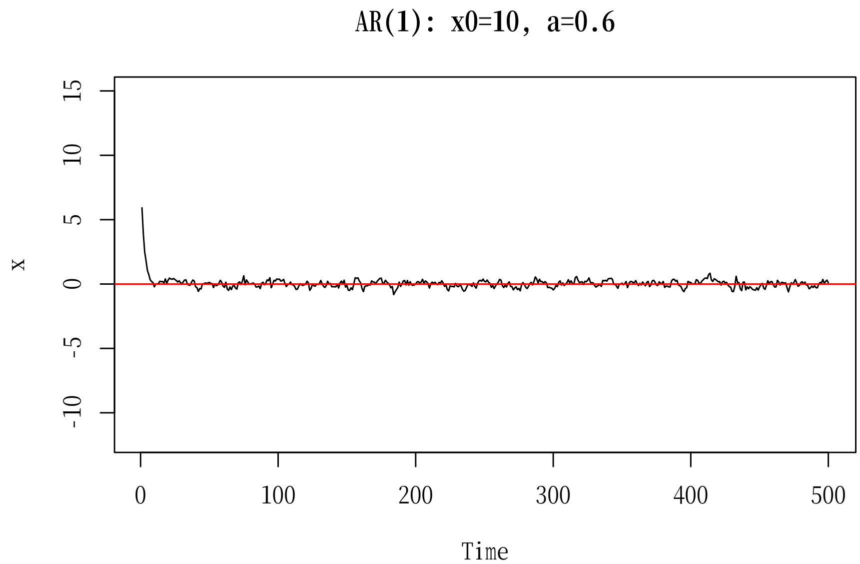

AR(1): x0=10, a=-0.1
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

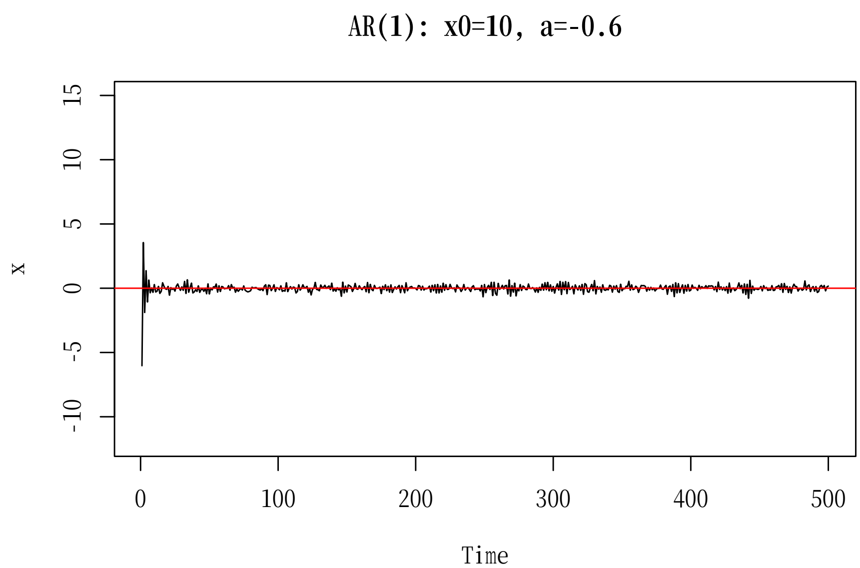

AR(1): x0=10, a=0.3
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

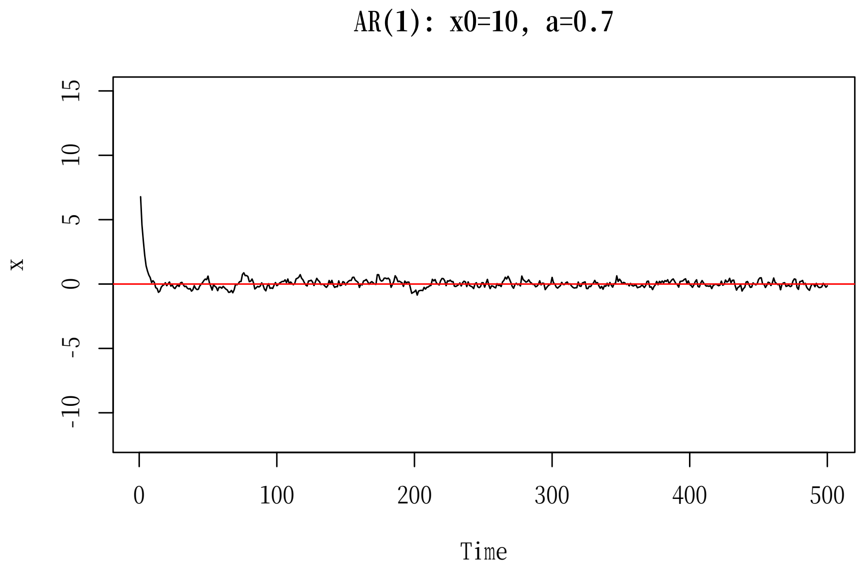

AR(1): x0=10, a=-0.3
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

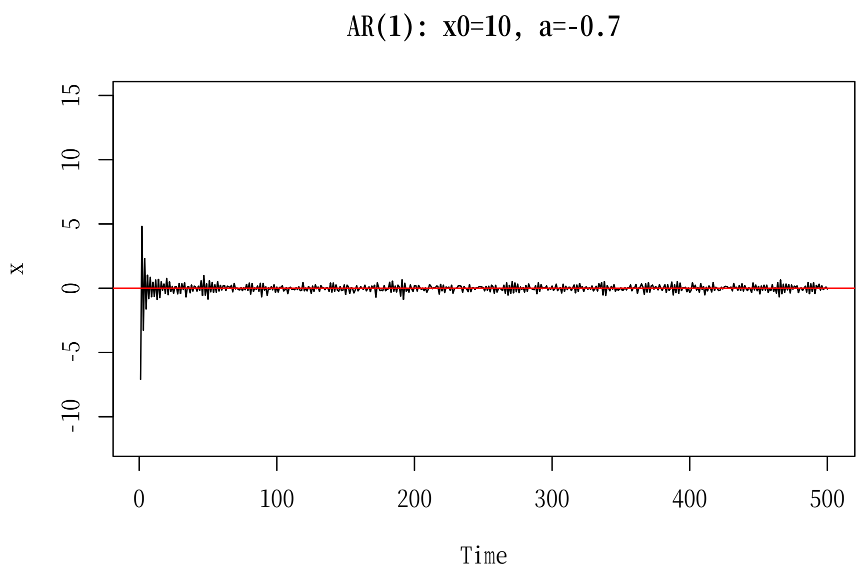

AR(1): x0=10, a=0.5
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

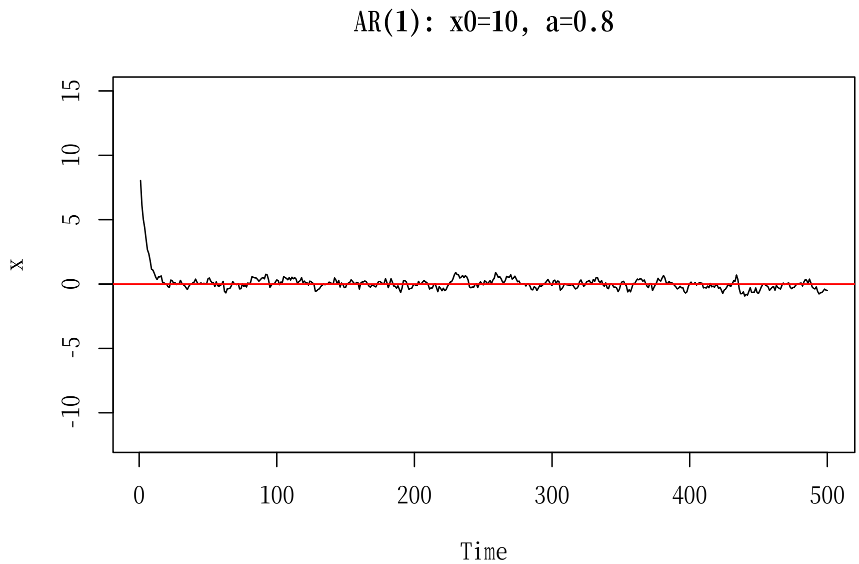

AR(1): x0=10, a=-0.5
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

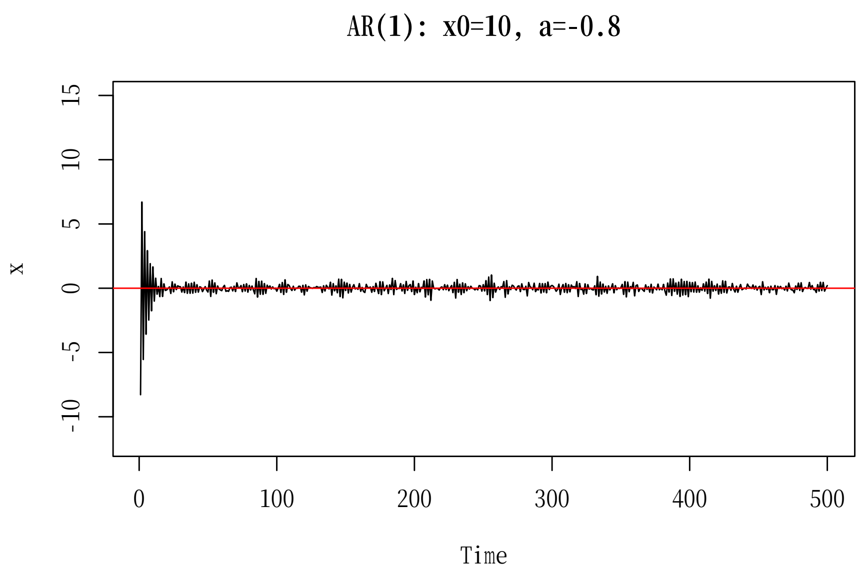

AR(1): x0=10, a=0.6
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

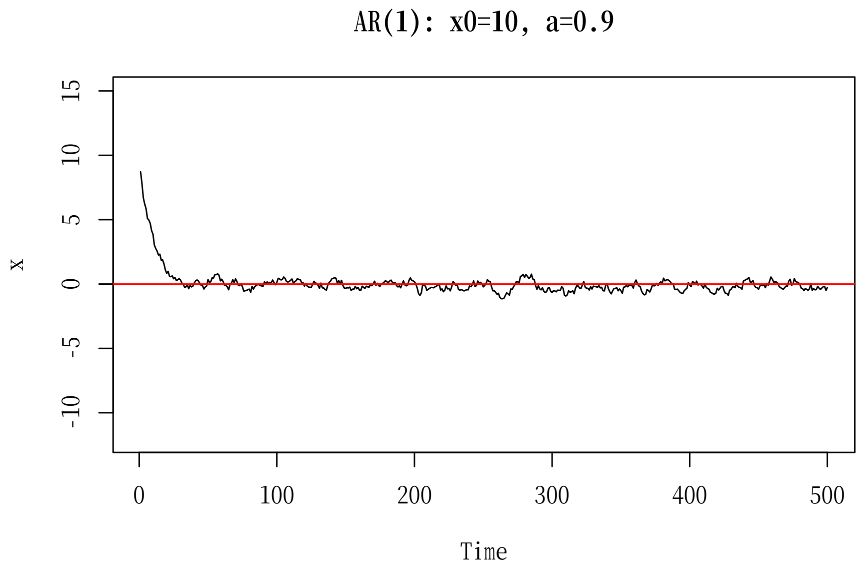

AR(1): x0=10, a=-0.6
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

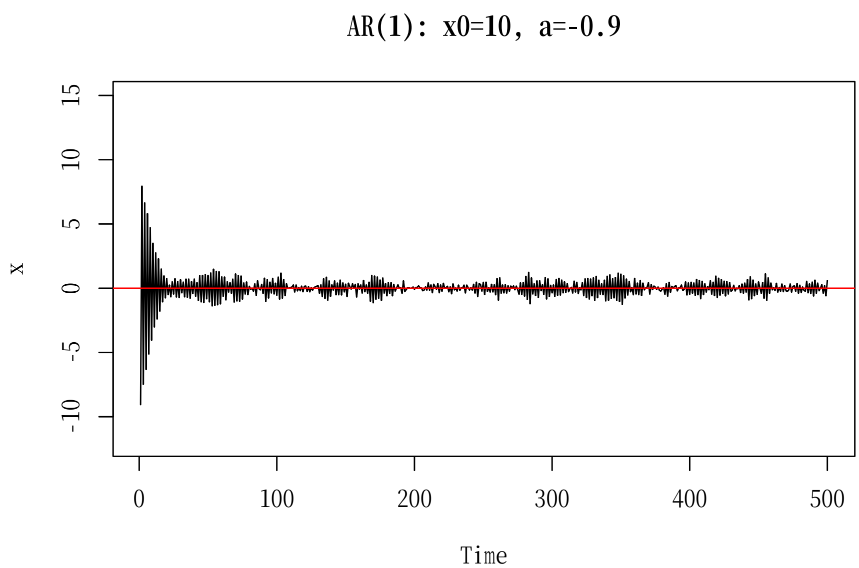

AR(1): x0=10, a=0.7
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

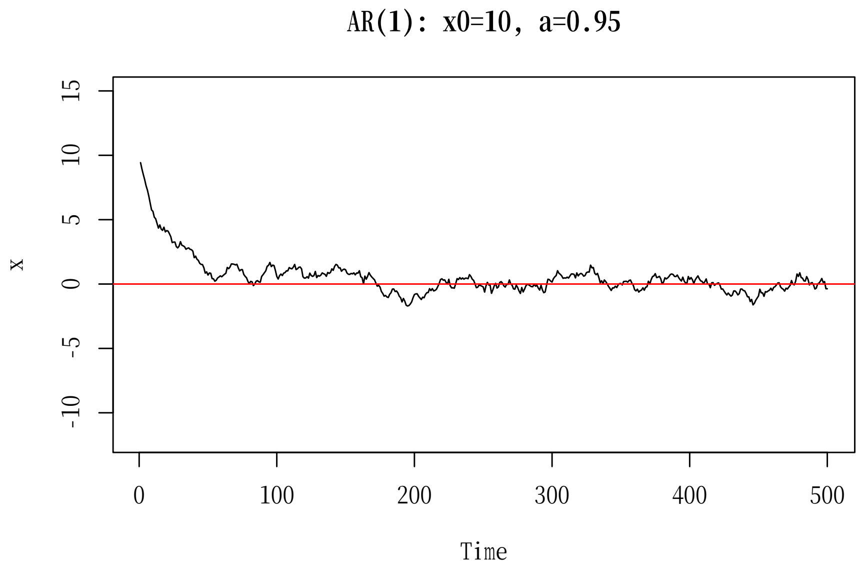

AR(1): x0=10, a=-0.7
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

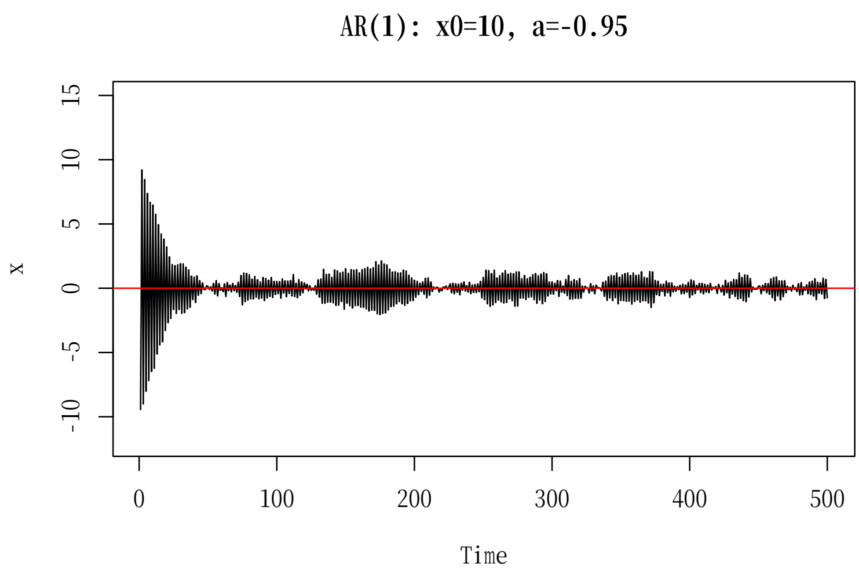

AR(1): x0=10, a=0.8
x
-10
-5
0
5
10
15
0
100
200
300
400
500
Time

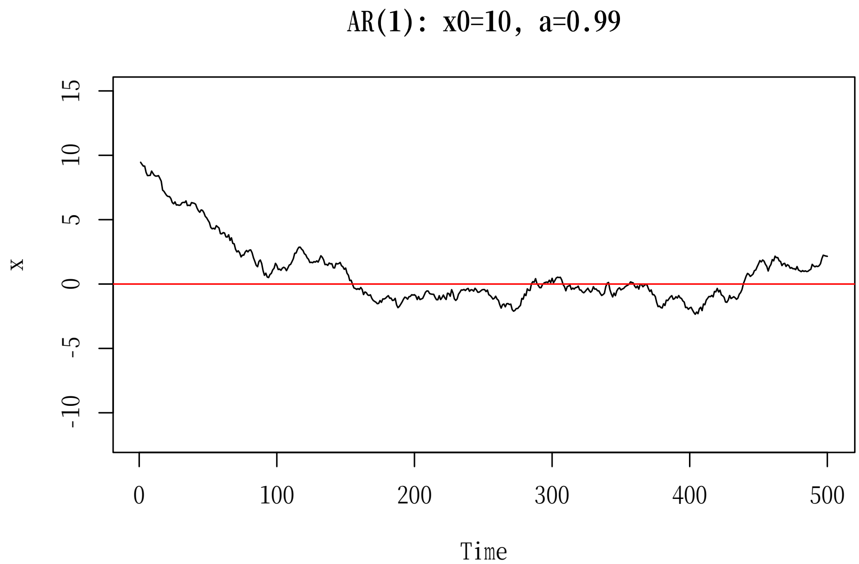

AR(1): x0=10, a=-0.8
x
-10
-5
0
5
10
15
0
100
200
300
400
500
Time

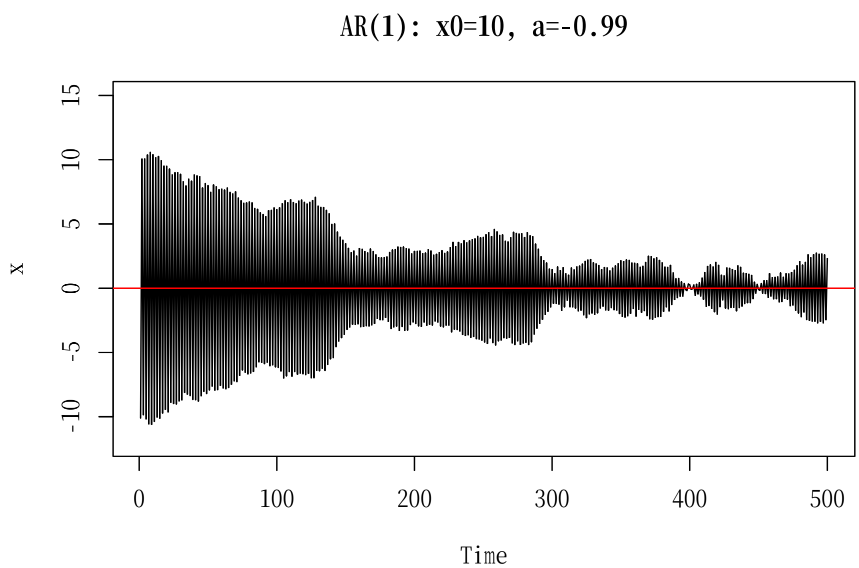

AR(1): x0=10, a=0.9
15
10
5
0
-5
-10
x
0
100
200
300
400
500
Time

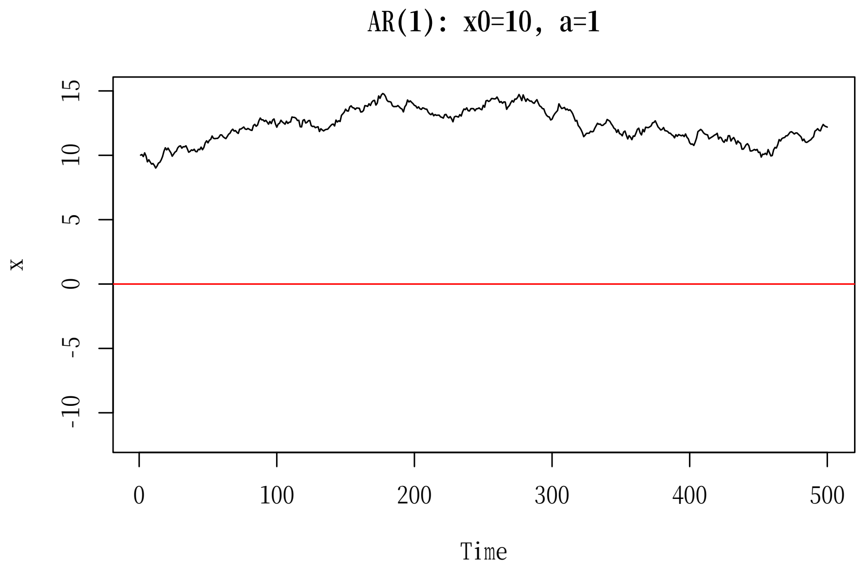

AR(1): x0=10, a=-0.9
15
10
5
0
-5
-10
x
0
100
200
300
400
500
Time

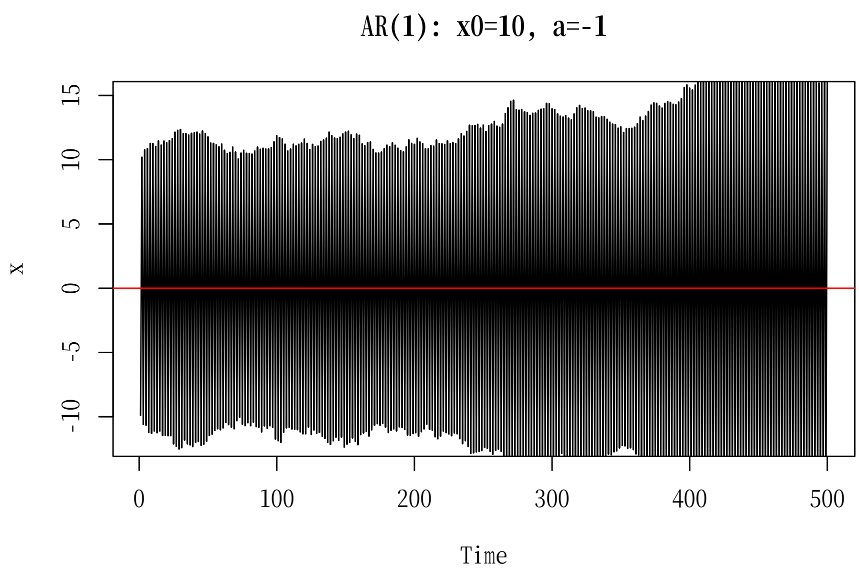

AR(1): x0=10, a=0.95
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

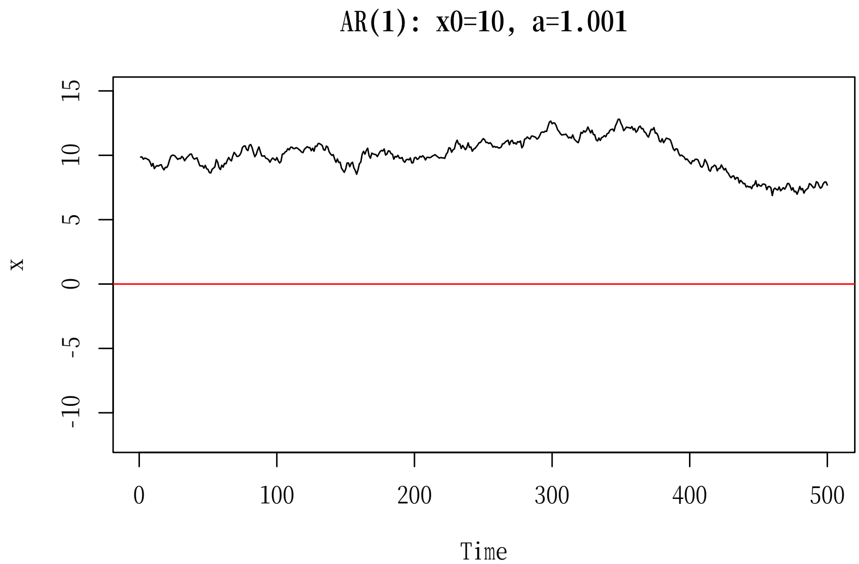

AR(1): x0=10, a=-0.95
x
15
10
5
0
-5
-10
0
100
200
300
400
500
Time

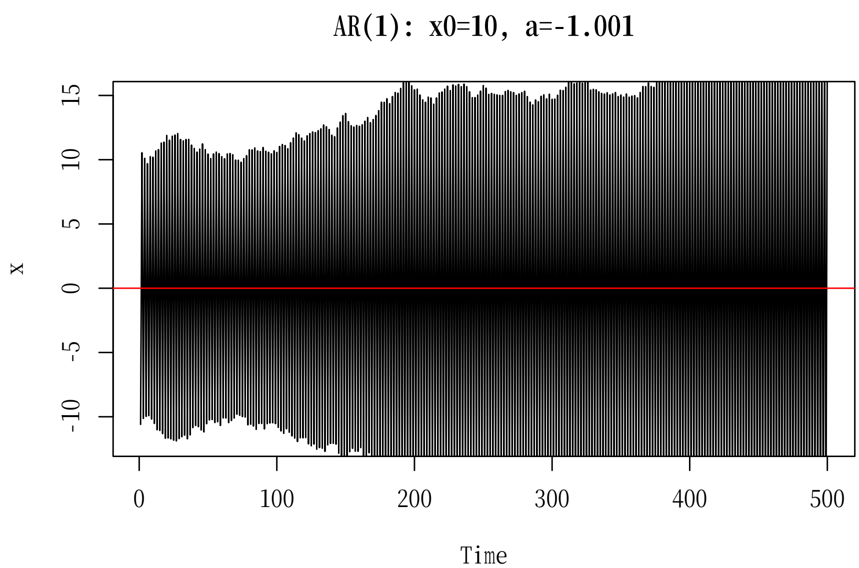

AR(1): x0=10, a=0.99
15
10
5
0
-5
-10
x
0
100
200
300
400
500
Time


AR(1): x0=10, a=-0.99
15
10
5
0
-5
-10
x
0
100
200
300
400
500
Time


AR(1): x0=10, a=1
x
Time


AR(1): x0=10, a=-1
x
Time


AR(1): x0=10, a=1.001
15
10
5
0
-5
-10
x
0
100
200
300
400
500
Time


AR(1): x0=10, a=-1.001
15
10
5
0
-5
-10
x
0
100
200
300
400
500
Time

8.1.1 AR(1) 的差分方程及平稳解

$A(z)=1-az$ 是差分方程(8.1)的特征多项式。$z_1=\frac{1}{a}$ 是特征根。稳定的充分必要条件是 $|a|<1$，或 $|z_1|>1$，即特征根都在单位圆外。

当 $|a|<1$ 时下面的线性序列有定义:

$$X_t=\sum_{j=0}^{\infty}a^j\varepsilon_{t-j},\quad t\in\mathbb{Z} \tag{8.2}$$

$$\begin{aligned}aX_{t-1}+\varepsilon_t=&a\sum_{j=0}^{\infty}a^j\varepsilon_{t-1-j}+\varepsilon_t\\=&\sum_{i=1}^{\infty}a^i\varepsilon_{t-i}+\varepsilon_t\quad(i=j+1)\\=&X_t\end{aligned}$$

于是平稳序列(8.2)是非齐次差分方程(8.1)的解，称为**平稳解**。

(8.1)的通解为

$$X_t=\sum_{j=0}^{\infty}a^j\varepsilon_{t-j},+\xi a^t,\quad t\in\mathbb{Z} \tag{8.3}$$

当 $t\to\infty$ 时(8.1)的所有解 a.s. 收敛到平稳解(8.2)。收敛速度是负指数速度 $|a|^t$。平稳解可以看成系统(8.1)处于稳定状态的情况。特征根 $\frac{1}{a}$ 离单位圆越远，稳定性越好。

8.2 一般 AR(p)

定义 8.1. 如果 $\{\varepsilon_t\}$ 是白噪声 WN$(0,\sigma^2)$, 实数 $a_1,a_2,\dots,a_p$ $(a_p\neq 0)$ 使得多项式 $A(z)$ 的零点都在单位圆外:

$$A(z)=1-\sum_{j=1}^{p}a_jz^j\neq 0,\ |z|\leq 1, \tag{8.4}$$

则称 p 阶差分方程

$$X_t=\sum_{j=1}^{p}a_jX_{t-j}+\varepsilon_t,\ \ t\in\mathbb{Z}, \tag{8.5}$$

是一个 p 阶自回归模型, 简称为 **AR(p) 模型**.

满足 AR(p) 模型(8.5)的平稳时间序列 $\{X_t\}$ 称为(8.5)的**平稳解**, 也称作 AR(p) 序列.

称 $a = (a_1, a_2, \cdots, a_p)^T$ 是 AR(p) 模型的自回归系数. 称条件(8.4)是**稳定性条件或最小相位条件**. $A(z)$ 称为模型(8.5)的特征多项式。模型可用推移算子写成

$$A(\mathscr{B})X_t = \varepsilon_t, \quad t \in \mathbb{Z} \tag{8.6}$$

8.3 平稳解和通解

8.3.1 AR(p) 的平稳解

由定理7.2可知，

$$A^{-1}(z) = \sum_{j=0}^{\infty} \psi_j z^j,$$

展开序列在 $|z| \leq \rho$ 收敛，其中 $\rho > 1$ 小于特征多项式的每一个根的模，$\psi_j = o(\rho^{-j})$。于是

$$Y_t = A^{-1}(\mathscr{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

是线性平稳列，有

$$A(\mathscr{B})Y_t = A(\mathscr{B})A^{-1}(\mathscr{B})\varepsilon_t = \varepsilon_t,$$

从而 $\{Y_t\}$ 是 AR(p) 模型的平稳解。反之，如果 $\{X_t\}$ 是平稳解，则根据定理7.2，

$$A^{-1}(\mathscr{B})[A(\mathscr{B})X_t] = A^{-1}(\mathscr{B})\varepsilon_t = Y_t,$$

即有 $X_t = Y_t$。所以

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{8.7}$$

这是 AR(p) 模型(8.5)的唯一的平稳解。$\{\psi_j\}$ 称为平稳序列 $\{X_t\}$ 的 Wold 系数，以负指数速度衰减，且特征根离单位圆越远，衰减速度越快。

定理 8.1. *(1)* 由(8.7)定义的时间序列 $\{X_t\}$ 是 *AR(p)* 模型(8.5)的唯一 *(a.s.* 意义*)* 平稳解；

(2) AR(p) 的模型的通解有如下形式

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t}, \quad t \in \mathbb{Z}. \tag{8.8}$$

8.3.2 Wold 系数的递推公式

记 $a_0 = -1$ 则 $A(z) = -\sum_{j=0}^{p} a_j z^j$，

$$1 = A(z)A^{-1}(z) = -\sum_{m=0}^{\infty} \left(\sum_{j=0}^{p} a_j \psi_{m-j} \right) z^m$$

故 $\psi_0 = 1$, $\sum_{j=0}^{p} a_j \psi_{m-j} = 0$, $m > 0$。于是

$$\begin{cases} \psi_0 = 1, & \\ \psi_m = \sum_{j=1}^{p} a_j \psi_{m-j}, & m = 1, 2, \ldots \\ \psi_m = 0, & m < 0 \end{cases}$$

用推移算子表示为

$$\psi_m = 0, m < 0; \quad \psi_0 = 1; \quad A(\mathscr{B})\psi_m = 0, m \geq 1.$$

8.3.3 通解与平稳解的关系

AR(p) 的通解 $\{Y_t\}$ 与平稳解有如下关系

$$|Y_t - X_t| = \left| \sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t} \right| = o(\rho^{-t}), \text{a.s.}, \ t \to \infty$$

其中 $1 < \rho < \min\{|z_j|\}$。

根离单位园越远，稳定下来的速度越快。可以用此事实作为模拟产生 AR(p) 序列的理论基础。

8.3.4 AR 序列的模拟

取 $x_{1-p} = \cdots = x_0 = 0$，生成 $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$. 迭代得到 $x_t = \varepsilon_t + \sum_{j=0}^{p} a_j x_{t-j}, j = 1, 2, \ldots, n_0 + n$。取 $y_t = x_{t+n_0}, t = 1, 2, \ldots, n$.

白噪声列可以用正态分布随机数来生成。

n_0 取 50 即可，但特征根接近单位圆时要取大的 n_0。

Chapter 9

AR(p) 序列的谱密度和 Yule-Walker 方程

9.1 AR(p) 序列的谱密度

9.1.1 AR(p) 序列的自协方差

因为 AR(p) 的平稳解为

$$X_t = A^{-1}(\mathscr{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

由线性平稳列性质知 $\{X_t\}$ 为零均值，自协方差函数为

$$\gamma_k = E(X_{t+k}X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \quad k = 0, 1, ... \tag{9.1}$$

设 $1 < \rho < \min\{|z_j|\}$，则 $\psi_j = o(\rho^{-j})$, 有

$$|\gamma_k| \le \sigma^2 \sum_{j=0}^{\infty} |\psi_j| \cdot |\psi_{j+k}| \le \sigma^2 (\sum_{j=0}^{\infty} |\psi_j|)(\sum_{l=k}^{\infty} |\psi_l|) \tag{9.2}$$

$$\le c_0 \sum_{l=k}^{\infty} \rho^{-l} \le c_1 \rho^{-k} \tag{9.3}$$

即 $\{\gamma_k\}$ 负指数衰减。

$\{X_t\}$ 序列前后的相关减小很快，称为时间序列的**短记忆性**。征根离单位圆越远 $\{\gamma_k\}$ 衰减越快。

9.1.2 AR(p) 的谱密度

由线性平稳列的谱密度公式得平稳解的谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \psi_j e^{ij\lambda} \right|^2$$

而 $\sum \psi_j z^j = 1/A(z)$ 所以

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{\left|A(e^{i\lambda})\right|^2} \tag{9.4}$$

$f(\lambda)$ 是一个恒正的偶函数。如果 $A(z)$ 有靠近单位圆的根 $\rho_j e^{i\lambda_j}$ 则 $|A(e^{i\lambda_j})|$ 会接近零，造成谱密度在 $\lambda = \lambda_j$ 处有一个峰值。

9.1.3 谱密度的自协方差函数反演公式

谱密度的定义是满足

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$$

的非负可积函数。上式是一个 Fourier 级数系数的公式 (差一个常数)。

在 $\{\gamma_k\}$ 满足一定条件下 $f(\lambda)$ 必存在，且可表成 $\{\gamma_k\}$ 的 Fourier 级数。

定理 9.1. *如果平稳序列 $\{X_t\}$ 的自协方差函数 $\{\gamma_k\}$ 绝对可和: $\sum |\gamma_k| < \infty$, 则 $\{X_t\}$ 有谱密度*

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-ik\lambda}. \tag{9.5}$$

由于谱密度是实值函数, 所以(9.5)还可以写成

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k \cos(k\lambda) = \frac{1}{2\pi} \left[\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(k\lambda) \right].$$

证明: 因为 $\{\gamma_k\}$ 绝对可和所以(9.5)右边绝对一致收敛, $f(\lambda)$ 连续。于是积分与级数可交换：

$$\int_{-\pi}^{\pi} f(\lambda) e^{ij\lambda}\, d\lambda = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k \int_{-\pi}^{\pi} e^{-i(k-j)\lambda}\, d\lambda = \gamma_j.$$

还要验证 $f(\lambda)$ 非负。若 $X_1, \dots, X_N$ 为 $\{X_t\}$ 的观测值，

$$I_N(\lambda) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N} X_j e^{ij\lambda}\right|^2, \quad \lambda \in [-\pi, \pi]$$

称为 $X_1, \dots, X_N$ 的**周期图**。令 $f_N(\lambda) = EI_N(\lambda)$, 则 $f_N(\lambda) \geq 0$，于是

$$\begin{aligned} 0 \leq & f_N(\lambda) = \frac{1}{2\pi N}\sum_{k=1}^{N}\sum_{j=1}^{N}\gamma_{k-j} e^{-i(k-j)\lambda} \\ =& \frac{1}{2\pi N}\sum_{m=1-N}^{N-1}(N-|m|)\gamma_m e^{-im\lambda} \qquad (*) \\ =& \frac{1}{2\pi}\sum_{m=1-N}^{N-1}\gamma_m e^{-im\lambda} - \frac{1}{2\pi N}\sum_{m=1-N}^{N-1}|m|\gamma_m e^{-im\lambda}. \end{aligned}$$

由 Kronecker 引理知后一项趋于 0, 于是

$$f(\lambda) = \lim_{N\to\infty} f_N(\lambda) \geq 0.$$

附注：(*) 式的二重求和的简化

$$\begin{aligned} &\frac{1}{2\pi N}\sum_{k=1}^{N}\sum_{j=1}^{N}\gamma_{k-j} e^{-i(k-j)\lambda} \\ =&\frac{1}{2\pi N}\sum_{k=1}^{N}\sum_{m=k-N}^{k-1}\gamma_m e^{-im\lambda} \quad (\text{令} m = k-j) \end{aligned}$$

交换 m 与 k 的求和次序。因为关于 m 的条件为

$$k - N \leq m \leq k-1$$

所以 $k \leq m + N$, $k \geq m+1$，求和变为

$$\frac{1}{2\pi N}\sum_{m=1-N}^{N-1}\sum_{k=\max(m+1,1)}^{k=\min(m+N,N)}\gamma_m e^{-im\lambda}$$

因为 $m \geq 0$ 时 k 的求和从 $m+1$ 到 N 有 $N-m$ 项，$m<0$ 时 k 的求和从 1 到 $m+N=N-|m|$ 有 $N-|m|$ 项，所以求和变为

$$\frac{1}{2\pi N}\sum_{m=1-N}^{N-1}(N-|m|)\gamma_m e^{-im\lambda}$$

附注：Kronecker 引理：设复数或实数级数 $\sum_{j=1}^n x_j$ 收敛，非负实数列 b_n 单调不减趋于 $+\infty$，则

$$\lim_{n\to\infty}\frac{1}{b_n}\sum_{j=1}^n b_j x_j = 0.$$

见B.2。

推论 9.1. *AR(p) 的平稳解序列 $\{X_t\}$ 有谱密度*

$$f(\lambda)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-ik\lambda}=\frac{\sigma^2}{2\pi}\frac{1}{|A(e^{i\lambda})|^2}.$$

9.2 Yule-Walker 方程

9.2.1 白噪声列与平稳解的关系

$A(\mathscr{B})X_t=\varepsilon_t$ 的平稳解为

$$X_t = A^{-1}(\mathscr{B})\varepsilon_t=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$$

对 $k\geq 1$ 由控制收敛定理或内积的连续性得

$$E(X_t\varepsilon_{t+k})=\sum_{j=0}^{\infty}\psi_j E(\varepsilon_{t-j}\varepsilon_{t+k})=0$$

即 X_t 与未来的输入不相关。

如果 $\{\varepsilon_t\}$ 是独立白噪声则 X_t 与未来的输入独立。

9.2.2 AR 序列的等价定义

设 $a_1,\dots,a_p$ 为实数，$\{\varepsilon_t\}$ 为 $\mathrm{WN}(0,\sigma^2)$, 若平稳列 $\{X_t, t\in\mathbb{Z}\}$ 满足

$$X_t = a_1X_{t-1}+\cdots+a_p x_{t-p}+\varepsilon_t,\ t\in\mathbb{Z},$$

且 $\varepsilon_{t+k}, k=1,2,...$ 与 $\{X_s : s \le t\}$ 互不相关，则称 $\{X_t\}$ 为 AR(p) 序列。

可以看出，两个定义是等价的。由前面的推导可知原始定义可以导出等价定义中的性质。反之，如果等价定义成立，$A(z) = 1 - a_1 z - \cdots - a_p z^p$ 一定满足最小相位性。

9.2.3 Yule-Walker 方程

在 AR(p) 模型

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_{t-p}$$

两边同时乘以 $X_{t-k}(k \ge 1)$ 后取期望，有

$$E(X_t X_{t-k}) = a_1 E(X_{t-1} X_{t-k}) + \cdots + a_p E(X_{t-p} X_{t-k}) + E(\varepsilon_t X_{t-k})$$

即有

$$\gamma_k = a_1 \gamma_{k-1} + \cdots + a_p \gamma_{k-p} + 0$$

所以，对 $k \ge 1$，有递推式

$$\gamma_k = a_1 \gamma_{k-1} + \cdots + a_p \gamma_{k-p}.$$

$k \ge 1$ 时，γ_k 满足齐次线性差分方程

$$A(\mathscr{B}) \gamma_k = 0.$$

将 $k = 1, 2, \ldots, p$ 的方程写成

$$\begin{aligned}
\gamma_1 =& a_1 \gamma_0 + a_2 \gamma_1 + \cdots + a_p \gamma_{p-1} \\
\gamma_2 =& a_1 \gamma_1 + a_2 \gamma_0 + \cdots + a_p \gamma_{p-2} \\
\vdots =& \vdots \\
\gamma_p =& a_1 \gamma_{p-1} + a_2 \gamma_{p-2} + \cdots + a_p \gamma_0
\end{aligned}$$

写成矩阵形式：

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_p \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \cdots & \gamma_{p-3} \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

记

$$\gamma_p = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_p \end{pmatrix}, \ \Gamma_p = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \cdots & \gamma_{p-3} \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix}, \ a_p = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

有

$$\Gamma_p a_p = \gamma_p$$

对 γ_0，

$$\begin{aligned} E[X_t \varepsilon_t] =& E[(a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t)\varepsilon_t] \\ =& 0 + \sigma^2 = \sigma^2. \end{aligned}$$

在 $X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t$ 两边同时乘以 X_t 后取期望，得

$$\begin{aligned} \gamma_0 =& a_1\gamma_1 + \cdots + a_p\gamma_p + E[X_t\varepsilon_t] \\ =& a_1\gamma_1 + \cdots + a_p\gamma_p + \sigma^2. \end{aligned}$$

上式另一证明为：

$$\begin{aligned} \gamma_0 =& E X_t^2 = E\left(\sum_{j=1}^p a_j X_{t-j} + \varepsilon_t\right)^2 \\ =& E\left(\sum_{j=1}^p a_j X_{t-j}\right)^2 + E\varepsilon_t^2 \\ =& a_p^T \Gamma_p a_p + \sigma^2 \\ =& a_p^T \gamma_p + \sigma^2 \\ =& a_1\gamma_1 + a_2\gamma_2 + \cdots + a_p\gamma_p + \sigma^2 \end{aligned}$$

于是

$$\begin{aligned} \gamma_k =& a_1\gamma_{k-1} + a_2\gamma_{k-2} + \cdots + a_p\gamma_{k-p}, \ k \geq 1 \\ \gamma_0 =& a_1\gamma_1 + a_2\gamma_2 + \cdots + a_p\gamma_p + \sigma^2. \end{aligned}$$

用推移算子写成

$$\begin{aligned} A(\mathscr{B})\gamma_k =& 0, \ k \geq 1 \\ A(\mathscr{B})\gamma_0 =& \sigma^2. \end{aligned}$$

对 $n \geq p$，令

$$\gamma_n = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_n \end{pmatrix}, \ a_n = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

仍有

$$\Gamma_n a_n = \gamma_n \tag{9.6}$$

在节22.1将可以看到，对于一般的平稳列 $\{X_t\}$，a_n 是用 $X_{t-1}, \dots, X_{t-n}$ 预测 X_t 时的最优线性预测系数。

为了证明(9.6)，对 $n \geq p$, 把 $X_t, X_{t+1}, \dots, X_{t+n-1}$ 的递推式写成矩阵形式得

$$\begin{pmatrix} X_t \\ X_{t+1} \\ \vdots \\ X_{t+n-1} \end{pmatrix} \tag{9.7}$$

$$= \begin{pmatrix} X_{t-1} & X_{t-2} & \dots & X_{t-n} \\ X_t & X_{t-1} & \dots & X_{t+1-n} \\ \vdots & \vdots & & \vdots \\ X_{t+n-2} & X_{t+n-3} & \dots & X_{t-1} \end{pmatrix} a_n + \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t+1} \\ \vdots \\ \varepsilon_{t+n-1} \end{pmatrix}, \tag{9.8}$$

在(9.8)两边同时乘上 X_{t-1} 后取数学期望, 利用 X_t 与未来输入的不相关性就可以证明(9.6)式。

定理 9.2 (Yule-Walker 方程). *$AR(p)$ 序列的自协方差函数满足*

$$\gamma_n = \Gamma_n a_n, \quad \gamma_0 = \gamma_n^T a_n + \sigma^2, \quad n \geq p, \tag{9.9}$$

即

$$\gamma_k = a_1\gamma_{k-1} + a_2\gamma_{k-2} + \dots + a_p\gamma_{k-p}, \quad k \geq 1 \tag{9.10}$$

$$A(\mathscr{B})\gamma_k = 0, \quad k \geq 1 \tag{9.11}$$

$$A(\mathscr{B})\gamma_0 = \gamma_0 - a_1\gamma_1 - a_2\gamma_2 - \dots - a_p\gamma_p = \sigma^2 \tag{9.12}$$

特别地，当 $n=p$ 时

$$\Gamma_p \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_p \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \cdots \\ \gamma_p \end{pmatrix} \tag{9.13}$$

记 $\phi_0=1, \phi_1=-a_1, \dots, \phi_p=-a_p$, 则 $A(z)=\sum_{j=0}^p \phi_j z^j$，AR 模型可写成 $\sum_{j=0}^p \phi_j X_{t-j}=\varepsilon_t$。

Yule-Walker 方程可写成

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_p \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \vdots & & \vdots \\ \gamma_p & \gamma_{p-1} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

9.3 自协方差函数的周期性

对 $k<0$ 定义 $\psi_k=0$。

推论 9.2. *AR(p) 序列的自协方差函数 $\{\gamma_k\}$ 满足和 AR(p) 模型 $A(\mathscr{B})X_t=\varepsilon_t$ 相应的差分方程*

$$\gamma_k-(a_1\gamma_{k-1}+a_2\gamma_{k-2}+\cdots+a_p\gamma_{k-p})=\sigma^2\psi_{-k}, \quad k\in\mathbb{Z}.$$

证明: $k\geq 0$ 时即定理结论。对 $k<0$，

$$\begin{aligned} &\gamma_k-(a_1\gamma_{k-1}+a_2\gamma_{k-2}+\cdots+a_p\gamma_{k-p}) \\ =&E\left[X_{t-k}(X_t-\sum_{j=1}^p a_j X_{t-j})\right] \\ =&E(X_{t-k}\varepsilon_t)=E\left[\sum_{j=0}^\infty \psi_j\varepsilon_{t-k-j}\varepsilon_t\right]=\sigma^2\psi_{-k}. \end{aligned}$$

设 $A(z)$ 有 p 个互异根 $z_j = \rho_j e^{i\lambda_j}, j = 1, \dots, p$，可以证明（略）

$$\gamma_t = A^{-1}(\mathscr{B})\sigma^2 \psi_{-t} \tag{9.14}$$

$$= \sigma^2 \sum_{j=1}^{p} c_j A^{-1}(z_j^{-1}) z_j^{-t} \tag{9.15}$$

$$= \sigma^2 \sum_{j=1}^{p} A_j \rho_j^{-t} \cos(\lambda_j t + \theta_j), \quad t \geq 0. \tag{9.16}$$

可见如果 $\{z_j\}$ 中有靠近单位圆的复根则 $\{\gamma_k\}$ 的衰减振荡特性会显现出来。

例 9.1. AR(4) 模型 1:

$$z_1, z_2 = 1.09 e^{\pm i\pi/3}, \quad z_3, z_4 = 1.098 e^{\pm i2\pi/3}$$

周期为 $2\pi/(\pi/3) = 6$ 和 $2\pi/(2\pi/3) = 3$。

AR(4) 模型 2:

$$z_1, z_2 = 1.264 e^{\pm i\pi/3}, \quad z_3, z_4 = 1.273 e^{\pm i2\pi/3}$$

AR(4) 模型 3:

$$z_1, z_2 = 1.635 e^{\pm i\pi/3}, \quad z_3, z_4 = 1.647 e^{\pm i2\pi/3}$$

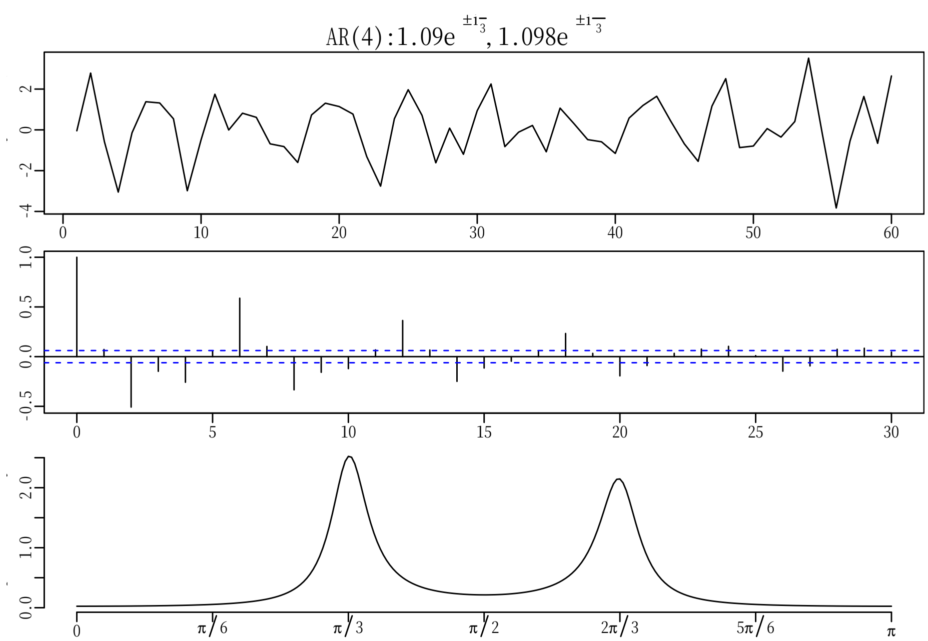
程序演示:

```r
library(polynom)
demo.ar.roots <- function(){
  n <- 1024
  rtlis <- list(c(complex(mod=1.09, arg=pi/3*c(1,-1)),
                  complex(mod=1.098, arg=pi*2/3*c(1,-1))),
                c(complex(mod=1.264, arg=pi/3*c(1,-1)),
                  complex(mod=1.273, arg=pi*2/3*c(1,-1))),
                c(complex(mod=1.635, arg=pi/3*c(1,-1)),
                  complex(mod=1.647, arg=pi*2/3*c(1,-1))),
                complex(mod=1.02, arg=pi/6*c(1,-1)),
                complex(mod=1.02, arg=pi/2*c(1,-1)),
```

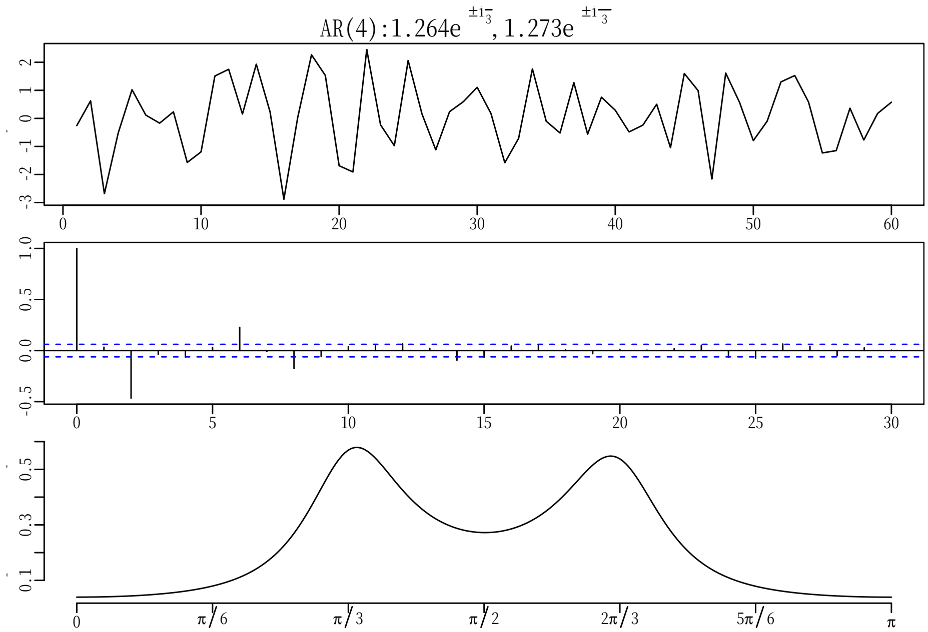
```
          c(complex(mod=1.05, arg=pi/6*c(1,-1)),
            complex(mod=1.05, arg=pi/2*c(1,-1))))
tits <- c("AR(4): 1.09exp(+-i pi/3), 1.098exp(+-i 2pi/3)",
          "AR(4): 1.264exp(+-i pi/3), 1.273exp(+-i 2pi/3)",
          "AR(4): 1.635exp(+-i pi/3), 1.647exp(+-i 2pi/3)",
          "AR(2): mod=1.02 arg=+-pi/6",
          "AR(2): mod=1.02 arg=+-pi/2",
          "AR(4): mod=1.05 arg=+-pi/6,+-pi/2")
tits <- c(
  expression(paste("AR(4):",
                   list(1.09*e^{phantom(.) %+-% i*frac(pi,3)},
                        1.098*e^{phantom(.) %+-% i*frac(2*pi,3)}) )),
  expression(paste("AR(4):",
                   list(1.264*e^{phantom(.) %+-% i*frac(pi,3)},
                        1.273*e^{phantom(.) %+-% i*frac(2*pi,3)}) )),
  expression(paste("AR(4):",
                   list(1.635*e^{phantom(.) %+-% i*frac(pi,3)},
                        1.647*e^{phantom(.) %+-% i*frac(2*pi,3)}) )),
  expression(paste("AR(2):",
                   list(1.02*e^{phantom(.) %+-% i*frac(pi,6)}) )),
  expression(paste("AR(2):",
                   list(1.02*e^{phantom(.) %+-% i*frac(pi,2)}) )),
  expression(paste("AR(4):",
                   list(1.05*e^{phantom(.) %+-% i*frac(pi,6)},
                        1.05*e^{phantom(.) %+-% i*frac(pi,2)}) ))
  )
oldpar <- par(mfrow=c(3,1), mar=c(2,2,0,0),
              mgp=c(2, 0.5, 0), oma=c(0,0,2,0))
on.exit(par(oldpar))
for(ii in seq(along=rtlis)){
  rt = rtlis[[ii]]
  y <- ar.gen(n, rt, sigma=1.0, by.roots=TRUE,
              plot.it=FALSE)
```

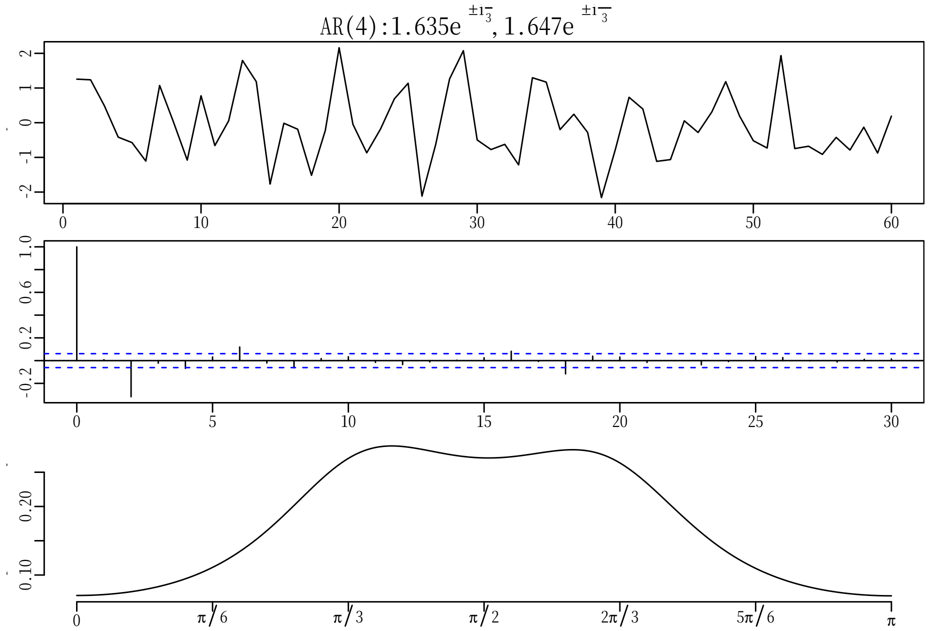
```
    plot(window(y, 1, 60))
    acf(y)
    ##spectrum(y, taper=0.2)
    ar.true.spectrum(attr(y, "a"), title="")
    mtext(tits[ii], side=3, outer=TRUE)
  }
}
demo.ar.roots()
```

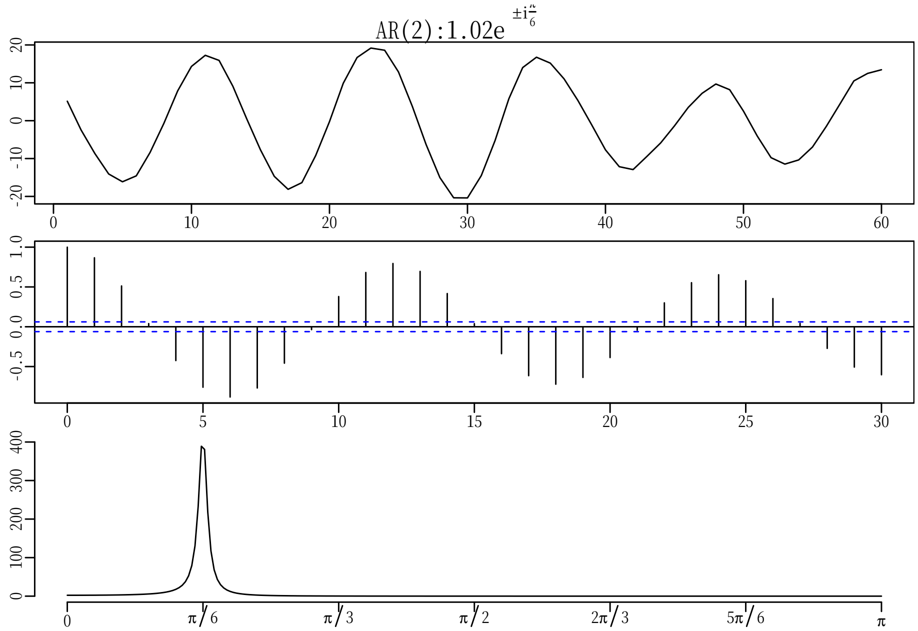
```
## Warning in polynomial(p): 强制改变时丢弃了虚数部分
```



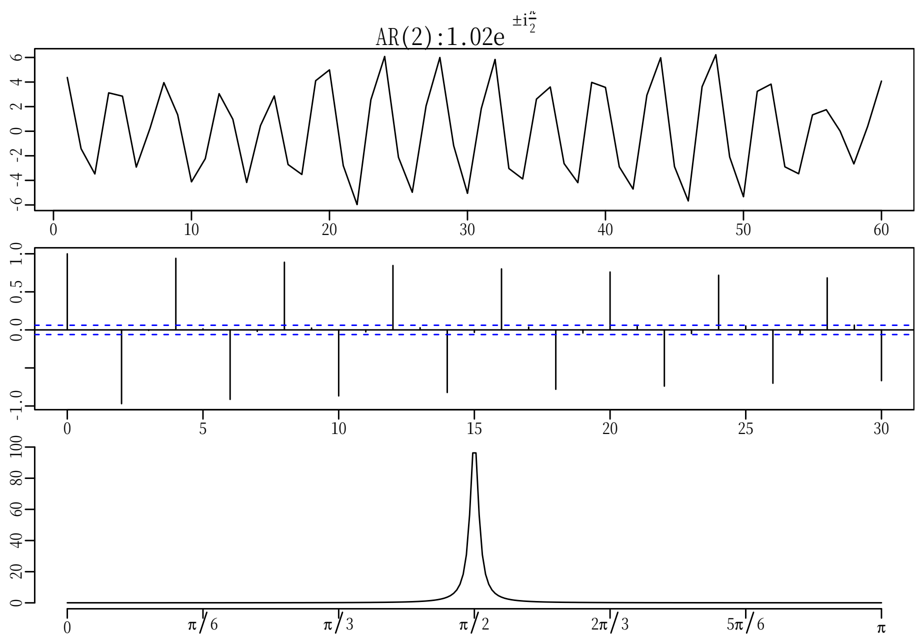

```
## Warning in polynomial(p): 强制改变时丢弃了虚数部分
```

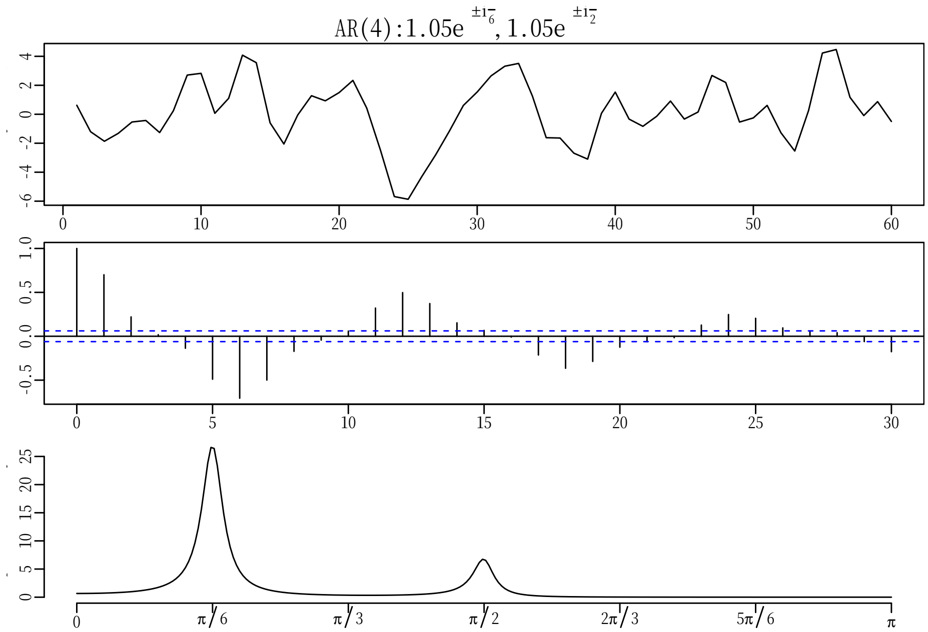

AR(4):1.264e, 1.273e


AR(4):1.635e, 1.647e



Warning in polynomial(p): 强制改变时丢弃了虚数部分





9.4 自协方差函数的正定性

AR(p) 平稳解唯一故自协方差函数可被自回归系数和白噪声方差唯一决定。反之，若 Γ_p 正定则根据 Yule-Walker 方程可以从 $\gamma_0, \gamma_1, \dots, \gamma_p$ 解出 $a_1, \dots, a_p, \sigma^2$:

$$a_p = \Gamma_p^{-1} \gamma_p, \quad \sigma^2 = \gamma_0 - \gamma_p^T \Gamma_p^{-1} \gamma_p. \tag{9.17}$$

定理 9.3. 设 Γ_n 是实值平稳序列 $\{X_t\}$ 的 n 阶自协方差矩阵, $\gamma_0 > 0$。

(1) 如果 $\{X_t\}$ 的谱密度 $f(\lambda)$ 存在, 则对 $n \geq 1$, Γ_n 正定;

(2) 如果 $\lim_{k\to\infty} \gamma_k = 0$, 则对 $n \geq 1$, Γ_n 正定。

引理 9.1. 对实平稳列 $\{X_t\}$, 设其自协方差阵为 Γ_n, $n \in \mathbb{N}$; 设其谱函数为 $F(\lambda)$。$\forall b = (b_1, \dots, b_n) \in \mathbb{C}^n$ 有

$$b^* \Gamma_n b = \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n} b_j e^{ij\lambda} \right|^2 dF(\lambda).$$

若 $\{X_t\}$ 有谱密度 $f(\lambda)$ 则

$$b^*\Gamma_n b = \int_{-\pi}^{\pi} \left|\sum_{j=1}^{n} b_j e^{ij\lambda}\right|^2 f(\lambda) d\lambda.$$

引理证明：

$$\begin{aligned}
b^*\Gamma_n b &= \sum_{j=1}^{n}\sum_{k=1}^{n} \bar{b}_j b_k \gamma_{k-j} \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} \bar{b}_j b_k \int_{-\pi}^{\pi} e^{i(k-j)\lambda} dF(\lambda) \\
&= \int_{-\pi}^{\pi} \sum_{j=1}^{n}\sum_{k=1}^{n} \bar{b}_j b_k e^{-ij\lambda} e^{ik\lambda} dF(\lambda) \\
&= \int_{-\pi}^{\pi} \overline{\left(\sum_{j=1}^{n} b_j e^{ij\lambda}\right)} \left(\sum_{k=1}^{n} b_k e^{ik\lambda}\right) dF(\lambda) \\
&= \int_{-\pi}^{\pi} \left|\sum_{j=1}^{n} b_j e^{ij\lambda}\right|^2 dF(\lambda).
\end{aligned}$$

定理证明：

(1) 对 $b = (b_1, \dots, b_n)^T$，$\sum_{j=1}^{n} b_j z^{j-1}$ 至多有 $n-1$ 个零点。$\gamma_0 = \int_{-\pi}^{\pi} f(\lambda) d\lambda > 0$, 于是

$$b^T\Gamma_n b = \int_{-\pi}^{\pi} \left|\sum_{j=1}^{n} b_j e^{ij\lambda}\right|^2 f(\lambda) d\lambda > 0.$$

(2) 用反证法。设 Γ_n 正定, $\det(\Gamma_{n+1}) = 0$ 和 $EX_t = 0$ (非零均值情况只要减去均值).

定义

$$X_n = (X_1, X_2, \dots, X_n)^T$$

对任何实向量 $b = (b_1, b_2, \dots, b_n)^T \neq 0$ 有

$$E(b^T X_n)^2 = b^T \Gamma_n b > 0,$$

且由 $|\Gamma_{n+1}| = 0$ 知存在 $a = (a_1, a_2, \dots, a_{n+1})^T \neq 0$, $a_{n+1} \neq 0$ 使得

$$E(a^T X_{n+1})^2 = a^T \Gamma_{n+1} a = 0.$$

于是

$$a^T X_{n+1} = a_1 X_1 + a_2 X_2 + \cdots + a_{n+1} X_{n+1} = 0$$

a.s. 成立, X_{n+1} 可以由 X_n 线性表示:

$$X_{n+1} = -\frac{a_n}{a_{n+1}} X_n - \frac{a_{n-1}}{a_{n+1}} X_{n-1} - \cdots - \frac{a_1}{a_{n+1}} X_1, \quad \text{a.s.},$$

利用 $\{X_t\}$ 的平稳性知道

$$X_t = -\frac{a_n}{a_{n+1}} X_{t-1} - \frac{a_{n-1}}{a_{n+1}} X_{t-2} - \cdots - \frac{a_1}{a_{n+1}} X_{t-n}, \quad \text{a.s.}, \quad t \in \mathbb{Z}.$$

递推知对任何 $k \geq 1$, X_{n+k} 可以由 $X_1, X_2, \dots, X_n$ 线性表示, 即有实向量 $\alpha \triangleq \alpha^{(k)} \triangleq (\alpha_1^{(k)}, \dots, \alpha_n^{(k)})^T$ 使得

$$X_{n+k} = (\alpha^{(k)})^T X_n.$$

X_{n+k} 被 X_n 线性表示，说明 X_{n+k} 与 $X_1, \dots, X_n$ 有强的相关，而定理假设是 $\gamma_k \to 0$，又说明 X_{n+k} 与 $X_1, \dots, X_n$ 的相关性要趋于零，这就会有矛盾，下面把矛盾严格表述。

用 $0 < \lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$ 表示 Γ_n 的特征值, 则有正交矩阵 B 使得

$$B\Gamma_n B^T = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

用 $|\alpha^{(k)}|$ 表示 $\alpha^{(k)}$ 的欧氏模, 则有

$$\begin{aligned}
\gamma_0 =& E X_{n+k}^2 = E((\alpha^{(k)})^T X_n)^2 = (\alpha^{(k)})^T \Gamma_n \alpha^{(k)} \\
=& ((\alpha^{(k)})^T B^T)(B\Gamma_n B^T)(B\alpha^{(k)}) \\
=& ((\alpha^{(k)})^T B^T)\Lambda(B\alpha^{(k)}) \\
\geq& \lambda_1 (B\alpha^{(k)})^T (B\alpha^{(k)}) = \lambda_1 |\alpha^{(k)}|^2.
\end{aligned}$$

即有 $|\alpha^{(k)}| \leq \sqrt{\gamma_0/\lambda_1} < \infty$.

另一方面

$$
\begin{aligned}
\gamma_0 =&E((\alpha^{(k)})^T X_n \cdot X_{n+k}) = (\alpha^{(k)})^T E(X_n X_{n+k}) \\
=&(\alpha^{(k)})^T (\gamma_{n+k-1}, \gamma_{n+k-2}, \cdots, \gamma_k)^T \\
\le& |\alpha^{(k)}| \left(\sum_{j=0}^{n-1} \gamma_{j+k}^2 \right)^{1/2} \\
\le& (\gamma_0/\lambda_1)^{1/2} \left(\sum_{j=0}^{n-1} \gamma_{k+j}^2 \right)^{1/2} \\
\to& 0. \quad (\text{当} k \to \infty)
\end{aligned}
$$

这与 $\gamma_0 > 0$ 矛盾, 故 $\det(\Gamma_{n+1}) = 0$ 不成立.

推论 9.3. 系数平方可和的线性平稳序列的自协方差阵总是正定的。

这是上面的定理与定理6.2的推论。

对平稳列 $\{X_n\}$ 的自协方差函数 $\{\gamma_j, j \in \mathbb{Z}\}$，如果 $\{X_n\}$ 自协方差阵总是正定的，则称自协方差函数 $\{\gamma_j, j \in \mathbb{Z}\}$ 是正定序列。

定理 9.4. 设实平稳列 $\{X_t\}$ 的谱函数 $F(\lambda)$ 是阶梯函数。如果 $F(\lambda)$ 恰好有 n 个跳跃点，则 Γ_n 正定而 Γ_{n+1} 退化。如果 $F(\lambda)$ 有无穷个跳跃点，则对任意 $n \ge 1$，Γ_n 都是正定的。

证明：$\forall b = (b_1, \dots, b_n)^T$, λ 的函数

$$
g(\lambda) = \sum_{k=1}^{n} b_k e^{ik\lambda}
$$

是函数

$$
h(z) = \sum_{k=1}^{n} b_k z^k = z \sum_{j=0}^{n-1} b_{j+1} z^j
$$

在 $z = e^{i\lambda}$ 的值，所以 $g(\lambda)$ 至多有 $n-1$ 个零点。当 $F(\lambda)$ 的跳跃点个数 $\ge n$ 时，

$$
\begin{aligned}
b^T \Gamma_n b =& \int_{-\pi}^{\pi} \left| \sum_{k=1}^{n} b_k e^{ik\lambda} \right|^2 \, dF(\lambda) \\
=& \int_{-\pi}^{\pi} |g(\lambda)|^2 \, dF(\lambda)
\end{aligned}
$$

关于 $dF(\lambda)$ 的积分等于跳跃高度乘以跳跃点处被积函数然后求和，这里被积函数只有至多 $n-1$ 个零点而跳跃点有 n 个以上，所以求和至少有一项非零，故积分为正值，即 Γ_n 正定。

如果 $G(\lambda)$ 恰好有如下的 n 个跳跃点：

$$-\pi < \lambda_1 < \cdots < \lambda_n \leq \pi$$

定义复数 $b_0, b_1, \ldots, b_n$ 为

$$\sum_{k=0}^{n} b_k e^{ik\lambda} = \prod_{j=1}^{n} (1 - e^{i(\lambda-\lambda_j)})$$

则 $b_0 = 1$，令 $b = (b_0, b_1, \ldots, b_n)^T \neq 0$，当 $G(\cdot)$ 是 $n/2$ 个频率的离散谱序列时，因为频率是相反数成对出现的，所以 b 为实向量。对 b 有

$$b^* \Gamma_{n+1} b = \int_{-\pi}^{\pi} \left| \sum_{k=0}^{n} b_k e^{ik\lambda} \right|^2 dF(\lambda)$$

因为函数 $\sum_{k=0}^{n} b_k e^{ik\lambda}$ 在 $F(\lambda)$ 的 n 个跳跃点处都等于零，所以上述积分为零，如果 b 为实向量，则 Γ_{n+1} 退化。

如果 b 为复向量，对零均值平稳列 $\{X_t\}$ 有

$$\mathrm{Var}(\sum_{j=0}^{n} b_j X_j) = E \left| \sum_{j=0}^{n} b_j X_j \right|^2 = b^* \Gamma_{n+1} b = 0$$

取 $a = \mathrm{Re}(b)$，则 $a^T \Gamma_{n+1} a = 0$，且 $a_0 = b_0 = 1$ 故 $a \neq 0$，Γ_{n+1} 退化。定理证毕。

9.5 时间序列的可完全预测性

有限个频率的离散谱序列的轨道具有周期性，可以用有限个历史值的线性组合无误差地预报整个序列。

对于方差有限的随机变量 $Y_1, Y_2, \cdots, Y_n$, 如果有不全为零的常数 $b_1, \ldots, b_n$ 使得

$$\mathrm{Var}(\sum_{j=1}^{n} b_j Y_j) = 0,$$

则称随机变量 $Y_1, Y_2, \cdots, Y_n$ 是**线性相关**的, 否则称为**线性无关**的.

线性相关时, 存在常数 b_0 使得 $\sum_{j=1}^n b_j Y_j = b_0$ a.s. 成立.
并且当 $b_n \neq 0$ 时, Y_n 可以由 $Y_1, Y_2, \cdots, Y_{n-1}$ 线性表示:

$$Y_n = a_0 + a_1 Y_{n-1} + \cdots + a_{n-1} Y_1$$

这时我们称 Y_n 可以由 $Y_1, Y_2, \cdots, Y_{n-1}$ **完全线性预测**.

对于平稳序列 $\{X_t\}$，$X_{t-1}, \dots, X_{t-n}$ 的一个带截距的线性组合为 $b_1 X_{t-1} + \dots b_n X_{t-n} - b_0$，这 n 个变量线性无关当且仅当

$$\begin{aligned}
&\mathrm{Var}(\sum_{j=1}^n b_j X_{t-j} - b_0) = \mathrm{Var}(\sum_{j=1}^n b_j X_{t-j}) \\
=& b\Gamma_n b > 0
\end{aligned}$$

即 Γ_n 正定。

反之，若 Γ_n 正定而 Γ_{n+1} 不满秩，则 X_t 可以被 $X_{t-1}, \dots, X_{t-n}$ 完全线性预测。

线性平稳列不能完全线性预测。

有限个频率成分的离散谱序列可完全线性预测。

Chapter 10

平稳序列的偏相关系数和 Levinson 递推公式

10.1 最优线性预测

10.1.1 有限个随机变量的最优线性预测

设 $X_1, X_2, \dots, X_n, Y$ 为随机变量。考虑估计问题

$$L(Y|X_1, \dots, X_n) \triangleq \operatorname*{argmin}_{\hat{Y}=a_0+a_1X_1+\cdots+a_nX_n} E(Y-\hat{Y})^2$$

称 $L(Y|X_1, \dots, X_n)$ 为 Y 关于 $X_1, \dots, X_n$ 的**最优线性预测**或者最优线性估计。$E(Y-\hat{Y})^2 = \|Y-\hat{Y}\|^2$ 称为预测的均方误差。$L(Y|X_1, \dots, X_n)$ 实际是用 $X_1, \dots, X_n$ 的带截距项的线性组合预测 Y 的均方误差最小的预测。

用 $\text{sp}(1, X_1, \dots, X_n)$ 表示由 $1, X_1, \dots, X_n$ 的线性组合组成的 L^2 的子空间。由 Hilbert 空间的投影理论，$L(Y|X_1, \dots, X_n)$ 是 Y 在 $\text{sp}(1, X_1, \dots, X_n)$ 上的投影。

下面用求多元函数最小值的方法推导 $L(Y|X_1, \dots, X_n)$ 的公式。记 $X = (X_1, \dots, X_n)^T$, 仅在 X 协方差阵正定情况下给出结果。

令 $\xi = X - EX$, $\eta = Y - EY$。设 $\Sigma_X \triangleq \text{Var}(X) = \text{Var}(\xi)$ 正定。记

$\Sigma_{X,Y} = \mathrm{Cov}(X,Y)$. 对 $a_0, a_1, \dots, a_n \in R$, 记 $a = (a_1, \dots, a_n)^T$，有

$$\begin{aligned}
&E\left(Y - (a_0 + a_1 X_1 + \cdots + a_n X_n)\right)^2 \\
=&E(\eta - (a_1\xi_1 + \dots a_n\xi_n))^2 + (EY - a_0 - a^T EX)^2
\end{aligned}$$

已知 $a_1, \dots, a_n$ 后取 $a_0 = EY - a^T EX$ 就可以使上式后一项为零，所以不妨设 $EX = 0$, $EY = 0$。

这时

$$\begin{aligned}
g(a) =&E(Y - (a_1X_1 + \dots a_nX_n))^2 = E(Y - a^T X)^2 \\
=&\mathrm{Var}(Y) + a^T\Sigma_X a - 2a^T\Sigma_{X,Y}, \\
\frac{\partial g(a)}{\partial a} =&2\Sigma_X a - 2\Sigma_{X,Y}, \\
\frac{\partial^2 g(a)}{\partial a \partial a^T} =&2\Sigma_X > 0.
\end{aligned}$$

令 $\frac{\partial g(a)}{\partial a} = 0$ 求得

$$a = \Sigma_X^{-1}\Sigma_{X,Y}$$

因为海色阵 $\frac{\partial^2 g(a)}{\partial a \partial a^T}$ 正定所以上式为 $g(a)$ 的唯一严格最小值点。

于是，

$$L(Y|X_1, X_2, \dots, X_n) = EY + \Sigma_{Y,X}\Sigma_X^{-1}(X - EX) \tag{10.1}$$

估计误差的最小值为

$$E(\eta - a\xi)^2 = \mathrm{Var}(Y) - \Sigma_{X,Y}^T\Sigma_X^{-1}\Sigma_{X,Y} \tag{10.2}$$

当 $|\Gamma_n| = 0$ 时（协方差阵不满秩时），最优线性估计也存在，但有无穷多个 (详见22)。

10.1.2 平稳列的最优线性预测

设 $\{X_t\}$ 为零均值平稳列。考虑用 $X_1,\dots,X_n$ 的线性组合预测 X_{n+1}。设 $\Gamma_n>0$，则

$$
\begin{aligned}
&L(X_{n+1}|X_n,X_{n-1},\dots,X_1)\\
=&\left[\mathrm{Var}(\begin{pmatrix}X_n\\ \vdots\\ X_1\end{pmatrix})^{-1}\mathrm{Cov}(\begin{pmatrix}X_n\\ \vdots\\ X_1\end{pmatrix},X_{n+1})\right]^T\begin{pmatrix}X_n\\ \vdots\\ X_1\end{pmatrix}\\
=&(\Gamma_n^{-1}\begin{pmatrix}\gamma_1\\ \gamma_2\\ \dots\\ \gamma_n\end{pmatrix})^T\begin{pmatrix}X_n\\ \dots\\ X_1\end{pmatrix}\\
\triangleq&a_{n1}X_n+a_{n2}X_{n-1}+\cdots+a_{nn}X_1\\
\triangleq&a_n^T(X_n,X_{n-1},\dots,X_1)^T
\end{aligned}
$$

Yule-Walker 方程:

$$
\Gamma_n\begin{pmatrix}a_{n1}\\ \vdots\\ a_{nn}\end{pmatrix}=\begin{pmatrix}\gamma_1\\ \vdots\\ \gamma_n\end{pmatrix}
$$

简记为

$$
\Gamma_n a_n=\gamma_n
$$

可以证明（见22），Y-W 方程总有解，当且仅当 Γ_n 满秩时解是唯一的。Y-W 方程的解 a_n 称为 $\{X_t\}$ 或 $\{\gamma_k\}$ 的 n 阶 Yuler-Walker 系数。

$$
L(X_{n+1}|X_n,\dots,X_1)=a_n^T(X_n,\dots,X_1)^T.
$$

由平稳性

$$
L(X_t|X_{t-1},\dots,X_{t-n})=a_n^T(X_{t-1},\dots,X_{t-n})^T.
$$

最小的线性预测方差为

$$\begin{aligned}\sigma_n^2 \triangleq & E(X_{n+1} - (a_{n1}X_n + \dots a_{nn}X_1))^2 \\ =& \mathrm{Var}(X_{n+1}) - \gamma_n^T \Gamma_n \gamma_n \\ =& \gamma_0 - \gamma_n^T a_n \\ =& \gamma_0 - a_{n1}\gamma_1 - \cdots - a_{nn}\gamma_n.\end{aligned}$$

由平稳性

$$E(X_t - (a_{n1}X_{t-1} + \dots a_{nn}X_{t-n}))^2 = \sigma_n^2.$$

10.2 最小相位性

如果 $\{\gamma_k\}$ 是某 AR(p) 序列的自协方差函数，则 p 阶的 Yuler-Walker 方程解出的 Yule-Walker 系数就是 AR 模型的自回归系数，所以满足如下的最小相位性：

$$A(z) = 1 - \sum_{j=1}^{p} a_j z^j \neq 0, \quad \text{对}|z| \leq 1$$

对于一般的平稳列有如下定理。

定理 10.1 (Y-W 系数的最小相位性). *如果实数列 $\gamma_k, k = 0, 1, \dots, n$ 使得*

$$\Gamma_{n+1} \triangleq \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_n \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-1} \\ \vdots & \vdots & & \vdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_0 \end{pmatrix} > 0$$

则解出的 n 阶 Yuler-Walker 系数 a_n 满足如下最小相位条件：

$$1 - \sum_{j=1}^{n} a_{nj} z^j \neq 0, \quad |z| \leq 1.$$

最小相位性就是以 a_n 为系数的 AR(p) 模型能表示成因果性线性平稳列的充分必要条件。

一般的线性平稳列的自协方差列正定，所以其任意 n 阶 Yuler-Walker 系数都满足最小相位条件。

10.3 Levinson 递推公式

定理 10.2 (Levinson 递推). 如果 Γ_{n+1} 正定，则 $\gamma_k, k=0,1,\dots,n$ 的 $1,2,\dots,n,n+1$ 阶 *Yuler-Walker* 系数 $\{a_{ij}, i=1,\dots,n+1, j=1,\dots,i\}$ 和均方误差 σ_k^2 可以如下递推计算：

$$\sigma_0^2 = \gamma_0 \tag{10.3}$$

$$a_{1,1} = \gamma_1/\gamma_0 \tag{10.4}$$

$$\sigma_k^2 = \sigma_{k-1}^2(1-a_{k,k}^2) \tag{10.5}$$

$$a_{k+1,k+1} = \frac{\gamma_{k+1} - a_{k,1}\gamma_k - a_{k,2}\gamma_{k-1} - \cdots - a_{k,k}\gamma_1}{\gamma_0 - a_{k,1}\gamma_1 - a_{k,2}\gamma_2 - \cdots - a_{k,k}\gamma_k} \tag{10.6}$$

$$a_{k+1,j} = a_{k,j} - a_{k+1,k+1}a_{k,k+1-j}, \quad 1 \le j \le k \tag{10.7}$$

其中

$$\sigma_k^2 = E(X_{k+1} - (a_{k,1}X_k + \cdots + a_{k,k}X_1))^2 \tag{10.8}$$

是用 $X_k, X_{k-1}, \dots, X_1$ 预测 X_{k+1} 的均方误差。

10.3.1 Levinson 公式的记忆方法

回忆 §9.2中的(9.10)和(9.12)

$$\gamma_k - \sum_{j=1}^p a_j \gamma_{k-j} = 0, \quad k \ge 1, \tag{10.9}$$

$$\sigma^2 = \gamma_0 - \sum_{j=1}^p a_j \gamma_{0-j}. \tag{10.10}$$

在(10.9)中将 k 替换成 $k+1$ 得

$$\gamma_{k+1} - \sum_{j=1}^p a_j \gamma_{k+1-j} = 0.$$

在 Levinson 递推的 $a_{k+1,k+1}$ 递推公式中可以将分子看成上式左边 $a_j = a_{k,j}$, $p=k$ 的情形，将分母看成是(10.10)中 $p=k$, $a_j = a_{k,j}$ 的情形。

$a_{k+1,j}, 1 \le j \le k$ 的公式可以写成如下的矩阵形式

$$\begin{pmatrix} a_{k+1,1} \\ \vdots \\ a_{k+1,k} \end{pmatrix} = \begin{pmatrix} a_{k,1} \\ \vdots \\ a_{k,k} \end{pmatrix} - a_{k+1,k+1} \begin{pmatrix} a_{k,k} \\ \vdots \\ a_{k,1} \end{pmatrix}.$$

关于 σ_k^2，记 $X_k = (X_k, X_{k-1}, \dots, X_1)^T$，有:

$$
\begin{aligned}
\sigma_k^2 =& E[X_{k+1} - a_k^T X_k]^2 \\
=& E[(X_{k+1} - a_k^T X_k) X_{k+1}] - E[(X_{k+1} - a_k^T X_k) a_k^T X_k] \\
=& \gamma_0 - a_k^T (\gamma_1 \ \dots \gamma_k)^T - 0 \quad (\text{根据 Y-W 方程}) \\
=& \gamma_0 - a_{k,1}\gamma_1 - \dots - a_{k,k}\gamma_k
\end{aligned}
$$

这是 $a_{k+1,k+1}$ 的递推公式的分母，所以 $a_{k+1,k+1}$ 的递推公式也可以写成

$$
a_{k+1,k+1} = \frac{\gamma_{k+1} - a_{k,1}\gamma_k - a_{k,2}\gamma_{k-1} - \dots - a_{k,k}\gamma_1}{\sigma_k^2}
$$

注意 σ_k^2 是用 k 个历史值预报第 $k+1$ 个的最小均方误差线性预测的均方误差。

10.3.2 Levinson 递推的计算顺序

用 Levinson 递推公式计算各阶 Yuler-Walker 系数和

$$
\sigma_k^2 = E[X_{k+1} - a_{k,1} X_k - a_{k,2} X_{k-1} - \dots - a_{k,k} X_1]^2
$$

次序应为

- 初值 (不用历史资料预报 X_1)：

$$
\sigma_0^2 = E[X_1 - 0]^2 = \gamma_0
$$

- $k+1 = 1$(用 X_1 预报 X_2):

$$
\begin{aligned}
a_{1,1} =& \gamma_1/\gamma_0 \\
\sigma_1^2 =& E[X_2 - a_{1,1}X_1]^2 \\
=& \sigma_0^2(1 - a_{1,1}^2)
\end{aligned}
$$

- $k+1 = 2$(用 X_1, X_2 预报 X_3):

$$
\begin{aligned}
a_{2,2} =& \frac{\gamma_2 - a_{1,1}\gamma_1}{\sigma_1^2} \\
a_{2,1} =& a_{1,1} - a_{2,2}a_{1,1} \\
\sigma_2^2 =& E[X_3 - a_{2,1}X_2 - X_{2,2}X_1]^2 \\
=& \sigma_1^2(1 - a_{2,2}^2)
\end{aligned}
$$

- $k+1=3$(用 X_1, X_2, X_3 预报 X_4):

$$\begin{aligned}
a_{3,3} =& \frac{\gamma_3 - a_{2,1}\gamma_2 - a_{2,2}\gamma_1}{\sigma_2^2} \\
a_{3,1} =& a_{2,1} - a_{3,3}a_{2,2} \\
a_{3,2} =& a_{2,2} - a_{3,3}a_{2,1} \\
\sigma_3^2 =& E[X_4 - a_{3,1}X_3 - a_{3,2}X_2 - a_{3,3}X_3]^2 \\
=& \sigma_2^2(1 - a_{3,3}^2)
\end{aligned}$$

- ………………

计算次序为

k	$a_{k,j}$					σ_k^2
0						σ_0^2
1	$a_{1,1}$					σ_1^2
2	$a_{2,2}$	$a_{2,1}$				σ_2^2
3	$a_{3,3}$	$a_{3,1}$	$a_{3,2}$			σ_3^2
4	$a_{4,4}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$		σ_4^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$

10.4 偏相关系数

定义 10.1 (偏相关系数). 如果 Γ_n 正定，称 $a_{n,n}$ 为 $\{X_t\}$ 或 $\{\gamma_k\}$ 的 n 阶偏 (自) 相关系数。序列 $\{a_{n,n}, n=1,2,\dots\}$ 称为 $\{X_t\}$ 或 $\{\gamma_k\}$ 的 n 阶偏 (自) 相关函数。

偏自相关是 X_1 和 X_{n+1} 之间的如下意义下的**偏相关系数**:

$$\begin{aligned}
a_{n,n} = \mathrm{Corr}[&X_1 - L(X_1|X_2,\dots,X_n), \\
&X_{n+1} - L(X_{n+1}|X_2,\dots,X_n)]
\end{aligned}$$

即 $a_{n,n}$ 为 X_1 和 X_{n+1} 扣除 $X_2,\dots,X_n$ 的线性影响后的相关系数。

设 $\{X_t\}$ 是 AR(p) 序列。其自协方差函数正定。由 Yule-Walker 方程(9.9)知其 n 阶 $(n \ge p)$Y-W 系数为

$$\begin{aligned}
a_n =& (a_1,\dots,a_p,0,\dots,0)^T && (10.11)\\
=& (a_{n,1},a_{n,2},\dots,a_{n,n})^T, \quad n \ge p && (10.12)
\end{aligned}$$

其偏相关系数满足

$$a_{n,n} = \begin{cases} a_p \neq 0, & n = p \\ 0, & n > p \end{cases} \tag{10.13}$$

称此性质为 AR(p) 序列的相关系数 p 后截尾。

反之，如果一个零均值平稳列偏相关系数 p 后截尾，则它必是 AR(p) 序列（见下面的定理）。

偏相关截尾条件隐含要求自协方差列正定。

定理 10.3 (AR 序列的偏相关函数条件)**.** *设零均值平稳列 $\{X_t\}$ 的自协方差函数 $\{\gamma_k\}$ 是正定序列，则 $\{X_t\}$ 是 AR(p) 序列的充分必要条件是，它的偏相关系数 $\{a_{n,n}\}$ p 后截尾。*

证明：

只要证明充分性。记 $(a_{p,1}, \dots, a_{p,p}) = (a_1, \dots, a_p)$，令 $\varepsilon_t = X_t - \sum_{j=1}^p a_j X_{t-j}$，只要证明 $\{\varepsilon_t\}$ 是白噪声。最小相位性由定理10.1给出。

记 $a_p = (a_{p,1}, \dots, a_{p,p}) = (a_1, \dots, a_p)$,由 Levinson 公式和 $a_{p+k,p+k} = 0 (k > 0)$ 得

$$\begin{aligned} a_{p+1,j} =& a_{p,j} - a_{p+1,p+1} a_{p,p+1-j} = a_j, && 1 \leq j \leq p \\ a_{p+k,j} =& a_{p+k-1,j} = \cdots = a_{p,j} = a_j, && k \geq 2, 1 \leq j \leq p \\ a_{p+k,j} =& a_{p+k-1,j} = 0 && p < j \leq p+k \end{aligned}$$

即 $n \geq p$ 时

$$a_n = (a_{n,1}, a_{n,2}, \dots, a_{n,n}) = (a_1, a_2, \dots, a_p, 0, \dots, 0)$$

注意 a_n 是 Y-W 方程的解，即

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}$$

可写成

$$\begin{aligned}\gamma_k =& a_1\gamma_{k-1} + a_2\gamma_{k-2} + \cdots + a_p\gamma_{k-p} \\ =& \sum_{j=1}^{p} a_j\gamma_{k-j}, \quad k \ge 1\end{aligned} \tag{*}$$

由定理10.1知 $A(z) = 1 - \sum_{j=1}^{p} a_j z^j$ 满足最小相位条件。

令

$$\varepsilon_t = X_t - \sum_{j=1}^{p} a_j X_{t-j}, \quad t \in \mathbb{Z}$$

则 $\{\varepsilon_t\}$ 是平稳序列，满足 $E\varepsilon_t = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma_p^2 > 0$ (因为 $\{\gamma_k\}$ 为正定序列所以 $\{X_t\}$ 不是可完全线性预测的)。

下面只要证明 $\{\varepsilon_t\}$ 是白噪声。$\forall t > s$ 有

$$\begin{aligned}E(\varepsilon_t X_s) =& E\left[\left(X_t - \sum_{j=1}^{p} a_j X_{t-j}\right) X_s\right] \\ =& \gamma_{t-s} - \sum_{j=1}^{p} a_j \gamma_{t-s-j} \\ =& 0 \qquad (\text{由 } (*))\end{aligned}$$

所以 $t > s$ 时

$$E(\varepsilon_t \varepsilon_s) = E\left[\varepsilon_t \left(X_s - \sum_{j=1}^{p} a_j X_{s-j}\right)\right] = 0$$

即 $\{\varepsilon_t\}$ 是 $\mathrm{WN}(0, \sigma_p^2)$，且 $a_1, a_2, \dots, a_p$ 满足最小相位条件。证毕。

10.5 本节内容的应用意义

- 有了观测样本 $x_1, x_2, \dots, x_N$ 可以估计样本自协方差函数:

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

- 有了 $\{\hat{\gamma}_k\}$ 可以计算各阶偏相关系数的估计 $\{a_{k,k}\}$。
- 如果发现样本偏相关系数呈现截尾性则可以拟合 AR 模型。

- 定理10.1保证当 $\hat{\Gamma}_{p+1}$ 正定时得到的模型系数满足最小相位条件。
- 最小相位条件保证系统是稳定的，预测有意义。
- 真实模型为 AR(p) 时 $\hat{\Gamma}_{p+1}$ a.s. 正定。

10.6 附录：最优线性预测的 Hilbert 空间投影意义

设 $X_1, X_2, \dots, X_n, Y$ 为随机变量。考虑估计问题

$$L(Y|X_1, \dots, X_n) \triangleq \underset{\hat{Y}=a_0+a_1X_1+\cdots+a_nX_n}{\operatorname{argmin}} E(Y-\hat{Y})^2$$

称 $L(Y|X_1, \dots, X_n)$ 为 Y 关于 $X_1, \dots, X_n$ 的**最优线性估计**。

上述最优线性预测问题等价于在 L^2 的闭子空间 $\text{sp}(1, X_1, \dots, X_n)$ 求一个与 Y 距离最近的元素，根据 Hilbert 空间投影的性质可知，$L(Y|X_1, \dots, X_n)$ 是 Y 在 $\text{sp}(1, X_1, \dots, X_n)$ 的投影。$L(Y|X_1, \dots, X_n)$ 满足的条件是

$$Y - L(Y|X_1, \dots, X_n) \perp \text{sp}(1, X_1, \dots, X_n)$$

设 $L(Y|X_1, \dots, X_n) = a_0 + a_1X_1 + \cdots + a_nX_n$，则

$$Y - a_0 + a_1X_1 + \cdots + a_nX_n \perp 1, X_1, \dots, X_n$$

也既是

$$\begin{aligned}
&E(Y - a_0 + a_1X_1 + \cdots + a_nX_n) = 0 \\
&EX_j(Y - a_0 + a_1X_1 + \cdots + a_nX_n) = 0, j = 1, \dots, n
\end{aligned}$$

记 $a = (a_1, \dots, a_n)^T$，$X = (X_1, \dots, X_n)^T$，$\Sigma_{XX} = \text{Var}(X)$，$\Sigma_{XY} = \text{Cov}(X, Y)$。估计问题可以写成

$$L(Y|X) \triangleq \underset{\hat{Y}=a_0+a^TX}{\operatorname{argmin}} E(Y-\hat{Y})^2$$

a_0 和 a 的充分必要条件是

$$\begin{aligned}
&a_0 = EY - a^TEX \\
&E[(Y - a_0 - a^TX)X^T] = 0
\end{aligned}$$

将第一式代入到第二式中，得

$$E[(Y - EY - a^T(X - EX))X^T] = 0$$

即

$$\begin{aligned}\mathrm{Cov}(Y,X) =& a^T\mathrm{Var}(X)\\ \mathrm{Var}(X)a =& \mathrm{Cov}(Y,X)\\ \Sigma_{XX}a =& \Sigma_{XY}\end{aligned}$$

由投影的存在性知 a 必有解，且 $\Sigma_{XX}>0$ 时

$$a=\Sigma_{XX}^{-1}\Sigma_{XY}$$

当 $|\Sigma_{XX}|=0$ 时 a 有无穷多解，但是得到的最佳线性预测都是同一个。

10.7 附录：最小相位性定理证明

来证明定理10.1的 Y-W 系数最小相位性。

用 I_n 表示 n 阶单位阵，用 0_n 表示元素都是 0 的 n 阶方阵。由平稳性，$(X_m,X_{m-1},\dots,X_{m-n})$ 与 $(X_{n+1},X_n,\dots,X_1)$ 有相同的协方差阵 Γ_{n+1}，所以用 $X_{m-1},\dots,X_{m-n}$ 对 X_m 作最优线性预测的公式为

$$\hat X_m \triangleq L(X_m|X_{m-1},\dots,X_{m-n})=\sum_{j=1}^n a_{nj}X_{m-j} \tag{10.14}$$

定义

$$V_m=X_m-\hat X_m=X_m-\sum_{j=1}^n a_{nj}X_{m-j} \tag{10.15}$$

则由 $\Gamma_{n+1}>0$ 可知 $EV_m^2\triangleq\sigma_n^2>0$。

由最优线性预测的性质（或者 L^2 中投影算子的性质）可知

$$E(V_mX_{m-j})=0,j=1,2,\dots,n$$

引入

$$Y_m=\begin{pmatrix}X_m\\X_{m-1}\\\vdots\\X_{m-n+1}\end{pmatrix}\quad V_m=\begin{pmatrix}V_m\\0\\\vdots\\0\end{pmatrix}$$

$$A=\begin{pmatrix}a_{n1}\ a_{n2}\ \cdots a_{n,n-1} & a_{nn}\\ I_{n-1} & 0\end{pmatrix}$$

则有

$$Y_m - AY_{m-1} = V_m \tag{10.16}$$

这称为 AR 模型的马氏扩张。于是有

$$\begin{aligned}
\begin{pmatrix} \sigma_n^2 & 0 \\ 0 & 0_{n-1} \end{pmatrix} &= E(V_m V_m^T) \\
=& E[V_m(Y_m - AY_{m-1})^T] \\
=& E\left[V_m Y_m^T\right] - E[V_m Y_{m-1}^T]A^T \\
=& E\left[V_m Y_m^T\right] \quad (\text{利用} E(V_m X_{m-j} = 0)) \\
=& E\left[(Y_m - AY_{m-1})Y_m^T\right] \\
=& E(Y_m Y_m^T) - AE(Y_{m-1}Y_m^T) \\
=& \Gamma_n - AE\left[Y_{m-1}(AY_{m-1} + V_m)^T\right] \\
=& \Gamma_n - AE(Y_{m-1}Y_{m-1}^T)A^T - AE(Y_{m-1}V_m^T) \\
=& \Gamma_n - A\Gamma_n A^T \quad (\text{再次利用} E(V_m X_{m-j} = 0))
\end{aligned} \tag{10.17}$$

设复数 z_0 使得 $\det(I_n - z_0 A) = 0$，方程即

$$\det\begin{pmatrix}
1 - a_{n1}z_0 & -a_{n2}z_0 & -a_{n3}z_0 & \cdots & -a_{n,n-1}z_0 & -a_{nn}z_0 \\
-z_0 & 1 & 0 & \cdots & 0 & 0 \\
0 & -z_0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & -z_0 & 1
\end{pmatrix}$$

$$=1 - a_{n1}z_0 - a_{n2}z_0^2 - \cdots - a_{nn}z_0^n = 0$$

当 $z_0 = 0$ 时行列式等于 1，所以如果 z_0 是方程的根则 $z_0 \neq 0$。要证明最小相位性，只要证明所有使得上面方程成立的 z_0 都满足 $|z_0| > 1$。

因为行列式 $\det(I_n - z_0 A) = 0$，必存在非零复向量 $\alpha^* = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ 使得

$$\alpha^*(I_n - z_0 A) = 0$$

即

$$\alpha^* A = z_0^{-1}\alpha^*$$

由矩阵 A 的结构，这可以写成

$$\begin{cases} a_{n1}\alpha_1 + \alpha_2 = z_0^{-1}\alpha_1 \\ a_{n2}\alpha_1 + \alpha_3 = z_0^{-1}\alpha_2 \\ \vdots \\ a_{n,n-1}\alpha_1 + \alpha_n = z_0^{-1}\alpha_{n-1} \\ a_{nn}\alpha_1 = z_0^{-1}\alpha_n \end{cases} \tag{10.18}$$

由此可知 $\alpha_1 \neq 0$，如果不然，则由(10.18) 递推可得 $0 = \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$，这与 α^* 非零矛盾。

利用 $\sigma_n^2 = EV_m^2 > 0$ 和(10.17) 及 $\alpha^* A = z_0^{-1}\alpha^*$ 有

$$\begin{aligned} 0 <& \alpha_1 \sigma_n^2 \alpha_1^* \\ =& \alpha^* \begin{pmatrix} \sigma_n^2 & 0 \\ 0 & 0_{n-1} \end{pmatrix} \alpha \\ =& \alpha^* \Gamma_n \alpha - \alpha^* A \Gamma_n A^T \alpha \\ =& \alpha^* \Gamma_n \alpha - |z_0|^{-2} \alpha^* \Gamma_n \alpha \\ =& (1 - |z_0|^{-2}) \alpha^* \Gamma_n \alpha \end{aligned}$$

由 Γ_n 的正定性知 $\alpha^*\Gamma_n\alpha > 0$，所以 $1 - |z_0|^{-2} > 0$，即有 $|z_0| > 1$。定理证毕。

10.8 附录：AR 序列的等价定义

定理 10.4 (AR 序列的等价定义)**.** *设 $a_1, \ldots, a_p$ 为实数，$a_p \neq 0$，$A(z) = 1 - a_1 z - \ldots a_p z^p$，$\{\varepsilon_t\}$ 为 WN(0, σ^2)，零均值平稳列 $\{X_t, t \in \mathbb{Z}\}$ 满足*

$$A(\mathscr{B})X_t = \varepsilon_t,\ t \in \mathbb{Z}$$

如果 $\varepsilon_{t+k}, k = 1, 2, \ldots$ 与 $\{X_s, s \leq t\}$ 互不相关，则 $A(z)$ 的根都在单位圆外，从而 $\{X_t\}$ 为 AR(p) 序列。

证明 $L(X_t|X_{t-1}, \ldots, X_{t-p})$ 即 X_t 到 $\mathrm{sp}(X_{t-1}, \ldots, X_{t-p})$ 的投影，由投影的线性性质及 ε_t 与 $X_{t-1}, \ldots, X_{t-p}$ 正交可得

$$L(X_t|X_{t-1}, \ldots, X_{t-p}) = a_1 X_{t-1} + \cdots + a_p X_{t-p} + 0$$

可知用 $X_{t-1},\dots,X_{t-p}$ 预测 X_t 的 Y-W 系数为 $(a_1,\dots,a_p)$，又

$$E(X_t - L(X_t|X_{t-1},\dots,X_{t-p}))^2 = E\varepsilon_t^2 = \sigma^2 > 0$$

所以 $X_t, X_{t-1},\dots,X_{t-p}$ 线性无关，有 $\Gamma_{p+1} > 0$。由定理10.1可知 $A(z)$ 满足最小相位性，因此 $\{X_t\}$ 是 AR(p) 序列。

10.9 Levinson 递推公式证明

来证明定理10.2。

$a_{1,1} = \gamma_1/\gamma_0$ 显然，设对 $k(k \le n-1)$ 已知

$$\begin{aligned}
a_{k+1,k+1} =& \frac{\gamma_{k+1} - a_{k,1}\gamma_k - a_{k,2}\gamma_{k-1} - \cdots - a_{k,k}\gamma_1}{\gamma_0 - a_{k,1}\gamma_1 - a_{k,2}\gamma_2 - \cdots - a_{k,k}\gamma_k} \\
a_{k+1,j} =& a_{k,j} - a_{k+1,k+1}a_{k,k+1-j}, \quad 1 \le j \le k
\end{aligned}$$

引入正交阵

$$T = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{pmatrix},$$

则 TA 的结果是将 A 的各行颠倒次序，AT 的结果是将 A 的各列颠倒次序，且 $T^T = T^{-1} = T$。由 Γ_k 的结构易见

$$T\Gamma_k T = \Gamma_k, \quad g_k = T\gamma_k = (\gamma_k,\dots,\gamma_1)^T.$$

再引入

$$\begin{aligned}
\Gamma_{k+1} =& \begin{pmatrix} \Gamma_k & g_k \\ g_k^T & \gamma_0 \end{pmatrix} = \begin{pmatrix} \Gamma_k & T\gamma_k \\ \gamma_k^T T & \gamma_0 \end{pmatrix}, \\
a_{k+1}^{[1:k]} =& (a_{k+1,1},\dots,a_{k+1,k})^T, \\
a_{k+1} =& \begin{pmatrix} a_{k+1}^{[1:k]} \\ a_{k+1,k+1} \end{pmatrix}, \\
\gamma_{k+1} =& \begin{pmatrix} \gamma_k \\ \gamma_{k+1} \end{pmatrix}.
\end{aligned}$$

利用上述记号将 $k+1$ 阶 Y-W 方程写成

$$\begin{pmatrix} \Gamma_k & g_k \\ g_k^T & \gamma_0 \end{pmatrix} \begin{pmatrix} a_{k+1}^{[1:k]} \\ a_{k+1,k+1} \end{pmatrix} = \begin{pmatrix} \gamma_k \\ \gamma_{k+1} \end{pmatrix}.$$

分别写成

$$\Gamma_k a_{k+1}^{[1:k]} + a_{k+1,k+1} g_k = \gamma_k, \tag{10.19}$$
$$g_k^T a_{k+1}^{[1:k]} + a_{k+1,k+1} \gamma_0 = \gamma_{k+1}. \tag{10.20}$$

由(10.19)得

$$\begin{aligned} a_{k+1}^{[1:k]} =& \Gamma_k^{-1} \gamma_k - a_{k+1,k+1} \Gamma_k^{-1} g_k & (10.21) \\ =& \Gamma_k^{-1} \gamma_k - a_{k+1,k+1} \Gamma_k^{-1} T \gamma_k & (10.22) \\ =& a_k - a_{k+1,k+1} (T \Gamma_k T)^{-1} T \gamma_k & (10.23) \\ =& a_k - a_{k+1,k+1} T \Gamma_k^{-1} \gamma_k & (10.24) \\ =& a_k - a_{k+1,k+1} T a_k. & (10.25) \end{aligned}$$

将(10.25)代入(10.20)，得

$$g_k^T (a_k - a_{k+1,k+1} T a_k) + a_{k+1,k+1} \gamma_0 = \gamma_{k+1},$$

即

$$g_k^T a_k - a_{k+1,k+1} g_k^T T a_k + a_{k+1,k+1} \gamma_0 = \gamma_{k+1},$$

化简得

$$\begin{aligned} a_{k+1,k+1}(\gamma_0 - a_k^T \gamma_k) =& \gamma_{k+1} - a_k^T g_k, \\ a_{k+1,k+1} =& \frac{\gamma_{k+1} - a_k^T g_k}{\gamma_0 - a_k^T \gamma_k} \\ =& \frac{\gamma_{k+1} - a_{k+1,1} \gamma_k - \cdots - a_{k+1,k} \gamma_1}{\gamma_0 - a_{k+1,1} \gamma_1 - \cdots - a_{k+1,k} \gamma_k}. \end{aligned}$$

这就证明了 $a_{k+1,k+1}$ 的公式，而(10.25)就是 $a_{k+1,j}$, $1 \leq j \leq k$ 的公式。

再来证明 σ_k 的公式。$\sigma_0^2 = \gamma_0$ 显然，记

$$X_k = (X_k, X_{k-1}, \ldots, X_1)^T,$$

则按定义有

$$
\begin{aligned}
\sigma_k^2 =& E[X_{k+1} - a_k^T X_k]^2 \\
=& E[(X_{k+1} - a_k^T X_k) X_{k+1}] - E[(X_{k+1} - a_k^T X_k) X_k^T a_k] \\
=& \gamma_0 - a_k^T \gamma_k - (\gamma_k^T - a_k^T \Gamma_k) a_k \\
=& \gamma_0 - a_k^T \gamma_k - 0 a_k \\
=& \gamma_0 - a_k^T \gamma_k.
\end{aligned}
$$

于是关于 σ_{k+1}^2 有

$$
\begin{aligned}
\sigma_{k+1}^2 =& \gamma_0 - a_{k+1} \gamma_{k+1} \\
=& \gamma_0 - ([a_{k+1}^{[1:k]}]^T, a_{k+1,k+1}) \begin{pmatrix} \gamma_k \\ \gamma_{k+1} \end{pmatrix} \\
=& \gamma_0 - [a_{k+1}^{[1:k]}]^T \gamma_k - a_{k+1,k+1} \gamma_{k+1} \\
=& \gamma_0 - a_k^T \gamma_k + a_{k+1,k+1} a_k^T T \gamma_k - a_{k+1,k+1} \gamma_{k+1} \\
=& \sigma_k^2 - a_{k+1,k+1} (\gamma_{k+1} - a_k^T g_k) \\
=& \sigma_k^2 - a_{k+1,k+1} [a_{k+1,k+1} (\gamma_0 - a_k^T g_k)] \\
=& \sigma_k^2 - a_{k+1,k+1}^2 \sigma_k^2 = \sigma_k^2 (1 - a_{k+1,k+1}^2).
\end{aligned}
$$

得证。

10.10 附录：离散谱序列的预测

考虑离散谱序列

$$X_t = A\cos(\omega t) + B\sin(\omega t), t \in \mathbb{Z},$$

其中 $0 < \omega < \pi$, A, B 是互不相关的零均值随机变量，$\mathrm{Var}(A) = \mathrm{Var}(B) = \sigma^2$。这是零均值平稳列，自协方差函数为

$$\gamma_k = \sigma^2 \cos(k\omega), k = 0, 1, 2, ...$$

考虑最优线性预测问题。

10.10.1 直接求解 Y-W 方程

为了用 X_1 预测 X_2，只要求解

$$\gamma_0 a_{11} = \gamma_1$$

即可得 $a_{11} = \cos\omega$，

$$L(X_2|X_1) = a_{11}X_1 = \cos\omega \cdot X_1$$

预测的均方误差为

$$\begin{aligned}
\sigma_2^2 =& E(X_2 - \cos\omega \cdot X_1)^2 \\
=& \sigma^2(\cos 2\omega - \cos\omega\cos\omega)^2 + \sigma^2(\sin 2\omega - \cos\omega\sin\omega)^2 \\
=& \sigma^2(\sin^4\omega + \cos^2\omega\sin^2\omega) \\
=& \sigma^2\sin^2\omega > 0
\end{aligned}$$

考虑用 X_2, X_1 预测 X_3 的问题。

$$\Gamma_2 = \begin{pmatrix} 1 & \cos\omega \\ \cos\omega & 1 \end{pmatrix}$$

$|\Gamma_2| = 1 - \cos^2\omega = \sin^2\omega > 0$，所以 $\Gamma_2 > 0$。求解方程

$$\begin{pmatrix} 1 & \cos\omega \\ \cos\omega & 1 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} \cos\omega \\ \cos 2\omega \end{pmatrix}$$

即

$$\begin{aligned}
& a_{21} + \cos\omega \cdot a_{22} = \cos\omega \\
& \cos\omega \cdot a_{21} + a_{22} = \cos 2\omega \\
& a_{22} = \cos 2\omega - \cos\omega \cdot a_{21} \\
& a_{21} + \cos\omega(\cos 2\omega - \cos\omega \cdot a_{21}) = \cos\omega \\
& a_{21} + \cos\omega\cos 2\omega - \cos^2\omega \cdot a_{21} = \cos\omega \\
& \sin^2\omega \cdot a_{21} = \cos\omega(1 - \cos 2\omega) = 2\cos\omega\sin^2\omega \\
& a_{21} = 2\cos\omega \\
& a_{22} = \cos 2\omega - \cos\omega \cdot a_{21} = \cos 2\omega - \cos\omega \cdot 2\cos\omega = -1
\end{aligned}$$

其中用到三角函数公式 $\cos 2\omega = 2\cos^2\omega - 1 = 1 - 2\sin^2\omega$。

于是 X_3 的最优线性预测为

$$\hat{X}_3 = a_{21}X_2 + a_{22}X_1 = 2\cos\omega \cdot X_2 - X_1$$

预测的均方误差为

$$
\begin{aligned}
\sigma_2^2 =& E(X_3 - 2\cos\omega \cdot X_2 + X_1)^2 \\
=& \sigma^2 \left(\cos 3\omega - 2\cos\omega\cos 2\omega + \cos\omega\right)^2 \\
& + \sigma^2 \left(\sin 3\omega - 2\cos\omega\sin 2\omega + \sin\omega\right)^2 \\
=& 0
\end{aligned}
$$

所以离散谱序列可完全线性预测。对任意 t，为了用 $X_{t-1}, \dots, X_{t-p}$ 预测 X_t，当 $p \geq 2$ 时有

$$
L(X_t | X_{t-1}, \dots, X_{t-p}) = 2\cos\omega \cdot X_{t-1} - X_{t-2}
$$

且 $L(X_t | X_{t-1}, \dots, X_{t-p}) = X_t$，预测误差为零。

事实上，Γ_3 为

$$
\begin{pmatrix}
1 & \cos\omega & \cos 2\omega \\
\cos\omega & 1 & \cos\omega \\
\cos 2\omega & \cos\omega & 1
\end{pmatrix}
$$

记 $b = \cos\omega$，则 $\cos 2\omega = 2b^2 - 1$，行列式为

$$
\begin{aligned}
& \begin{vmatrix}
1 & b & 2b^2-1 \\
b & 1 & b \\
2b^2-1 & b & 1
\end{vmatrix} \\
=& \begin{vmatrix}
1 & b & 2b^2-1 \\
0 & 1-b^2 & 2b-2b^2 \\
0 & 2b-2b^2 & -4b^4+4b^2
\end{vmatrix} \\
=& (1-b^2)^2 \begin{vmatrix}
1 & 2b \\
2b & 4b^2
\end{vmatrix} \\
=& 0
\end{aligned}
$$

10.10.2 用 Levinson 递推求解 Y-W 方程

$\gamma_0 = \sigma^2$, $\gamma_1 = \sigma^2 \cos\omega$, $\gamma_2 = \sigma^2 \cos 2\omega$。

按照 Levinson 递推公式的递推次序依次计算：

$$
\begin{aligned}
&\text{初值：}\\
\sigma_0^2 =&\gamma_0 = \sigma^2\\
k+1=1\text{：}&\\
a_{11} =&\frac{\gamma_1}{\gamma_0} = \cos\omega\\
\sigma_1^2 =&E[X_2 - L(X_2|X_1)]^2\\
=&\sigma_0^2[1-a_{11}^2] = \sigma^2\sin^2\omega\\
k+1=2\text{：}&\\
a_{22} =&\frac{\gamma_2 - a_{11}\gamma_1}{\sigma_1^2}\\
=&\frac{\cos 2\omega - \cos\omega\cos\omega}{\sin^2\omega}\\
=&-1\\
a_{21} =&a_{11} - a_{22}a_{11}\\
=&\cos\omega[1-(-1)] = 2\cos\omega\\
\sigma_2^2 =&E[X_3 - L(X_3|X_2, X_1)]^2\\
=&\sigma_1^2[1-a_{22}^2]\\
=&\sigma^2\sin^2\omega[1-(-1)^2]\\
=&0
\end{aligned}
$$

10.11 附录：Y-W 方程反解讨论

Y-W 方程中如果 $\Gamma_{p+1} > 0$ 则 $a_1, \dots, a_p, \sigma^2$ 唯一确定, 其中 $a_1, \dots, a_p$ 满足最小相位条件（见定理10.1），$\sigma^2 > 0$。

反过来，如果 $a_1, \dots, a_p$ 满足最小相位条件，$\sigma^2 > 0$, Y-W 方程中的 $\gamma_0, \gamma_1, \dots, \gamma_p$ 是否唯一确定？

参考：谢衷洁《时间序列分析》P.189 定理 4.4。定理说明，给定某平稳列的前 $p+1$ 个自协方差 $\gamma_0, \gamma_1, \dots, \gamma_p$, 必存在 AR($p$) 序列使其前 $p+1$ 个自协方差函数等于这 $p+1$ 个，模型参数由 Y-W 解出。见习题 6.1.2。另外，该参考书定理 4.5 说明在前 $p+1$ 个自协方差函数等于给定的这 $p+1$ 个的所有平稳列中，AR(p) 模型的一步预测误差达到最大，从而信息量最大。

更一般地，对非可完全线性预测平稳列 $\{X_t\}$ 的自协方差列 $\{\gamma_k\}$，有各阶 Y-W 方程：

$$\begin{aligned}\Gamma_n a_n =& \gamma_n \\ \gamma_0 - a_n^T \gamma_n =& \sigma_n^2\end{aligned}$$

其中 $\gamma_n = (\gamma_1, \dots, \gamma_n)^T$。假设 $a_n = (a_{n1}, a_{n2}, \dots, a_{nn})^T$ 和 $\sigma_n^2 > 0$ 给定，a_n 满足最小相位条件，则满足上述 Y-W 方程的 $\gamma_0, \gamma_1, \dots, \gamma_n$ 是否唯一确定？

对 $n = 1$，显然

$$\begin{aligned}\gamma_0 =& \frac{\sigma^2}{1 - a_1^2} \\ \gamma_1 =& a_1 \gamma_0\end{aligned}$$

唯一。

对 $n = 2$，方程为

$$\begin{aligned}\gamma_0 a_1 + \gamma_1 a_2 =& \gamma_1 \\ \gamma_1 a_1 + \gamma_0 a_2 =& \gamma_2 \\ \gamma_0 - a_1 a_1 \gamma_1 - a_2 \gamma_2 =& \sigma^2\end{aligned}$$

消元得

$$\begin{aligned}\gamma_0 =& \sigma^2 / \left(1 - a_2^2 - \frac{(1 + a_2) a_1^2}{1 - a_2}\right) \\ \gamma_1 =& \frac{a_1}{1 - a_2} \gamma_0 \\ \gamma_2 =& a_1 \gamma_1 + a_2 \gamma_0\end{aligned}$$

对 $n = 3$，因为 $\gamma_0 > 0$，$\gamma_k = \rho_k \gamma_0$，所以如果能由 a_1, a_2, a_3 决定 ρ_1, ρ_2, ρ_3 则可由

$$\sigma_3^2 = \gamma_0 - a_1 \gamma_1 - a_2 \gamma_2 - -a_3 \gamma_3 = \gamma_0 (1 - a_1 \rho_1 - a_2 \rho_2 - a_3 \rho_3)$$

解出 γ_0。把

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_2 & \gamma_1 & \gamma_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}$$

两边除以 γ_0 并写成关于 ρ_1, ρ_2, ρ_3 的方程，得

$$\begin{pmatrix} a_2 - 1 & 0 & a_3 \\ a_1 + a_3 & -1 & 0 \\ a_2 & a_1 & -1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} = - \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

很难判别此三元一次方程组的系数矩阵是否满秩。可以计算其行列式为

$$a_1^2 a_3 + a_1 a_3^2 + a_2 a_3 + a_2 - 1$$

这个矩阵可能有不满秩的情况，例如当

$$A(z) = 1 - 1.8z + 1.775789z^2 - 0.9z^3$$

时，此矩阵行列式为零，且 $A(z)$ 满足最小相位条件。

Chapter 11

AR 模型举例

11.1 AR(1) 模型

$|a| < 1$, AR(1) 模型

$$
\begin{aligned}
X_t =& a X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z} \\
\{\varepsilon_t\} \sim& \mathrm{WN}(0, \sigma^2)
\end{aligned}
$$

有平稳解

$$
X_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j}
$$

自协方差函数

$$
\begin{aligned}
\gamma_0 =& \sigma^2 \sum_{j=0}^{\infty} a^{2j} = \frac{\sigma^2}{1-a^2} \\
\gamma_k =& a \gamma_{k-1} = \cdots = a^k \gamma_0, \quad k \geq 1
\end{aligned} \tag{5.2}
$$

自相关系数

$$
\rho_k = \frac{\gamma_k}{\gamma_0} = a^k \tag{11.1}
$$

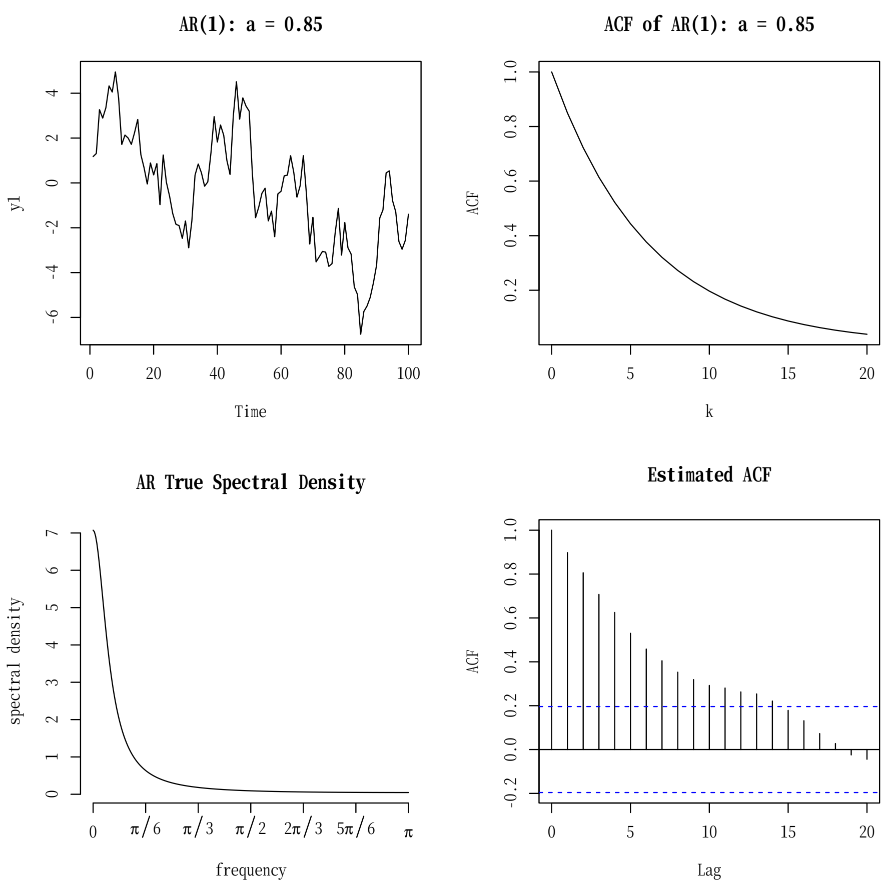
谱密度

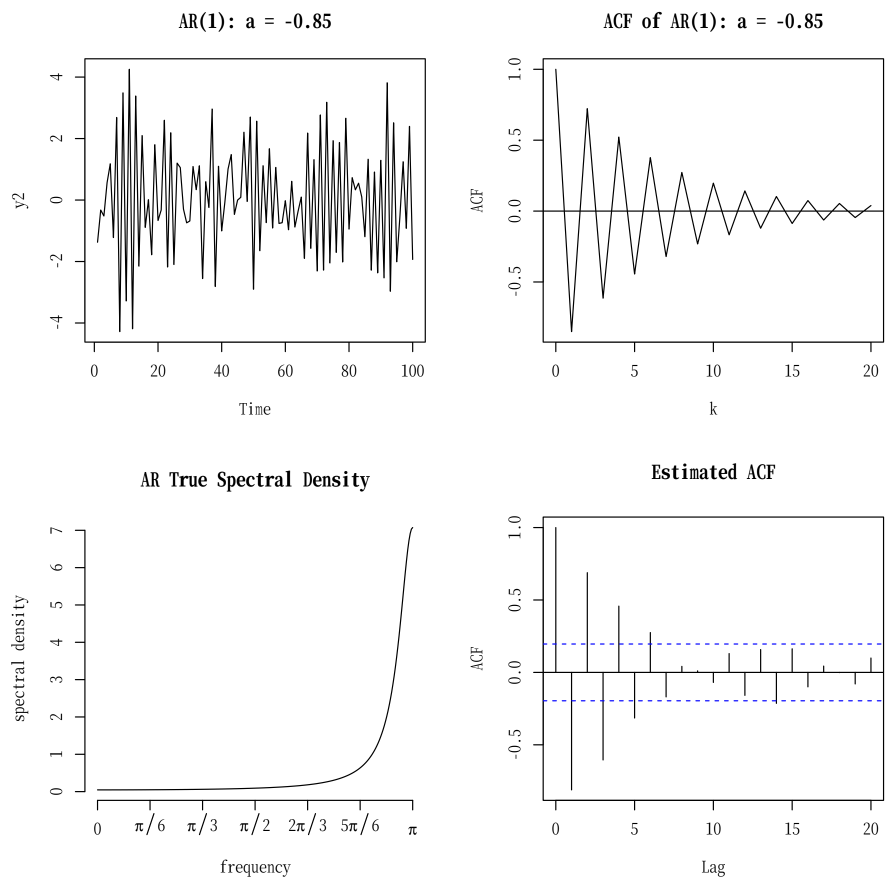
$$
\begin{aligned}
f(\lambda) =& \frac{\sigma^2}{2\pi} \frac{1}{|1 - a e^{i\lambda}|^2} && (11.2)\\
=& \frac{\sigma^2}{2\pi} [1 + a^2 - 2a\cos\lambda]^{-1}, \quad \lambda \in [-\pi, \pi] && (11.3)
\end{aligned}
$$

演示: $a = \pm 0.85$ 时序列的演示。

```
demo.ar1.example <- function(){
  oldpar <- par(mfrow=c(2,2)); on.exit(par(oldpar))
  n <- 100
  a1 <- 0.85

  y1 <- ar.gen(n, a1, sigma=1.0, plot.it=FALSE)
  y2 <- ar.gen(n, -a1, sigma=1.0, plot.it=FALSE)
  plot(y1, main="AR(1): a = 0.85")
  acf1 <- exp(seq(from=0, by=log(a1), length=21))
  plot(0:20, acf1, type="l",
       xlab="k", ylab="ACF",
       main="ACF of AR(1): a = 0.85")
  ar.true.spectrum(attr(y1, "a"))
  #plot(y1, type="n", axes=F, xlab="", ylab="")
  acf(y1, main="Estimated ACF")

  plot(y2, main="AR(1): a = -0.85")
  acf2 <- (-a1)^seq(from=0, length=21)
  plot(0:20, acf2, type="l",
       xlab="k", ylab="ACF",
       main="ACF of AR(1): a = -0.85")
  abline(h=0)
  ar.true.spectrum(attr(y2, "a"))
  #plot(y1, type="n", axes=F, xlab="", ylab="")
  acf(y2, main="Estimated ACF")
  invisible()
}
```

```r
demo.ar1.example()
```





比较:

表 11.1: AR(1) 正负系数的比较

	$a = 0.85$	$a = -0.85$
(1)	数据表现出趋势性, 相邻的数据差别不大;	数据上下摆动，趋势性不明显;
(2)	(1) 中的现象在 $\{\rho_k\}$ 得到体现: 相邻随机变量正相关	(1) 中的现象在 $\{\rho_k\}$ 得到体现: 相邻随机变量负相关
(3)	ρ_k 单调减少趋于 0	(3) ρ_k 正负交替趋于 0

	$a=0.85$	$a=-0.85$
(4)	谱密度的能量集中在低频, $f(\lambda)<f(0),\lambda\in(0,\pi]$, 数据无周期现象, 周期 $T=\frac{2\pi}{0}=\infty$	谱密度能量集中在高频, $f(\lambda)<f(\pi),\lambda\in[0,\pi)$, 数据有周期现象, 周期 $T=\frac{2\pi}{\pi}=2$
(5)	偏相关系数 $a_{1,1}=0.85$, $a_{k,k}=0$, 当 $k>1$	偏相关系数 $a_{1,1}=-0.85$, $a_{k,k}=0$, 当 $k>1$
(6)	随 a 接近于 0, 以更快的速度收敛到 0	上述性质随 a 接近 -1 变得更明显, 随 a 接近 0 变得不明显

11.2 AR(2) 模型

11.2.1 稳定域和允许域

AR(2) 模型为

$$
\begin{aligned}
X_t =& a_1X_{t-1}+a_2X_{t-2}+\varepsilon_t, \quad \varepsilon_t\sim \mathrm{WN}(0,\sigma^2)\\
A(z) =& 1-a_1z-a_2z^2\neq 0, |z|\leq 1
\end{aligned}
$$

稳定性条件为：

$$a_2\pm a_1<1, \quad |a_2|<1$$

稳定性条件的证明比较长，区分复根与实根两种情况；复根时，求复根的模大于 1 的条件，比较容易得到充要条件为

$$|a_1|<2,\ -1<a_2<0,\ a_2<-\frac{1}{4}a_1^2.$$

实根的情况，可以分 $a_2>0$ 和 $a_2<0$ 讨论，每种情况中继续分 $a_1\geq 0$ 和 $a_1<0$ 讨论，即分四种情况讨论。最终可以得到稳定性条件。

设 $A(z)$ 的根为 $z_1=b_1e^{i\lambda_1}$, $z_2=b_2e^{i\lambda_2}$。由 Y-W 方程

$$
\begin{aligned}
\rho_0 =& 1,\\
\rho_1 =& \frac{a_1}{1-a_2}\\
\rho_k =& a_1\rho_{k-1}+a_2\rho_{k-2}, \quad k\geq 1
\end{aligned}
$$

$a_{11}=\rho_1$, $a_{2,2}=a_2$, $a_{k,k}=0(k\geq 3)$。

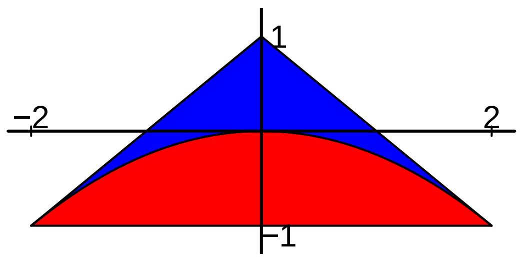


图 11.1: AR(2) 稳定性条件 (蓝色：实根；红色：复根)

$$\mathscr{A} = \{(a_1, a_2) : a_2 \pm a_1 < 1, |a_2| < 1\}$$

称为 AR(2) 的**稳定域**。

从 Y-W 方程可以用 ρ_1, ρ_2 表示 a_1, a_2:

$$a_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}, \quad a_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$$

反之有

$$\rho_1 = \frac{a_1}{1-a_2}, \quad \rho_2 = a_2 + \frac{a_1^2}{1-a_2}$$

$(a_1, a_2) \in \mathscr{A} \iff (\rho_1, \rho_2) \in$

$$\mathscr{C} = \{(\rho_1, \rho_2) : \rho_1^2 < (1+\rho_2)/2, |\rho_1| < 1, |\rho_2| < 1\}$$

$\mathscr{C}$ 称为 AR(2) 的**允许域**。

11.2.2 特征根与系数关系

特征根与系数有如下关系

$$1 - a_1 z - a_2 z^2 = (1 - \frac{z}{z_1})(1 - \frac{z}{z_2})$$

$$a_1 = \frac{1}{z_1} + \frac{1}{z_2}, \quad a_2 = -\frac{1}{z_1}\frac{1}{z_2}$$

$$z_1 z_2 = \frac{1}{(-a_2)}, \quad z_1 + z_2 = \frac{a_1}{(-a_2)}$$

谱密度为

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|a_2||e^{i\lambda} - b_1 e^{i\lambda_1}|^2 \cdot |e^{i\lambda} - b_2 e^{i\lambda_2}|^2}$$

11.2.3 AR(2) 实特征根时的表现

- 当 $a_1^2 + 4a_2 \geq 0$ 时 (蓝色区域) 特征方程有两个实根。
- 当 $a_2 > 0$ 时 z_1 和 z_2 异号，设 $z_1 < -1, z_2 > 1$。$f(\lambda)$ 在 0 和 π 处有峰值。由 $z_1 + z_2 = \frac{(-a_1)}{a_2}$ 知
  - $a_1 > 0$ 时 $z_1 + z_2 < 0$, 正根 z_2 离单位圆更近，$f(\lambda)$ 在 0 点最大。ρ_k 都为正数，震荡衰减。$\{X_t\}$ 表现出相邻点的正相关。
  - $a_1 < 0$ 时 $z_1 + z_2 > 0$，负根 z_1 离单位圆更近，$f(\lambda)$ 在 π 点最大。ρ_k 主要呈现正负交替衰减。$\{X_t\}$ 也表现出正负振荡。
  - $a_1 = 0$ 时 $z_1 = -z_2$，离单位圆一样近，$f(\lambda)$ 在 $0, \pi$ 一样高。ρ_k 都非负，表现出振荡衰减。$\{X_t\}$ 表现出正负振荡。
- 当 $a_2 < 0$ 有复根或实根；为实根时 z_1 与 z_2 同号，与 a_1 同号。
- $a_1 > 0$ 时有两个正根, $f(\lambda)$ 只在 0 点有峰值。ρ_k 单调衰减。$\{X_t\}$ 表现出相邻点的正相关。
- $a_1 < 0$ 时有两个负根, $f(\lambda)$ 只在 π 点有峰值。ρ_k 正负交替衰减。$\{X_t\}$ 表现出正负振荡。

11.2.4 AR(2) 虚特征根时的表现

z_1, z_2 是虚根 $\iff a_1^2 + 4a_2 < 0$(红色区域)。

$$z_1, z_2 = be^{\pm i\lambda_0}, \quad b > 1, \lambda_0 \neq 0, \pi$$

由(9.16)得

$$\rho_k = \frac{\cos(k\lambda_0 + \theta_0)}{b^k \cos(\theta_0)}, \quad k \geq 0 \tag{11.4}$$

这时 $\{\rho_k\}$ 振荡衰减，振荡角频率为 λ_0。

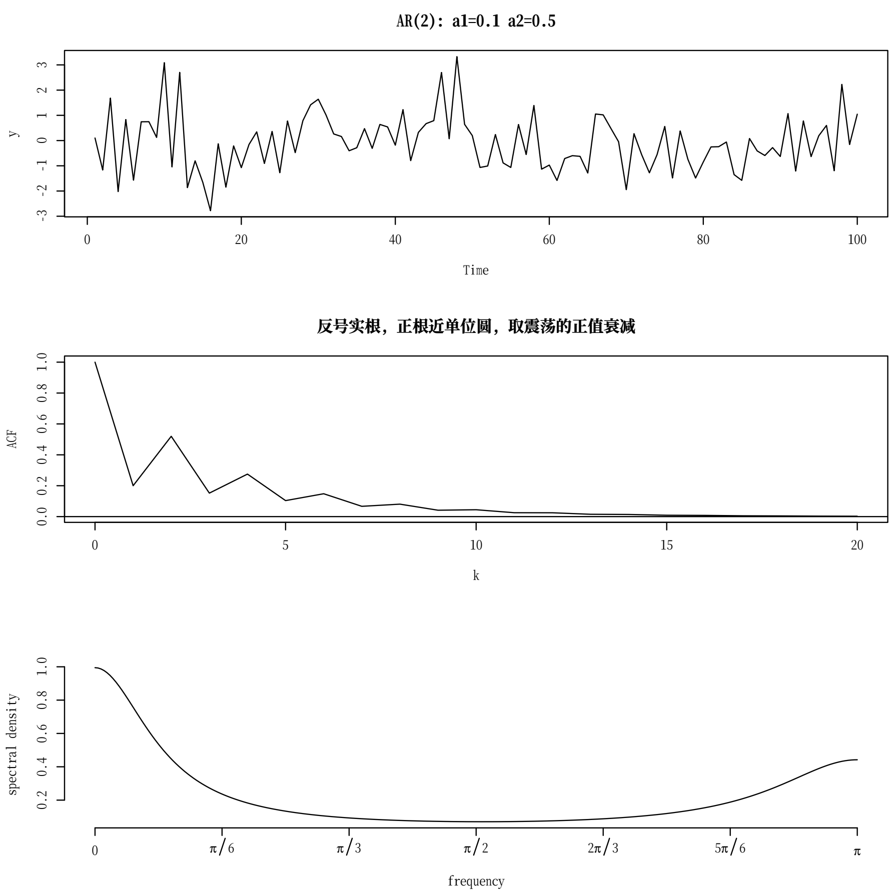
$\{X_t\}$ 也呈现出在频率 λ_0 处的振荡。

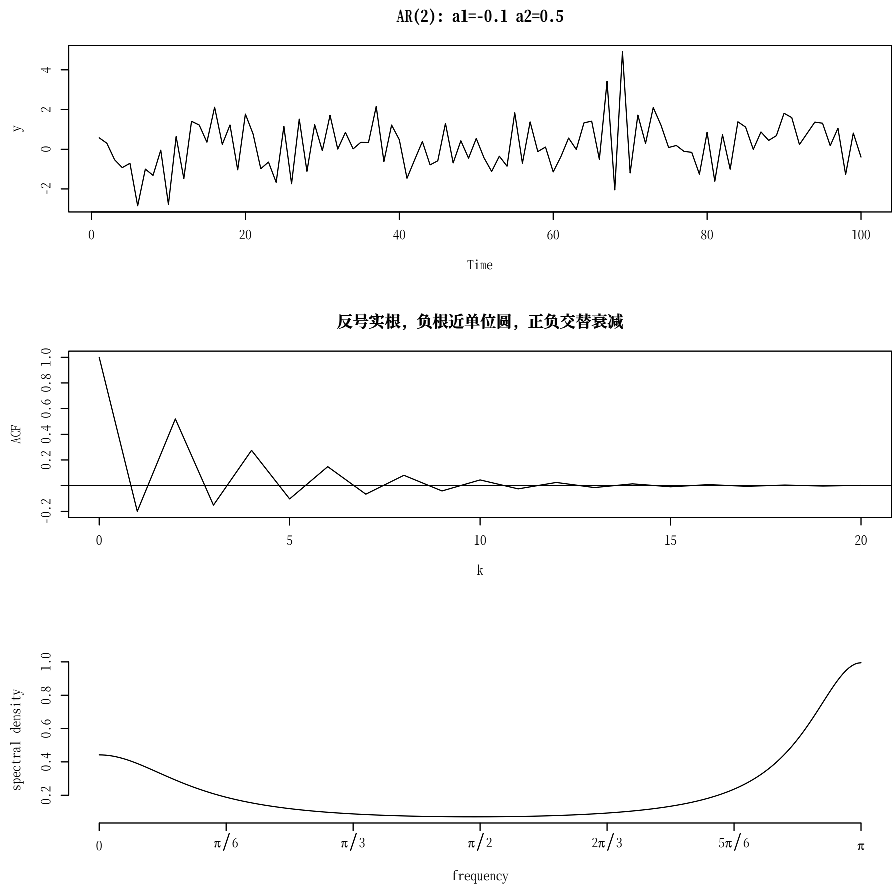
用 R 对 AR(2) 的各种情形作模拟图如下。

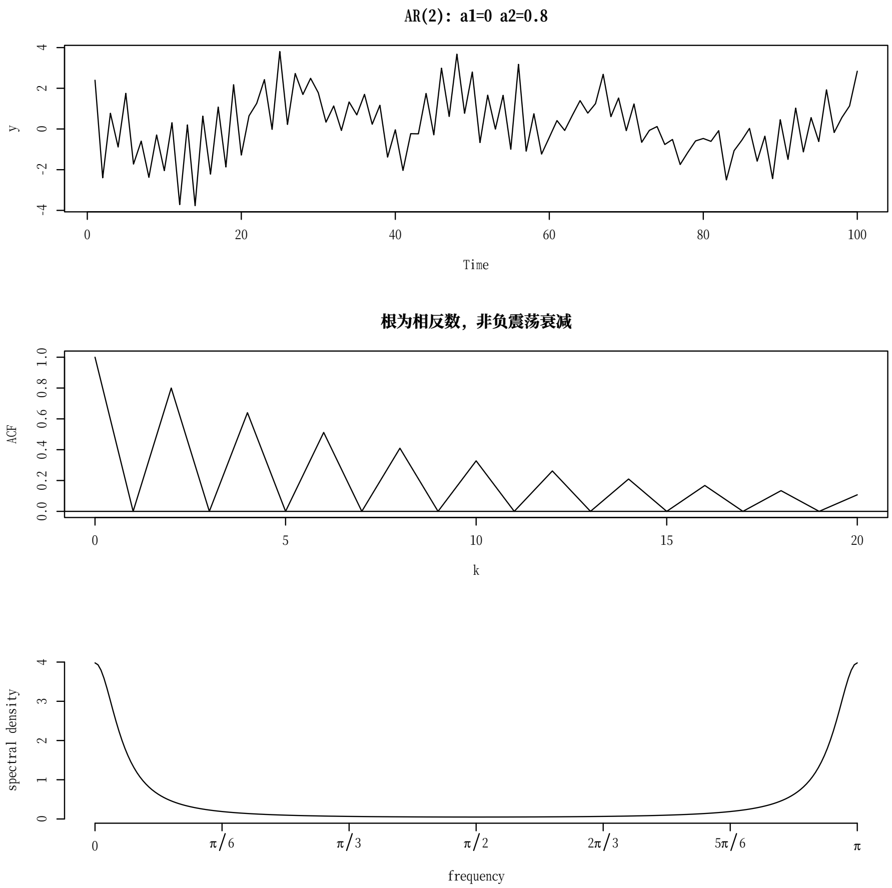
```
demo.ar2.example <- function(){
  oldpar <- par(mfrow=c(3,1)); on.exit(par(oldpar))
  n <- 100
```

```
acomplex <- cbind(2/1.1*cos(c(pi/6, pi/2, 2*pi/3)),
                  -1/1.1^2)
alist <- rbind(c(0.1, 0.5),
               c(-0.1, 0.5),
               c(0, 0.8),
               c(1, -0.1),
               c(-1, -0.1),
               acomplex)
tlist.y <- paste("AR(2): a1=", round(alist[,1],3),
                 " a2=", round(alist[,2],3), sep="")
tlist.rho <- c(" 反号实根，正根近单位圆，取震荡的正值衰减",
               " 反号实根，负根近单位圆，正负交替衰减",
               " 根为相反数，非负震荡衰减",
               " 两正根，单调正值衰减",
               " 两负根，正负交替衰减",
               " 共轭复根，以周期 12 震荡衰减",
               " 共轭复根，以周期 4 震荡衰减",
               " 共轭复根，以周期 3 震荡衰减")

m <- 20
acfv <- numeric(m+1)
acfv[1] <- 1
for(ii in seq(nrow(alist))){
  y <- ar.gen(n, alist[ii,], sigma=1.0, plot.it=FALSE)
  a1 <- alist[ii,1];
  a2 <- alist[ii,2];
  acfv[2] <- a1 / (1 - a2)
  for(jj in seq(3, m+1)){
    acfv[jj] <- a1*acfv[jj-1] + a2*acfv[jj-2]
  }
  plot(y, main=tlist.y[ii], type="l")
  plot(0:m, acfv, type="l",
       xlab="k", ylab="ACF",
```

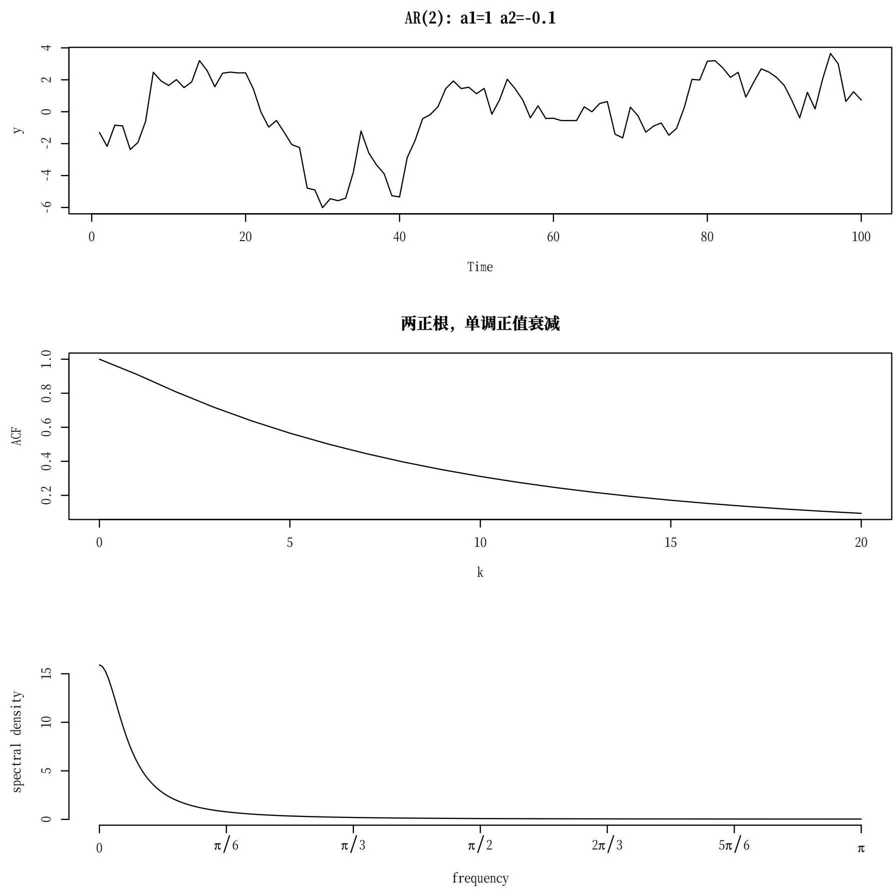
```
         main=tlist.rho[ii])
    abline(h=0)
    ar.true.spectrum(attr(y, "a"), title="")
  }
  invisible()
}
demo.ar2.example()
```

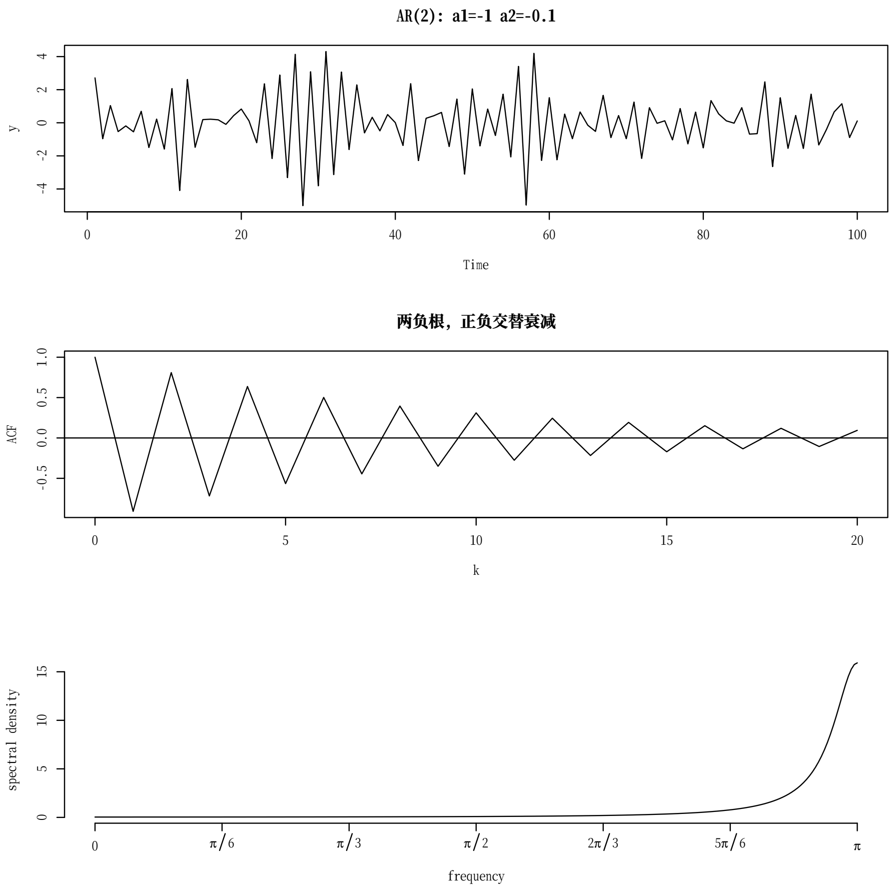


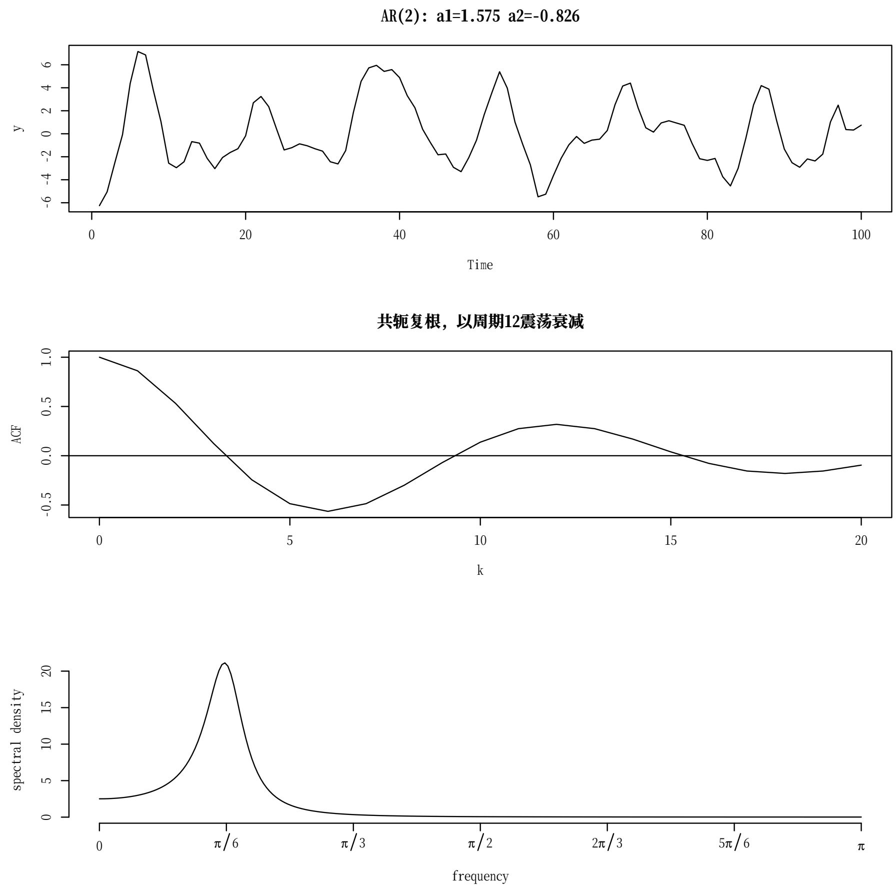

AR(2): a1=-0.1 a2=0.5
y
Time
反号实根，负根近单位圆，正负交替衰减
ACF
k
spectral density
frequency
0
π/6
π/3
π/2
2π/3
5π/6
π

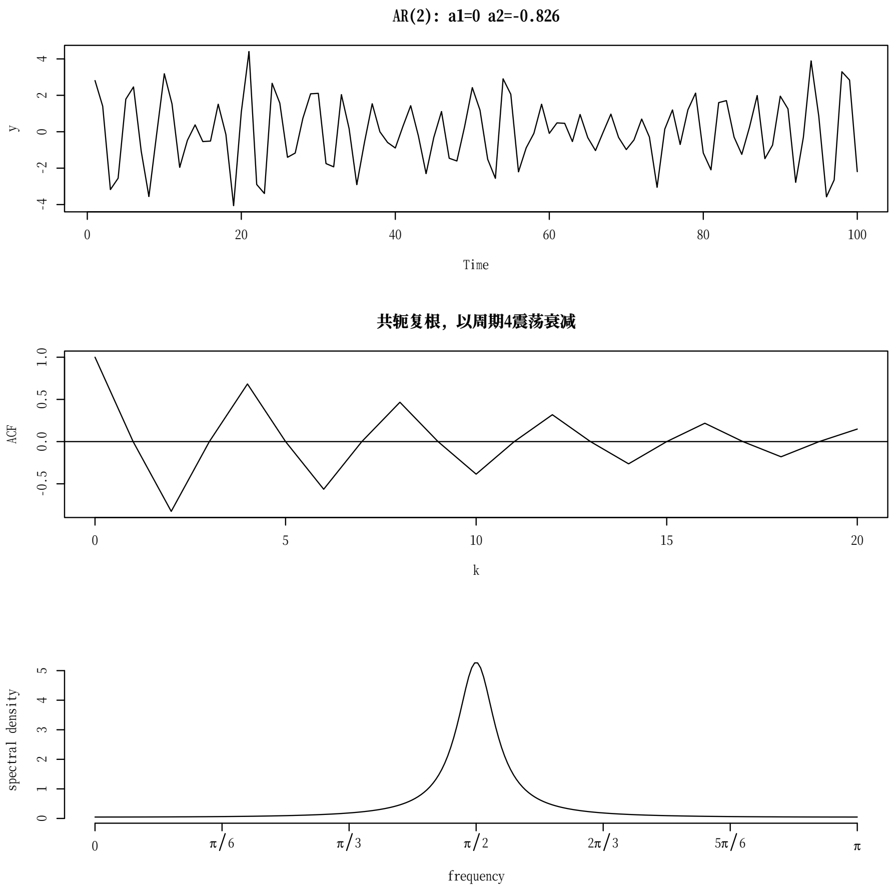


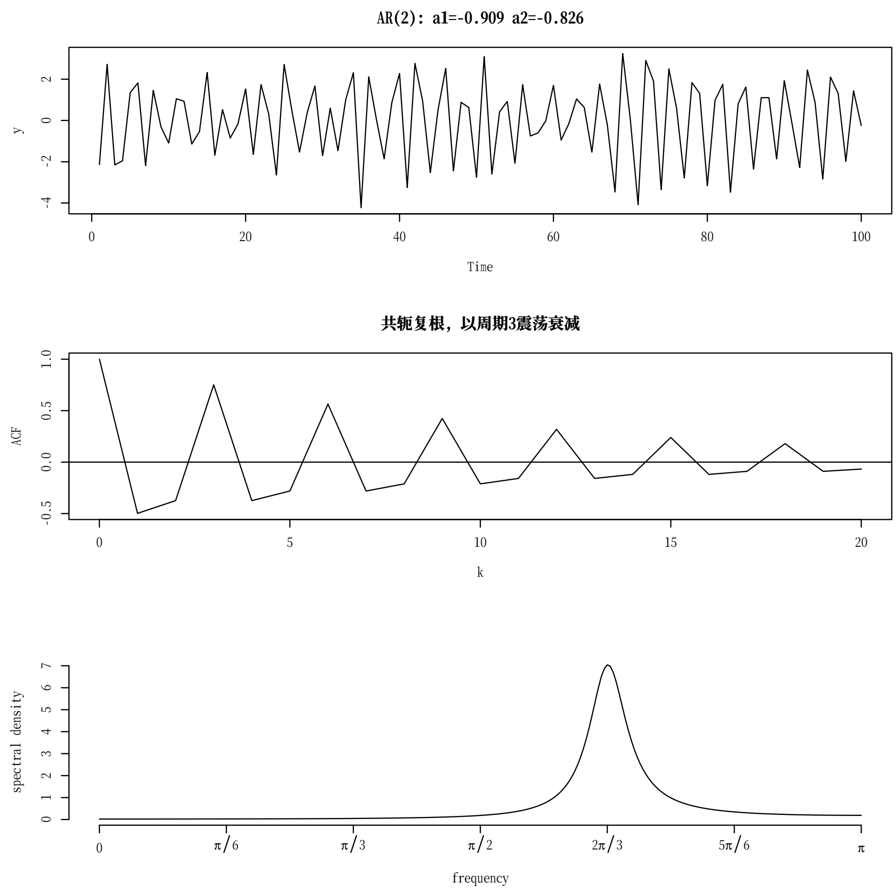
AR(2): a1=0 a2=0.8
y
Time
根为相反数，非负震荡衰减
ACF
k
spectral density
frequency


AR(2): a1=1 a2=-0.1
y
Time
两正根，单调正值衰减
ACF
k
spectral density
frequency


AR(2): a1=-1 a2=-0.1
y
Time
两负根，正负交替衰减
ACF
k
spectral density
frequency


AR(2): a1=1.575 a2=-0.826
y
Time
共轭复根，以周期12震荡衰减
ACF
k
spectral density
frequency
0
π/6
π/3
π/2
2π/3
5π/6
π


AR(2): a1=0 a2=-0.826
y
Time
共轭复根，以周期4震荡衰减
ACF
k
spectral density
frequency



11.3 附录：AR 模型有关的一些 R 函数

11.3.1 模拟和基本统计

`arima.sim()` 可以用来模拟生成来自 AR(p) 模型的数据。例如，对如下 AR(2)：

$$X_t = X_{t-1} - 0.1X_{t-2} + \varepsilon_t,\ \varepsilon_t \sim \text{WN}(0, 0.2^2)$$

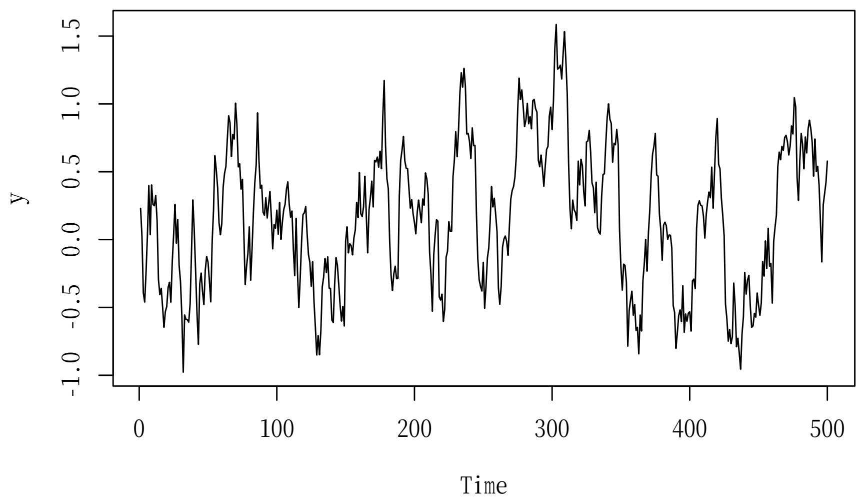
可以用如下程序生成模拟数据，长度 500：

```
set.seed(2)
y = arima.sim(model=list(ar=c(1, -0.1)), n=500, sd=0.2)
head(y)
```

```
## [1]  0.23049850  0.02289326 -0.39333224 -0.46021578 -0.23371005  0.04015748
```

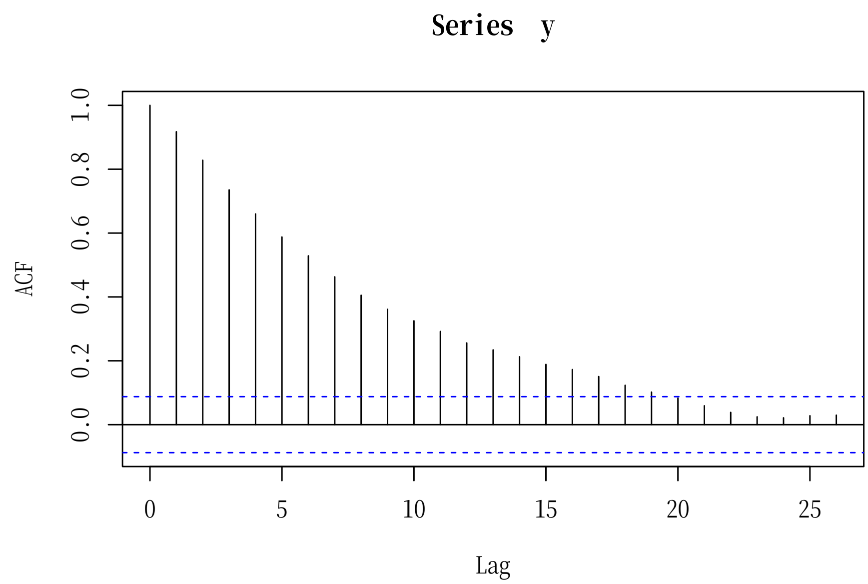
对时间序列数据用 plot() 作时间序列图：

```r
plot(y)
```



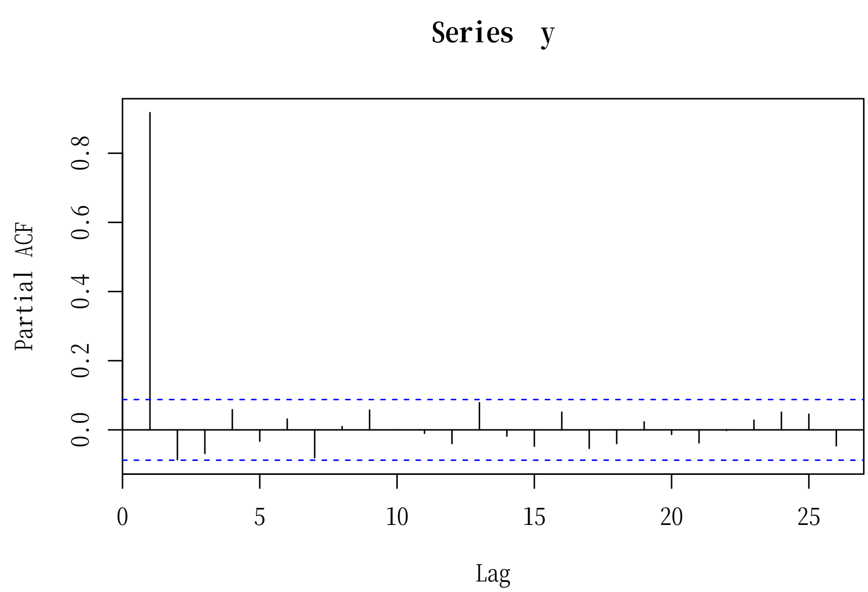
用 acf() 作时间序列的自相关函数图：

```r
acf(y)
```



用 `pacf()` 作时间序列的偏相关函数图：

```
pacf(y)
```



下面是自定义的一个 AR(p) 模拟程序，允许自己指定特征多项式的根作为模型参数，可以指定模拟初值，并且不要求模型必须平稳：

```
## simulate AR(p) model.
## a is vector a_1, \dots, a_p
## Model is X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \epsilon_t
## \Var(\epsilon_t) = sigma^2
## This version use the ``filter'' function.
ar.gen <- function(n, a, sigma=1.0, by.roots=FALSE,
                   plot.it=FALSE, n0=1000,
                   x0=numeric(length(a))){
  if(by.roots){
    require(polynom)
    cf <- Re(coef(poly.calc(a)))
    cf <- cf / cf[1]
    a <- -cf[-1]
  }
  n2 <- n0 + n
  p <- length(a)
  eps <- rnorm(n2, 0, sigma)
  x2 <- filter(eps, a, method="recursive", side=1, init=x0)
  x <- x2[(n0+1):n2]
  x <- ts(x)
  attr(x, "model") <- "AR"
  attr(x, "a") <- a
  attr(x, "sigma") <- sigma
  if(plot.it) plot(x)
  x
}
```

11.3.2 特征多项式的根

对多项式 $A(z) = a_0 + a_1 z + \cdots + a_p z^p$，函数 `polyroot(c(a0, a1, ..., ap))` 求多项式的所有复根。所以特征多项式为 $A(z) = 1 - a_1 z - \cdots - a_p z^p$ 的

AR(p) 模型的所有特征根可以用 `polyroot` 得到。比如，上面的 AR(2) 模型的所有特征根：

```
polyroot(c(1, -c(1, -0.1)))
```

```
## [1] 1.127017-1.614657e-21i 8.872983+1.614657e-21i
```

求特征根的模，平稳的 AR 模型的特征多项式的模大于 1：

```
abs(polyroot(c(1, -c(1, -0.1))))
```

```
## [1] 1.127017 8.872983
```

11.3.3 模型的理论谱密度图

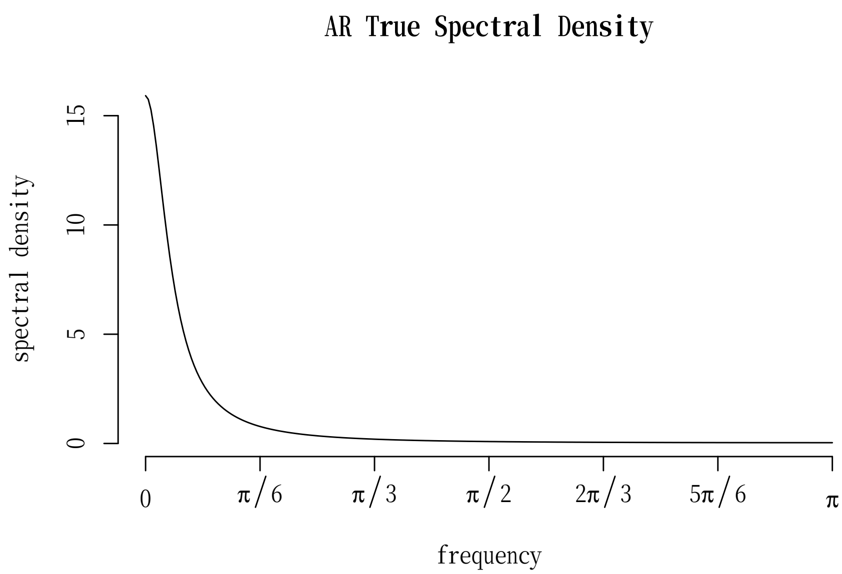
下面的程序输入 AR 的系数，绘制理论谱密度图形：

```
## AR theoretical spectral density plot given AR coefficients
ar.true.spectrum <- function(a, ngrid=256, sigma=1, plot.it=TRUE,
                             title="AR True Spectral Density"){
  p <- length(a)
  freqs <- seq(from=0, to=pi, length=ngrid)
  spec <- numeric(ngrid)
  for(ii in seq(ngrid)){
    spec[ii] <- sigma^2 / (2*pi) /
      abs(1 - sum(complex(mod=a, arg=freqs[ii]*seq(p))))^2
  }
  if(plot.it){
    plot(freqs, spec, type='l',
         main=title,
         xlab="frequency", ylab="spectral density",
         axes=FALSE)
    axis(2)
    axis(1, at=(0:6)/6*pi,
         labels=c(0, expression(pi/6),
           expression(pi/3), expression(pi/2),
           expression(2*pi/3), expression(5*pi/6), expression(pi)))
```

```
  }
  invisible(list(frequencies=freqs, spectrum=spec,
              ar.coefficients=a, sigma=sigma))
}
```

例如，对上面的 AR(2) 模型做理论谱密度图：

```
ar.true.spectrum(c(1, -0.1))
```



11.3.4 AR 模型估计

R 中 `ar()` 可以从数据估计 AR 模型，还可以自动确定阶 p。

`arima()` 函数指定模型阶后估计模型。一般的时间序列不一定是零均值的，所以估计函数同时估计模型的均值。

如：

```
set.seed(2)
y = 100 + arima.sim(model=list(ar=c(1, -0.1)), n=500, sd=0.2)
arm01 <- arima(y, order=c(2, 0, 0))
```

```r
print(arm01)
```

```
##
## Call:
## arima(x = y, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       0.9998  -0.0901   100.1638
## s.e.  0.0445   0.0445     0.0989
##
## sigma^2 estimated as 0.04126:  log likelihood = 86.59,  aic = -165.18
```

其中 `order=c(2,0,0)` 表示拟合 AR(2) 模型。结果中 `intercept` 是模型期望值 EX_t 的估计。

Chapter 12

滑动平均模型

12.1 模型引入

平稳序列 $\{X_t\}$ 的自协方差函数若满足 $\gamma_q \neq 0, \gamma_k = 0, k > q$，则称 $\{X_t\}$ 是q**步相关的**。有限项的线性平稳列具有 q 步相关性，称为**滑动平均模型**。

例 12.1. 考虑每隔 2 小时记录的化学反应数据时间序列 $\{x_t, t = 1, \dots, 197\}$ 的模型。

读入数据并作序列图（见图12.1）：

```
x <- scan(textConnection("
17.0   16.6   16.3    16.1   17.1   16.9   16.8   17.4   17.1   17.0   16.7   17.4   17.2
17.4   17.4   17.0    17.3   17.2   17.4   16.8   17.1   17.4   17.4   17.5   17.4   17.6
17.4   17.3   17.0    17.8   17.5   18.1   17.5   17.4   17.4   17.1   17.5   17.7   17.4
17.8   17.6   17.5    16.5   17.8   17.3   17.3   17.1   17.4   16.9   17.3   17.6   16.9
16.7   16.8   16.8    17.2   16.8   17.6   17.2   16.6   17.1   16.9   16.6   18.0   17.2
17.3   17.0   16.9    17.3   16.8   17.3   17.4   17.7   16.8   16.9   17.0   16.9   17.0
16.6   16.7   16.8    16.7   16.4   16.5   16.4   16.6   16.5   16.7   16.4   16.4   16.2
16.4   16.3   16.4    17.0   16.9   17.1   17.1   16.7   16.9   16.5   17.2   16.4   17.0
17.0   16.7   16.2    16.6   16.9   16.5   16.6   16.6   17.0   17.1   17.1   16.7   16.8
16.3   16.6   16.8    16.9   17.1   16.8   17.0   17.2   17.3   17.2   17.3   17.2   17.2
```

```
17.5   16.9   16.9    16.9   17.0   16.5   16.7   16.8   16.7   16.7   16.6   16.5
16.7   16.7   16.9    17.4   17.1   17.0   16.8   17.2   17.2   17.4   17.2   16.9
17.0   17.4   17.2    17.2   17.1   17.1   17.1   17.4   17.2   16.9   16.9   17.0
16.9   17.3   17.8    17.8   17.6   17.5   17.0   16.9   17.1   17.2   17.4   17.5
17.0   17.0   17.0    17.2   17.3   17.4   17.4   17.0   18.0   18.2   17.6   17.8
17.2   17.4
"), quiet=TRUE)
x <- ts(x)
plot(x, main="Chemical reaction time series")
```

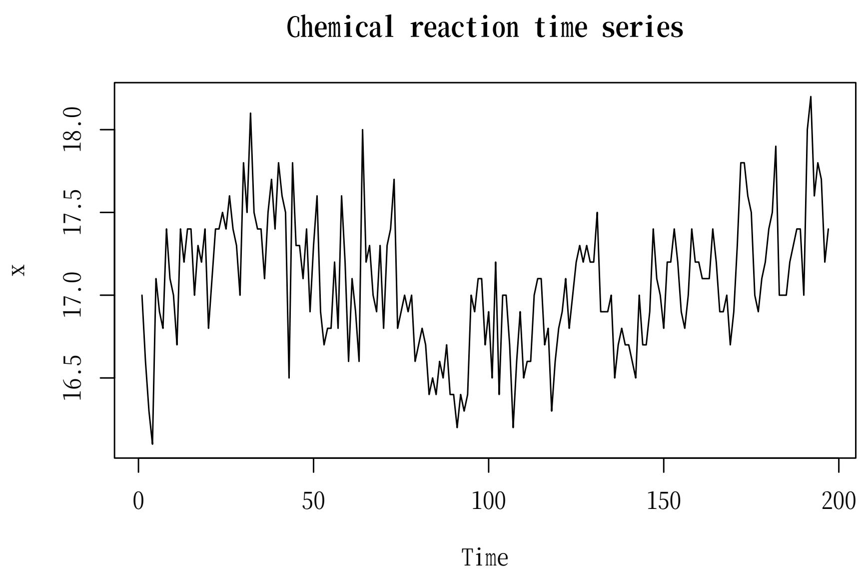


图 12.1: 某化学试验溶液浓度的时间序列

序列看起来不像是均值恒定的。一阶差分得

$$y_t = x_t - x_{t-1}, t = 2, \dots, 197$$

差分后的序列看起来像是平稳的。见图12.2。

```r
y <- diff(x)
plot(y, main="First order difference")
```

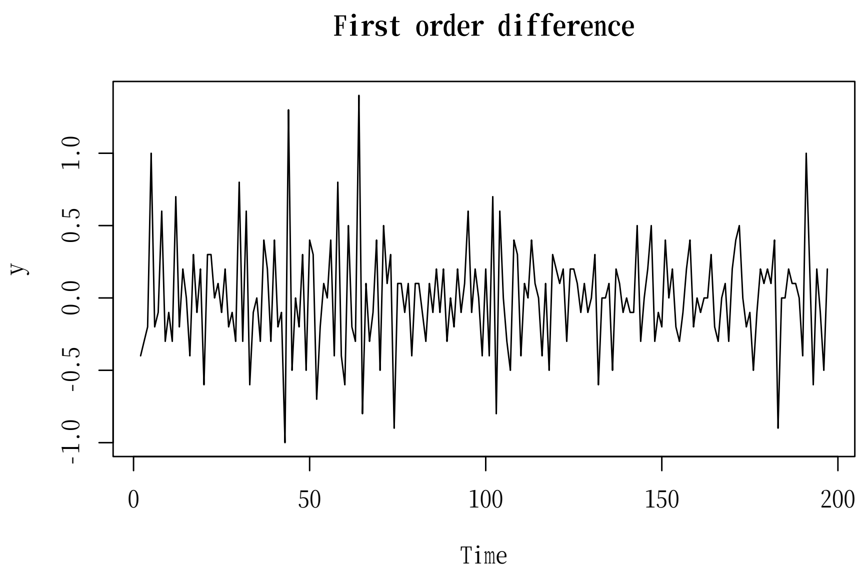


图 12.2: 浓度差分序列

$\{y_t\}$ 的样本自相关系数列呈现截尾性，ACF 在滞后 1 处比较显著，$k > 1$ 时都很小，见图12.3。

```r
acf(y)
```

对差分序列，可以拟合如下模型

$$Y_t = \varepsilon_t + \hat{b}\varepsilon_{t-1}, t \in \mathbb{Z} \tag{12.1}$$

模型特点是 $\{\gamma_k\}$ 1 步截尾。

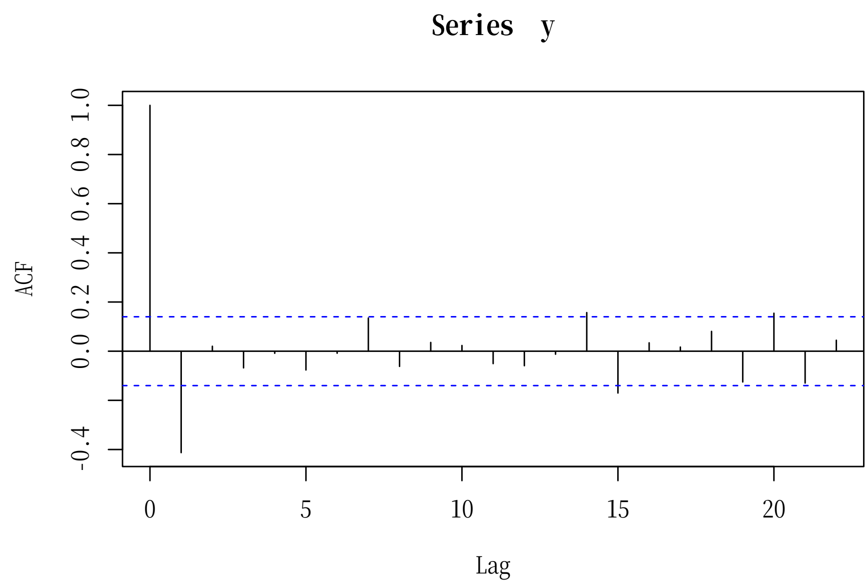


图 12.3: 浓度差分序列 ACF

12.2 MA(q) 模型和 MA(q) 序列

定义 12.1 (滑动平均模型)**.** 设 $\{\varepsilon_t\}$ 是 $\mathrm{WN}(0,\sigma^2)$, 如果实数 $b_1, b_2, \dots, b_q$ ($b_q \neq 0$) 使得

$$B(z) = 1 + \sum_{j=1}^{q} b_j z^j \neq 0,\ |z| < 1,$$

则称

$$X_t = \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{12.2}$$

是q **阶滑动平均模型**, 简称为 **MA**(q) **模型**。

模型(12.2)中的 $\{X_t\}$ 显然是平稳列。称由(12.2)决定的平稳序列 $\{X_t\}$ 是**滑动平均序列**, 简称为 **MA**(q) **序列**.

如果进一步要求多项式 $B(z)$ 在单位圆上也没有零点: $B(z) \neq 0$ 当 $|z| \leq 1$, 则称(12.2)是**可逆的 MA**(q) **模型**, 称相应的平稳序列是**可逆的 MA**(q) **序列**.

12.3 MA 的特征

低阶的 MA 与 AR 相比较光滑 (滑动平均)，振荡较轻。稳定性较好。高阶的 MA 可以模拟 AR 的特征。

用推移算子把模型写为

$$X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z} \tag{12.3}$$

对于可逆 MA，$B^{-1}(z)$ 有 Taylor 展式

$$B^{-1}(z) = \sum_{j=0}^{\infty} \phi_j z^j, \quad |z| \leq 1 + \delta\ (\delta > 0)$$

所以

$$\varepsilon_t = B^{-1}(\mathscr{B}) X_t = \sum_{j=0}^{\infty} \phi_j X_{t-j} \tag{12.4}$$

12.4 MA 序列的自协方差函数

记 $b_0 = 1$，则对 MA(q) 序列有 $EX_t = 0$，

$$\gamma_k = E(X_t X_{t+k}) \tag{12.5}$$

$$= \begin{cases} \sigma^2 \sum_{j=0}^{q-k} b_j b_{j+k}, & 0 \leq k \leq q \\ 0, & k > q \end{cases} \tag{12.6}$$

12.5 MA 序列的谱密度

定理 12.1. *MA(q) 序列 $\{X_t\}$ 的自协方差函数是 q 步截尾的：*

$$\gamma_q = \sigma^2 b_q \neq 0, \quad \gamma_k = 0,\ |k| > q. \tag{12.7}$$

并且有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} |B(e^{i\lambda})|^2 \tag{12.8}$$

$$= \frac{1}{2\pi} \sum_{k=-q}^{q} \gamma_k e^{-ik\lambda}, \quad \lambda \in [-\pi, \pi]. \tag{12.9}$$

(12.9)的第一条可以用线性平稳列的定理6.2或用谱密度定义直接证明。第二条用自协方差绝对可和时谱密度公式 (定理9.1) 得到也可以用谱密度定义得到。

12.6 MA(q) 序列的充要条件

MA(q) 序列是自协方差函数 q 步截尾的，反之若平稳列 $\{X_t\}$ 自协方差函数 q 步截尾则其必为 MA(q) 序列。

定理 12.2 (自协方差截尾的充要性)**.** *设零均值平稳序列 $\{X_t\}$ 有自协方差函数 $\{\gamma_k\}$, 则 $\{X_t\}$ 是 MA(q) 序列的充分必要条件是*

$$\gamma_q \neq 0,\ \gamma_k = 0,\ |k| > q.$$

只需要证明充分性。证明需要一个复变函数引理。

引理 12.1. *设实常数 $\{c_j\}$ 使得 $c_q \neq 0$ 和*

$$g(\lambda) = \frac{1}{2\pi} \sum_{j=-q}^{q} c_j e^{-ij\lambda} \geq 0, \lambda \in [-\pi, \pi],$$

则有惟一的实系数多项式:

$$B(z) = 1 + \sum_{j=1}^{q} b_j z^j \neq 0,\ |z| < 1,\ b_q \neq 0. \tag{12.10}$$

使得

$$g(\lambda) = \frac{\sigma^2}{2\pi} |B(e^{i\lambda})|^2.$$

这里 σ^2 为某个正常数.

证明时，令 $G(z) = \sum_{j=-q}^{q} c_j z^{j+q}$ 则 $G(z)$ 的 $2q$ 个根中 z_1 是根必有 z_1^{-1} 也是根。$z_1 \neq \pm 1$ 时 $z_1^{-1} \neq z_1$。把这样成对的根只取其中一个且要求不在单位圆内即可组成 $B(.)$。详见谢衷洁《时间序列分析》(谢衷洁, 1990) P89–P92。

定理12.2证明:

由自协方差绝对可和时谱密度公式得

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-q}^{q} \gamma_k e^{-ik\lambda}$$

由引理12.1，

$$f(\lambda) = \frac{\sigma^2}{2\pi} |B(e^{-i\lambda})|^2, \quad B(z) \text{ 单位园内没有根.}$$

如果 $B(z)$ 在单位圆内和单位圆上都没有根，即 $f(\lambda)$ 在 $[0,\pi]$ 上恒正，则可定义

$$\varepsilon_t = B^{-1}(\mathscr{B})X_t = \sum_{j=0}^{\infty} h_j X_{t-j}$$

其中 $B^{-1}(z) = \sum_{j=0}^{\infty} h_j z^j$ 是 $B^{-1}(z)$ 在 $|z| < 1+\delta$ 的泰勒展开式，$B(z)$ 的根的模都大于 $1+\delta$。泰勒展开系数 $\{h_j\}$ 绝对可和，所以 $\{\varepsilon_t\}$ 是平稳列，且 $E\varepsilon_t = 0$，用线性滤波的谱密度公式（见定理6.5）可得 $\{\varepsilon_t\}$ 的谱密度为

$$\begin{aligned} f_\varepsilon(\lambda) =& \left|\sum_{j=0}^{\infty} h_j e^{ij\lambda}\right|^2 f(\lambda) \\ =& \left|B(e^{i\lambda})\right|^{-2} \frac{\sigma^2}{2\pi} |B(e^{-i\lambda})|^2 \\ =& \frac{\sigma^2}{2\pi} \end{aligned}$$

这是白噪声的谱密度，由谱密度与自协方差函数的关系可知 $\{\varepsilon_t\}$ 是 WN(0, σ^2)，所以 $f(\lambda)$ 在 $[0,\pi]$ 上恒正时定理的充分性得证。

对于 $B(z)$ 单位园上可能有根的一般情况可以用 Hilbert 空间预测的方法证明。(见 (Brockwell & Davis, 1987) §3.2.1, P.89, 但那里的 MA 没有根的条件；或参考 (谢衷洁, 1990) P92.)。

12.7 最小序列

定义 12.2 (最小序列)**.** 设 $\{X_t :\ t \in \mathbb{Z}\}$ 是平稳序列. 用 H_x 表示 $\{X_t\}$ 产生的 Hilbert 空间, 用 $H_x(s)$ 表示由 $\{X_t : t \neq s\}$ 产生的 Hilbert 空间. 如果

$$H_x \neq H_x(s)$$

对某个 $s \in \mathbb{Z}$ 成立, 则称 $\{X_t\}$ 是**最小序列**.

对平稳序列，可以证明对某个 s 成立 $H_x(s) \neq H_x$ 则对所有的 s 成立 $H_x(s) \neq H_x$。说明最小序列的每个 X_t 都含有其他 X_s 中没有的信息。可完全线性预测的平稳序列 (某 Γ_n 不满秩) 非最小序列。

定理 12.3 (最小序列的谱密度条件). 设平稳序列 $\{X_t\}$ 有谱密度 $f(\lambda)$, 则 $\{X_t\}$ 是最小序列的充分必要条件是

$$\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)} < \infty. \tag{12.11}$$

证明参见 Kolgomorov 著，郑绍濂、陶宗英等译等译的“希尔伯脱空间中的平稳序列”，上海科学技术出版社 1963。

由以上定理可见谱密度连续恒正的平稳列是最小序列。于是，AR(p) 序列是最小序列。可逆的 MA(q) 序列的谱密度连续有正下界所以是最小序列。

单位圆上有根的 MA(q) 序列其谱密度 $f(\lambda)$ 中有

$$|1 - e^{i(\lambda-\lambda_j)}|^2 = O(|\lambda-\lambda_j|^2)$$

成分所以其倒数不可积，因此不可逆的 MA(q) 序列不是最小序列。

平稳的最小序列一定不能 n 步完全线性预测，所以其自协方差列正定。

12.8 MA(q) 系数的递推计算

MA(q) 序列的系数 $(b_1, b_2, \dots, b_q)$ 及 σ^2 可以被 $\gamma_0, \gamma_1, \dots, \gamma_q$ 唯一确定。可以用如下文献方法计算模型参数：李雷 (1991), 平稳过程的状态空间模型及随机实现算法时多元情况的解决，北京大学硕士论文。

记

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{q\times q}, \quad C = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{q\times 1} \tag{12.12}$$

$$\Omega_k = \begin{pmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_k \\ \gamma_2 & \gamma_3 & \cdots & \gamma_{k+1} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_q & \gamma_{q+1} & \cdots & \gamma_{q+k-1} \end{pmatrix}, \quad \gamma_q = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_q \end{pmatrix}$$

则有：

$$b_q = \frac{1}{\sigma^2}(\gamma_q - A\Pi C),\ \sigma^2 = \gamma_0 - C^\tau \Pi C, \tag{12.13}$$

其中

$$\Pi = \lim_{k\to\infty} \Omega_k \Gamma_k^{-1} \Omega_k^T. \tag{12.14}$$

公式(12.13), (12.14)为以后用观测样本估计 MA(q) 模型的参数打下了基础.

12.9 MA(q) 模型举例

12.9.1 MA(1)

可逆 MA(1) 模型为

$$X_t = \varepsilon_t + b\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathrm{WN}(0,\sigma^2), \quad |b|<1$$

自协方差和自相关

$$\begin{cases} \gamma_0 = & \sigma^2(1+b^2) \\ \gamma_1 = & \sigma^2 b \\ \gamma_k = & 0, \quad k \ge 2 \end{cases}$$

$$\begin{cases} \rho_1 = & \frac{b}{1+b^2} \\ \rho_k = & 0, \quad k \ge 2 \end{cases}$$

谱密度

$$\begin{aligned} f(\lambda) =& \frac{\sigma^2}{2\pi}|1+be^{i\lambda}|^2 \\ =& \frac{\sigma^2}{2\pi}(1+b^2+2b\cos\lambda), \quad \lambda \in [-\pi,\pi] \end{aligned}$$

偏相关系数不截尾:

$$a_{k,k} = -\frac{(-b)^k(1-b^2)}{(1-b^{2k+2})}, \quad k \ge 1$$

逆表示

$$\varepsilon_t = \sum_{j=0}^{\infty}(-b)^j X_{t-j}$$

12.9.2 MA(1) 实例

先提供几个自定义 R 函数。

绘制 MA 理论谱密度曲线的函数，模拟 MA 序列的函数，从自协方差函数利用递推方法估计模型参数的函数：

```
## MA theoretical spectrum given MA coefficients
ma.true.spectrum <- function(a, ngrid=256, sigma=1,
                             tit="True MA Spectral Density",
                             plot.it=TRUE){
  p <- length(a)
  freqs <- seq(from=0, to=pi, length=ngrid)
  spec <- numeric(ngrid)
  for(ii in seq(ngrid)){
    spec[ii] <- 1 + sum(complex(mod=a, arg=freqs[ii]*seq(p)))
  }
  spec = sigma^2 / (2*pi) * abs(spec)^2
  if(plot.it){
    plot(freqs, spec, type='l',
         main=tit,
         xlab="frequency", ylab="spectrum",
         axes=FALSE)
    axis(2)
    axis(1, at=(0:6)/6*pi,
         labels=c(0, expression(pi/6),
           expression(pi/3), expression(pi/2),
           expression(2*pi/3), expression(5*pi/6), expression(pi)))
    box()
  }
  invisible(list(frequencies=freqs, spectrum=spec,
            ma.coefficients=a, sigma=sigma))
}

## simulate MA(q) model.
```

```
## a is vector a_1, \dots, a_q
## Model is X_t = eps_t + a_1 eps_{t-1} + \dots + a_q eps_{t-q}
## \Var(\epsilon_t) = sigma^2
## This version uses the filter function.
ma.gen <- function(n, a, sigma=1.0, by.roots=FALSE,
                   plot.it=FALSE){
  if(by.roots){
    require(polynom)
    cf <- Re(c(poly.calc(a)))
    cf <- cf / cf[1]
    a <- cf
  }
  q <- length(a)
  n2 <- n+q
  eps <- rnorm(n2, 0, sigma)
  x2 <- filter(eps, c(1,a), method="convolution", side=1)
  x <- x2[(q+1):n2]
  x <- ts(x)
  attr(x, "model") <- "MA"
  attr(x, "b") <- a
  attr(x, "sigma") <- sigma
  if(plot.it) plot(x)
  x
}

## Given \gamma_0, \gamma_1, \dots, \gamma_q,
## Solve MA(q) coefficients b_1, \dots, b_q, \sigma^2
## Using Li Lei's algorithm.
## Input: gms -- \gamma_0, \gamma_1, \dots, \gamma_q
ma.solve <- function(gms, k=100){
  q <- length(gms)-1
  if(q==1){
    rho1 <- gms[2] / gms[1]
```

```
    b <- (1 - sqrt(1 - 4*rho1^2))/(2*rho1)
    s2 <- gms[1] / (1 + b^2)
    return(list(b=b, s2=s2))
  }

  A <- matrix(0, nrow=q, ncol=q)
  for(j in seq(2,q)){
    A[j-1,j] <- 1
  }
  cc <- numeric(q); cc[1] <- 1
  gamma0 <- gms[1]
  gammas <- numeric(q+k)
  gammas[1:(q+1)] <- gms
  gamq <- gms[-1]
  Gammak <- matrix(0, nrow=k, ncol=k)
  for(ii in seq(k)){
    for(jj in seq(k)){
      Gammak[ii,jj] <- gammas[abs(ii-jj)+1]
    }
  }

  Omk <- matrix(0, nrow=q, ncol=k)
  for(ii in seq(q)){
    for(jj in seq(k)){
      Omk[ii,jj] <- gammas[ii+jj-1+1]
    }
  }
  PI <- Omk %*% solve(Gammak, t(Omk))

  s2 <- gamma0 - c(t(cc) %*% PI %*% cc)
  b <- 1/s2 * c(gamq - A %*% PI %*% cc)
  return(list(b=b, s2=s2))
}
```

一个系数 $b_1 > 0$ 的 MA(1) 模型为：

$$X_t = \varepsilon_t + 0.6\varepsilon_{t-1}, \ \{\varepsilon_t\} \sim \text{i.i.d N}(0,1)$$

下面产生了长度 $n = 100$ 的模拟数据，时间序列图见图12.4，理论谱密度见图12.5。

```
n <- 100
x1 <- ma.gen(n, a=0.6)
plot(x1, main="MA(1) Series: b=0.6")
```

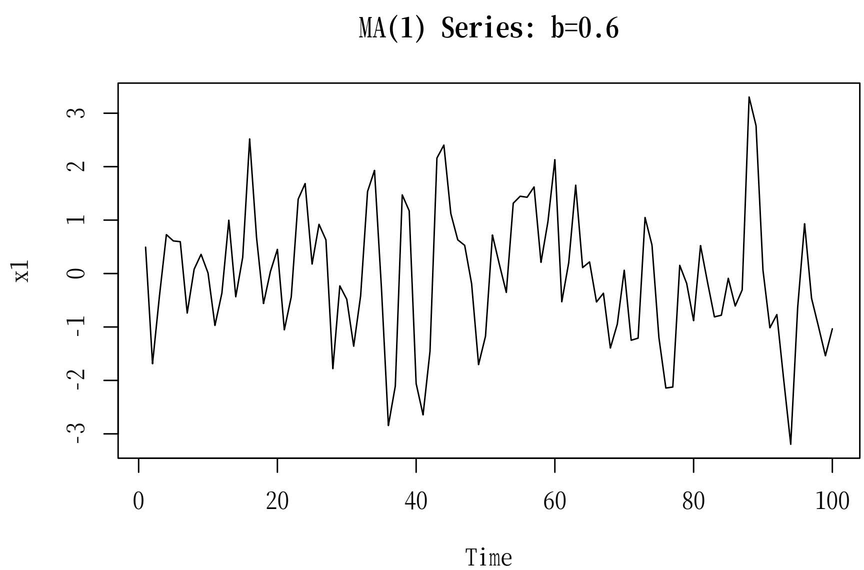


图 12.4: 正系数 MA(1) 模型模拟序列

```
ma.true.spectrum(a=0.6, tit="MA(1) Spectral Density: b=0.6")
```

一个系数 $b_1 < 0$ 的 MA(1) 模型为：

$$X_t = \varepsilon_t - 0.6\varepsilon_{t-1}, \ \{\varepsilon_t\} \sim \text{i.i.d N}(0,1)$$

下面产生了长度 $n = 100$ 的模拟数据，时间序列图见图12.6，理论谱密度见图12.7。

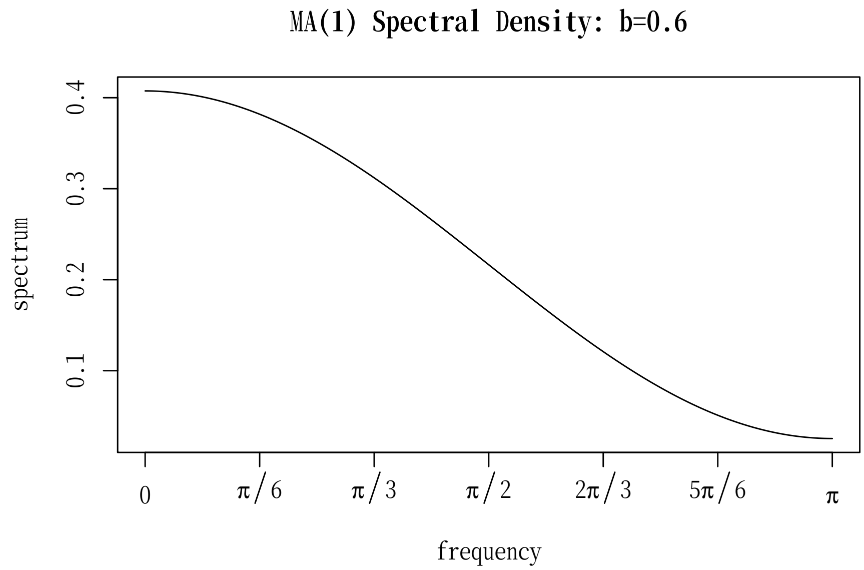


图 12.5: 正系数 MA(1) 模型的理论谱密度

```
x1 <- ma.gen(n, a=-0.6)
plot(x1, main="MA(1) Series: b=-0.6")
```

```
ma.true.spectrum(a=-0.6, tit="MA(1) Spectral Density: b=-0.6")
```

与 AR 模型相比，低阶 MA 模型的谱都很平缓。$b_1 > 0$ 时谱密度峰出现在低频，$b_1 < 0$ 是谱密度峰出现在高频。

12.9.3 MA(2)

可逆 MA(2)

$$X_t = \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2},\ t \in \mathbb{Z}$$

$$B(z) = 1 + b_1 z + b_2 z^2 \neq 0,\ |z| \leq 1$$

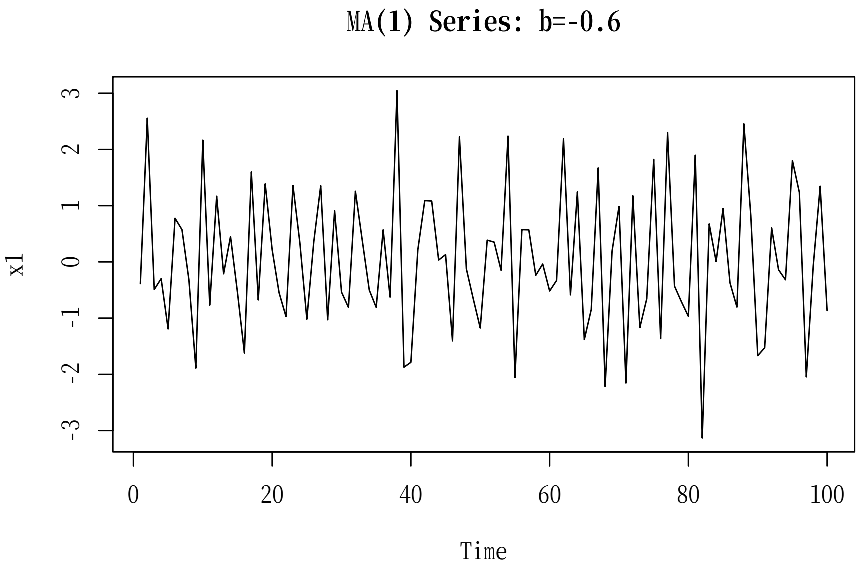


图 12.6: 负系数 MA(1) 模型模拟序列

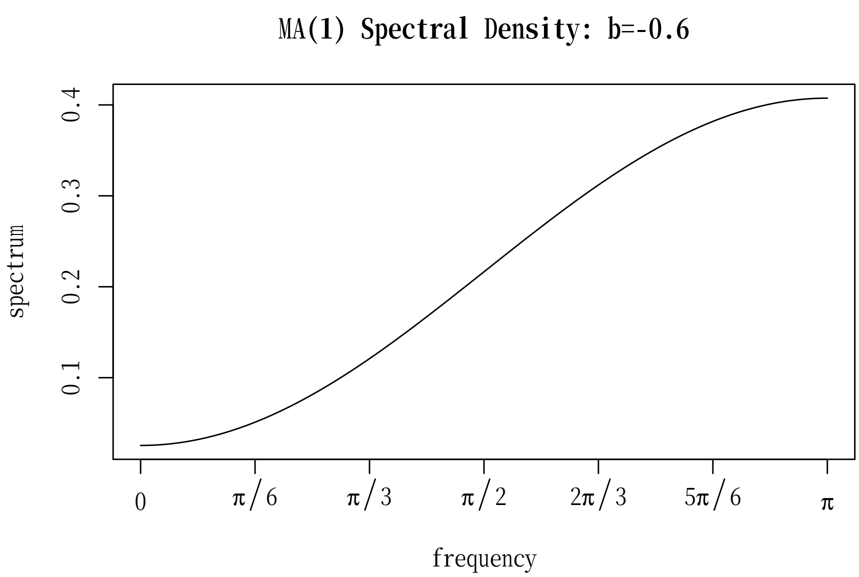


图 12.7: 负系数 MA(1) 模型的理论谱密度

可逆域:

$$
\begin{aligned}
&\{(b_1, b_2) : B(z) \neq 0, |z| \leq 1\} \\
=&\{(b_1, b_2) : b_2 \pm b_1 > -1, |b_2| < 1\}
\end{aligned}
$$

自协方差

$$
\begin{aligned}
\gamma_0 =&\sigma^2(1 + b_1^2 + b_2^2) \\
\gamma_1 =&\sigma^2(b_1 + b_1 b_2) \\
\gamma_2 =&\sigma^2 b_2 \\
\gamma_k =&0, \quad k > 2
\end{aligned}
$$

自相关系数

$$
\begin{aligned}
\rho_1 =&\frac{b_1 + b_1 b_2}{1 + b_1^2 + b_2^2}, \quad \rho_2 = \frac{b_2}{1 + b_1^2 + b_2^2} \\
\rho_k =&0, \quad k > 2
\end{aligned}
$$

谱密度

$$
f(\lambda) = \frac{\sigma^2}{2\pi} \left|1 + b_1 e^{i\lambda} + b_2 e^{i\,2\lambda}\right|^2
$$

12.9.4 MA(2) 实例

MA(2) 的一个实际例子：

$$
X_t = \varepsilon_t - 0.36\varepsilon_{t-1} + 0.85\varepsilon_{t-2}
$$

特征根为 $1.084652 e^{\pm i 1.374297}$。

如下程序求特征根、特征根的模和辐角：

```r
z <- polyroot(c(1, -0.36, 0.85))
Mod(z)
```

```
## [1] 1.084652 1.084652
```

```r
Arg(z)
```

```
## [1]  1.374297 -1.374297
```

自协方差函数：

$$
\begin{aligned}
\gamma_0 &= \sigma^2(1 + b_1^2 + b_2^2) && = 7.4084 \\
\gamma_1 &= \sigma^2(b_1 + b_1 b_2) && = -2.664 \\
\gamma_2 &= \sigma^2 b_2 && = 3.4 \\
\gamma_k & && = 0, \quad k > 2
\end{aligned}
$$

自相关函数：

$$(\rho_1, \rho_2) = (-0.3596, 0, 4589))$$

下面产生了长度 $n = 120$ 的模拟数据，时间序列图见图12.8，理论谱密度见图12.9。

```r
n <- 120
sigma <- 2
b <- c(-0.36, 0.85)
x <- ma.gen(n, sigma=2, a=b)
plot(x, main="MA(2) Series: b=(-0.36,0.85)")
```

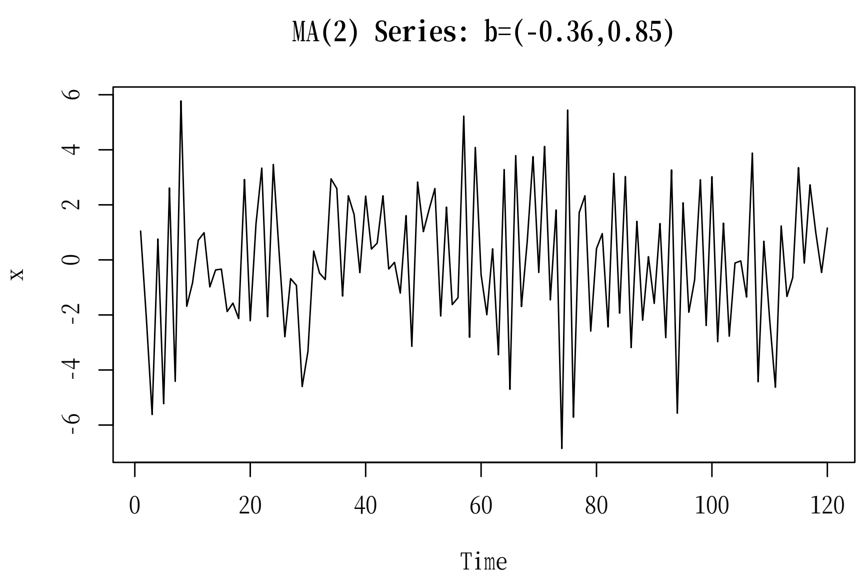


图 12.8: MA(2) 模型模拟数据

```
ma.true.spectrum(a=b, tit="MA(2) Spectral Density: b=(-0.36,0.85)")
```

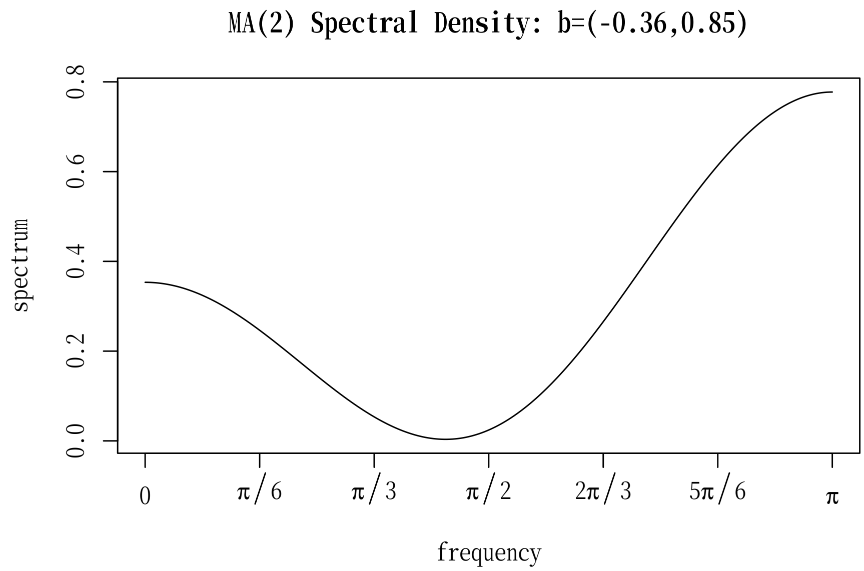


图 12.9: MA(2) 模型理论谱密度

下面的程序尝试利用前面的矩阵递推计算方法，首先计算理论自协方差函数值：

```
sigma <- 2
b <- c(-0.36, 0.85)
gms <- numeric(3)
gms[1] <- sigma^2 * (1 + b[1]^2 + b[2]^2)
gms[2] <- sigma^2 * (b[1] + b[1]*b[2])
gms[3] <- sigma^2 * b[2]
names(gms) <- 0:2
cat("======== Autocovariances ==========\n")
```

```
## ======== Autocovariances ==========
```

```
print(gms)
```

```
##       0       1       2
```

```
## 7.4084 -2.6640 3.4000
```

下面利用递推极限估计模型参数，极限 k 值分别取 6, 12, ……，100。

```
ks <- c(6, 12, 20, 30, 40, 50, 60, 100)
nks <- length(ks)
results <- matrix(0, nrow=nks, ncol=3)
rownames(results) <- ks
for(ii in seq(along=ks)){
  k <- ks[ii]
  res <- ma.solve(gms, k=k)
  results[ii,] <- c(res$b[1], res$b[2], res$s2)
}
cat("======== Solving MA coefficients ========\n")
```

```
## ======== Solving MA coefficients ========
```

```
cat("True coefficients: ", b, "\n")
```

```
## True coefficients:  -0.36 0.85
```

```
cat("True white noise variance =", sigma^2, "\n")
```

```
## True white noise variance = 4
```

```
cat("Coefficients estimates using true autocovariance and k=6,12,...:\n")
```

```
## Coefficients estimates using true autocovariance and k=6,12,...:
```

```
print(results)
```

```
##          [,1]      [,2]     [,3]
## 6   -0.3367354 0.7514921 4.524332
## 12  -0.3527358 0.8233976 4.129233
## 20  -0.3587046 0.8421336 4.037364
## 30  -0.3597471 0.8486942 4.006155
## 40  -0.3599178 0.8497082 4.001374
## 50  -0.3599929 0.8499442 4.000262
## 60  -0.3599973 0.8499899 4.000047
```

```
## 100 -0.3600000 0.8500000 4.000000
```

可见极限 k 值取到 100 时逼近精度已经基本在小数点后第 8 位。当然，如果自协方差函数时估计值，估计误差会主要来自数据的随机性。

12.10 附录：充要条件

设平稳列 $\{X_t\}$ 有恒正谱密度，则 $\{X_t\}$ 是可逆 MA $\iff$ $\{X_t\}$ 自协方差 q 后截尾 $\iff$ 谱密度可以写成

$$f(\lambda) = \frac{\sigma^2}{2\pi}|B(e^{-i\lambda})|^2$$

(其中 $B(\cdot)$ 根都在单位圆外)。见谢衷洁《时间序列分析》(谢衷洁, 1990) P89 定理 2.10。当自协方差 q 后截尾时必为 MA 序列（不一定可逆），见 (谢衷洁, 1990) P92 定理 2.11，这一点不需要假设有谱密度 (自协方差截尾推出有谱密度)。单位圆上有复根的话必为共轭出现且有偶数重。

自协方差 q 后截尾必为广义 MA(q)(无根条件)，证明见 (Brockwell & Davis, 1987) §3.2 Proposition 3.2.1. 用了新息分解，正交分解来证明。$b_1, b_2, \dots, b_q$ 是关于新息张成的空间上的投影系数。

Chapter 13

自回归滑动平均模型

13.1 ARMA(p,q) 模型及其平稳解

定义 13.1 (ARMA 模型)**.** 设 $\{\varepsilon_t\}$ 是 WN$(0,\sigma^2)$, 实系数多项式 $A(z)$ 和 $B(z)$ 没有公共根, 满足 $b_0=1$, $a_pb_q\neq 0$ 和

$$A(z)=1-\sum_{j=1}^{p}a_jz^j\neq 0, |z|\leq 1, \tag{13.1}$$

$$B(z)=\sum_{j=0}^{q}b_jz^j\neq 0,\ |z|<1, \tag{13.2}$$

就称差分方程：

$$X_t=\sum_{j=1}^{p}a_jX_{t-j}+\sum_{j=0}^{q}b_j\varepsilon_{t-j},\quad t\in\mathbb{Z}, \tag{13.3}$$

是一个**自回归滑动平均模型**, 简称为 ARMA(p,q) 模型. 称满足(13.3)的平稳序列 $\{X_t\}$ 为**平稳解**或 **ARMA(p,q) 序列**.

模型写成

$$A(\mathscr{B})X_t=B(\mathscr{B})\varepsilon_t,\quad t\in\mathbb{Z} \tag{13.4}$$

$A^{-1}(z)B(z)$ 在 $|z| < \rho$ 解析 ($1 < \rho < \min\{|z_j|\}$, $\{z_j\}$ 为 $A(z)$ 的所有根)，可以 Taylor 展开

$$\Psi(z) \triangleq A^{-1}(z)B(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \le \rho \tag{13.5}$$

易见 $\psi_j = o(\rho^{-j})$,

$$A^{-1}(\mathscr{B})B(\mathscr{B})\varepsilon_t = \Psi(\mathscr{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

是线性平稳列。两边用 $A(\mathscr{B})$ 作用，根据7中线性滤波的逆的定理7.2知

$$A(\mathscr{B})\Psi(\mathscr{B})\varepsilon_t = A(\mathscr{B})A^{-1}(\mathscr{B})B(\mathscr{B})\varepsilon_t = B(\mathscr{B})\varepsilon_t$$

即 $\Psi(\mathscr{B})\varepsilon_t$ 是 ARMA(p,q) 模型(13.3)的解。

反之，若 $\{Y_t\}$ 是(13.3)的一个平稳解，在(13.3)两边作用 $A^{-1}(\mathscr{B})$ 即得

$$A^{-1}(\mathscr{B})A(\mathscr{B})Y_t = Y_t = A^{-1}(\mathscr{B})B(\mathscr{B})\varepsilon_t = \Psi(\mathscr{B})\varepsilon_t$$

即

$$X_t = A^{-1}(\mathscr{B})B(\mathscr{B})\varepsilon_t = \Psi(\mathscr{B})\varepsilon_t \tag{13.6}$$

是 ARMA(p,q) 模型(13.3)的唯一平稳解。

称(13.6)中的 $\{\psi_j\}$ 为 $\{X_t\}$ 的 Wold 系数。

定理 13.1. *由(13.6)定义的平稳序列 $\{X_t\}$ 是 ARMA(p,q) 模型(13.3)的惟一平稳解.*

模型(13.3)的任意解可以写成

$$Y_t = X_t + \sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} V_{l,j} t^l \rho_j^{-t} \cos(\lambda_j t - \theta_{l,j}), t \in \mathbb{Z}, \tag{13.7}$$

其中 $\{X_t\}$ 为平稳解(13.6)，$z_1, z_2, \ldots, z_k$ 为 $A(z)$ 的全体互不相同的零点, $z_j = \rho_j e^{i\lambda_j}$ 有重数 $r(j)$. 随机变量 $V_{j,l}, \theta_{l,j}$ 由 $Y_0 - X_0, Y_1 - X_1, \ldots, Y_{p-1} - X_{p-1}$ 惟一决定.

13.2 ARMA 模型的模拟生成

(13.7)中的 $\{Y_t\}$ 与 $\{X_t\}$ 当 $t \to \infty$ 时无限接近：

$$|Y_t - X_t| \leq \sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} |V_{l,j}| t^l \rho_j^{-t}, t \to \infty \tag{13.8}$$

可以据此模拟 ARMA 模型: 取初值 $Y_{-(p-1)} = \dots = Y_{-1} = Y_0 = 0$，递推得

$$Y_t = \sum_{j=1}^{p} a_j Y_{t-j} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}, t = 1, 2, \dots, m+n$$

当 m 较大时取后一段 $Y_t, t = m+1, m+2, \dots, m+n$ 作为 ARMA(p,q) 模型的模拟数据。当 $A(z)$ 有靠近单位圆的根时 m 要取得较大。

13.3 ARMA(p,q) 序列的自协方差函数

13.3.1 用 Wold 系数表示

$\{\gamma_k\}$ 可由 Wold 系数表示:

$$\gamma_k = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \quad k = 0, 1, 2, \dots \tag{13.9}$$

由于 $\psi_j = o(\rho^{-j}), j \to \infty$，由(13.9)可得 $\gamma_k = o(\rho^{-j}), j \to \infty$。

13.3.2 Wold 系数递推公式

记 $b_j = 0, j < 0$ 或 $j > q$, $b_0 = 1$；$\psi_j = 0, j < 0$。由参数 $a_p = (a_1, \dots, a_p)^T$, $b_p = (b_1, \dots, b_q)^T$ 计算 $\{\psi_j\}$ 时可以递推

$$\psi_j = \begin{cases} 1, & j = 0, \\ b_j + \sum_{k=1}^{p} a_k \psi_{j-k}, & j = 1, 2, \dots \end{cases} \tag{13.10}$$

证明:

记 $A(z)=1-\sum_{j=1}^{p} a_j z^j=\sum_{j=0}^{p}\phi_j z^j$。注意

$$\begin{aligned}A(z)\Psi(z)=&\sum_{k=0}^{p}\phi_k z^k\sum_{j=0}^{\infty}\psi_j z^j\\=&\sum_{j=0}^{\infty}\sum_{k=0}^{p}\phi_k\psi_{j-k}z^j=B(z)\end{aligned}$$

比较系数得

$$\begin{aligned}\sum_{k=0}^{p}\phi_k\psi_{j-k}=&b_j,\quad j\geq 1\\\psi_j=&\sum_{k=1}^{p}a_k\psi_{j-k}+b_j,\quad j\geq 1\end{aligned}$$

即(13.10)成立。

13.4 ARMA(p,q) 模型的可识别性

我们将证明: 由 ARMA(p,q) 模型的自协方差函数 $\{\gamma_k\}$ 可以决定 ARMA(p,q) 模型的参数

$$(a^T,b^T,\sigma^2)=(a_1,\dots,a_p,b_1,\dots,b_q,\sigma^2)$$

引理 13.1. *设 $\{X_t\}$ 是(13.3)的平稳解. 如果又有白噪声 $\{\eta_t\}$ 和实系数多项式 $C(\mathscr{B})$,$D(\mathscr{B})$ 使得*

$$C(\mathscr{B})X_t=D(\mathscr{B})\eta_t,\quad t\in\mathbb{Z},$$

成立, 则 $C(z)$ 的阶数 $\geq p$, $D(z)$ 的阶数 $\geq q$.

这主要因为我们要求多项式 $A(z)$ 和 $B(z)$ 互素。证明略。

13.4.1 ARMA 序列的 Y-W 方程

ARMA 模型的平稳解为

$$X_t=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$$

所以

$$E(\varepsilon_{t+k}X_t)=0,\quad k>0$$

类似 AR 模型可推导 ARMA 模型的 Y-W 方程: 在模型方程两边同乘以 X_{t-k} 求期望得

$$E(X_t X_{t-k}) = \sum_{j=1}^{p} a_j E(X_{t-j} X_{t-k}) + \sum_{j=0}^{q} b_j E(\varepsilon_{t-j} X_{t-k})$$

即

$$\begin{aligned}
\gamma_k &= \sum_{j=1}^{p} a_j \gamma_{k-j} + \sum_{j=0}^{q} b_j E(\varepsilon_{t-j} \sum_{l=0}^{\infty} \psi_l \varepsilon_{t-k-l}) \\
&= \sum_{j=1}^{p} a_j \gamma_{k-j} + \sum_{j=0}^{q} b_j \psi_{j-k} \sigma^2, \quad k \in \mathbb{Z}
\end{aligned}$$

当 $k > q$ 时 $\psi_{j-k} = 0, j = 0, 1, \dots, q$，上式为

$$\gamma_k = \sum_{j=1}^{p} a_j \gamma_{k-j}, \quad k \geq q+1$$

总之

$$\gamma_k - \sum_{j=1}^{p} a_j \gamma_{k-j} = \begin{cases} \sigma^2 \sum_{j=\max(0,k)}^{q} b_j \psi_{j-k}, & k < q \\ \sigma^2 b_q, & k = q \\ 0 & k > q \end{cases} \tag{13.11}$$

对 $k > q$ 的 Y-W 方程可以写成矩阵形式：

$$\begin{bmatrix} \gamma_{q+1} \\ \gamma_{q+2} \\ \vdots \\ \gamma_{q+p} \end{bmatrix} = \begin{bmatrix} \gamma_q & \gamma_{q-1} & \cdots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q-p+2} \\ \vdots & \vdots & & \vdots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \tag{13.12}$$

对比 AR(p) 的 Y-W 方程，相当于 Γ_p 的 (i, j) 元素写成 γ_{i-j} 后给所有 $\gamma_\cdot$ 的下标都加上 q。

把系数矩阵记为 $\Gamma_{p,q}$:

$$\begin{aligned}\Gamma_{p,q} =&(\gamma_{|q+i-j|})_{i,j=1,2,\dots,p}\\ =&\begin{bmatrix} \gamma_q & \gamma_{q-1} & \cdots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q-p+2} \\ \vdots & \vdots & & \vdots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q \end{bmatrix}\end{aligned}$$

只要 $\Gamma_{p,q}$ 可逆则可解出 $a_1, \dots, a_p$。

解出 $a_1, \dots, a_p$ 后令

$$Y_t = A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z}$$

则 $\{Y_t\}$ 是一个 MA(q) 序列，其自协方差函数为 q 步截尾，且

$$\begin{aligned}\gamma_y(k) =&E(Y_t Y_{t-k})\\ =&\sum_{j=0}^{p}\sum_{l=0}^{p}\phi_j\phi_l E(X_{t-j}X_{t-l})\\ =&\sum_{j=0}^{p}\sum_{l=0}^{p}\phi_j\phi_l\gamma_{k+l-j}, \quad 0 \le k \le q\end{aligned}$$

可以用 §12的方法唯一解出 $b_1, \dots, b_q, \sigma^2$。

于是，只要 $\Gamma_{p,q}$ 可逆, 则 ARMA(p, q) 序列的自协方差函数和 ARMA(p, q) 模型的参数 (a_p^T, b_q^T, σ^2) 相互惟一决定。

13.4.2 ARMA 模型中 AR 部分的参数求解

如果 $\Gamma_{p,q}$ 可逆则由(13.12)可以解出 $a_1, \dots, a_p$。

定理 13.2. *设 $\{\gamma_k\}$ 为 ARMA(p,q) 序列 $\{X_t\}$ 的自协方差函数列, 则 $m \ge p$ 时 $\Gamma_{m,q}$ 可逆。*

证明:

用反证法然后由引理13.1导出矛盾。

设 $\Gamma_{m,q}(m\times m$ 矩阵) 不满秩，则存在 $\beta=(\beta_0,\beta_1,\dots,\beta_{m-1})^T\neq 0$ 使得 $\Gamma_{m,q}\beta=0$，即

$$\sum_{l=0}^{m-1}\beta_l\gamma_{q+k-l}=0,\quad k=0,1,\dots,m-1 \tag{13.13}$$

注意当 $k\geq m$ 时 $q+k-l>q$，所以这时 $\gamma_{q+k-l}=\sum_{j=1}^{p}a_j\gamma_{q+k-l-j}$，所以取 $k=m$ 有

$$\begin{aligned}\sum_{l=0}^{m-1}\beta_l\,\gamma_{q+k-l}=&\sum_{l=0}^{m-1}\beta_l\sum_{j=1}^{p}a_j\gamma_{q+k-l-j}\\=&\sum_{j=1}^{p}a_j\sum_{l=0}^{m-1}\beta_l\,\gamma_{q+(k-j)-l}\\=&0\quad(\text{由 (13.12) 及}0\leq k-j=m-j\leq m-1)\end{aligned}$$

递推得上式当 $k>m$ 时也成立。因此

$$\sum_{l=0}^{m-1}\beta_l\,\gamma_{q+k-l}=0,\quad k\geq 0.$$

令 $Y_t=\sum_{l=0}^{m-1}\beta_lX_{t-l}$ 则 $\{Y_t\}$ 是零均值平稳列，利用

$$E(Y_tX_{t-q-k})=\sum_{l=0}^{m-1}\beta_l\gamma_{q+k-l}=0,\quad k\geq 0$$

可知 $\{Y_t\}$ 的自协方差 $q-1$ 步截尾:

$$E(Y_tY_{t-q-k})=0,\quad k\geq 0$$

所以 $\{Y_t\}$ 是 $\mathrm{MA}(q')(q'\leq q-1)$ 序列，存在 $\{\eta_t\}\sim\mathrm{WN}(0,s^2)$ 使得

$$\sum_{l=0}^{m-1}\beta_lX_{t-l}=\sum_{j=0}^{q'}d_j\eta_{t-j}$$

与引理13.1矛盾。

13.4.3 ARMA 模型的一个充分条件

定理 13.3. 设零均值平稳序列 $\{X_t\}$ 有自协方差函数 $\{\gamma_k\}$. 又设实数 $a_1, a_2, \cdots, a_p$ $(a_p \neq 0)$ 使得 $A(z) = 1 - \sum_{j=1}^p a_j z^j$ 满足最小相位条件, 另外

$$\gamma_k - \sum_{j=1}^p a_j \gamma_{k-j} = \begin{cases} c \neq 0, & k = q, \\ 0, & k > q, \end{cases} \tag{13.14}$$

则 $\{X_t\}$ 是一个 $ARMA(p', q')$ 序列, 其中 $p' \leq p$, $q' \leq q$. \

证明: 设 $Y_t = A(\mathscr{B})X_t = X_t - \sum_{j=1}^p a_j X_{t-j}$. 则 $\{Y_t\}$ 是零均值平稳序列, 满足

$$E(Y_t X_{t-k}) = \gamma_k - \sum_{j=1}^p a_j \gamma_{k-j} = \begin{cases} c \neq 0, & k = q, \\ 0, & k > q. \end{cases}$$

所以有

$$\begin{aligned} \gamma_y(k) =& E(Y_t Y_{t-k}) = E[Y_t(X_{t-k} - \sum_{j=1}^p a_j X_{t-k-j})] \\ =& \begin{cases} c \neq 0, & k = q, \\ 0, & k > q. \end{cases} \end{aligned}$$

说明 $\{Y_t\}$ 的自协方差函数是 q 后截尾的.

由定理12.2知道, $\{Y_t\}$ 为一个 MA(q) 序列, 即存在单位圆内没有根的 q 阶实系数多项式 $B(z)$ 使得 $B(0) = b_0 = 1$ 和

$$A(\mathscr{B})X_t = Y_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z}, \tag{13.15}$$

其中 $\{\varepsilon_t\}$ 是 WN$(0, \sigma^2)$.

如果 $A(z)$ 和 $B(z)$ 没有公因子, 上述模型就是所需要的 ARMA(p, q) 模型. 否则设公因子是 $C(z)$, 则有 $A(z) = C(z)A'(z)$, $B(z) = C(z)B'(z)$. 这时(13.15)变成

$$C(\mathscr{B})A'(\mathscr{B})X_t = C(\mathscr{B})B'(\mathscr{B})\varepsilon_t.$$

两边乘 $C^{-1}(\mathscr{B})$(显然 $C(z)$ 也满足最小相位条件) 后得到所需 ARMA(p', q') 模型:

$$A'(\mathscr{B})X_t = B'(\mathscr{B})\varepsilon_t.$$

13.5 ARMA 序列的谱密度

由于 ARMA 序列的 $\{\gamma_k\}$ 绝对可和，以及平稳解的线性序列表达式，可得 ARMA(p,q) 序列 (2.6) 有谱密度

$$f(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-ik\lambda} \tag{13.16}$$

$$= \frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty}\psi_j e^{ij\lambda}\right|^2 = \frac{\sigma^2}{2\pi}\left|\frac{B(e^{i\lambda})}{A(e^{i\lambda})}\right|^2 \tag{13.17}$$

形如(13.17)的谱密度被称为**有理谱密度**.

13.6 可逆 ARMA 序列

定义 13.2. 在 ARMA(p,q) 模型的定义 13.1 中, 如果进一步要求 $B(z)$ 在单位圆上无根:

$$B(z) = 1 + \sum_{j=1}^{q} b_j z^j \neq 0, |z| \leq 1 \tag{13.18}$$

则称 ARMA(p,q) 模型(13.18)为可逆的 ARMA 模型, 称相应的平稳解为可逆的 ARMA(p,q) 序列.

从定理12.3(最小序列的谱密度条件）知道可逆的 ARMA(p,q) 序列是最小序列.

对于可逆的 ARMA(p,q) 模型(13.18)，由于 $B^{-1}(z)A(z)$ 在 $\{z : |z| \leq \rho\}$ $(\rho > 1)$ 内解析, 所以有 Taylor 展式：

$$B^{-1}(z)A(z) = \sum_{j=0}^{\infty}\varphi_j z^j, \quad |z| \leq \rho, \tag{13.19}$$

其中 $|\varphi_j| = o(\rho^{-j})$, 当 $j \to \infty$, 从而可以定义 $B^{-1}(\mathscr{B})A(\mathscr{B}) = \sum_{j=0}^{\infty}\varphi_j\mathscr{B}^j$. 在(13.19)两边乘以 $B^{-1}(\mathscr{B})$, 得到：

$$\varepsilon_t = B^{-1}(\mathscr{B})A(\mathscr{B})X_t = \sum_{j=0}^{\infty}\varphi_j X_{t-j}, \quad t \in \mathbb{Z}. \tag{13.20}$$

(13.20)是(13.6)的逆转形式, 表明可逆 ARMA(p,q) 序列和它的噪声序列可以相互线性表示.

13.7 ARMA 模型例子

先给出一些有关的自定义 R 函数。

```
## ARMA theoretical spectrum given ARMA coefficients
arma.true.spectrum <- function(a, b, ngrid=256, sigma=1,
                               tit="True ARMA Spectral Density",
                               plot.it=TRUE){
  p <- length(a)
  q <- length(b)
  freqs <- seq(from=0, to=pi, length=ngrid)
  spec1 <- numeric(ngrid)
  spec2 <- numeric(ngrid)
  for(ii in seq(ngrid)){
    spec1[ii] <- 1 + sum(complex(mod=b, arg=freqs[ii]*seq(q)))
    spec2[ii] <- 1 - sum(complex(mod=a, arg=freqs[ii]*seq(p)))
  }
  spec = sigma^2 / (2*pi) * abs(spec1)^2 / abs(spec2)^2
  if(plot.it){
    plot(freqs, spec, type='l',
         main=tit,
         xlab="frequency", ylab="spectrum",
         axes=FALSE)
    axis(2)
    axis(1, at=(0:6)/6*pi,
         labels=c(0, expression(pi/6),
           expression(pi/3), expression(pi/2),
           expression(2*pi/3), expression(5*pi/6), expression(pi)))
  }
  box()
  invisible(list(frequencies=freqs, spectrum=spec,
       ar.coefficients=a, ma.coefficients=b,
       sigma=sigma))
}
```

```
## simulate ARMA(p, q) model.
## a is vector a_1, \dots, a_p
## b is vector b_1, \dots, b_q
## Model is
##   X_t = a_1 X_{t-1} + \dots + a_p X_{t-p}
##         + \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q}
## \Var(\epsilon_t) = sigma^2
arma.gen <- function(n, a, b, sigma=1.0,
                     by.roots.ar=FALSE, by.roots.ma=FALSE,
                     plot.it=FALSE, n0=1000,
                     x0=numeric(length(a))){
  n2 <- n0 + n
  p <- length(a)
  ## first generate n0+n MA(q) series
  eps <- ma.gen(n2, b, sigma=sigma,
                by.roots=by.roots.ma,
                plot.it=FALSE)
  b <- attr(eps, "b")
  if(by.roots.ar){
    require(polynom)
    cf <- Re(c(poly.calc(a)))
    cf <- cf / cf[1]
    a <- -cf[-1]
  }

  ##set.seed(1)
  x2 <- filter(eps, a, method="recursive", side=1, init=x0)
  x <- x2[(n0+1):n2]
  x <- ts(x)
  attr(x, "model") <- "ARMA"
  attr(x, "a") <- a
  attr(x, "b") <- b
```

```r
  attr(x, "sigma") <- sigma
  if(plot.it) plot(x)
  x
}

## Wold coefficients for the ARMA model
arma.Wold <- function(n, a, b=numeric(0)){
  p <- length(a)
  q <- length(b)
  arev <- rev(a)
  psi <- numeric(n)
  psi[1] <- 1
  for(j in seq(n-1)){
    if(j <= q) bj=b[j]
    else bj=0
    psis <- psi[max(1, j+1-p):j]
    np <- length(psis)
    if(np < p) psis <- c(rep(0,p-np), psis)
    psi[j+1] <- bj + sum(arev * psis)
  }

  psi
}

## Calculate theoretical autocovariance function
## of ARMA model using Wold expansion
arma.gamma.by.Wold <- function(n, a, b=numeric(0), sigma=1){
  nn <- n + 100
  psi <- arma.Wold(nn, a, b)
  gam <- numeric(n)
  for(ii in seq(0, n-1)){
    gam[ii+1] <- sum(psi[1:(nn-ii)] * psi[(ii+1):nn])
  }
```

```
  gam <- (sigma^2) * gam
  gam
}
arma.gamma <- arma.gamma.by.Wold

## Solve ARMA parameters given autocovariance functions
arma.solve <- function(gms, p, q){
  Gpq <- matrix(0, nrow=p, ncol=p)
  for(ii in seq(p)) for(jj in seq(p)){
    Gpq[ii,jj] <- gms[abs(q + ii - jj)+1]
  }
  gs <- gms[(q+1+1):(q+p+1)]
  a <- solve(Gpq, gs)
  aa <- c(-1, a)

  gys <- numeric(q+1)
  for(k in seq(0, q)){
    gys[k+1] <- sum(c(outer(0:p,0:p,
                            function(ii,jj) aa[ii+1] * aa[jj+1]
                            * gms[abs(k+jj-ii)+1])))
  }

  res <- ma.solve(gys)
  b <- res$b
  sigma <- sqrt(res$s2)

  list(a=a, b=b, sigma=sigma)
}

## simulate MA(q) model.
## a is vector a_1, \dots, a_q
## Model is X_t = eps_t + a_1 eps_{t-1} + \dots + a_q eps_{t-q}
## \Var(\epsilon_t) = sigma^2
```

```r
## This version uses the filter function.
ma.gen <- function(n, a, sigma=1.0, by.roots=FALSE,
                   plot.it=FALSE){
  if(by.roots){
    require(polynom)
    cf <- Re(c(poly.calc(a)))
    cf <- cf / cf[1]
    a <- cf
  }
  q <- length(a)
  n2 <- n+q
  eps <- rnorm(n2, 0, sigma)
  x2 <- filter(eps, c(1,a), method="convolution", side=1)
  x <- x2[(q+1):n2]
  x <- ts(x)
  attr(x, "model") <- "MA"
  attr(x, "b") <- a
  attr(x, "sigma") <- sigma
  if(plot.it) plot(x)
  x
}

## Given \gamma_0, \gamma_1, \dots, \gamma_q,
## Solve MA(q) coefficients b_1, \dots, b_q, \sigma^2
## Using Li Lei's algorithm.
## Input: gms -- \gamma_0, \gamma_1, \dots, \gamma_q
ma.solve <- function(gms, k=100){
  q <- length(gms)-1
  if(q==1){
    rho1 <- gms[2] / gms[1]
    b <- (1 - sqrt(1 - 4*rho1^2))/(2*rho1)
    s2 <- gms[1] / (1 + b^2)
    return(list(b=b, s2=s2))
```

```
  }

  A <- matrix(0, nrow=q, ncol=q)
  for(j in seq(2,q)){
    A[j-1,j] <- 1
  }
  cc <- numeric(q); cc[1] <- 1
  gamma0 <- gms[1]
  gammas <- numeric(q+k)
  gammas[1:(q+1)] <- gms
  gamq <- gms[-1]
  Gammak <- matrix(0, nrow=k, ncol=k)
  for(ii in seq(k)){
    for(jj in seq(k)){
      Gammak[ii,jj] <- gammas[abs(ii-jj)+1]
    }
  }

  Omk <- matrix(0, nrow=q, ncol=k)
  for(ii in seq(q)){
    for(jj in seq(k)){
      Omk[ii,jj] <- gammas[ii+jj-1+1]
    }
  }
  PI <- Omk %*% solve(Gammak, t(Omk))

  s2 <- gamma0 - c(t(cc) %*% PI %*% cc)
  b <- 1/s2 * c(gamq - A %*% PI %*% cc)
  return(list(b=b, s2=s2))
}
```

13.7.1 ARMA(4,2) 例子

ARMA(4,2):

$$\begin{aligned} a_1 &= -0.9, \quad a_2 = -1.4, \\ a_3 &= -0.7, \quad a_4 = -0.6; \\ b_1 &= 0.5, \quad b_2 = -0.4. \end{aligned} \tag{13.21}$$

$A(z)$ 的根为 $1.1380e^{\pm 2.2062i}$, $1.1344e^{\pm 1.4896i}$, $B(z)$ 的两个实根为 $2.3252, -1.0752$。此时间序列有两个频率成分，对应于 AR 部分的特征根的辐角。

模拟生成长度 $n = 80$ 的样本（结果见图13.1）：

```
set.seed(1)
n <- 80
a <- c(-0.9, -1.4, -0.7, -0.6)
b <- c(0.5, -0.4)

x <- arma.gen(n, a, b, plot.it=F)

ts.plot(x, main="ARMA(4,2) Series")
```

对模拟数据估计 ACF 并作图（结果见图13.2）：

```
acf(x)
```

模型的理论谱密度（结果见图13.3）：

```
arma.true.spectrum(a, b)
```

模型的理论自协方差函数（结果见图13.4）：

```
ng <- 21
gams <- arma.gamma(ng, a, b, sigma=1)
plot(0:(ng-1), gams, type="h",
     main="Theoretical gamma by Wold",
     xlab="k", ylab=expression(gamma[k]))
abline(h=0)
```

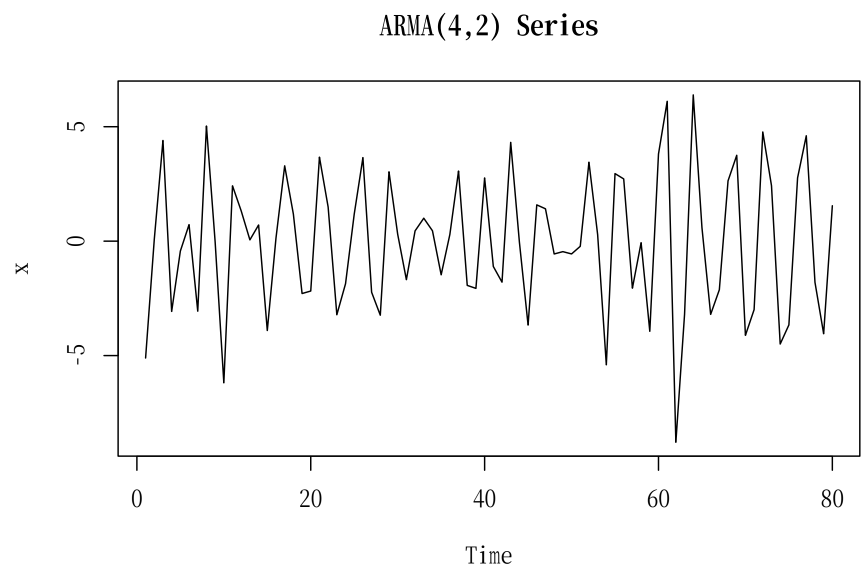


图 13.1: ARMA(4,2) 模拟数据

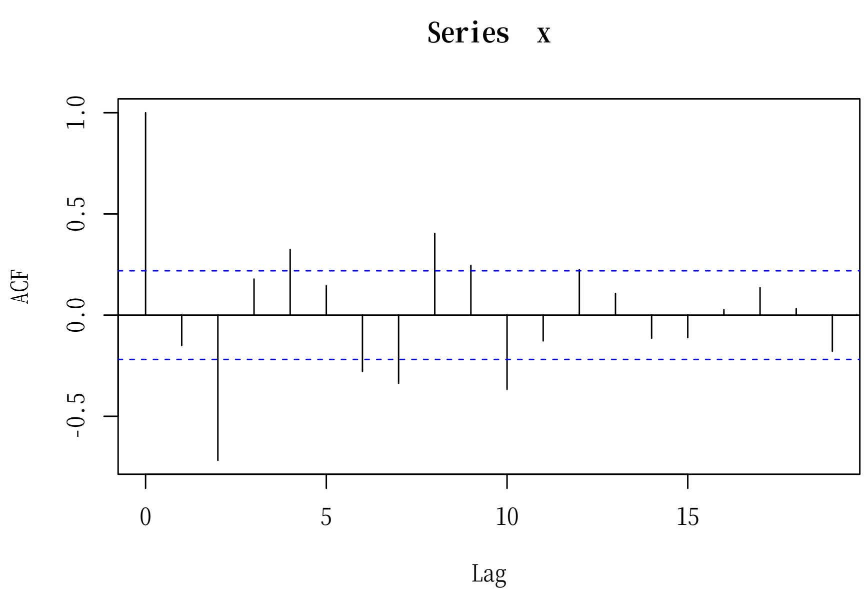


图 13.2: ARMA(4,2) 数据的 ACF

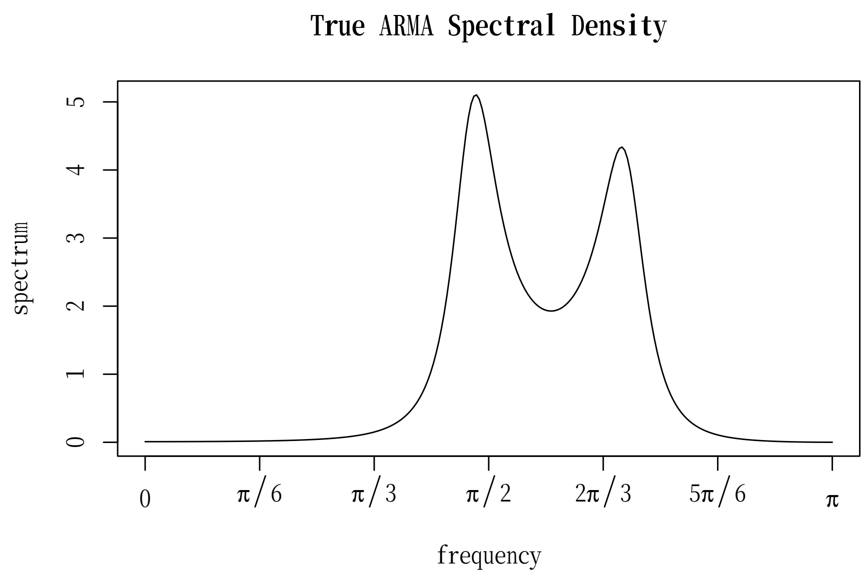


图 13.3: ARMA(4,2) 数据的理论谱密度

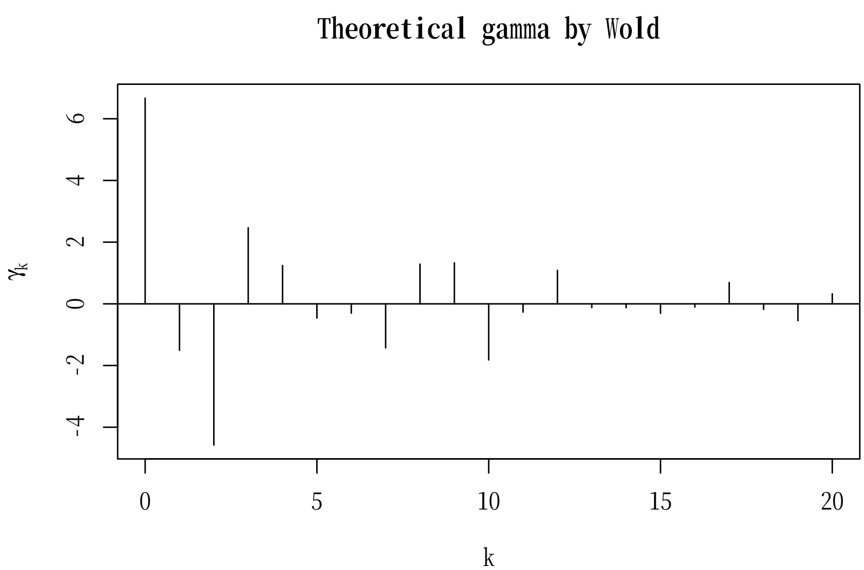


图 13.4: ARMA(4,2) 数据的理论自协方差函数

```r
cat("\n====== Autocovariances ========\n")
```

```
##
## ====== Autocovariances ========
```

```r
names(gams) <- 0:(ng-1)
print(round(gams, 4))
```

```
##       0       1       2       3       4       5       6       7       8       9
##  6.6708 -1.5078 -4.5792  2.4672  1.2433 -0.4630 -0.3035 -1.4293  1.2894  1.3309
##      10      11      12      13      14      15      16      17      18      19
## -1.8203 -0.2699  1.0861 -0.1239 -0.1279 -0.3097 -0.1071  0.6939 -0.1810 -0.5477
##      20
##  0.3249
```

下面从理论自协方差函数反解模型参数。

```r
cat("\n===== Solve ARMA from ACV ======\n")
```

```
##
## ===== Solve ARMA from ACV ======
```

```r
res <- arma.solve(gams, p=4, q=2)
print(res)
```

```
## $a
## [1] -0.9 -1.4 -0.7 -0.6
##
## $b
## [1]  0.4999999 -0.4000000
##
## $sigma
## [1] 1
```

用样本中估计的自协方差函数反解模型参数并与用理论值反解的结果对比：

```r
gams2 <- acf(x, type="covariance", plot=FALSE)$acf
res2 <- arma.solve(gams2, p=4, q=2)
print(res2)
```

```
## $a
## [1] -1.0102537 -1.5340719 -0.8871690 -0.7297601
##
## $b
## [1] 0.63530816 0.03771917
##
## $sigma
## [1] 1.358952
```

13.7.2 ARMA(2,2) 例子

设 $\{X_t\} \sim \text{ARMA}(2,2)$, 已知

$$(\gamma_0, \gamma_1, \cdots, \gamma_4) = (4.61,\ -1.06,\ 0.29,\ 0.69,\ -0.12)$$

要反解 ARMA 参数。

```
gams3 <- c(4.61, -1.06, 0.29, 0.69, -0.12)
res3 <- arma.solve(gams3, p=2, q=2)
print(res3)
```

```
## $a
## [1]  0.08939301 -0.62648682
##
## $b
## [1] -0.3334025  0.8157936
##
## $sigma
## [1] 2.002966
```

```
z1 <- polyroot(c(1, -res3[["a"]]))
cat("AR roots: ", Mod(z1[1]), "exp(+- i ", Arg(z1[1]), ")", "\n")
```

```
## AR roots:  1.263409 exp(+- i  1.514296 )
```

```
z2 <- polyroot(c(1, res3[["b"]]))
cat("MA roots: ", Mod(z2[1]), "exp(+- i ", Arg(z2[1]), ")", "\n")
```

```
## MA roots:  1.107159 exp(+- i  1.385167 )
```

解出的模型为

$$X_t = 0.0894X_{t-1} - 0.6265X_{t-2} + \varepsilon_t - 0.3334\varepsilon_{t-1} + 0.8158\varepsilon_{t-2},\ \{\varepsilon_t\} \sim \text{WN}(0, 2.0030^2)$$

AR 部分的特征根为 $1.26e^{\pm i1.51}$, MA 部分的特征根为 $1.11e^{\pm i1.39}$。

13.8 附录：ARAM(1,1) 例子

对 ARMA(1,1)

$$X_t = aX_{t-1} + \varepsilon_t + b\varepsilon_{t-1}$$

由 Wold 系数递推公式得 $\psi_0 = 1$,

$$\begin{aligned}
\psi_1 =& b + a\psi_{1-1} = a + b \\
\psi_j =& a\psi_{j-1} = \cdots = a^{j-1}\psi_1 = a^{j-1}(a+b),\ j = 2, 3, \dots
\end{aligned}$$

于是 ARMA(1,1) 的平稳解可以写成

$$X_t = \varepsilon_t + (a+b)\sum_{j=1}^{\infty} a^{j-1}\varepsilon_{t-j}$$

从 Wold 系数计算协方差函数得

$$
\begin{aligned}
\gamma_0 =& \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma^2 \left\{ 1 + \sum_{j=1}^{\infty} (a+b)^2 a^{2(j-1)} \right\} \\
=& \sigma^2 \left\{ 1 + \frac{(a+b)^2}{1-a^2} \right\} = \sigma^2 \frac{1+2ab+b^2}{1-a^2} \\
\gamma_1 =& \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} = \sigma^2 \left\{ \psi_1 + \sum_{j=1}^{\infty} a(a+b)^2 a^{2(j-1)} \right\} \\
=& \sigma^2 \frac{(a+b)(1+ab)}{1-a^2} = \sigma^2 \frac{a+b+a^2b+ab^2}{1-a^2} \\
\gamma_k =& \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} = \sigma^2 a \sum_{j=0}^{\infty} \psi_j \psi_{j+k-1} = a\gamma_{k-1} = \cdots \\
=& a^{k-1}\gamma_1 = a^{k-1}\sigma^2 \frac{(a+b)(1+ab)}{1-a^2}
\end{aligned}
$$

也可以利用 YW 方程计算协方差函数。对 $k = 0, 1, 2, \dots$，模型方程两边同乘以 X_{t-k} 取期望，因为 $E\varepsilon_t X_{t-k} = 0$ 对 $k \geq 1$，有

$$
\begin{aligned}
\gamma_0 =& a\gamma_1 + E(X_t\varepsilon_t) + bE(X_t\varepsilon_{t-1}) \\
\gamma_1 =& a\gamma_0 + 0 + bE(X_{t-1}\varepsilon_{t-1}) \\
\gamma_k =& a\gamma_{k-1} + 0 + b \cdot 0, k = 2, 3, \dots
\end{aligned}
$$

由 Wold 表示可知 $E(X_t\varepsilon_{t-j}) = \sigma^2\psi_j$，所以

$$
\begin{aligned}
\gamma_0 =& a\gamma_1 + \sigma^2[1 + b(a+b)] \\
\gamma_1 =& a\gamma_0 + \sigma^2 b
\end{aligned}
$$

将第二式代入第一式就可求解 γ_0, γ_1 为

$$
\begin{aligned}
\gamma_0 =& \sigma^2 \frac{1+2ab+b^2}{1-a^2}, \\
\gamma_1 =& \sigma^2 \frac{a+b+a^2b+ab^2}{1-a^2} \\
\gamma_k =& a\gamma_{k-1} = a^{k-1}\gamma_1, k = 2, 3, \dots
\end{aligned}
$$

自相关函数为

$$
\begin{aligned}
\rho_1 =& \frac{(a+b)(1+ab)}{1+2ab+b^2} \\
\rho_k =& a\rho_{k-1} = a^{k-1}\rho_1, k = 2, 3, \dots
\end{aligned}
$$

下面求出 ARMA(1,1) 的偏相关函数前几项。

$$a_{11} = \rho_1 = \frac{(a+b)(1+ab)}{1+2ab+b^2}$$

a_{22} 满足方程

$$\begin{pmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$$

两边除以 γ_0 得

$$\begin{pmatrix} \rho_0 & \rho_1 \\ \rho_1 & \rho_0 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$$

其中 $\rho_0 = 1$, $\rho_2 = a\rho_1$，即

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$$

由 Cramer 法则，

$$a_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & a\rho_1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1(a-\rho_1)}{1-\rho_1^2}$$

类似地，

$$a_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1(a-\rho_1)^2}{1+2a\rho_1^3-\rho_1^2(2+a^2)}$$

13.9 附录：ARMA 模型的谱条件

ARMA 模型的谱表示:

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{B(e^{-i\lambda})}{A(e^{-i\lambda})} dZ_\varepsilon(\lambda)$$

其中 $\frac{B(z)}{A(z)}$ 叫做 ARMA 模型的极大解析函数。

AMRA 模型的谱密度:

设 $A(\cdot)$, $B(\cdot)$ 根都在单位圆外，$\{\varepsilon_t\}$ 是 $\text{WN}(0, \sigma^2)$, 则平稳列 $\{X_t\}$ 是可逆 ARMA 模型

$$A(\mathscr{B}) X_t = B(\mathscr{B}) \varepsilon_t$$

的充分必要条件是它有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{B(e^{-i\lambda})}{A(e^{-i\lambda})} \right|^2$$

见 (谢衷洁, 1990) P.76 定理 2.4。

证明

必要性教材中已证明。设充分性条件成立，这时设

$$\frac{A(z)}{B(z)} = \sum_{j=0}^\infty d_j z^j,\ |z| \leq \rho_2\ (\rho_2 > 1)$$

令

$$\eta_t = \sum_{j=0}^\infty d_j X_{t-j}$$

则

$$A(\mathscr{B}) X_t = B(\mathscr{B}) \eta_t$$

且 $\{\eta_t\}$ 的谱密度为

$$\begin{aligned}
f_\eta(\lambda) =& \left| \sum_{j=0}^\infty d_j e^{-ij\lambda} \right|^2 \cdot \frac{\sigma^2}{2\pi} \left| \frac{B(e^{-i\lambda})}{A(e^{-i\lambda})} \right|^2 \\
=& \left| \frac{A(e^{-i\lambda})}{B(e^{-i\lambda})} \right|^2 \cdot \frac{\sigma^2}{2\pi} \left| \frac{B(e^{-i\lambda})}{A(e^{-i\lambda})} \right|^2 \\
=& \frac{\sigma^2}{2\pi}
\end{aligned}$$

即 $\{\eta_t\}$ 为白噪声，所以 $\{X_t\}$ 为可逆 ARMA 序列。

13.10 附录：ARMA 模型与 Hilbert 空间

参考 (谢衷洁, 1990) P.78。

考虑可逆 ARMA 模型

$$A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t$$

设

$$\begin{aligned}H_X =&\mathcal{L}\{X_t : t \in \mathbb{Z}\}\\H_\varepsilon =&\mathcal{L}\{\varepsilon_t : t \in \mathbb{Z}\}\\H_X(t) =&\mathcal{L}\{X_s : s \le t,\ s \in \mathbb{Z}\}\\H_\varepsilon(t) =&\mathcal{L}\{\varepsilon_s : s \le t,\ s \in \mathbb{Z}\}\end{aligned}$$

其中 $\mathcal{L}$ 表示线性闭包为 Hilbert 空间。则 $\varepsilon_t \in H_X(t)$, $\varepsilon_t \in H_X(t) \ominus H_X(t-1)$, $\{\varepsilon_t/\sigma, t \in \mathbb{Z}\}$ 是 H_X 的一组完备标准正交基。把 X_t 用标准正交基展开成 Wold 表示

$$X_t = \sigma \sum_{j=0}^{\infty} \psi_j \frac{\varepsilon_{t-j}}{\sigma}$$

展开的系数 $\sigma\psi_j$ 是 H_X 中 X_t 对正交基 $\{\varepsilon_t/\sigma, t \in \mathbb{Z}\}$ 的广义傅立叶系数

$$\sigma\psi_j =< X_t, \varepsilon_{t-j}/\sigma >,\ j = 0, 1, 2, ...$$

可逆 ARMA 模型中的白噪声 ε_t 是新息 (Wold 序列)：

$$\varepsilon_t = X_t - L(X_t|H_X(t-1))$$

(参考 (谢衷洁, 1990) P.81 定理 2.6).

设可逆 ARMA 模型 Wold 表示为

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

逆表示为

$$\varepsilon_t = \sum_{j=0}^{\infty} d_j X_{t-j}$$

则 $\{d_j\}$ 可解为

$$\begin{aligned}d_0 =&1\\d_j =&-\sum_{k=1}^{j} \psi_k d_{j-k},\ j = 1, 2, ...\end{aligned}$$

这是因为

$$\left(\sum_{k=0}^{\infty} d_k z^k\right)\left(\sum_{l=0}^{\infty} \psi_l z^l\right) = 1$$

即

$$\sum_{j=0}^{\infty}\left(\sum_{k=0}^{j}\psi_k d_{j-k}\right)z^j = 1$$

Chapter 14

广义 ARMA 模型和 ARIMA 模型介绍

14.1 广义 ARMA 模型

设 $A(z) = 1 - \sum_{j=1}^{p} a_j z^j$, $B(z) = 1 + \sum_{j=1}^{q} b_j z^j$ 是两个没有公共根的实系数多项式, $a_p b_q \neq 0$, $\{\varepsilon_t\}$ 是 $\mathrm{WN}(0, \sigma^2)$. 如果不对 $A(z)$, $B(z)$ 的根做任何其他限制, 则称差分方程:

$$A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z}, \tag{14.1}$$

为**广义 ARMA**(p,q) **模型**, 称满足(14.1)的 $\{X_t\}$ 为广义 ARMA(p,q) 序列.

如果 $A(z)$ 在单位圆上有根, 可以证明(14.1)没有平稳解.

如果 $A(z)$ 在单位圆上没有根, 则有 $0 < \rho_1 < 1 < \rho_2$ 使得复变函数 $B(z)/A(z)$ 在圆环

$$D = \{z : \ \rho_1 \leq |z| \leq \rho_2\} \tag{14.2}$$

内解析. 于是 $B(z)/A(z)$ 有 Laurent 级数展开

$$A^{-1}(z)B(z) = \sum_{j=-\infty}^{\infty} c_j z^j, \ z \in D, \tag{14.3}$$

存在 $\rho > 1$ 使 $c_j = o(\rho^{-|j|}), j \to \pm\infty$。

这样, 从(14.1)可以得到惟一的平稳解

$$X_t = A^{-1}(\mathscr{B})B(\mathscr{B})\varepsilon_t = \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \tag{14.4}$$

如果 $A(z)$ 在单位圆内有根, 由(14.4)定义的平稳序列是白噪声的双边无穷滑动和。这个平稳序列某种意义下是不合理的, 因为 t 时的观测受到了 t 以后的干扰的影响.

再由差分方程的理论知道, 这时(14.1)的其他解都随着时间的增加而加速振荡. 为此, 人们把这时的广义 ARMA 模型称为“爆炸模型”。

14.2 ARIMA(p, d, q) 模型

ARIMA(p, d, q) 模型是 AR 部分有单位特征根（即 1）的广义 ARMA 模型，d 为单位特征根的重数。除了单位根以外，要求 AR 部分根都在单位圆外，MA 部分单位圆内没有根。

ARIMA(p, d, q) 序列是 d 阶差分后服从 ARMA(p, q) 模型的非平稳时间序列。设 d 是一个正整数, 如果

$$Y_t = (1-\mathscr{B})^d X_t = \sum_{k=0}^{d} C_d^k (-1)^k X_{t-k}, \quad t \in \mathbb{Z}, \tag{14.5}$$

是一个 ARMA(p, q) 序列, 其模型 MA 部分的特征多项式没有等于 1 的特征根，则称 $\{X_t\}$ 是一个求和自回归滑动平均 (p, d, q) 序列. 简称为 ARIMA(p, d, q) 序列,
其中 C_d^k 是二项式系数.

例 14.1. ARIMA$(p, 1, q)$。

这时 $d = 1$,

$$Y_t = (1-\mathscr{B})X_t = X_t - X_{t-1}$$

为一个 ARMA(p, q) 序列.

给定初值 X_0 后, 有

$$
\begin{aligned}
X_t &= X_{t-1} + Y_t && (14.6)\\
&= X_{t-2} + Y_{t-1} + Y_t && (14.7)\\
&= \cdots && (14.8)\\
&= X_0 + Y_1 + Y_2 + \cdots + Y_t && (14.9)\\
&= X_0 + \sum_{j=1}^{t} Y_j,\ t \geq 1. && (14.10)
\end{aligned}
$$

(14.10)是

$$(1-\mathscr{B})X_t = Y_t$$

的解，也是其通解。

如果 $\{Y_t\}$ 是相应的 ARMA(p,q) 模型的通解，则(14.10)也是相应的 ARIMA$(p,1,q)$ 模型的通解。所以, 求和 ARIMA$(p,1,q)$ 序列不是平稳序列.

对正整数 d, 求和 ARIMA(p,d,q) 序列也都不是平稳序列.

例 14.2. ARIMA$(p,2,q)$。

这时 $d=2$,

$$Y_t = (1-\mathscr{B})^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

为一个 ARMA(p,q) 序列.

给定初值 X_{-1}, X_0 后, 有

$$X_t - X_{t-1} = X_{t-1} - X_{t-2} + Y_t,\ t \geq 1$$

两边对 $t = 1, 2, \dots, n_1$ 求和得

$$X_{n_1} - X_0 = X_{n_1-1} - X_{-1} + \sum_{i=1}^{n_1} Y_i$$

整理后得

$$X_{n_1} - X_{n_1-1} = X_0 - X_{-1} + \sum_{i=1}^{n_1} Y_i$$

两边再对 $n_1 = 1, 2, \dots, t$ 求和，得

$$X_t - X_0 = t(X_0 - X_{-1}) + \sum_{n_1=1}^{t} \sum_{i=1}^{n_1} Y_i,\ t \geq q$$

或

$$X_t = X_0 + t(X_0 - X_{-1}) + \sum_{n_1=1}^{t} \sum_{i=1}^{n_1} Y_i,\ t \geq q \tag{14.11}$$

一般地，设 $\{X_t\}$ 为 ARIMA(p,d,q) 序列，$(1-\mathscr{B})^d X_t = Y_t$，模型通解为

$$X_t = C_0 + C_1 t + \cdots + C_{d-1} t^{d-1} + \sum_{n_{d-1}=1}^{t} \cdots \sum_{n_1=1}^{n_2} \sum_{j=1}^{n_1} Y_j.$$

实际问题中有许多数据经过一或两次差分后会稳定下来. 差分运算是对数据进行预处理的常用方法之一.

14.3 单位根过程

ARIMA$(p,1,q)$ 模型称为单位根过程，相应的时间序列被称为单位根序列。

单位根过程与有一个 AR 部分特征根 $|z_j| > 1$ 但 $|z_j|$ 十分接近于 1 的平稳 ARMA 序列很难区分。

单位根过程与带有线性趋势的模型不同。单位根过程数据没有固定走势，可以称为随机趋势。带有线性趋势的模型减去线性趋势后就平稳了，单位根过程减去任何线性趋势都不平稳。

ARIMA$(0,1,0)$ 模型是最简单的单位根过程，也可以称为随机游动。随机游动 $p_t = p_{t+1} + \varepsilon_t$ 与固定趋势加扰动 $Y_t = a + bt + X_t$(其中 $\{X_t\}$ 平稳) 都能呈现出缓慢的趋势变化。区别在于：

- 随机游动的方差是线性增长的，固定趋势的观测值方差不变；
- 随机游动的扰动的影响是永久的，固定趋势的扰动的影响仅在一个时刻（如果扰动 X_t 是白噪声）或者很短时间（如果是扰动 X_t 是线性时间序列）；
- 随机游动的趋势没有固定方向，固定趋势的变化形状是固定的；

- 固定趋势模型 Y_t 减去一个固定的回归函数 $Y = a + bt$ 就可以变成平稳列，随机游动减去任意的非随机函数都不能变平稳，可以用差分运算变成平稳。

下面用随机模拟方法演示随机游动（单位根过程）与固定趋势模型。设随机游动模型为

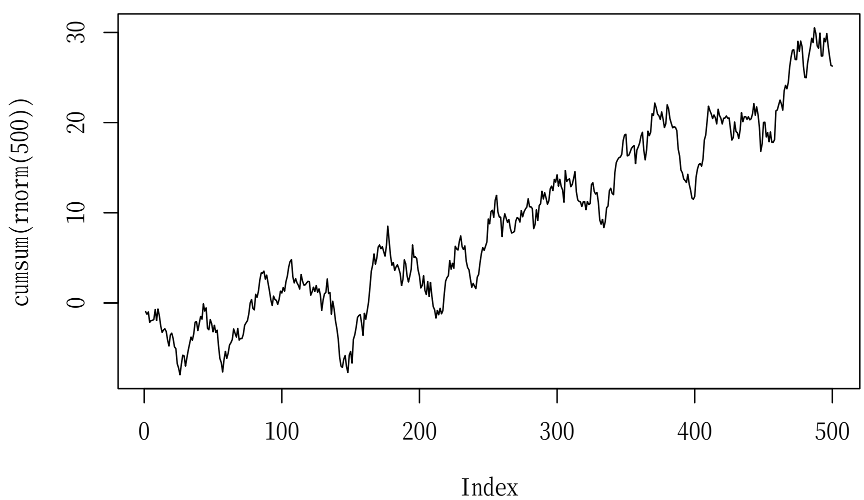
$$Y_t = Y_{t-1} + \varepsilon_t,\ \varepsilon_t \text{iid N}(0,1)$$

固定趋势模型为

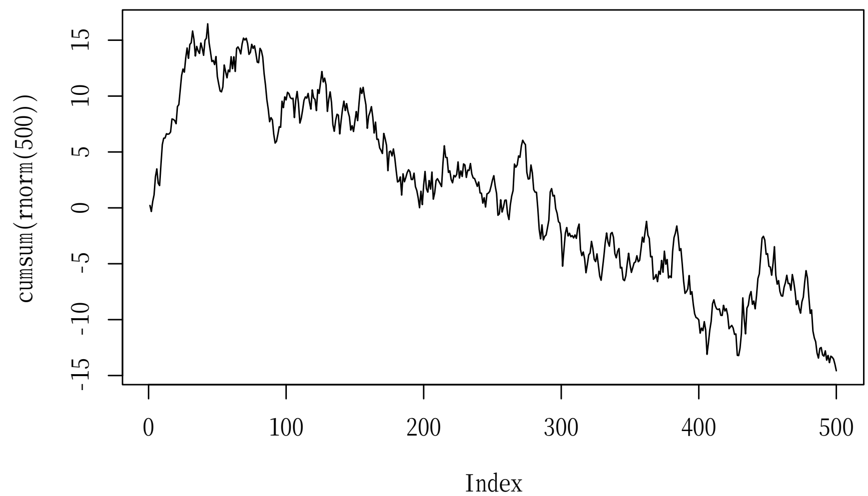
$$Y_t = 10 + 0.01t + \varepsilon_t,\ \varepsilon_t \text{iid N}(0,1)$$

随机游动的模拟：

```
set.seed(3)
plot(cumsum(rnorm(500)), type="l")
```




```
set.seed(4)
plot(cumsum(rnorm(500)), type="l")
```

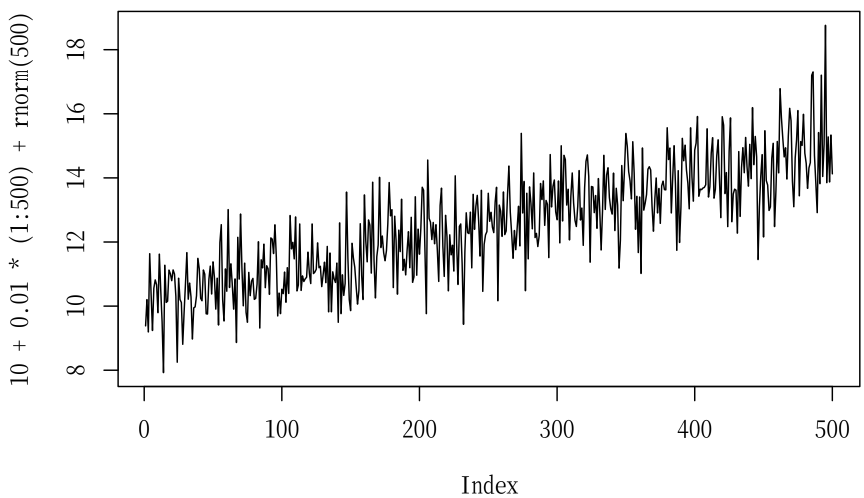



```r
set.seed(9)
plot(cumsum(rnorm(500)), type="l")
```

单位根过程的水平是随机变化的，也没有固定的方向。

固定趋势模型的模拟：

```
set.seed(1)
plot(10 + 0.01*(1:500) + rnorm(500), type="l")
```



固定趋势模型的水平是有规律的。

14.4 分数差分 ARFIMA(p,d,q) 模型

ARMA 序列自协方差函数负指数衰减，是短记忆的。离散谱序列自协方差不衰减到 0，是长记忆的。

其它的平稳序列如何区分长记忆还是短记忆？若存在 $d < \frac{1}{2}$, 使得

$$\gamma_k \sim k^{2d-1},\ k \to \infty$$

则称相应的序列为**长记忆序列**。即 $\gamma_k \sim \frac{1}{k^\alpha}, \alpha > 0$, γ_k 以幂函数速度趋于零。

对于 $d \neq 0$, $d \in (-0.5, 0.5)$, $(1-z)^{-d}$ 有 Taylor 展开公式

$$(1-z)^{-d} = \sum_{j=0}^{\infty} \pi_j z^j,\ |z| \leq 1, \tag{14.12}$$

其中

$$\pi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} = \frac{d(d+1)\dots(d+j-1)}{j!}, \quad j = 0, 1, \dots$$

用 Stirling 公式

$$\Gamma(x) \sim \sqrt{2\pi} e^{1-x}(x-1)^{x-0.5},\ x \to +\infty$$

可证明

$$\pi_j \sim j^{d-1},\ j \to +\infty$$

所以 $\{\pi_j\}$ 平方可和。

定义线性平稳序列

$$X_t = (1-\mathscr{B})^{-d}\varepsilon_t = \sum_{j=0}^{\infty} \pi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \tag{14.13}$$

则 $\{X_t\}$ 是模型

$$(1-\mathscr{B})^d X_t = \varepsilon_t, \quad t \in \mathbb{Z}, \tag{14.14}$$

的惟一平稳解. 人们称(14.14)是 ARFIMA$(0, d, 0)$ 模型.

谱密度为

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty} \pi_j e^{ij\lambda}\right|^2 = \frac{\sigma^2}{2\pi}\left|1 - e^{i\lambda}\right|^{-2d} \tag{14.15}$$

$$= \frac{\sigma^2}{2\pi}\frac{1}{|2\sin(\lambda/2)|^{2d}}, \quad \lambda \in [-\pi, \pi] \tag{14.16}$$

对 ARFIMA(0,d,0) 序列,

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda = \frac{\sigma^2}{\pi}\int_0^{\pi} \frac{\cos(k\lambda)}{[2\sin(\lambda/2)]^{2d}} d\lambda \tag{14.17}$$

$$= \frac{\sigma^2\Gamma(1-2d)\Gamma(k+d)}{\Gamma(k-d+1)\Gamma(d)\Gamma(1-d)},\ k \in \mathbb{N}_+ \tag{14.18}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\Gamma(k+d)\Gamma(k-d)}{\Gamma(k+1-d)\Gamma(d)} \tag{14.19}$$

对 $k=1$ 有

$$d = \frac{\rho_1}{1+\rho_1}$$

由 Sterling 公式可证明

$$\gamma_k \sim k^{2d-1}\frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}, \quad \text{当 } k\to\infty. \tag{14.20}$$

即长记忆。

当 $d\in(-0.5,0)$ 时

$$\sum_{k=0}^{\infty}|\gamma_k| < \infty.$$

当 $d\in(0,0.5)$ 时

$$\sum_{k=0}^{\infty}|\gamma_k| = \infty.$$

用 Levinson 递推公式和归纳法可得 ARFIMA(0,d,0) 的偏相关函数满足

$$a_{k,k} = \frac{d}{k-d}, k=1,2,\dots$$

也可以用于 d 的估计。

类似可定义 ARFIMA(p,d,q) 模型。一般 ARFIMA(p,d,q) 的讨论略过。

14.5 附录：ARIMA 不平稳证明

ARIMA(p,d,q) 模型没有平稳解。

当 $d=1$ 时，问题为，$\{\xi_t\}$ 是 ARMA(p,q) 序列, MA 部分的特征多项式没有等于 1 的根，$\{X_t\}$ 满足

$$X_t - X_{t-1} = \xi_t \tag{14.21}$$

来证明(14.21)没有平稳解。这个结论不能推广到对任意的平稳列 $\{\xi_t\}$ 都成立，因为取 $\{X_t\}$ 为白噪声列，$\xi_t = X_t - X_{t-1}$ 是一个 MA(1) 序列，这时 $\{X_t\}$ 平稳。

归纳地，如果结论对 $d=1$ 成立，则 $d=2$ 时，若 X_t 是如下 ARIMA$(p,2,q)$ 的平稳解：

$$A(\mathscr{B})(1-\mathscr{B})^2 X_t = B(\mathscr{B})\varepsilon_t$$

令 $Y_t = X_t - X_{t-1}$，则 Y_t 是如下 ARIMA$(p,1,q)$ 的平稳解：

$$A(\mathscr{B})(1-\mathscr{B})Y_t = B(\mathscr{B})\varepsilon_t$$

由归纳法假设这不可能。所以只要证明 ARIMA$(p,1,q)$ 没有平稳解则 ARIMA(p,d,q) 都没有平稳解。

情形 1:

当 $\{\xi_t\}$ 为 WN$(0,\sigma^2)$ 时，因

$$X_t - X_0 = \sum_{k=1}^{t} \xi_t$$

所以

$$\mathrm{Var}(X_t - X_0) = t\sigma^2$$

因此 $\{X_t\}$ 不平稳。

情形 2:

设 X_t 满足模型

$$A(\mathscr{B})X_t = \varepsilon_t$$

其中 $\{\varepsilon_t\}$ 为白噪声列，$A(1)=0$，则可分解 $A(z)=(1-z)A_1(z)$，若有平稳列 X_t 使得 $A(\mathscr{B})X_t=\varepsilon_t$，令 $Y_t = A_1(\mathscr{B})X_t$，$Y_t$ 也平稳，且 $Y_t - Y_{t-1} = \varepsilon_t$，由情形 1 推出矛盾。

情形 3

当 $\{\xi_t\}$ 是可逆 ARMA(p,q) 序列时, 设其自协方差函数为 $\{\gamma_k\}$，这时 $\{\xi_t\}$ 有连续谱密度 $f(\lambda), \lambda \in [-\pi,\pi]$，且有正下界 $c>0$。于是

$$\begin{aligned}
&\mathrm{Var}(X_t - X_0) = \mathrm{Var}(\sum_{k=1}^{t} \xi_t) = \sum_{k=1}^{t}\sum_{j=1}^{t} \gamma_{k-j} \\
=&\sum_{k=1}^{t}\sum_{j=1}^{t} \int_{-\pi}^{\pi} e^{i(k-j)\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \left| \sum_{k=1}^{t} e^{ik\lambda} \right|^2 f(\lambda) d\lambda \\
\geq& c\int_{-\pi}^{\pi} \left| \sum_{k=1}^{t} e^{ik\lambda} \right|^2 d\lambda = c\sum_{k=1}^{t}\sum_{j=1}^{t} \int_{-\pi}^{\pi} e^{i(k-j)\lambda} d\lambda \\
=& c\sum_{k=1}^{t} 2\pi \quad (\text{仅 } k-j=0 \text{ 时积分不为零}) = 2\pi c t
\end{aligned}$$

所以 $\{X_t\}$ 不平稳。

情形 4

当 $\{\xi_t\}$ 为一般的 ARMA(p,q) 时：

$$A(\mathscr{B})\xi_t = B(\mathscr{B})\varepsilon_t$$

设其自协方差函数为 $\{\gamma_k\}$，这时 $\{X_t\}$ 满足如下广义 ARMA 模型：

$$A(\mathscr{B})(1-\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t$$

按问题条件知 $B(1) \neq 0$。于是

$$\begin{aligned}
\operatorname{Var}(X_t - X_0) = \operatorname{Var}(\sum_{k=1}^{t} \xi_t) = \sum_{k=1}^{t}\sum_{j=1}^{t} \gamma_{k-j} \\
= \sum_{k=1-t}^{t-1} (t-|k|)\gamma_k \\
= t\sum_{k=1-t}^{t-1} \gamma_k - \sum_{k=1-t}^{t-1} |k|\gamma_k
\end{aligned}$$

令 $t \to +\infty$, 注意到 ARMA 序列 $\{\xi_t\}$ 的自协方差函数 $\{\gamma_k\}$ 负指数衰减所以 $\sum_{k=-\infty}^{\infty} |k|\gamma_k$ 绝对收敛，而

$$\sum_{k=-\infty}^{\infty} \gamma_k = 2\pi f(0) = 2\pi \frac{|B(1)|^2}{|A(1)|^2} > 0$$

所以 $\operatorname{Var}(X_t - X_0) \to +\infty$, $t \to +\infty$，$\{X_t\}$ 非平稳。

对于 AR 部分特征多项式单位圆上有复根的情况，复根必共轭成对出现，且重数必为偶数重，否则不可能组成实系数多项式。这种情形的具体证明有待补充。

Part III

模型估计

Chapter 15

均值的估计

零均值的 AR, MA, ARMA 模型的参数可以由自协方差函数唯一确定。有了样本之后，可以先估计均值和自协方差函数，然后由均值和自协方差函数解出模型参数。均值和自协方差可以用矩估计法求。还要考虑相合性、渐近分布、收敛速度等问题。

设 $x_1, x_2, \dots, x_N$ 是平稳列 $\{X_t\}$ 的观测。$\mu = EX_t$ 的点估计为

$$\hat{\mu} = \bar{x}_N = \frac{1}{N} \sum_{k=1}^{N} x_k \tag{15.1}$$

把观测样本看成随机样本时记做大写的 $X_1, X_2, \dots, X_N$。

15.1 相合性

设统计量 $\hat{\theta}_N$
是 θ 的估计. 在统计学中有如下的定义

- 如果 $E\hat{\theta}_N = \theta$, 则称 $\hat{\theta}_N$ 是 θ 的**无偏估计**.
- 如果当 $N \to \infty$, $E\hat{\theta}_N \to \theta$, 则称 $\hat{\theta}_N$ 是 θ 的**渐近无偏估计**.
- 如果 $\hat{\theta}_N$ 依概率收敛到 θ, 则称 $\hat{\theta}_N$ 是 θ 的**相合估计**.
- 如果 $\hat{\theta}_N$ a.s. 收敛到 θ, 则称 $\hat{\theta}_N$ 是 θ 的**强相合估计**.

一般情况下, 无偏估计比有偏估计来得好. 对于由 (15.1)定义的 $\bar{X}_N$, 有

$$E\bar{X}_N = \frac{1}{N}\sum_{k=1}^{N} EX_k = \frac{1}{N}\sum_{k=1}^{N}\mu = \mu.$$

所以, $\bar{X}_N$ 是均值 μ 的无偏估计.

好的估计量起码应当是相合的. 否则, 估计量不收敛到要估计的参数, 它无助于实际问题的解决. 对于平稳序列 $\{X_t\}$, 如果它的自协方差函数 $\{\gamma_k\}$ 收敛到零, 则

$$E(\bar{X}_N - \mu)^2 = E\Big[\frac{1}{N}\sum_{k=1}^{N}(X_k - \mu)\Big]^2 \tag{15.2}$$

$$=\frac{1}{N^2}E[\sum_{j=1}^{N}\sum_{k=1}^{N}(X_k-\mu)(X_j-\mu)] = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\gamma_{k-j} \tag{15.3}$$

$$=\frac{1}{N^2}\sum_{j=1}^{N}\sum_{m=1-j}^{N-j}\gamma_m \quad (\text{令} m = k-j) \tag{15.4}$$

$$=\frac{1}{N^2}\sum_{m=-N+1}^{N-1}\sum_{j=\max(1-m,1)}^{\min(N-m,N)}\gamma_m = \frac{1}{N^2}\sum_{m=-N+1}^{N-1}(N-|m|)\gamma_m \tag{15.5}$$

$$\leq\frac{1}{N}\sum_{m=-N}^{N}|\gamma_m| \to 0, \quad \text{当} N \to \infty. \tag{15.6}$$

即 $\bar{X}_N$ 均方收敛到 μ. 证明用到：数列趋于零其平均值也趋于零，见B.1(Stolz定理的推论)。

利用切比雪夫不等式

$$\Pr(|\bar{X}_N - \mu| \geq \delta) \leq \frac{E(\bar{X}_N-\mu)^2}{\delta^2} \to 0, \quad (\delta > 0)$$

得到 $\bar{X}_N$ 依概率收敛到 μ. 于是 $\bar{X}_N$ 是 μ 的相合估计.

定理 15.1. *设平稳序列 $\{X_t\}$ 有均值 μ 和自协方差函数 $\{\gamma_k\}$, 则*

- *$\bar{X}_N$ 是 μ 的无偏估计.*
- *如果 $\gamma_k \to 0$, 则 $\bar{X}_N$ 是 μ 的相合估计.*
- *如果 $\{X_t\}$ 还是严平稳遍历序列, 则 $\bar{X}_N$ 是 μ 的强相合估计.*

第三条结论利用第4章的遍历定理4.1可得。

任何强相合估计一定是相合估计.

由第2章定理2.3可知线性平稳列的自协方差趋于零，于是由上述定理可知线性平稳列的均值估计是相合估计。

ARMA 序列是线性平稳列，所以其均值估计是相合估计。

15.2 中心极限定理

若 $X_1, X_2, \ldots, X_N$ iid $\sim (\mu, \sigma^2)$, 则 $\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2)$. 可以据此计算 μ 的 95% 置信区间。

$$[\bar{X}_N - 1.96\sigma/\sqrt{N},\ \bar{X}_N + 1.96\sigma/\sqrt{N}]. \tag{15.7}$$

其中的 1.96 也经常用 2 近似代替。

定理 15.2. 设 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$. 线性平稳序列 $\{X_t\}$ 由

$$X_t = \mu + \sum_{k=-\infty}^{\infty} \psi_k \varepsilon_{t-k}, \quad t \in \mathbb{Z}, \tag{15.8}$$

定义, 其中 $\{\psi_k\}$ 平方可和. 如果 $\{X_t\}$ 的谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{k=-\infty}^{\infty} \psi_k e^{-ik\lambda} \right|^2 \tag{15.9}$$

在 $\lambda = 0$ 连续, 并且 $f(0) \neq 0$, 则当 $N \to \infty$ 时,

$$\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{d} N(0, 2\pi f(0))$$

证明出自 Hall, P. and Heyde C.C.(1980), Martingale Limit Theory and Its Applications, Academic Press, 推论 5.2。

当 $\{\psi_k\}$ 绝对可和时, $f(\lambda)$ 的傅里叶级数表示是绝对一致收敛的，这时 $f(\lambda)$ 连续.

推论 15.1. 在定理*15.2*条件下，如果 $\sum_{k=-\infty}^{\infty} |\psi_k| < \infty$ 和 $\sum_{k=-\infty}^{\infty} \psi_k \neq 0$ 成立, 则当 $N \to \infty$ 时

$$\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{d} N(0, 2\pi f(0))$$

并且

$$2\pi f(0) = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j. \tag{15.10}$$

中心极限定理成立时可以构造 μ 的渐近置信区间或对 μ 作假设检验。

15.3 收敛速度

相合的估计量的渐近性质除了是否服从中心极限定理外, 还包括这个估计量的收敛速度. 收敛速度的描述方法之一是所谓的重对数律. 重对数律成立时, 得到的收敛速度的阶数一般是

$$O\left(\sqrt{\frac{2\ln\ln N}{N}}\right).$$

除了个别的情况, 这个阶数一般不能再被改进.

定理 15.3. 设 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0,\sigma^2)$, 线性平稳序列 $\{X_t\}$ 由(15.8)定义, 谱密度 $f(0) \neq 0$. 当以下的条件之一成立时:

- 当 $k \to \pm\infty$, $\psi_{|k|}$ 以负指数阶收敛于 0;
- 谱密度 $f(\lambda)$ 在 $\lambda = 0$ 连续, 并且 $E|\varepsilon_t|^r < \infty$ 对某个 $r > 2$ 成立,

则有重对数律

$$\limsup_{N\to\infty} \sqrt{\frac{N}{2\ln\ln N}}(\bar{X}_N - \mu) = \sqrt{2\pi f(0)}, \quad a.s. \tag{15.11}$$

$$\liminf_{N\to\infty} \sqrt{\frac{N}{2\ln\ln N}}(\bar{X}_N - \mu) = -\sqrt{2\pi f(0)}, \quad a.s. \tag{15.12}$$

易见重对数率满足时 $(\bar{X}_n - \mu)\cdot o(1) = o(\sqrt{\frac{\ln\ln N}{N}})$, $\sqrt{\frac{N}{2\ln\ln N}}(\bar{X}_n - \mu)/o(1)$ 不收敛。

15.4 $\bar{X}_N$ 的模拟计算

令

$$A(z) = (1 - \rho e^{i\theta}\cdot z)(1 - \rho e^{-i\theta}\cdot z)$$

考虑 AR(2) 模型

$$\begin{aligned} A(\mathscr{B})X_t =& \varepsilon_t \\ X_t =& 2\rho\cos\theta X_{t-1} - \rho^2 X_{t-2} + \varepsilon_t \end{aligned}$$

为模拟方便设 $\{\varepsilon_t\}$ iid $\sim \mathrm{N}(0, \sigma^2)$。

$$\bar{x}_N = \frac{1}{N}\sum_{t=1}^N x_t, \quad \bar{\varepsilon}_N = \frac{1}{N}\sum_{t=1}^N \varepsilon_t$$

模型方程两边求平均

$$\begin{aligned} \bar{x}_N =& 2\rho\cos\theta \frac{1}{N}\sum_{k=0}^{N-1} x_k - \rho^2 \frac{1}{N}\sum_{k=-1}^{N-2} x_k + \bar{\varepsilon}_N \\ =& 2\rho\cos\theta \bar{X}_N - \rho^2 \bar{X}_N + \bar{\varepsilon}_N \\ & + 2\rho\cos\theta \frac{1}{N}(x_0 - x_N) - \rho^2 \frac{1}{N}(x_{-1} + x_0 - x_{N-1} - x_{N-2}) \\ \approx& 2\rho\cos\theta \bar{X}_N - \rho^2 \bar{X}_N + \bar{\varepsilon}_N \\ \bar{x}_N \approx& \frac{1}{A(1)}\bar{\varepsilon}_N = \frac{1}{1 - 2\rho\cos\theta + \rho^2}\bar{\varepsilon}_N \end{aligned}$$

N 较大时 $\bar{x}_N$ 和 $\bar{\varepsilon}_t$ 成正比。

为了观察 $N \to \infty$ 时 $\bar{x}_N$ 的收敛可以模拟 L 个值然后观察 $\bar{x}_N, N = n_0, n_0 + 1, \dots, L$ 的变化。

为了研究固定 N 情况下 $\bar{X}_N$ 的精度以至于抽样分布，可以进行 M 次独立的随机模拟，得到 M 个 $\bar{X}_N$ 的观测值。这种方法对于难以得到估计量的理论分布的情况是很有用的。

下面是具体的模拟程序和结果。模拟生成 AR(2) 序列。

下面的函数针对变化的样本量 n 比较 $\bar{X}_n$ 与 $\bar{\varepsilon}_n$ 的估计，作 $\bar{X}_n$ 与 $\bar{\varepsilon}_n$ 对 n 的曲线图。

```
estmean.plot <- function(x, eps, subt=""){
  barx <- cumsum(x)/seq(L)
  bareps <- cumsum(eps)/seq(L)
  yr <- range(c(barx, bareps+0.1))
```

```
  plot(bareps+0.1, type="l",
       col="red",
       ylim=c(-0.3, 0.3),
       main=paste("Effect of sample size N",
         subt, sep="\n"),
       xlab="N", ylab="Avgs")
  lines(barx)
  abline(h=c(0, 0.1))
  legend(0.7*L, 0.3, lty=c(1,1),
         col=c("black","red"),
         legend=c("AR(2)", "WN"))
}
```

生成 4 个不同的 AR(2) 模型的数据，最大长度 $L = 2000$，比较不同样本量时均值 $\bar{X}_n$ 与 $\bar{\varepsilon}_n$, 作图时对 $\bar{\varepsilon}_n$ 加了 0.1。四个 AR(2) 模型的特征根分别为:

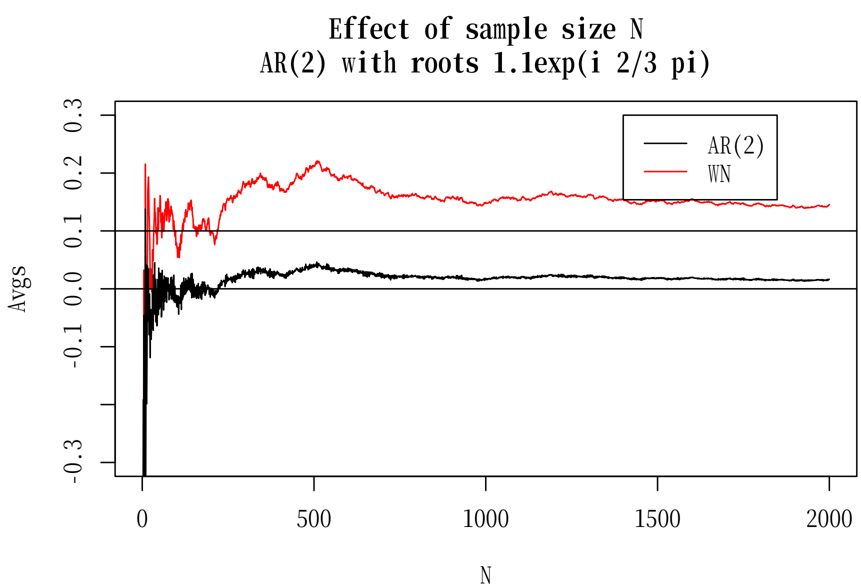
$$1.1e^{i\frac{2\pi}{3}}, 4.0e^{i\frac{2\pi}{3}}, 4.0e^{i\frac{5\pi}{6}}, 1.5e^{i\frac{\pi}{6}}$$

```
## 特征根的倒数
roots <- complex(mod=c(1/1.1, 1/4, 1/4, 1/1.5),
                 arg=c(pi*2/3, pi*2/3, pi*5/6, pi/6))
## 各 AR(2) 模型的 $a_1$ 和 $a_2$ 系数
coefs <- cbind(2*Re(roots), -Mod(roots)^2)
## $\bar X_n$ 与 $\bar\varepsilon_n$ 的比值倍数
multi <- apply(coefs, 1, function(a) 1/(1 - a[1] - a[2]))
cat("Magnification of bar X to bar eps:\n", round(multi, 2), "\n")
```

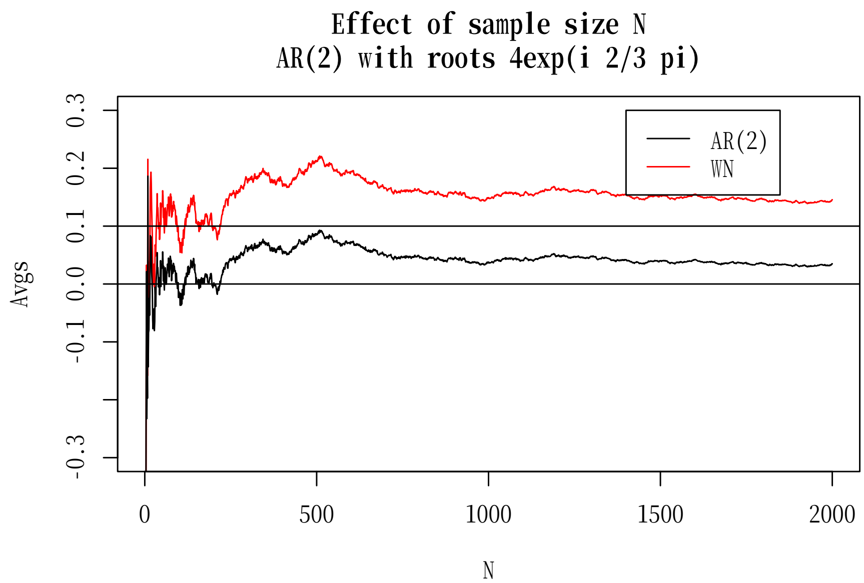
```
## Magnification of bar X to bar eps:
##  0.37 0.76 0.67 3.45
```

```
L <- 2000
sigma <- 1

n0 <- 200
n2 <- n0 + L
```

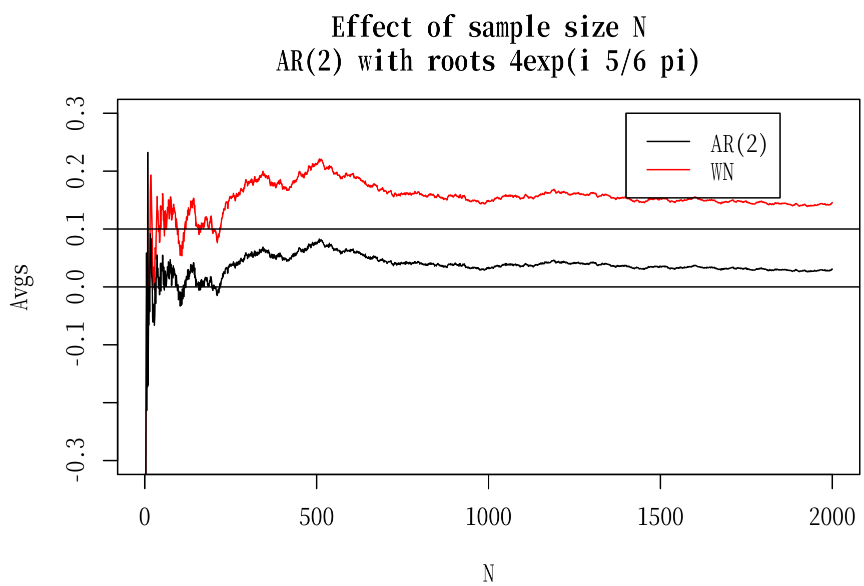
```
eps <- rnorm(n2, 0, sigma)
xx <- filter(eps, coefs[1,,drop=T], method="recursive", side=1)
x1 <- xx[(n0+1):n2]
xx <- filter(eps, coefs[2,,drop=T], method="recursive", side=1)
x2 <- xx[(n0+1):n2]
xx <- filter(eps, coefs[3,,drop=T], method="recursive", side=1)
x3 <- xx[(n0+1):n2]
xx <- filter(eps, coefs[4,,drop=T], method="recursive", side=1)
x4 <- xx[(n0+1):n2]

estmean.plot(x1, eps[(n0+1):n2], subt="AR(2) with roots 1.1exp(i 2/3 pi)")
```



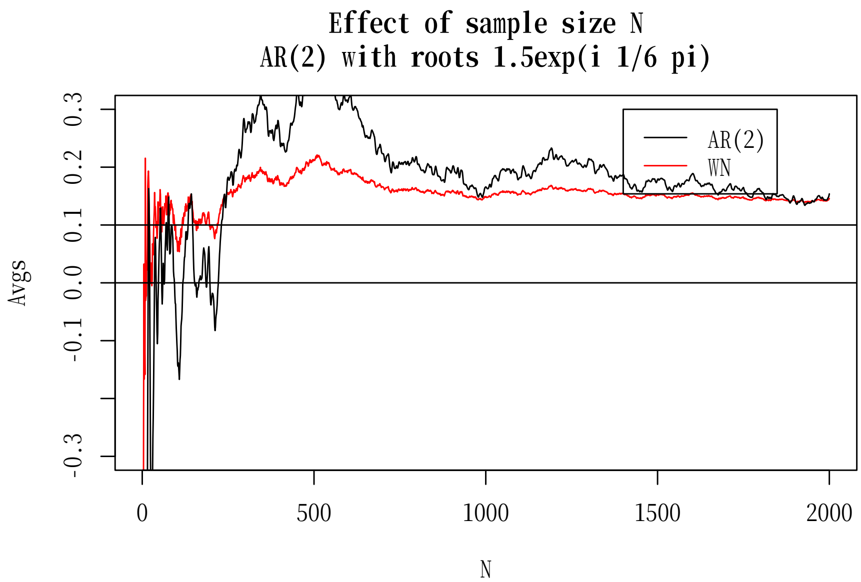

```
estmean.plot(x2, eps[(n0+1):n2], subt="AR(2) with roots 4exp(i 2/3 pi)")
```




```r
estmean.plot(x3, eps[(n0+1):n2], subt="AR(2) with roots 4exp(i 5/6 pi)")
```



```
estmean.plot(x4, eps[(n0+1):n2], subt="AR(2) with roots 1.5exp(i 1/6 pi)")
```



下面比较不同样本量时 $\bar{X}$ 的极限分布。对每一个样本量 n，重复模拟产生 $M = 500$ 条轨道的模拟序列。用了上面的第三个和第四个模型，作某个样本量下的多条轨道的均值直方图时仅用了第三个模型。

```
ns <- c(10, 20, 50, 100, 200, 1000)
M <- 500
nn <- length(ns)
bx3s <- matrix(0, nrow=M, ncol=nn)
bx4s <- matrix(0, nrow=M, ncol=nn)
bes <- matrix(0, nrow=M, ncol=nn)
for(mm in seq(M)){
  eps <- rnorm(n2, 0, sigma)
  xx <- filter(eps, coefs[3,,drop=T], method="recursive", side=1)
  x3 <- xx[(n0+1):n2]
  xx <- filter(eps, coefs[4,,drop=T], method="recursive", side=1)
  x4 <- xx[(n0+1):n2]
  barx3 <- cumsum(x3)/seq(L)
```

```
  barx4 <- cumsum(x4)/seq(L)
  bx3s[mm,] <- barx3[ns]
  bx4s[mm,] <- barx4[ns]
  bare <- cumsum(eps[(n0+1):n2])/seq(L)
  bes[mm,] <- bare[ns]
}
avgse <- apply(bes, 2, mean)
stdse <- apply(bes, 2, sd)
avgsx3 <- apply(bx3s, 2, mean)
stdsx3 <- apply(bx3s, 2, sd)
avgsx4 <- apply(bx4s, 2, mean)
stdsx4 <- apply(bx4s, 2, sd)
res <- rbind(avgse, stdse, avgsx3, stdsx3, avgsx4, stdsx4)
rownames(res) <- c("Avg.bareps", "Std.bareps",
                   "Avg.barx3", "Std.barx3",
                   "Avg.barx4", "Std.barx4")
colnames(res) <- ns
cat("\nSampling distribution by repeated sampling. ===\n")
```

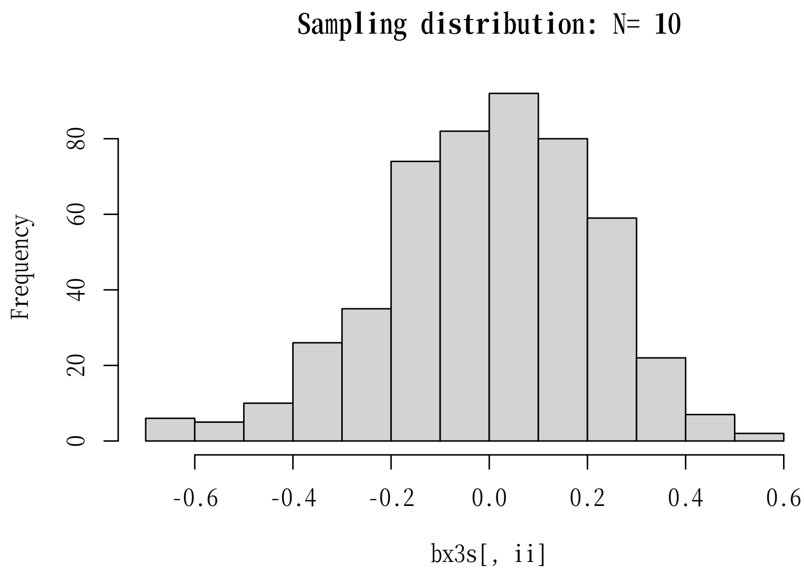
```
##
## Sampling distribution by repeated sampling. ===
```

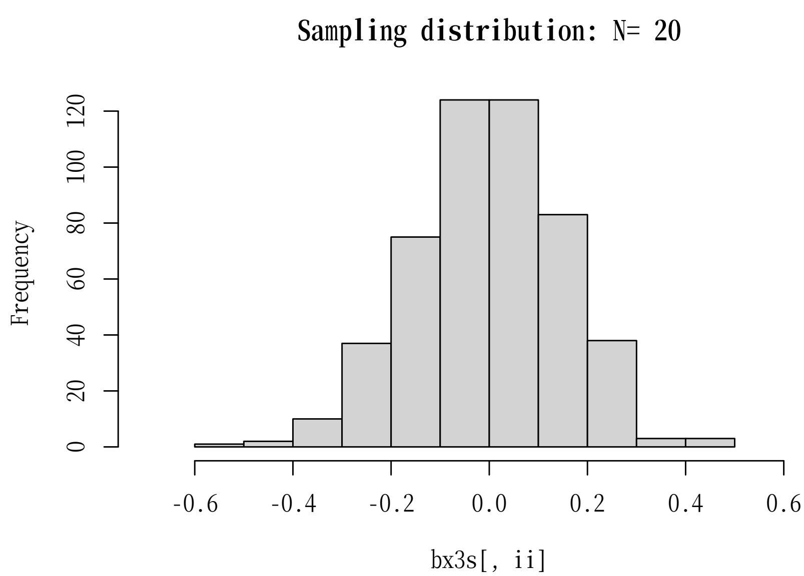
```
print(round(res, 4))
```

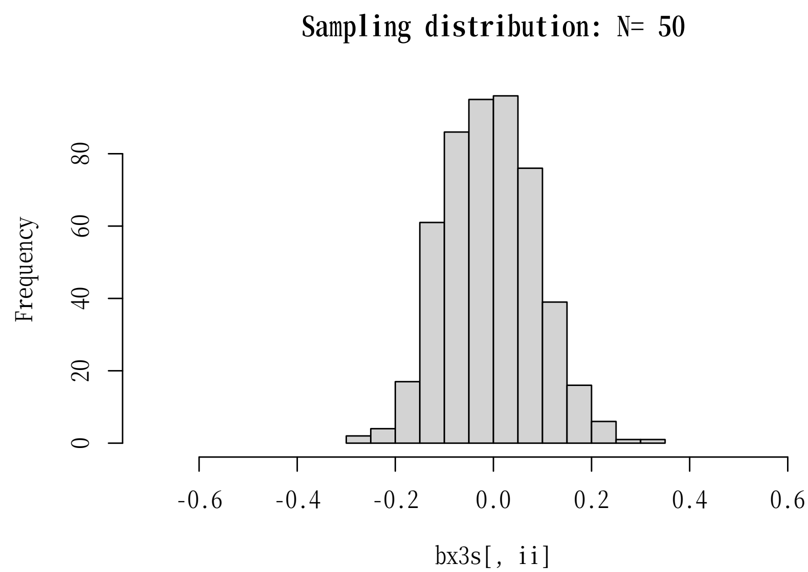
```
##                 10      20      50     100     200   1000
## Avg.bareps  0.0041 -0.0048 -0.0090 -0.0004  0.0001 0.0008
## Std.bareps  0.3048  0.2158  0.1388  0.0995  0.0699 0.0311
## Avg.barx3  -0.0007 -0.0040 -0.0063 -0.0003 -0.0001 0.0005
## Std.barx3   0.2177  0.1506  0.0945  0.0672  0.0467 0.0208
## Avg.barx4   0.0125 -0.0181 -0.0319 -0.0026  0.0012 0.0028
## Std.barx4   1.0265  0.7270  0.4746  0.3398  0.2427 0.1072
```

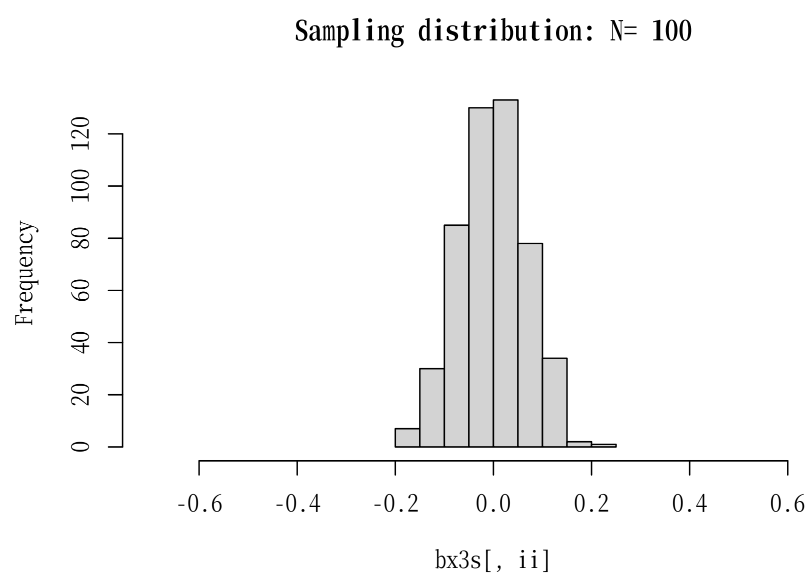
```
for(ii in seq(ns)){
  hist(bx3s[,ii], main=paste("Sampling distribution: N=", ns[ii]),
       xlim=c(-0.7,0.7))
```

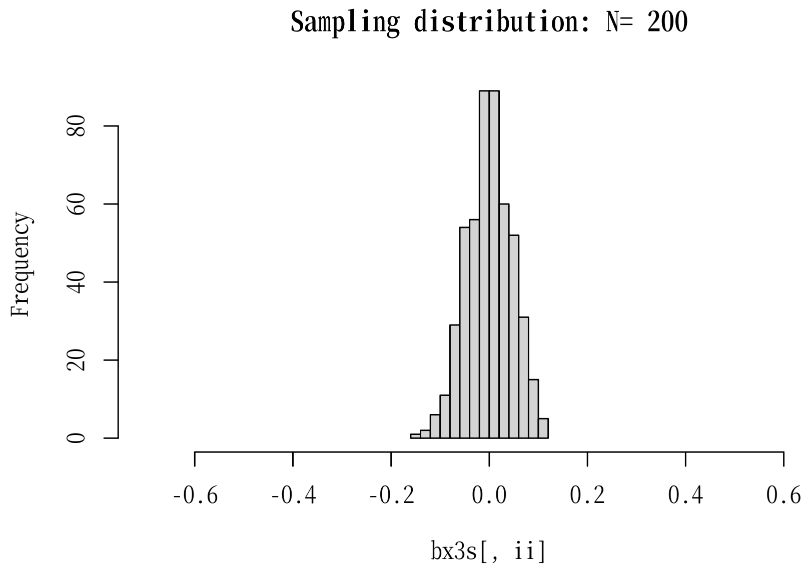
```
}
```

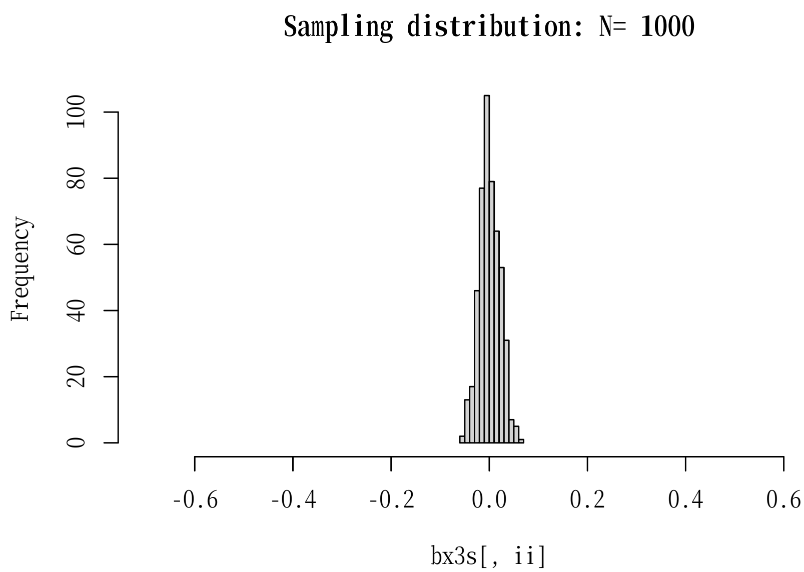

Sampling distribution: N= 10
Frequency
0
20
40
60
80
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
bx3s[, ii]


Sampling distribution: N= 20
Frequency
0
20
40
60
80
100
120
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
bx3s[, ii]


Sampling distribution: N= 50
Frequency
0
20
40
60
80
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
bx3s[, ii]


Sampling distribution: N= 100
Frequency
0
20
40
60
80
100
120
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
bx3s[, ii]


Sampling distribution: N= 200
Frequency
0
20
40
60
80
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
bx3s[, ii]


Sampling distribution: N= 1000
Frequency
0
20
40
60
80
100
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
bx3s[, ii]

Chapter 16

自协方差函数的估计

16.1 自协方差估计公式及正定性

$$
\begin{aligned}
\hat\gamma_k =& \frac{1}{N}\sum_{j=1}^{N-k}(x_j - \bar x_N)(x_{j+k} - \bar x_N), \quad 0 \leq k \leq N-1, \\
\hat\gamma_{-k} =& \hat\gamma_k
\end{aligned}
\tag{16.1}
$$

样本自相关系数 (ACF) 估计为

$$
\hat\rho_k = \frac{\hat\gamma_k}{\hat\gamma_0}, \quad |k| \leq N-1 \tag{16.2}
$$

k 较大时参与平均的项减少所以 $\hat\gamma_k$ 估计误差会随 k 增大而变大。

估计 γ_k 一般不使用除以 $N-k$ 的估计形式:

$$
\hat\gamma_k = \frac{1}{N-k}\sum_{j=1}^{N-k}(x_j - \bar x_N)(x_{j+k} - \bar x_N) \tag{16.3}
$$

因为：

- 我们不对大的 k 计算 $\hat\gamma_k$;
- 更重要的是只有除以 N 的估计式才是正定的。

只要观测 $x_1, x_2, \dots, x_N$ 不全相同则

$$\hat{\Gamma}_N = (\hat{\gamma}_{k-j})_{k,j=1,2,\dots,N}$$

正定。

令 $y_j = x_j - \bar{x}_N$，记

$$A = \begin{pmatrix} 0 & \cdots & 0 & y_1 & y_2 & \cdots & y_{N-1} & y_N \\ 0 & \cdots & y_1 & y_2 & y_3 & \cdots & y_N & 0 \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ y_1 & \cdots & y_{N-1} & y_N & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{16.4}$$

则

$$\hat{\Gamma}_N = \frac{1}{N} A A^T$$

只要 y_j 不全是零则 A 满秩。

事实上，设 $y_1 = \cdots = y_{k-1} = 0, y_k \neq 0$，则 A 矩阵左面会出现一个以 y_k 值开始非零的斜面，显然是满秩的。故 $x_1, \dots, x_N$ 不全相同时 $\hat{\Gamma}_N$ 正定。$\hat{\Gamma}_n (1 \leq n \leq N)$ 作为 $\hat{\Gamma}_N$ 的主子式也是正定的。

16.2 $\hat{\gamma}_k$ 的相合性

定理 16.1. 设平稳序列的样本自协方差函数 $\hat{\gamma}_k$ 由(16.3)或(16.3)定义.

(1) 如果当 $k \to \infty$ 时 $\gamma_k \to 0$, 则对每个确定的 k, $\hat{\gamma}_k$ 是 γ_k 的渐近无偏估计:

$$\lim_{N \to \infty} E\hat{\gamma}_k = \gamma_k.$$

(2) 如果 $\{X_t\}$ 是严平稳遍历序列, 则对每个确定的 k, $\hat{\gamma}_k$ 和 $\hat{\rho}_k$ 分别是 γ_k 和 ρ_k 的强相合估计:

$$\lim_{N \to \infty} \hat{\gamma}_k = \gamma_k \ , a.s., \quad \lim_{N \to \infty} \hat{\rho}_k = \rho_k \ , a.s.$$

证明

下面只对由(16.3)定义的样本自协方差函数证明, 对由(16.3)定义的 $\hat{\gamma}_k$ 的证明是一样的.

设 $\mu = EX_1$, 则 $\{Y_t\} = \{X_t - \mu\}$ 是零均值的平稳序列. 利用

$$\bar{Y}_N = \frac{1}{N}\sum_{j=1}^{N} Y_j = \bar{X}_N - \mu$$

得到

$$\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-k}(Y_j - \bar{Y}_N)(Y_{j+k} - \bar{Y}_N) \tag{16.5}$$

$$= \frac{1}{N}\sum_{j=1}^{N-k}[Y_jY_{j+k} - \bar{Y}_N(Y_{j+k} + Y_j) + \bar{Y}_N^2]. \tag{16.6}$$

利用(15.6)得到 $E\bar{Y}_N^2 = E(\bar{X}_N - \mu)^2 \to 0$. 利用 Schwarz 不等式得到

$$E|\bar{Y}_N(Y_{j+k} + Y_j)| \le [E\bar{Y}_N^2 E(Y_{j+k} + Y_j)^2]^{1/2} \le [E\bar{Y}_N^2 \cdot 4\gamma_0]^{1/2} \to 0.$$

所以当 $N \to \infty$,

$$E\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-k} E(Y_jY_{j+k}) + o(1) = \frac{N-k}{N}\gamma_k + o(1) \to \gamma_k.$$

强相合性的证明: 用遍历定理得到

$$\begin{aligned}
&\bar{Y}_N \to EY_1 = 0, \text{a.s.},\\
&\frac{1}{N}\sum_{j=1}^{N-k}(Y_{j+k} + Y_j) = \frac{1}{N}\left(\sum_{j=1}^{N} Y_j - \sum_{j=1}^{k} Y_j + \sum_{j=1}^{N-k} Y_j\right)\\
=&\bar{Y}_N - \frac{1}{N}\sum_{j=1}^{k} Y_j + \frac{N-k}{N}\bar{Y}_{N-k} \to 0, \text{a.s.},
\end{aligned}$$

于是, 从(16.6)式可知

$$\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-k} Y_jY_{j+k} + o(1) \to E(Y_1Y_{1+k}) = \gamma_k, \ \text{a.s.}$$

16.3 $\hat{\gamma}_k$ 的渐近分布

只考虑线性序列。设 $\{\varepsilon_t\}$ 是 4 阶矩有限的独立同分布的 $\mathrm{WN}(0,\sigma^2)(\sigma^2>0)$, 实数列 $\{\psi_k\}$ 平方可和. 线性平稳序列

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{16.7}$$

$\{X_t\}$ 有自协方差函数

$$\gamma_k = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} \tag{16.8}$$

$\{X_t\}$ 有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} \psi_j e^{ij\lambda} \right|^2. \tag{16.9}$$

设自协方差函数列 $\{\gamma_k\}$ 平方可和。设 $\{W_t\}$ 为独立同分布 N(0,1)。令

$$\mu_4 = E\varepsilon_1^4, \quad M_0 = \frac{1}{\sigma^2}(\mu_4 - \sigma^4)^{1/2} > 0$$

定义正态时间序列

$$\xi_j = (M_0\gamma_j)W_0 + \sum_{t=1}^{\infty}(\gamma_{t+j}+\gamma_{t-j})W_t, \quad j \geq 0 \tag{2.11}$$

$$R_j = \sum_{t=1}^{\infty}(\rho_{t+j}+\rho_{t-j}-2\rho_t\rho_j)W_t, \quad j \geq 1, \tag{16.10}$$

定理 16.2. *设 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0,\sigma^2)$, 满足 $\mu_4 = E\varepsilon_1^4 < \infty$. 如果线性平稳序列(16.7)的谱密度(16.9)平方可积:*

$$\int_{-\pi}^{\pi} [f(\lambda)]^2 \, d\lambda < \infty,$$

则对任何正整数 h, 当 $N \to \infty$ 时, 有以下结果

(1) $\sqrt{N}(\hat{\gamma}_0-\gamma_0, \hat{\gamma}_1-\gamma_1, \dots, \hat{\gamma}_h-\gamma_h)$ 依分布收敛到 $(\xi_0, \xi_1, \dots, \xi_h)$.

(2) $\sqrt{N}(\hat{\rho}_1-\rho_1, \hat{\rho}_2-\rho_2, \dots, \hat{\rho}_h-\rho_h)$ 依分布收敛到 $(R_1, R_2, \dots, R_h)$

可以据此构造 γ_k 和 ρ_k 的近似的区间估计和近似的假设检验。

16.3.1 MA(q) 序列的自相关截尾的检验

对 MA(q) 序列 $\{X_t\}$, 利用定理16.2可知, 只要 $m > q$: $\sqrt{N}\hat{\rho}_m$ 依分布收敛到 R_m 的分布。

$$R_m = \sum_{t=1}^{\infty}(\rho_{t+m} + \rho_{t-m} - 2\rho_t\rho_m)W_t, \quad m \geq q+1$$

注意 $m \geq q+1$ 时 $\rho_m = 0, \rho_{t+m} = 0$, ρ_{t-m} 中的 $t-m$ 应属于 $[-q, q]$，所以令 $l = t-m$ 有

$$R_m = \sum_{l=-q}^{q} \rho_l W_{l+m}$$

R_m 为期望为 0, 方差为 $1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_q^2$ 的正态分布.

在假设 H_0: $\{X_t\}$ 是 MA(q) 下, 对 $m > q$ 有

$$\Pr\left(\frac{\sqrt{N}|\hat{\rho}_m|}{\sqrt{1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_q^2}} \geq 1.96\right) \approx 0.05.$$

令

$$T_q(m) = \frac{\sqrt{N}\hat{\rho}_{m+q}}{\sqrt{1 + 2\hat{\rho}_1^2 + 2\hat{\rho}_2^2 + \cdots + 2\hat{\rho}_q^2}}, \ m > q$$

则在 H_0 下 $T_q(m)$ 近似服从标准正态分布，可以用 $\{|T_q(m)| > 1.96\}$ 作为 H_0 的水平为 0.05 的近似检验否定域。

例 16.1. 现在用 $\{X_t\}$ 表示第12章例12.1中差分后的化学浓度数据, 在 H_0: $\{X_t\}$ 是 MA(q) 下, 分别对 $q = 0, 1$ 计算出 $T_q(m)$, $m = 1, 2, \ldots, 6$。

$m=$	1	2	3	4	5	6
$q=0$	-5.778	0.281	-0.951	-0.121	-1.071	-0.116
$q=1$	0.243	-0.821	-0.104	-0.925	-0.100	1.631

在 $q = 0$ 的假设下,$|T_0(1)| = 5.778 > 1.96$, 所以应当否定 $q = 0$.

也可以计算检验的近似 p 值，如 $q = 0$, $m = 1$ 时

$$p = P(|Z| \geq |-5.778|),$$

其中 Z 表示标准正态分布随机变量 p 值越小, 数据提供的否定原假设的依据越充分. 这里的 p 值基本为 0，所以应该拒绝 $q=0$ 的假设。

取 $q=1$ 时 $|T_1(m)|<1.96(1\le m\le 6)$, 所以不能拒绝 $\{X_t\}$ 是 MA(1) 的假设.

16.3.2 谱密度平方可积的充要条件

对于实际工作者来讲谱密度平方可积的条件通常很难验证, 于是希望能把定理16.2中谱密度平方可积的条件改加在自协方差函数 $\{\gamma_k\}$ 的收敛速度上.

定理 16.3. *对任一平稳序列 $\{X_t\}$, 它的自协方差函数平方可和的充分必要条件是它的谱密度平方可积.*

这个结论主要是利用实变函数论中 Fourier 级数的理论，只有证明 $f(\lambda)\ge 0$ 时用了周期图 (如定理9.1的证明，那里 $\{\gamma_k\}$ 绝对可和)。证明略。

推论 16.1. *设 $\{\varepsilon_t\}$ 是独立同分布的白噪声 $WN(0,\sigma^2)$, 满足 $\mu_4=E\varepsilon_t^4<\infty$. 如果线性平稳序列(16.7)的自协方差函数平方可和: $\sum_k \gamma_k^2<\infty$, 则定理16.2中的结论成立.*

16.3.3 ψ_k 快速收敛条件下的中心极限定理

定理 16.2 要求白噪声的方差有 4 阶矩. 下面关于线性平稳序列的样本自相关系数的中心极限定理不要求噪声项的 4 阶矩有限.

定理 16.4. *设 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0,\sigma^2)$, 线性平稳序列 $\{X_t\}$ 由(16.7)定义. 如果自协方差函数 $\{\gamma_k\}$ 平方可和, 并且对某个常数 $\alpha>0.5$,*

$$m^\alpha \sum_{|k|\ge m}\psi_k^2 \ \to\ 0, \quad m\to\infty, \tag{16.11}$$

则对任何正整数 h, 当 $N\to\infty$ 时

$$\sqrt{N}(\hat\rho_1-\rho_1,\hat\rho_2-\rho_2,\ldots,\hat\rho_h-\rho_h)$$

依分布收敛到

$$(R_1,R_2,\cdots,R_h).$$

ARMA 序列的 $\{\psi_j\}$ 满足(16.11)所以 ARMA 序列的白噪声列是独立同分布序列时定理16.4结论成立。

16.3.4 关于独立同分布列的中心极限定理

推论 16.2. 如果 $\{X_t\}$ 是独立同分布的白噪声,

$$\hat{\rho}_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x}_N)(x_{t+k} - \bar{x}_N)}{\sum_{t=1}^{N}(x_t - \bar{x}_N)^2}$$

是样本自相关系数, 则对任何正整数 h

(1)

$$\sqrt{N}(\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_h)$$

依分布收敛到多元标准正态分布 $N(0, I_h)$. 这里, I_h 是 $h \times h$ 的单位矩阵.

(2) 如果 $\mu_4 = EX_t^4 < \infty$, 则

$$\sqrt{N}(\hat{\gamma}_0 - \sigma^2, \hat{\gamma}_1, \dots, \hat{\gamma}_h)$$

依分布收敛到

$$\sigma^2(M_0 W_0, W_1, \dots, W_h).$$

证明:

对白噪声，$\gamma_0 = \sigma^2$,

$$\begin{aligned}
R_j =& \sum_{t=1}^{\infty}(\rho_{t+j} + \rho_{t-j} - 2\rho_t\rho_j)W_t = W_j, j \geq 1 \\
\xi_j =& (M_0\gamma_j)W_0 + \sum_{t=1}^{\infty}(\gamma_{t+j} + \gamma_{t-j})W_t \\
=& \gamma_0 W_j = \sigma^2 W_j \\
\xi_0 =& (M_0\gamma_0)W_0 + \sum_{t=1}^{\infty}(\gamma_{t+0} + \gamma_{t-0})W_t \\
=& M_0\gamma_0 W_0 = \sigma^2 M_0 W_0
\end{aligned}$$

定理16.4的条件满足。第二条满足推论16.1的条件。

16.4 模拟计算

首先用图形表示 N 不同时 $\hat{\gamma}_k$ 的误差。然后重复 $M = 400$ 次计算 400 个 $\hat{\gamma}_k$ 的标准差 (称为标准误差)。发现 N 增大时标准误差减小。误差随 N 减小的速度为 $N^{-1/2}$。根离单位圆近的模型其估计标准误差大。

利用 AR(2) 模型进行模拟，选择两个不同的 AR(2) 模型，特征根分别为：

$$1.8e^{\pm i2.13},\ 1.1e^{\pm i2.13}$$

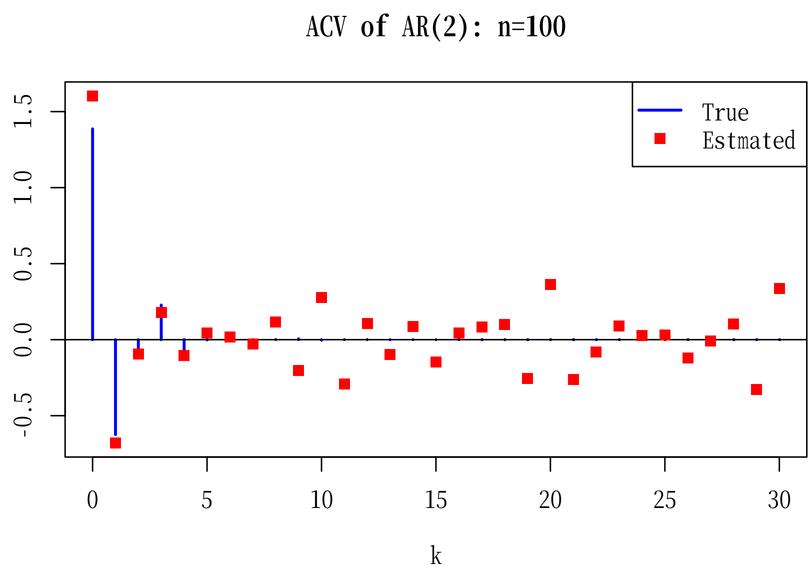
首先考察样本量的变化对 $\hat{\gamma}_k$ 估计的影响。计算 $\hat{\gamma}_k, k = 0, 1, \ldots, 30$。这里仅使用第一个模型，即特征根离单位圆较远的那个模型。

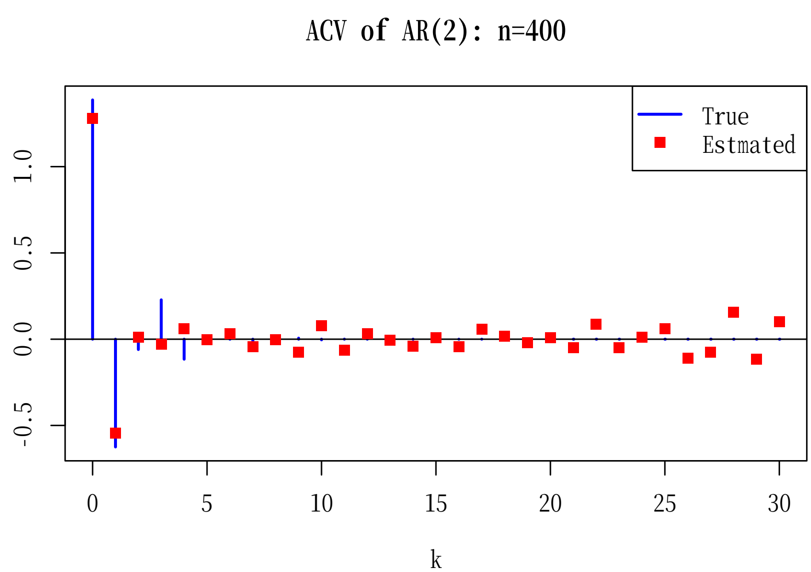
```
ar.gen <- function(n, a, sigma=1.0, by.roots=FALSE,
                   plot.it=FALSE, n0=1000,
                   x0=numeric(length(a))){
  if(by.roots){
    cf <- Re(coef(poly.calc(a)))
    cf <- cf / cf[1]
    a <- -cf[-1]
  }
  n2 <- n0 + n
  p <- length(a)
  eps <- rnorm(n2, 0, sigma)
  x2 <- filter(eps, a, method="recursive", side=1, init=x0)
  x <- x2[(n0+1):n2]
  x <- ts(x)
  attr(x, "model") <- "AR"
  attr(x, "a") <- a
  attr(x, "sigma") <- sigma
  if(plot.it) plot(x)
  x
}
arma.Wold <- function(n, a, b=numeric(0)){
  p <- length(a)
  q <- length(b)
```

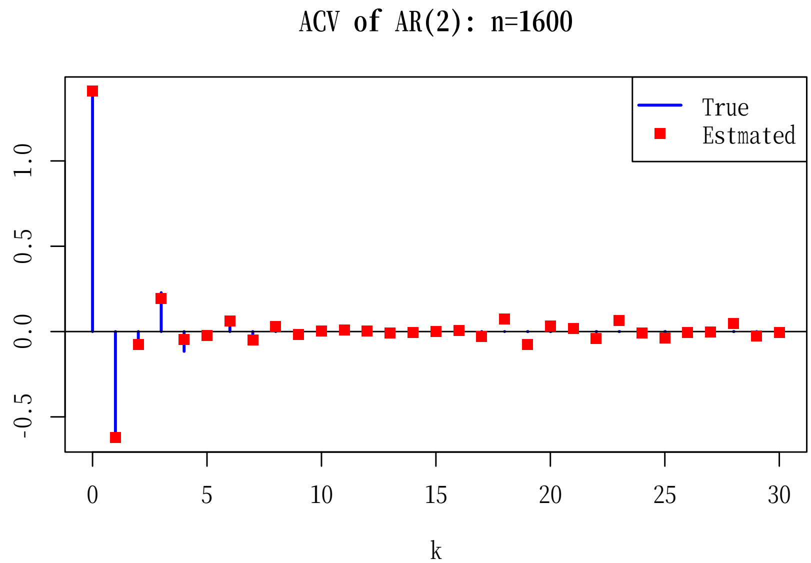
```
  arev <- rev(a)
  psi <- numeric(n)
  psi[1] <- 1
  for(j in seq(n-1)){
    if(j <= q) bj=b[j]
    else bj=0
    psis <- psi[max(1, j+1-p):j]
    np <- length(psis)
    if(np < p) psis <- c(rep(0,p-np), psis)
    psi[j+1] <- bj + sum(arev * psis)
  }

  psi
}
arma.gamma.by.Wold <- function(n, a, b=numeric(0), sigma=1){
  nn <- n + 100
  psi <- arma.Wold(nn, a, b)
  gam <- numeric(n)
  for(ii in seq(0, n-1)){
    gam[ii+1] <- sum(psi[1:(nn-ii)] * psi[(ii+1):nn])
  }
  gam <- (sigma^2) * gam
  gam
}
arma.gamma <- arma.gamma.by.Wold
```

```
roots <- complex(mod=c(1/1.8, 1/1.1),
                 arg=c(2.13, 2.13))
coefs <- cbind(2*Re(roots), -Mod(roots)^2)
sigma <- 1
ngam <- 31

L <- 1600
```

```
x <- ar.gen(L, a=coefs[1,])
true.gam <- arma.gamma.by.Wold(ngam, a=coefs[1,],
                               sigma=sigma)

for(nn in c(100, 400, 1600)){
  est.gam <- acf(x[1:nn], lag.max=ngam-1, type="covariance",
                 plot=F)$acf
  plot(0:(ngam-1), true.gam,
       type="h", col='blue', lwd=2,
       main=paste("ACV of AR(2): n=", nn, sep=""),
       ylim=range(c(true.gam, est.gam)),
       xlab="k", ylab="")
  abline(h=0)
  points(0:(ngam-1), est.gam, pch=15, col='red')

  legend("topright", pch=c(NA, 15), lty=c(1, 0),
         lwd=c(2,1), col=c("blue", "red"),
         legend=c("True", "Estmated"))
}
```


ACV of AR(2): n=100
True
Estmated
-0.5
0.0
0.5
1.0
1.5
0
5
10
15
20
25
30
k


ACV of AR(2): n=400
True
Estmated
-0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
k



下面对两个 AR(2) 模型模拟产生多条 $(M = 400)$ 轨道，针对不同的样本量 n 估计 $\hat\gamma_k$。模拟函数：

```
demo.estacv.se <- function(a){
  M <- 400
  ns <- c(1600, 400, 100, 25)
  L <- max(ns)
  ngam <- 6
  stderrs <- matrix(0, nrow=length(ns), ncol=ngam)
  colnames(stderrs) <- 0:(ngam-1)
  rownames(stderrs) <- ns
  gams <- as.list(ns)
  for(ii in seq(along=gams)){
    gams[[ii]] <- matrix(0, nrow=M, ncol=ngam)
  }
  true.gam <- arma.gamma.by.Wold(ngam, a=a,
                                 sigma=sigma)
  for(mm in seq(M)){
    x <- ar.gen(L, a=a)
```

```
    for(ig in seq(along=ns)){
      n <- ns[ig]
      est.gam <- acf(x[1:n], lag.max=ngam-1,
                     type="covariance",
                     plot=F)$acf
      gams[[ig]][mm,] <- est.gam
    }
  }
  rats <- numeric(length(ns))
  rats[length(rats)] <- NA
  for(ig in seq(along=ns)){
    stderrs[ig,] <- apply(gams[[ig]], 2, sd)
  }
  for(ig in seq(along=ns)){
    if(ig<length(ns))
      rats[ig] <- mean(stderrs[ig+1]/stderrs[ig])
  }
  cat("Roots of characteristic polynomial: ",
      abs(1/roots[1]), "*Exp(", Arg(1/roots[1]), " i)\n", sep="")
  cat("True gamma's\n")
  print(round(true.gam[0:(ngam-1)],2))
  cat("Standard deviation of \\hat\\gamma_k ===\n")
  print(round(cbind(stderrs, "ratio"=rats),2))
}
```

首先对特征根离单位圆比较远的那个 AR(2) 模型进行模拟:

```
a <- coefs[1,]
demo.estacv.se(a)
```

```
## Roots of characteristic polynomial: 1.8*Exp(-2.13 i)
## True gamma's
## [1]  1.39 -0.62 -0.06  0.23 -0.12
## Standard deviation of \hat\gamma_k ===
##         0    1    2    3    4    5 ratio
```

```
## 1600 0.06 0.04 0.04 0.04 0.04 0.04  1.95
## 400  0.12 0.08 0.08 0.09 0.09 0.09  1.95
## 100  0.23 0.16 0.15 0.16 0.16 0.16  1.96
## 25   0.45 0.30 0.28 0.32 0.30 0.29    NA
```

其次对特征根离单位圆比较近的那个 AR(2) 模型进行模拟:

```
a <- coefs[2,]
demo.estacv.se(a)
```

```
## Roots of characteristic polynomial: 1.8*Exp(-2.13 i)
## True gamma's
## [1]  4.37 -2.31 -1.39  3.25 -1.99
## Standard deviation of \hat\gamma_k ===
##         0    1    2    3    4    5 ratio
## 1600 0.36 0.20 0.16 0.35 0.24 0.16  2.01
## 400  0.73 0.39 0.32 0.70 0.47 0.32  1.90
## 100  1.39 0.76 0.61 1.32 0.90 0.59  1.93
## 25   2.67 1.47 1.04 2.32 1.66 0.91    NA
```

16.5 附录：均值和自协方差估计的强大数律

当平稳列 X_t 的自协方差函数绝对可和或以幂率收敛时，对 $\forall\lambda \in [-\pi, \pi]$，有

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n X_j e^{-ij\lambda} = 0, \text{ a.s.}$$

其中 $\lambda = 0$ 时即样本均值服从强大数律。详见谢衷洁《时间序列分析》PP53–57 定理 1.16 及推论。何书元教材中定理 4.1.1 是均方收敛的结论。

当零均值正态实值平稳列 X_t 的自协方差函数以幂率收敛时，有

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n X_{k+j} X_j = \gamma_k, \text{ a.s.}$$

详见谢衷洁《时间序列分析》PP58–60 定理 1.17,1.18 及推论。关于自协方差的条件还可以减弱到谱函数连续，从而自协方差绝对可和或有谱密度时也可以。何书元教材的条件为严平稳遍历。

Chapter 17

白噪声检验

17.1 白噪声的 χ^2 检验

若 $\{X_t\}$ 是独立同分布的白噪声，根据推论16.2, N 足够大时

$$\sqrt{N}(\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_m)$$

服从 iid 标准正态分布。于是

$$\hat{\chi}^2(m) \triangleq N(\hat{\rho}_1^2 + \hat{\rho}_2^2 + \cdots + \hat{\rho}_m^2)$$

近似服从 $\chi^2(m)$ 分布。

要检验 $H_0 : \{X_t\}$ 为独立同分布白噪声，可以用否定域

$$\{\hat{\chi}^2(m) > \lambda_\alpha(m)\}$$

其中 $\lambda_\alpha(m)$ 为 $\chi^2(m)$ 的右侧 α 分位数。检验的 p 值为

$$\Pr(\chi^2(m) > \hat{\chi}^2(m)) = 1 - F_m(\hat{\chi}^2(m))$$

其中 $F_m(.)$ 为 $\chi^2(m)$ 分布函数。虽然 H_0 要求 $\{X_t\}$ 独立同分布，此方法可以用作一般性的白噪声检验。

上述检验叫做 **Box-Pierce 检验**。

类似可以使用统计量

$$\hat{\chi}^2_{\text{LB}}(m) \triangleq N(N-1)\sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{N-k}$$

在 H_0 下也近似服从 χ^2_m 分布。相应的检验称为 **Ljung-Box 检验**。

17.1.1 白噪声检验的模拟比较

R 的 `Box.test()` 函数提供了 Box-Pierce 检验和 Ljung-Box 检验功能。R 的 portes 包提供了多种一元和多元白噪声检验功能。

下面用一些非白噪声的序列数据来考察 LB 检验的功效。首先选用 AR(2) 序列，对 AR(2) 序列，特征根离单位圆越远，序列越表现得像是白噪声，检验功效越差；特征根离单位圆较近时则白噪声检验功效较高。特征根的模分别取为 $1000, 10, 6, 4, 3, 2, 1.5$，辐角为 1.13。样本量 $N = 400$, 卡方检验的平方和项数 $m = 5$ 和 $m = 20$。使用 Ljung-Box 检验。输出结果是不同的特征根模 ρ 对应的检验功效，分别取 $m = 5$ 和 $m = 20$。

```
mods <- c("1000", "10", "6", "4", "3", "2", "1.5")
revroots <- complex(mod=1/c(1000, 10, 6, 4, 3, 2, 1.5),
                  arg=1.13)
coefs <- cbind(2*Re(revroots), -Mod(revroots)^2)
nmodels <- length(revroots)
sigma <- 1
N <- 400       ## sample size
nrep <- 1000    ## number repetitions
alpha <- 0.05 ## test level
power <- numeric(nmodels)
res <- data.frame(rho=mods, pv5=0, pv20=0)

set.seed(1)
for(jj in seq(nmodels)){
  p5 <- 0
  p20 <- 0
  for(ii in seq(nrep)){
    #x <- ar.gen(n=N, a=coefs[jj,], sigma=sigma)
```

```
    x <- arima.sim(model=list(ar=coefs[jj,]), n=N, rand.gen=function(n, ...) rnorm(n, 0, sig
    p5 <- p5 + (Box.test(x, lag=5, type="Ljung")$p.value < alpha)
    p20 <- p20 + (Box.test(x, lag=20, type="Ljung")$p.value < alpha)
  }
  ## correct rate
  res[jj, "pv5"] <- 100 * p5 / nrep
  res[jj, "pv20"] <- 100 * p20 / nrep
}
##print(res, quote=F)
knitr::kable(res)
```

rho	pv5	pv20
1000	5.0	4.0
10	18.3	12.4
6	52.0	31.6
4	92.0	68.6
3	99.9	94.4
2	100.0	100.0
1.5	100.0	100.0

对于 MA 模型，不能使用特征根来选择与白噪声的差别大小。考虑 MA(1) 模型，b_1 越接近于 0，序列越接近于白噪声。

```
bs <- c(0.001, 0.1, 0.15, 0.2, 0.3, 0.5, 0.8, 0.9)
nmodels <- length(bs)
sigma <- 1
N <- 400       ## sample size
nrep <- 1000   ## number repetitions
alpha <- 0.05 ## test level
power <- numeric(nmodels)
res <- data.frame(b=bs, pv5=0, pv20=0)

set.seed(1)
for(jj in seq(nmodels)){
  p5 <- 0
```

```r
  p20 <- 0
  for(ii in seq(nrep)){
    x <- arima.sim(model=list(ma=bs[jj]), n=N, rand.gen=function(n, ...) rnorm(n,
    p5 <- p5 + (Box.test(x, lag=5, type="Ljung")$p.value < alpha)
    p20 <- p20 + (Box.test(x, lag=20, type="Ljung")$p.value < alpha)
  }
  ## correct rate
  res[jj, "pv5"] <- 100 * p5 / nrep
  res[jj, "pv20"] <- 100 * p20 / nrep
}
##print(res, quote=F)
knitr::kable(res)
```

b	pv5	pv20
0.001	4.8	5.3
0.100	25.8	15.8
0.150	57.5	35.2
0.200	87.5	61.9
0.300	100.0	96.8
0.500	100.0	100.0
0.800	100.0	100.0
0.900	100.0	100.0

关于 $\hat{\chi}^2(m)$ 中项数 m 的选取：$m=5$ 比 $m=20$ 有效。注意以 ARMA 模型为例，当 k 较大时 ρ_k 已经很小，所以 $\hat{\rho}_k^2$ 贡献不大，取太大的 m 容易使检验不敏感。

17.2 样本自相关置信区间检验法

当 $\{X_t\}$ 为独立同分布白噪声时 $\sqrt{N}(\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_m)$ 近似 m 维标准正态分布。

$$\Pr(\sqrt{N}|\hat{\rho}_k| > 1.96) \approx 0.05$$

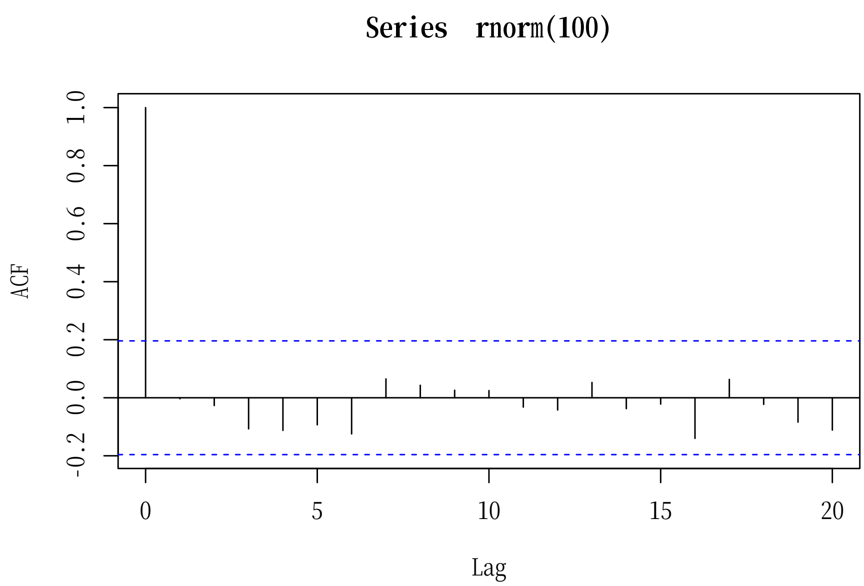
如果超过 5% 的 $|\hat{\rho}_j| \geq 1.96/\sqrt{n}$ 可否定 $H_0 : \{X_t\}$ 为独立同分布白噪声。与 χ^2 检验理由类似，m 不应取太大。

对 WN(0,1) 的检验通常不显著，检验成功。对 MA(1) 的检验如果取 $m = 20$ 则很可能不显著，使检验失败，因为一般只有 $\hat\rho_1$ 超过界限。对 AR 的检验一般是显著的，检验成功，因为其相关系数不截尾。

下面模拟演示。

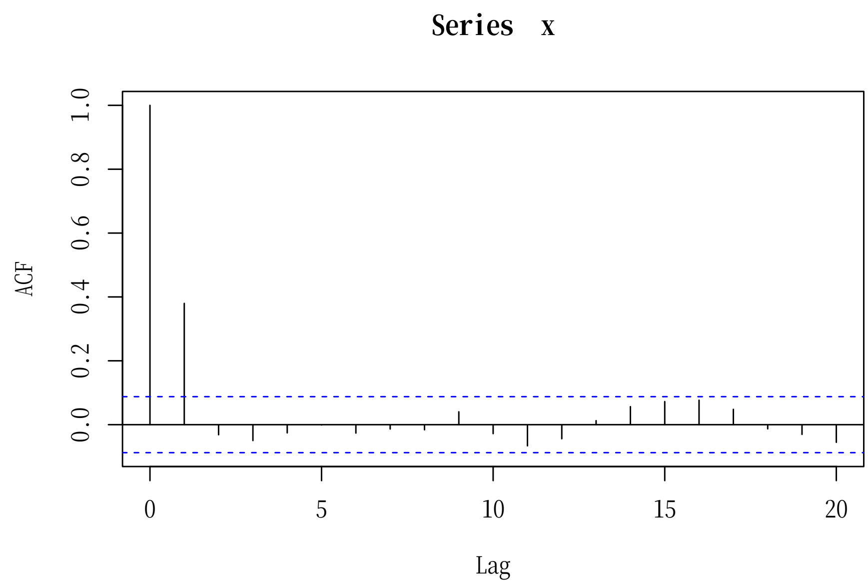
白噪声的 ACF:

```
set.seed(1)
acf(rnorm(100), lag.max=20)
```



MA(1) 当 $b = 0.5$ 时的 ACF:

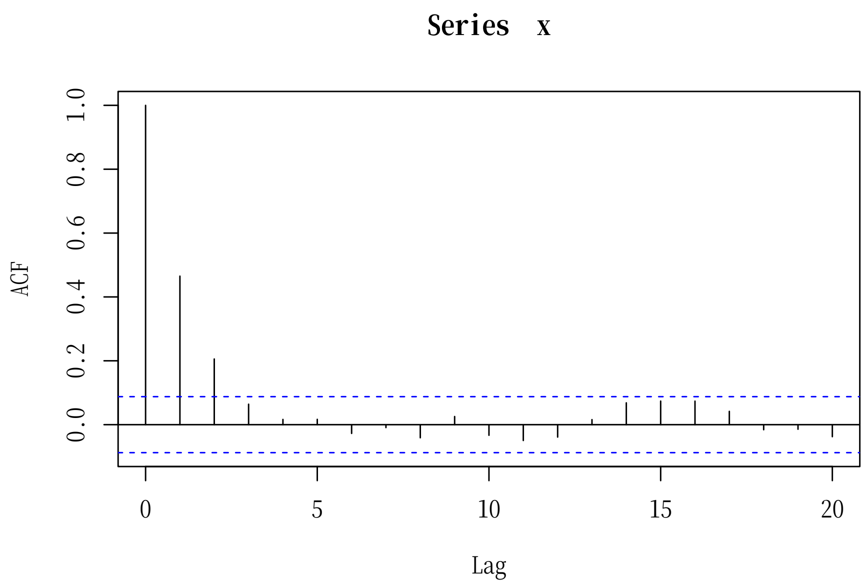
```
set.seed(1)
x <- arima.sim(model=list(ma=c(0.5)), n=500)
acf(x, lag.max=20)
```



$\hat{\rho}_1$ 到 $\hat{\rho}_{20}$ 中只有 $\hat{\rho}_1$ 超界。不能拒绝白噪声假设。

AR(1) 当 $a = 0.5$ 时的 ACF:

```
set.seed(1)
x <- arima.sim(model=list(ar=c(0.5)), n=500)
acf(x, lag.max=20)
```



$\hat{\rho}_1$ 和 $\hat{\rho}_2$ 都明显超界。

利用 $\sqrt{N}(\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_m)$ 近似独立标准正态做白噪声检验不如卡方检验，因为利用标准正态检验仅看超界的个数，而卡方检验还查看超界的量。在查看 ACF 时，如果某个 $\hat{\rho}_k$ 超界较多，也应怀疑非白噪声。

17.3 白噪声卡方检验的自由度调整

如果输入的序列 $\{X_t\}$ 是原始序列，计算 Box-Pierce 统计量或者 Ljung-Box 统计量，在白噪声零假设下统计量近似服从 $\chi^2(m)$ 分布。

自定义的 Ljung-Box 统计量计算函数：

```
LB_stat <- function(x, m = round(log(lenth(x)))){
  if(m < 1) m <- 1
  x <- as.vector(x)
  n <- length(x)
  rhos <- acf(x, lag.max = m, plot = FALSE)$acf[-1]
  stat <- n*(n-1)*sum(rhos^2 / (n - (1:m)))
  stat
```

```
}
```

下面是做一组模拟：

```
simwn1 <- function(n = 1024, m = 10, mean = 100, sd = 10){
  x <- rnorm(n, mean, sd)
  LB_stat(x, m)
}
```

上面是模拟一次并计算 LB 统计量的函数。n 是样本量，m 是卡方统计量中使用的自相关系数的个数，mean 和 sd 是白噪声的均值和标准差。

写一个重复模拟多次，计算许多个 LB 卡方统计量，并估计卡方统计量自由度的函数：

```
sim <- function(nrep = 1E4, fun = simwn1, n=1024, m=10){
  resm <- replicate(
    nrep, fun(n=n, m=m)  )
  est.df <- mean(resm)
  cat(" 平方和中的项数 = ", m, " 模拟估计自由度 = ", est.df, "\n")
}
```

上述程序中 nrep 是重复模拟次数，fun 是一次模拟的函数。模拟一万次，估计自由度：

```
set.seed(101)
sim(1E4, fun = simwn1, n=1024, m=10)
```

```
## 平方和中的项数 =  10  模拟估计自由度 =  9.976223
```

如果要检验的数据不是原始数据，而是 ARIMA 建模的残差，则自由度需要减去估计的系数个数（不包括截距项），而且自由度的误差较大。

下面模拟一个 $\phi_1 = 0.9$ 的 AR(1) 模型，并重复一万次，估计 LB 卡方统计量的实际自由度。AR(1) 的系数 ϕ_1 可以用 ρ_1 估计。

```
set.seed(101)
simar1 <- function(n = 1024, m = 10, mean = 100, sd = 10){
  x <- mean + sd*arima.sim(model = list(ar = c(0.9)), n = n)
```

```
  phi1 <- acf(x, lag.max=1, plot = FALSE)$acf[2]
  resi <- filter(x, c(1, -phi1), method = "convolution", sides = 1)[-1]
  LB_stat(resi, m)
}

sim(1E4, fun = simar1)
```

```
## 平方和中的项数 =  10  模拟估计自由度 =  9.151771
```

可见模拟得到的自由度估计近似等于平方和中项数减去模型系数个数，但误差较大，而且系数个数中不计入截距项或者序列均值。

下面模拟一个 ARMA(2,1) 的残差白噪声检验。自由度应该减去 3。

```
set.seed(101)
simarma21 <- function(
  n = 1024, m = 10,
  mean = 100, sd = 10){
  x <- mean + sd*arima.sim(
    model = list(ar = c(1.6, -0.64), ma=c(0.4)),
    n = n)
  mod <- arima(x, order=c(2,0,1), include.mean=TRUE)
  resi <- residuals(mod)
  LB_stat(resi, m)
}

sim(1E4, fun = simarma21)
## 平方和中的项数 =  10  模拟估计自由度 =  7.272344
```

在讨论条件异方差问题时，可以对 ARMA 模型的残差的平方进行白噪声检验。这时，检验统计量的自由度也需要调整吗？略修改上面的 ARMA(2,1) 模拟检验程序进行探索：

```
set.seed(101)
simarma21B <- function(
  n = 1024, m = 10,
```

```
  mean = 100, sd = 10){
  x <- mean + sd*arima.sim(
    model = list(ar = c(1.6, -0.64), ma=c(0.4)),
    n = n)
  mod <- arima(x, order=c(2,0,1), include.mean=TRUE)
  resi2 <- as.vector(residuals(mod))^2
  LB_stat(resi2, m)
}

sim(1E4, fun = simarma21B)
## 平方和中的项数 =  10  模拟估计自由度 =  9.862679
```

从这个模拟看出，一旦对 ARMA 的残差计算了平方，再计算 LB 检验就不需要调整自由度了。

Chapter 18

概率极限

18.1　几乎必然收敛

定理 18.1. 随机向量序列 ξ_n *a.s.* 收敛到 ξ，$g(\cdot)$ 为 $\mathbb{R}^n \to \mathbb{R}$ 的连续函数，则 $g(\xi_n)$ *a.s.* 收敛到 $g(\xi)$。

如果 $g(\cdot)$ 定义是开集或者闭集 G 上的连续函数，ξ_n 和 ξ 都取值于 G，则定理结论也成立。

18.2　依概率收敛

定理 18.2. 设 $\{a_n\}$ 为实数列，则 $a_n \overset{P}{\to} a$ 等价于 $\lim_{n\to\infty} a_n = a$。

证明充分性。设 $\lim_n a_n = a$，则 $\forall \delta > 0$, $\exists N$ 使得 $n > N$ 时 $|a_n - a| \leq \delta$。所以

$$\lim_n P(|a_n - a| > \delta) = \lim_n 0 = 0.$$

必要性。设 $a_n \overset{\text{P}}{\to} a$。$\forall \delta > 0$, 记 $p_n = P(|a_n - a| > \delta)$, 则 $\lim_n p_n = 0$。但是 p_n 只能取 0 或者 1，所以 $\{p_n\}$ 中只有有限个 1，所以存在 N 使得 $n > N$ 时 $P(|a_n - a| > \delta) = 0$, 即 $n > N$ 时 $|a_n - a| \leq \delta$，即 $\lim_n a_n = a$。

定理 18.3. 设 $\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则

$$\xi_n + \eta_n \xrightarrow{P} \xi + \eta.$$

证明任给定 $\varepsilon > 0$。

$$\begin{aligned}
&P(|(\xi_n+\eta_n)-(\xi+\eta)| \geq \varepsilon) \\
\leq &P(|\xi_n-\xi|+|\eta_n-\eta| \geq \varepsilon) \\
\leq &P(|\xi_n-\xi| \geq \frac{\varepsilon}{2} \text{ 或} |\eta_n-\eta| \geq \frac{\varepsilon}{2}) \\
\leq &P(|\xi_n-\xi| \geq \frac{\varepsilon}{2}) + P(|\eta_n-\eta| \geq \frac{\varepsilon}{2}) \\
\to &0, \quad (n \to \infty)
\end{aligned}$$

定理 18.4. 设 $\xi_n \xrightarrow{P} \xi$, a 为常数, 则

$$a\xi_n \xrightarrow{P} a\xi.$$

证明 $a = 0$ 时显然。当 $a \neq 0$ 时，$\forall \varepsilon > 0$,

$$\begin{aligned}
&P(|a\xi_n - a\xi| \geq \varepsilon) \\
= &P(|\xi_n - \xi| \geq \frac{\varepsilon}{|a|}) \to 0 \quad (n \to \infty)
\end{aligned}$$

引理 18.1. 设 $\{\xi_n\}$, $\{\eta_n\}$ 为两个随机序列, a, b, c 为常数, 若 $\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则 $a\xi_n + b\eta_n + c \xrightarrow{P} a\xi + b\eta + c$。

定理 18.5. 设 $\xi_n \xrightarrow{P} a$, a 为常数，函数 $g(\cdot)$ 在 a 连续，则

$$g(\xi_n) \xrightarrow{P} g(a).$$

证明 $\forall \varepsilon > 0$, $\exists \delta > 0$ 使得 $|x-a| < \delta$ 时 $|g(x)-g(a)| < \varepsilon$。于是

$$|g(x)-g(a)| \geq \varepsilon \Longrightarrow |x-a| \geq \delta$$

于是

$$P(|g(\xi_n)-g(a)| \geq \varepsilon) \leq P(|\xi_n - a| \geq \delta) \to 0, \quad (n \to \infty)$$

例如，若 $\xi_n \xrightarrow{\text{P}} a$, 则

$$\xi_n^2 \xrightarrow{\text{P}} a^2$$
$$1/\xi_n \xrightarrow{\text{P}} 1/a \quad (\text{只要} a \neq 0)$$
$$\sqrt{\xi_n} \xrightarrow{\text{P}} \sqrt{a} \quad (\text{只要} a \geq 0)$$

定理 18.6. 设 $\xi_n \xrightarrow{P} \xi$, $g(\cdot)$ 为连续函数，则

$$g(\xi_n) \xrightarrow{P} g(\xi).$$

证明参考 Tucker, H.G.(1967), A Graduate Course in Probability, New York: Academic Press.

定理 18.7. 设 $\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则

$$\xi_n \eta_n \xrightarrow{P} \xi\eta.$$

证明

$$\xi_n \eta_n = \frac{1}{2}\left[\xi_n^2 + \eta_n^2 - (\xi_n - \eta_n)^2\right]$$

其中 $\xi_n^2 \xrightarrow{\text{P}} \xi^2$, $\eta_n^2 \xrightarrow{\text{P}} \eta^2$, $\xi_n - \eta_n \xrightarrow{\text{P}} \xi - \eta$, $(\xi_n - \eta_n)^2 \xrightarrow{\text{P}} (\xi - \eta)^2$, 所以

$$\xi_n \eta_n \xrightarrow{\text{P}} \frac{1}{2}\left[\xi^2 + \eta^2 - (\xi - \eta)^2\right] = \xi\eta$$

18.3 依分布收敛

定理 18.8. 设 $\xi_n \xrightarrow{P} \xi$, 则 $\xi_n \xrightarrow{d} \xi$。

证明设 $\xi_n \sim F_n(\cdot)$, $\xi \sim F(\cdot)$, 设 x 为 $F(\cdot)$ 的一个连续点。对任意 $\epsilon > 0$,

$$\begin{aligned}
F_n(x) =& P(\xi_n \leq x) \\
=& P\left(\xi_n \leq x,\ |\xi_n - \xi| < \epsilon\right) + P\left(\xi_n \leq x,\ |\xi_n - \xi| \geq \epsilon\right) \\
\leq& P(\xi \leq x + \epsilon) + P(|\xi_n - \xi| \geq \epsilon)
\end{aligned}$$

于是

$$\overline{\lim_{n\to\infty}} F_n(x) \le F(x+\epsilon).$$

另一方面，

$$\begin{aligned} P(\xi_n > x) =& P(\xi_n > x,\ |\xi_n - \xi| < \epsilon) + P(\xi_n > x,\ |\xi_n - \xi| \ge \epsilon) \\ \le& P(\xi > x - \epsilon) + P(|\xi_n - \xi| \ge \epsilon), \\ \overline{\lim_{n\to\infty}} [1 - F_n(x)] \le& 1 - F(x - \epsilon), \\ \underline{\lim_{n\to\infty}} F_n(x) \ge& F(x - \epsilon), \end{aligned}$$

总之有

$$F(x-\epsilon) \le \underline{\lim_{n\to\infty}} F_n(x) \le \overline{\lim_{n\to\infty}} F_n(x) \le F(x+\epsilon),$$

令 $\epsilon \to 0+$ 则可知

$$\lim_{n\to\infty} F_n(x) = F(x).$$

定理 18.9. *若 a 是常数，$\xi_n \xrightarrow{d} a$，则 $\xi_n \xrightarrow{P} a$。*

证明记

$$F(x) = \begin{cases} 1, & x \ge a \\ 0, & x < a \end{cases}$$

则

$$P(\xi_n \le x) \to F(x),\ \forall x \ne a.$$

$\forall \delta > 0$,

$$\begin{aligned} & P(|\xi_n - a| > \delta) \\ =& P(\xi_n > a + \delta) + P(\xi_n < a - \delta) \\ =& 1 - P(\xi_n \le a + \delta) + P(\xi_n < a - \delta) \\ \le& 1 - P(\xi_n \le a + \delta) + P(\xi_n \le a - \delta) \\ \to& 0 \quad (n \to \infty) \end{aligned}$$

依分布收敛不一定依概率收敛。比如，设 $X \sim \mathrm{N}(0,1)$，则 $-X$ 与 X 同分布。令

$$X_n = \begin{cases} X, & n \text{ 为偶数}, \\ -X, & n \text{ 为奇数}, \end{cases}$$

则 $\{X_n\}$ 依分布收敛到 X，但是不依概率收敛到 X。

概率质量函数 (PMF) 的收敛性与分布函数收敛性不同。例如，取 $\xi_n = 2 + \frac{1}{n}$，则 ξ_n 的 PMF 为

$$p_n(x) = \begin{cases} 1, & x = 2 + \frac{1}{n}, \\ 0, & \text{其它} \end{cases},$$

且

$$\lim_n p_n(x) = 0,\ x \in (-\infty, \infty),$$

但是 ξ_n 的分布函数趋于 $\xi = 2$ 的分布函数。

如果 ξ_n 的密度函数 $p_n(x) \to p(x)$，$p(x)$ 为 ξ 的密度函数，$p_n(x)$ 有可积的上界，则根据控制收敛定理可知，ξ_n 依分布收敛到 ξ。

定理 18.10. 设 $\xi_n \overset{d}{\to} \xi$, $\eta_n \overset{P}{\to} 0$, 则

$$\xi_n + \eta_n \overset{d}{\to} \xi.$$

证明设 x_0 是 ξ 的分布函数 $F(x)$ 的连续点。对于 $\delta > 0$，由

$$\begin{aligned} & P(\xi_n + \eta_n \le x_0) \\ =& P(\xi_n + \eta_n \le x_0,\ \eta_n \le -\delta) + P(\xi_n + \eta_n \le x_0,\ \eta_n > -\delta) \\ \le& P(\eta_n \le -\delta) + P(\xi_n \le x_0 + \delta) \\ \le& P(|\eta_n| \ge \delta) + P(\xi_n \le x_0 + \delta) \end{aligned}$$

可知

$$\begin{aligned} & P(\xi_n + \eta_n \le x_0) - F(x_0) && (18.1) \\ \le& [P(\xi_n \le x_0 + \delta) - F(x_0 + \delta)] + [F(x_0 + \delta) - F(x_0)] + P(|\eta_n| \ge \delta) && (18.2) \end{aligned}$$

另一方面，

$$
\begin{aligned}
&P(\xi_n+\eta_n\le x_0)\\
\ge &P(\xi_n+\eta_n\le x_0,\ \eta_n\le\delta)\\
\ge &P(\xi_n+\delta\le x_0,\ \eta_n\le\delta)\\
= &P(\xi_n+\delta\le x_0)-P(\xi_n+\delta\le x_0,\ \eta_n>\delta)\\
\ge &P(\xi_n\le x_0-\delta)-P(\eta_n>\delta)\\
\ge &P(\xi_n\le x_0-\delta)-P(|\eta_n|>\delta)
\end{aligned}
$$

于是

$$
\begin{aligned}
&P(\xi_n+\eta_n\ge x_0)-F(x_0) &(18.3)\\
\ge &P(\xi_n\le x_0-\delta)-F(x_0-\delta)+F(x_0-\delta)-F(x_0)-P(|\eta_n|>\delta) &(18.4)
\end{aligned}
$$

任给定 $\varepsilon>0$，取 $\delta_1>0$ 足够小使得

$$F(x_0+\delta_1)-F(x_0)<\frac{\varepsilon}{3}$$

且 $x_0+\delta_1$ 是 $F(x)$ 的连续点（由 Lebesgue 定理，单调函数几乎处处可微，从而在任意小的区间上都有连续点）。由于 $\xi_n\overset{\mathrm{d}}{\to}\xi$, $\eta_n\overset{\mathrm{P}}{\to}0$, 存在 n_1 使得 $\forall n\ge n_1$，

$$
\begin{aligned}
P(\xi_n\le x_0+\delta)-F(x_0+\delta)<&\frac{\varepsilon}{3}\\
P(|\eta_n|\ge\delta)<&\frac{\varepsilon}{3}
\end{aligned}
$$

由(18.2)式, 当 $n\ge n_1$ 时

$$P(\xi_n+\eta_n\le x_0)-F(x_0)<\varepsilon$$

再取 $\delta_2>0$ 使

$$F(x_0)-F(x_0-\delta_2)<\frac{\varepsilon}{3}$$

且 $x_0-\delta_2$ 是 $F(x)$ 的连续点，存在 $n_2\ge n_1$ 使得 $\forall n\ge n_2$ 有

$$
\begin{aligned}
P(\xi_n\le x_0-\delta)-F(x_0-\delta)>&-\frac{\varepsilon}{3}\\
P(|\eta_n|\ge\delta)<&\frac{\varepsilon}{3}
\end{aligned}
$$

由(18.4)可得当 $n\ge n_2$ 时

$$P(\xi_n+\eta_n\le x_0)-F(x_0)>-\varepsilon$$

于是当 $n \geq n_2$ 时

$$|P(\xi_n + \eta_n \leq x_0) - F(x_0)| < \varepsilon$$

即有

$$\lim_{n\to\infty} P(\xi_n + \eta_n \leq x_0) = F(x_0)$$

即

$$\xi_n + \eta_n \xrightarrow{\mathrm{d}} \xi.$$

定理 18.11. *设 $\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{P} 1$, 则 $\eta_n\xi_n \xrightarrow{d} \xi$。*

证明与定理18.10类似。

定理 18.12. *设 $\xi_n \xrightarrow{d} \xi$, $g(\cdot)$ 是定义在 ξ 的支撑集上的连续函数，则*

$$g(\xi_n) \xrightarrow{d} g(\xi).$$

证明略。例如，$\xi_n \xrightarrow{\mathrm{d}} \mathrm{N}(0,1)$，则 $\xi_n^2 \xrightarrow{\mathrm{d}} \chi^2(1)$。

定理 18.13. *设 $\xi_n \xrightarrow{d} \xi$, $A_n \xrightarrow{P} a$, $B_n \xrightarrow{P} b$, 则*

$$A_n + B_n\xi_n \xrightarrow{d} a + b\xi.$$

证明略。

定理 18.14. *设取正整数整数值的随机变量序列 $\{N_n\}$ 满足 $N_n \leq N_{n+1}, \forall n$, ξ_n 依分布收敛到 ξ。*

(1) 如果 $\{N_n\}$ 与 $\{\xi_n\}$ 相互独立，则 ξ_{N_n} 依分布收敛到 ξ;

(2) 如果存在 $c > 0$ 使得 $N_n/n \to c$, a.s.，则 ξ_{N_n} 依分布收敛到 ξ。

其中的第二条称为 Anscombe 定理。证明略。

定理 18.15. *设随机变量序列 $\{\xi_n\}$ 依分布收敛到 ξ，$p > 0$，存在非负随机变量 η 使得 $E\eta < \infty$ 且 $|\xi_n| \leq \eta, \forall n$，则 $E\xi_n^p \to E\xi^p$。*

定理 18.16. 设随机变量序列 $\{\xi_n\}$ 依分布收敛到 ξ，$p>0$，$E\xi_n^p$ 满足如下的一致可积性条件：

$$\lim_{c\to\infty}\sup_{n\geq 1}E\left\{|\xi_n|^p 1_{\{|\xi_n|^p>c\}}\right\}=0$$

则 $E\xi_n^p\to E\xi^p$。

证明略。

定理 18.17. 设 $\{F_1,F_2,\dots,\}$ 是分布函数列，F_n 依分布收敛到 F，若对某一个 $\delta>0$ 和 $p_0>0$，数列 $\{\int_R|x|^{p_0+\delta}dF_n(x):n\geq 1\}$ 有界，则对任意 $p\in[0,p_0]$ 和整数 $k\in(0,p_0]$，当 $n\to\infty$ 时有

$$\int_R x^k dF_n(x)\to\int_R x^k dF(x),\quad \int_R|x|^p dF_n(x)\to\int_R|x|^p dF(x)$$

见朱成熹《测度论基础》（科学出版社 1983 年）P126 的推论。即二阶矩有界加依分布收敛推出一阶矩收敛；三阶矩有界加依分布收敛推出二阶矩收敛。

反例：设 $X_n\sim 2n\mathrm{U}(0,\frac{1}{n})$，则 $X_n\to 0$, a.s.。$EX_n\equiv 1$ 而 $E0=0$。

依分布收敛可以推广到更一般的情形。设 (S,ρ) 是距离空间，$\mathscr{B}(S)$ 是包含其中的开集的最小 σ 代数，若 Q_n 和 Q 是 $(S,\mathscr{B}(S))$ 中的概率测度，满足

$$\lim_{n\to\infty}\int_S g(s)\,dQ_n(s)=\int_S g(s)\,dQ(s)$$

对任意定义在 S 上的实值连续有界函数 g 成立，则称 Q_n 弱收敛到 Q。

X 是概率空间 $(\Omega,\mathscr{F},P)$ 到可测空间 $(S,\mathscr{B}(S))$ 的可测函数，称为随机元。X 导出了 S 中的测度 $Q(A)=P(X^{-1}(A))$。若 $P(X_n^{-1}(\cdot))$ 弱收敛到 $P(X^{-1}(\cdot))$，则称 X_n 弱收敛到 X（或依分布收敛）。

18.4 概率母函数

若 X 为取非负整数值的离散型随机变量，$P(X=j)=p_j$, 则

$$P(t)=Et^X=\sum_{j=0}^{\infty}t^j p_j$$

在 $t\in[-1,1]$ 一致绝对收敛，称 $P(t)$ 为 X 的概率母函数。

定理 18.18. 设取值为非负整数的随机变量 X 有概率母函数 $P(t) = Et^X$, $t \in [-1, 1]$, 和分布列 $\{p_j\}$，则分布列与概率母函数相互唯一决定。

分布列决定概率母函数显然；反之的证明略。参见 (李贤平, 2010) 节 4.4, (何书元, 2006) 节 5.1。

$$
\begin{aligned}
EX =& P'(1) \\
E(X(X-1)) =& P''(1) \\
\text{Var}(X) =& EX^2 - (EX)^2 = P''(1) + P'(1) - [P'(1)]^2
\end{aligned}
$$

18.5 矩母函数

对随机变量 X，如果存在 $h > 0$ 使得

$$
M(t) = Ee^{tX}
$$

在 $t \in (-h, h)$ 存在，则称 $M(t)$ 为 X 的**矩母函数**。条件也可以放松到在 $t \in [0, h)$ 存在，或者在 $t \in (-h, 0]$ 存在。

矩母函数存在时，

$$
\begin{aligned}
\frac{d}{dt} M(t) =& E(X e^{tX}), \quad \frac{d}{dt} M(0) = EX \\
\frac{d^2}{dt^2} M(t) =& E(X^2 e^{tX}), \quad \frac{d^2}{dt^2} M(0) = E(X^2) \\
& \cdots\cdots \\
\frac{d^n}{dt^n} M(t) =& E(X^n e^{tX}), \quad \frac{d^n}{dt^n} M(0) = E(X^n)
\end{aligned}
$$

这也是“矩母函数”的名称来源。

定理 18.19. 设随机变量 X 有矩母函数 $M(t) = Ee^{tX}$, $t \in (-h, h)$, 和分布函数 $F(x)$，则分布函数与矩母函数相互唯一决定。

分布函数决定矩母函数显然，矩母函数决定分布函数的证明略。

定理 18.20. 设 ξ_n 有矩母函数 $M_n(t) = Ee^{t\xi_n}$, $t \in (-h, h)$, ξ 有矩母函数 $M(t) = Ee^{t\xi}$, $t \in [-h_1, h_1]$, $0 < h_1 \leq h$, 若 $\lim_{n\to\infty} M_n(t) = M(t)$, $\forall t \in [-h_1, h_1]$, 则 $\xi_n \xrightarrow{d} \xi$。

证明略。

18.6 特征函数

概率母函数与矩母函数存在时都可以决定分布，对于研究独立随机变量和的分布有好的性质，但都不是对所有随机变量存在的。特征函数具有相同的优良性，而且对于所有的随机变量都存在，只不过用到复数，数学上较复杂。

对随机变量 X，定义

$$\phi(t) = Ee^{iX} = E\cos(tX) + i\,E\sin(tX),\ t \in \mathbb{R}$$

称 $\phi(t)$ 为 X 的**特征函数**。特征函数存在唯一。

定理 18.21 (逆转公式)**.** *如果 X 的分布函数 $F(x)$ 在 a, b 连续，则*

$$F(b) - F(a) = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{ita} - e^{itb}}{it}\phi(t)\,dt$$

证明略。

定理 18.22. *随机变量的分布函数与特征函数相互唯一决定。*

特征函数的性质：

定理 18.23. *设 X 为随机变量，$\phi(t) = Ee^{itX}$，则*

(1) $\phi(0) = 1$, $|\phi(t)| \le 1$, $\phi(-t) = \overline{\phi(t)}$。

(2) $\phi(t)$ 在 $(-\infty, \infty)$ 一致连续。

(3) 如果 EX^k 存在，则

$$\phi^{(k)}(t) = i^k E(X^k e^{itX}), \quad \phi^{(k)}(0) = i^k E(X^k)$$

(4) 非负定性：对任意复数 $a_1, a_2, \dots, a_n$，和实数 $t_1, t_2, \dots, t_n$，都有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}\phi(t_k - t_j)a_k\bar{a}_j \ge 0$$

(5) 设随机变量 $X_1, X_2, \dots, X_n$ 相互独立，如果 X_j 的特征函数为 $\phi_j(t)$，令 $Y = \sum_{j=1}^n X_j$，则 Y 的特征函数为

$$\phi_Y(t) = \prod_{j=1}^n \phi_j(t).$$

见 (何书元, 2006) 节 5.2。

正态分布 $\mathrm{N}(\mu, \sigma^2)$ 的特征函数为

$$\phi(t) = \exp\{i\mu t - \frac{1}{2}\sigma^2 t^2\}.$$

定理 18.24 (连续性定理)**.** 设 ξ_n 的特征函数为 $\phi_n(t)$，ξ 的特征函数为 $\phi(t)$，则 $\xi_n \xrightarrow{d} \xi$ 的充分必要条件是

$$\lim_{n\to\infty} \phi_n(t) = \phi(t),\ \forall t \in (-\infty, \infty)$$

证明略。

对随机向量 X, 定义其特征函数为

$$\phi(t) = E e^{it^T X},\ t \in \mathbb{R}^n$$

特征函数有定义。

定理 18.25 (随机向量特征函数性质)**.** 设 $X = (X_1, \dots, X_n)^T$，X 的特征函数为 $\phi(t)$, 分量 X_j 的特征函数为 $\phi_j(t)$。则

(1) $\phi(t)$ 与 X 的分布函数相互唯一决定；

(2) $X_1, X_2, \dots, X_n$ 相互独立当且仅当

$$\phi(t) = \phi_1(t_1)\phi_2(t_2)\dots\phi_n(t_n),\ \forall t = (t_1, t_2, \dots, t_n) \in \mathbb{R}^n.$$

(3) 设 $\{\xi_k\}$ 为随机向量序列, ξ_k 的特征函数为 $\phi_k(t)$, 如果 $\phi_k(t)$ 收敛到在 $t = 0$ 连续的函数 $g(t)$, 则 $g(t)$ 是某个随机向量 ξ 的特征函数，且对任意 $a \in \mathbb{R}^n$，有 $a^T\xi_k \xrightarrow{d} a^T\xi$。

18.7 中心极限定理

定理 18.26. 设 $\{\xi_n\}$ 为独立同分布的随机变量序列, $\mu = E\xi_1$, $\sigma^2 = Var(\xi_1) < \infty$, $\bar{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$, 则

$$\frac{\xi_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1).$$

定理 18.27. 设 $\{\xi_n\}$ 为独立同分布的随机变量序列, $\mu = E\xi_1$, $\sigma^2 = Var(\xi_1) < \infty$, $\bar{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$, 设 $\hat{\sigma}_n^2$ 依概率收敛到 σ^2, 则

$$\frac{\xi_n - \mu}{\hat{\sigma}_n/\sqrt{n}} \xrightarrow{d} N(0,1).$$

18.8 依概率有界

设 $\{\xi_n\}$ 是随机变量序列，如果对任意 $\varepsilon > 0$，存在正数 M，使得

$$\sup_n P(|\xi_n| > M) \le \varepsilon$$

就称随机变量序列 $\{\xi_n\}$ 是**依概率有界的**，记做 $\xi_n = O_p(1)$。对单个的随机变量 ξ，$\forall \varepsilon > 0$，显然存在 $M > 0$ 使得 $P(|\xi| > M) \le \varepsilon$，一个序列的依概率有界是要求对整个序列，给定 $\varepsilon > 0$ 后有共同的 M 使得 $P(|\xi_n| > M) \le \varepsilon$ 同时成立。

设 $\{c_n\}$ 是非零常数列，如果 $\{\xi_n/c_n\} = O_p(1)$，就称 $\xi_n = O_p(c_n)$。设随机变量序列 $\eta_n \ne 0$，若 $\{\xi_n/\eta_n\} = O_p(1)$ 则称 $\xi_n = O_p(\eta_n)$。

依概率有界的**等价定义**：称 $\{\xi_n\}$ 依概率有界，若 $\forall \varepsilon > 0, \exists M > 0$ 和 N 使得当 $n \ge N$ 时

$$P(|\xi_n| \le M) \ge 1 - \varepsilon.$$

事实上，当原定义条件成立时显然此等价定义的条件也成立。若此等价定义条件成立，则对 $n \ge N$ 有

$$P(|\xi_n| \le M) \ge 1 - \varepsilon, \quad P(|\xi_n| > M) < \varepsilon.$$

对 $j = 1, 2, \dots, N$，存在 $M_j > 0$ 使得

$$P(|\xi_j| > M_j) < \varepsilon$$

令 $M' = \max(M, M_1, M_2, \ldots, M_N)$，则

$$\begin{aligned} P(|\xi_n| > M') <& \varepsilon, \ \forall n \in \mathbb{N}_+, \\ \sup_{n \in \mathbb{N}_+} P(|\xi_n| > M') \le& \varepsilon, \end{aligned}$$

满足原定义。

若 $\xi_n \xrightarrow{\text{Pr}} 0$ $(n \to \infty)$ 则记 $\xi_n = o_p(1)$。设 $\{c_n\}$ 是非零常数列，如果 $\{\xi_n / c_n\} = o_p(1)$，就称 $\xi_n = o_p(c_n)$。设随机变量序列 $\eta_n \neq 0$，若 $\{\xi_n / \eta_n\} = o_p(1)$ 则称 $\xi_n = o_p(\eta_n)$。

定理 18.28. 若 $\xi_n = o_p(c_n)$ 则 $\xi_n = O_p(c_n)$。

证明按依概率收敛定义，$\forall \delta > 0, \forall \varepsilon > 0$，存在 N 使 $n > N$ 时

$$P(|\xi_n / c_n| > \delta) < \varepsilon$$

取 $M \ge \delta$ 使得

$$P(\max_{1 \le n \le N} |\xi_n / c_n| > M) < \varepsilon,$$

则

$$P(|\xi_n / c_n| > M) < \varepsilon, \quad n = 1, 2, \ldots$$

即 $\xi_n / c_n = O_p(1)$.

定理 18.29. 如果 $\xi_n = O_p(1)$ 而 $c_n \to \infty$ 则 $\xi_n / c_n = o_p(1)$。

证明 $\forall \varepsilon > 0, \forall \delta > 0$，$\exists M > 0$ 使

$$P(|\xi_n| > M) < \varepsilon, \ n \in \mathbb{N}_+.$$

$\exists N$ 使 $n > N$ 时 $c_n > M / \delta$, 于是

$$P\left(\left|\frac{\xi_n}{c_n}\right| > \delta\right) = P(|\xi_n| > |c_n| \delta) \le P(|\xi_n| > M) < \varepsilon.$$

定理 18.30. 对随机变量 ξ，$\xi = O_p(1)$.

定理 18.31. 若存在非负随机变量 ξ 使得 $|\xi_n| \le \xi$ $a.s.$ 则 $\xi_n = O_p(1)$。

证明由 $P(|\xi_n| \le \xi) = 1$，$\forall \varepsilon > 0$, $\exists M > 0$ 使 $P(\xi > M) < \varepsilon$，于是 $P(|\xi_n| > M) \le P(\xi > M) < \varepsilon$, $\forall n \in \mathbb{N}_+$。

定理 18.32. 若存在常数 $d > 0$, $c > 0$ 使得使得 $\sup_n E|\xi_n|^d \le c$, 则 $\xi_n = O_p(1)$。

证明由马尔可夫不等式，对任意 $M > 0$ 有

$$P(|\xi_n| > M) \le \frac{E|\xi_n|^d}{M^d} \to 0 \ (M \to \infty)$$

所以对任意 $\epsilon > 0$，存在 $M > 0$ 使得

$$P(|\xi_n| > M) < \epsilon, \ \forall n$$

定理 18.33. 若 $\{\xi_n\}$ 同分布，则 $\xi_n = O_p(1)$。

证明 $\forall \varepsilon > 0$, $\exists M > 0$ 使 $P(|\xi_1| > M) < \varepsilon$。由同分布性知 $P(|\xi_t| > M) = P(|\xi_1| > M) < \varepsilon$。

定理 18.34.

$$\begin{aligned} O_p(1) \pm O_p(1) =& O_p(1) \\ O_p(1) \cdot O_p(1) =& O_p(1) \\ O_p(1) \pm o_p(1) =& O_p(1) \\ O_p(1) \cdot o_p(1) =& o_p(1). \end{aligned}$$

证明仅证明 $O_p(1) \cdot o_p(1) = o_p(1)$。设 $\xi_n = O_p(1)$, $\eta_n = o_p(1)$。对任意给定的 $\delta > 0$ 和 $\epsilon > 0$, 存在 M, 使得

$$\sup_n P(|\xi_n| > M) < \epsilon.$$

于是

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} P(|\xi_n\eta_n| > \delta)\\
\le&\overline{\lim_{n\to\infty}} P(|\xi_n\eta_n| > \delta, |\xi_n| \le M) + \overline{\lim_{n\to\infty}} P(|\xi_n\eta_n| > \delta, |\xi_n| > M)\\
\le&\overline{\lim_{n\to\infty}} P(|\eta_n| > \frac{\delta}{M}) + \overline{\lim_{n\to\infty}} P(|\xi_n| > M)\\
\le&\epsilon,
\end{aligned}$$

即 $\lim_n P(|\xi_n\eta_n| > \delta) = 0$, $\xi_n\eta_n = o_p(1)$。

定理 18.35. 若 ξ_n 依概率收敛到 ξ 则 $\{\xi_n\} = O_p(1)$。

证明 $\xi_n = \xi + (\xi_n - \xi) = O_p(1) + o_p(1)$。

定理 18.36. 设 $\xi_n \xrightarrow{d} \xi$, 则 $\xi_n = O_p(1)$。

证明设 ξ 分布函数为 $F(x)$，$\forall\varepsilon > 0$，存在 $M > 0$ 且 M 和 $-M$ 为 $F(x)$ 的连续点，使得

$$\begin{aligned}
\underline{\lim_{n\to\infty}} P(|\xi_n| \le M) =& \lim_{n\to\infty} P(\xi_n \le M) - \underline{\lim_{n\to\infty}} P(\xi_n < -M)\\
\ge& \lim_{n\to\infty} P(\xi_n \le M) - \lim_{n\to\infty} P(\xi_n \le -M)\\
=&F(M) - F(-M) \ge 1 - \frac{\varepsilon}{2} > 1 - \varepsilon
\end{aligned}$$

于是 $\exists N$，当 $n \ge N$ 时

$$\begin{aligned}
\inf_{m\ge n} P(|\xi_m| \le M) \ge& 1 - \varepsilon\\
P(|\xi_n| \le M) \ge& 1 - \varepsilon
\end{aligned}$$

由 O_p 等价定义可知 $\xi_n = O_p(1)$。

定理 18.37. 设 $\xi_n = O_p(1)$，$\eta_n/\xi_n = o_p(1)$, 则 $\eta_n = o_p(1)$。

证明

$$\eta_n = o_p(1) \cdot O_p(1) = o_p(1).$$

也可以直接证明如下。

$\forall \delta > 0$, $\forall \varepsilon > 0$。由 $\xi_n = O_p(1)$ 可知存在 $M > 0$ 使得

$$P(|\xi_n| > M) < \frac{\varepsilon}{2},\ \forall n \in \mathbb{N}_+.$$

由 $\eta_n/\xi_n = o_p(1)$ 可知存在 $N > 0$ 使得 $n > N$ 时

$$P(|\eta_n/\xi_n| > \delta/M) < \frac{\varepsilon}{2},$$

于是 $n > N$ 时

$$\begin{aligned}
&P(|\eta| > \delta) \\
=&P(|\eta| > \delta,\ |\xi_n| \le M) + P(|\eta| > \delta,\ |\xi_n| > M) \\
\le&P(|\eta_n/\xi_n| > \delta/M) + P(|\xi_n| > M) \\
<&\varepsilon
\end{aligned}$$

即 $\eta_n = o_p(1)$。

18.9 Delta 方法

定理 18.38. 设随机变量序列 $\{\xi_n\}$ 有极限分布

$$\sqrt{n}(\xi_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

函数 $g(x)$ 在 θ 处可微，$g'(\theta) \neq 0$。则

$$\sqrt{n}(g(\xi_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2(g'(\theta))^2).$$

证明由泰勒公式

$$g(\xi_n) = g(\theta) + g'(\theta)(\xi_n - \theta) + \eta_n$$

其中

$$\begin{aligned}\eta_n =& h(\xi_n-\theta)\\ h(x) =& g(x+\theta)-g(\theta)-g'(\theta)x\\ h'(0) =& 0\end{aligned}$$

由后面的引理18.2可知 $\eta_n = o_p(|\xi_n-\theta|)$，于是

$$\sqrt{n}(g(\xi_n)-g(\theta)) = \sqrt{n}g'(\theta)(\xi_n-\theta)+o_p(\sqrt{n}|\xi_n-\theta|)$$

前一项依分布收敛到 $\mathrm{N}(0,\sigma^2(g'(\theta))^2)$, 后一项中 $\sqrt{n}|\xi_n-\theta| = O_p(1)$ 所以是 $o_p(1)$ 的，于是结果可得。

引理 18.2. *设函数 $h(x)$ 在 $x=0$ 处可微，$h'(0)=0$。若 $\xi_n = o_p(1)$ 则 $h(\xi_n) = o_p(|\xi_n|)$。*

引理证明 $\forall\varepsilon>0$, $\forall\delta>0$，由 $h'(0)=0$ 可知存在 $\delta_1>0$ 使得对任意 $0<|x|\le\delta_1$ 有

$$\left|\frac{h(x)}{x}\right| < \delta$$

于是

$$\begin{aligned}& P\left(\left|\frac{h(\xi_n)}{\xi_n}\right|>\delta\right)\\ =& P\left(\left|\frac{h(\xi_n)}{\xi_n}\right|>\delta,\ |\xi_n|\le\delta_1\right)+P\left(\left|\frac{h(\xi_n)}{\xi_n}\right|>\delta,\ |\xi_n|>\delta_1\right)\\ =& P\left(\left|\frac{h(\xi_n)}{\xi_n}\right|>\delta,\ |\xi_n|>\delta_1\right)\\ \le& P(|\xi_n|>\delta_1)\to 0 \quad (n\to\infty)\end{aligned}$$

引理 18.3.

$$\lim_{n\to\infty}\left(1+\frac{b}{n}+o(\frac{1}{n})\right)^{cn} = e^{bc}.$$

18.10 随机向量的极限

定义 18.1. 设 $\{\xi_n\}$ 为随机向量序列，ξ 为随机向量，如果对任意 $\delta>0$ 都有

$$\lim_{n\to\infty}P(|\xi_n-\xi|>\delta) = 0,$$

则称 ξ_n 依概率收敛到 ξ，记作 $\xi_n \xrightarrow{\mathrm{P}} \xi$。

定义中 $|\xi_n - \xi|$ 两边的竖线代表欧式长度。

定理 18.39. 设 $\xi_n = (\xi_{n1}, \dots, \xi_{nm})^T$, $\xi = (\xi_1, \dots, \xi_m)^T$, 则 $\xi_n \xrightarrow{P} \xi$ 当且仅当 $\xi_{nj} \xrightarrow{P} \xi_j$, $j = 1, \dots, m$。

定义 18.2. 设 $\{\xi_n\}$ 为随机向量序列，ξ_n 分布函数为 $F_n(x)$, ξ 为随机向量，有分布函数 $F(x)$, 如果对 $F(\cdot)$ 的任意连续点 x 均有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

则称 ξ_n 依分布收敛到 ξ 或依分布收敛到 $F(\cdot)$，记作 $\xi_n \xrightarrow{\mathrm{d}} \xi$ 或 $\xi_n \xrightarrow{\mathrm{d}} F(\cdot)$。

定理 18.40. 设 $\{\xi_n\}$ 为随机向量序列，ξ 为随机向量，$\xi_n \xrightarrow{d} \xi$, 函数 $g(x)$ 是定义于 ξ 的支撑集上的连续函数，则 $g(\xi_n)$ 依分布收敛于 $g(\xi)$。

推论：随机向量依分布收敛，则相应分量依分布收敛。

定理 18.41. 设 ξ_n 有矩母函数 $M_n(t) = Ee^{t^T\xi_n}$, ξ 有矩母函数 $M(t) = Ee^{t^T\xi}$, 若 $\lim_{n\to\infty} M_n(t) = M(t)$, $\|t\| \leq h(h > 0)$, 则 $\xi_n \xrightarrow{d} \xi$。

定理 18.42 (随机向量的中心极限定理)**.** 设独立同分布随机向量序列 $\{\xi_n\}$ 具有共同的期望 μ 和协方差阵 Σ, Σ 正定，设共同的矩母函数 $M(t)$ 在 0 的一个开邻域存在，令

$$\eta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - \mu) = \sqrt{n}(\bar{\xi} - \mu),$$

则 η_n 依分布收敛到 $N_m(0, \Sigma)$ 分布。

定理 18.43. 设 m 维随机向量序列 $\{\xi_n\}$ 渐近 $N_m(\mu, \Sigma)$ 分布，A, b 为非随机的矩阵和向量，则 $A\xi_n + b$ 渐近 $N_m(A\mu + b, A\Sigma A^T)$ 分布。

定理 18.44. 设 m 维随机向量序列 $\{\xi_n\}$ 满足

$$\sqrt{n}(\xi_n - \mu_0) \xrightarrow{d} N_m(0, \Sigma),$$

设 $g(x)$ 为一个 $\mathbb{R}^m$ 到 $\mathbb{R}^k$ 的变换 $(k \le m)$, 把各个一阶偏导数组成一个矩阵

$$B = \left(\frac{\partial g_i(x)}{\partial x_j} \right)_{\substack{i=1,\dots,k; \\ j=1,\dots,m}},$$

设在 μ_0 的某个邻域内, B 的各个元素连续且 B 不等于零矩阵, 记 B 在 $x = \mu_0$ 处的值为 B_0, 则

$$\sqrt{n}(g(\xi_n) - g(\mu_0)) \xrightarrow{d} N_m(0, B_0 \Sigma B_0^T).$$

Chapter 19

AR 模型的参数估计

对时间序列观测数据建立模型可以更深入了解数据的内在规律，也有利于预报。估计模型的一般指导原则是其他条件相同时较简单的模型更有利。建立模型，要兼顾对观测到的数据的拟合情况和模型的可推广性。下面讨论相合性时，由第4章定理4.2可知，ARMA 模型中白噪声项如果是独立同分布的则该平稳列为严平稳遍历。

19.1 依概率有界

定义 19.1 (依概率有界). 设 $\{\xi_n\}$ 是时间序列，$\{c_n\}$ 是非零常数列，如果对任意 $\varepsilon > 0$，存在正数 M，使得

$$\sup_n P(|\xi_n| > M) \leq \varepsilon$$

就称时间序列 $\{\xi_n\}$ 是**依概率有界的**，记做 $\xi_n = O_p(1)$。如果 $\{\xi_n / c_n\} = O_p(1)$，就称 $\xi_n = O_p(c_n)$。

依概率有界就是除去一个概率任意小的集合后序列有界。若 $\{\xi_n / c_n\} \stackrel{\text{Pr}}{\to} 0$ 则称 $\xi_n = o_p(c_n)$。

依概率有界的等价定义为：称 $\{\xi_n\}$ 依概率有界，若 $\forall \varepsilon > 0, \exists M > 0$ 和 N 使得当 $n \geq N$ 时

$$P(|\xi_n| \leq M) \geq 1 - \varepsilon.$$

事实上，当原定义条件成立时显然此等价定义的条件也成立。

若此等价定义条件成立，则对 $n \geq N$ 有

$$P(|\xi_n| \leq M) \geq 1 - \varepsilon, \quad P(|\xi_n| > M) < \varepsilon.$$

对 $j = 1, 2, \dots, N$，存在 $M_j > 0$ 使得

$$P(|\xi_j| > M_j) < \varepsilon$$

令 $M' = \max(M, M_1, M_2, \dots, M_N)$，则

$$\begin{aligned} P(|\xi_n| > M') &< \varepsilon, \ \forall n \in \mathbb{N}_+, \\ \sup_{n \in \mathbb{N}_+} P(|\xi_n| > M') &\leq \varepsilon, \end{aligned}$$

满足原定义。

引理 19.1. 若 $\xi_n = o_p(c_n)$ 则 $\xi_n = O_p(c_n)$。

事实上，按依概率收敛定义，$\forall \delta > 0, \forall \varepsilon > 0$，存在 N 使 $n > N$ 时

$$\mathrm{P}(|\xi_n / c_n| > \delta) < \varepsilon$$

取 $M \geq \delta$ 使得

$$\mathrm{P}(|\xi_n / c_n| > M) \leq \varepsilon, \quad n = 1, 2, \dots, N$$

于是

$$\mathrm{P}(|\xi_n / c_n| > M) \leq \varepsilon, \quad n = 1, 2, \dots$$

即 $\xi_n / c_n = O_p(1)$.

引理 19.2. 如果 $\xi_n = O_p(1)$ 而 $c_n \to \infty$ 则 $\xi_n / c_n = o_p(1)$。

证明略。

对随机变量 ξ，$\xi = O_p(1)$.

引理 19.3. 若 $\{\xi_n\}$ 同分布，则 $\xi_n = O_p(1)$。

事实上，$\forall \varepsilon > 0, \exists M > 0$ 使 $\mathrm{P}(|\xi_1| > M) < \varepsilon$。由同分布性知 $\mathrm{P}(|\xi_t| > M) = \mathrm{P}(|\xi_1| > M) < \varepsilon$。

引理 19.4.

$$\begin{aligned} O_p(1) \pm O_p(1) =& O_p(1) \\ O_p(1) \cdot O_p(1) =& O_p(1) \\ O_p(1) \pm o_p(1) =& O_p(1) \\ O_p(1) \cdot o_p(1) =& o_p(1). \end{aligned}$$

引理 19.5. *若 ξ_n 依概率收敛到 ξ 则 $\{\xi_n\} = O_p(1)$。*

事实上 $\xi_n = \xi + (\xi_n - \xi) = O_p(1) + o_p(1)$。另外，若 ξ_n 依分布收敛到 ξ 则 $\{\xi_n\} = O_p(1)$。

引理 19.6. *若存在非负随机变量 ξ 使得 $|\xi_n| \leq \xi$ $a.s.$ 则 $\xi_n = O_p(1)$。*

由 $P(|\xi_n| \leq \xi) = 1$，$\forall \varepsilon > 0, \exists M > 0$ 使 $P(\xi > M) < \varepsilon$，于是 $P(|\xi_n| > M) \leq P(\xi > M) < \varepsilon, \forall n \in \mathbb{N}$。

更多的讨论见第18章。

19.2 Yule-Walker 估计

对 AR(p) 模型:

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t \tag{19.1}$$

需要估计系数 $a_1, a_2, \ldots, a_p$ 和白噪声方差 σ^2。先假定 p 已知。

如果 AR(p) 序列的理论自协方差函数 $\{\gamma_k\}$ 已知则由 Yule-Walker 方程可以解出 $a_1, \ldots, a_p$ 和 σ^2, 见 §9.2定理9.2.

设 $x_1, x_2, \ldots, x_N$ 为观测数据，建模前一般先中心化:

$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \ldots, N$$

中心化以后，可以据此估计样本自协方差函数：

$$\hat{\gamma}_k = \frac{1}{N} \sum_{j=1}^{N-k} y_j y_{j+k}, k = 0, 1, \ldots, N-1 \tag{19.2}$$

在 Y-W 方程中用 $\hat{\gamma}_k$ 代替 γ_k 可以解出 $\hat{a}_1, \dots, \hat{a}_p, \hat{\sigma}^2$:

$$\hat{\Gamma}_p \hat{a} = \hat{\gamma}_p \tag{19.3}$$

$$\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{a}^T \hat{\gamma}_p \tag{19.4}$$

由 §16.1的讨论，如果 $x_1, \dots, x_N$ 不全相同则 $\hat{\Gamma}_{p+1}$ 正定，Y-W 方程的解存在唯一。由 §10.2定理10.1可知这样解出的 $\hat{a}_1, \dots, \hat{a}_p$ 满足最小相位条件。

为了避免计算 $\hat{\Gamma}_p$ 的逆可以用 Levinson 递推公式解 Y-W 方程得到 $\hat{a}_1, \dots, \hat{a}_p$ 和 $\hat{\sigma}^2$。这样得到的 AR(p) 参数估计叫做 Y-W 估计或 Levinson 估计。优点是保证最小相位性，计算简单。

由 §9.4定理9.3知 AR(p) 的 Γ_p 正定，根据线性代数中求解线性方程组的 Cramer 法则可知 $a_1, \dots, a_p, \sigma^2$ 是 $\gamma_0, \gamma_1, \dots, \gamma_p$ 的连续函数。如果 $\hat{\gamma}_k$ 是 γ_k 的强相合估计，则 $\hat{a}_1, \dots, \hat{a}_p, \hat{\sigma}^2$ 也是相应参数的强相合估计。要使 $\hat{\gamma}_k$ 强相合，由 §16.2定理16.1, 一个充分条件是白噪声 $\{\varepsilon_t\}$ 独立同分布。这时 ARMA 序列严平稳遍历。

定理 19.1. *如果 AR(p) 模型*(19.1)*中的 $\{\varepsilon_t\}$ 是独立同分布的 WN(0,σ^2), $E\varepsilon_t^4 < \infty$, 则当 $N \to \infty$ 时*

- *(1) $\hat{\sigma}^2 \to \sigma^2$, $\hat{a}_j \to a_j$, a.s., $1 \le j \le p$.*
- *(2) $\sqrt{N}(\hat{a}_1 - a_1, \dots, \hat{a}_p - a_p)^T$ 依分布收敛到 p 维正态分布 $N(0, \sigma^2 \Gamma_p^{-1})$。*
- *(3) $\sqrt{N} \sup\limits_{1 \le j \le p} |\hat{a}_j - a_j| = O(\sqrt{\ln \ln N})$, a.s., $\sqrt{N}|\hat{\sigma}^2 - \sigma^2| = O(\sqrt{\ln \ln N})$, a.s.*

证明: (1) 由 §16.2定理16.1知 $\hat{\gamma}_k$ 强相合，再根据 a_k, σ^2 为 $\{\gamma_k\}$ 的连续函数知 $\hat{a}_k, \hat{\sigma}^2$ 强相合。

(2) 证明略 (详见 (Brockwell & Davis, 1987))。

(3) 证明略。

当定理成立时参数估计有近似正态分布，可以用来作置信区间和假设检验。设 σ_{jj} 为 $\sigma^2 \Gamma_p^{-1}$ 的第 $j \times j$ 元素，则 $\hat{a}_j$ 近似服从 N($a_j, \sigma_{jj}/N$)，于是 a_j 的近似 95% 置信区间为

$$[\hat{a}_j - 1.96\sqrt{\sigma_{jj}}/\sqrt{N},\ \hat{a}_j + 1.96\sqrt{\sigma_{jj}}/\sqrt{N}]$$

其中 σ_{jj} 可以用 $\hat{\sigma}^2\hat{\Gamma}_p^{-1}$ 的元素 $\hat{\sigma}_{jj}$ 代替。

19.3 最小二乘估计

19.3.1 估计方程

最小二乘估计计算简单，不需要计算自协方差。最小二乘估计使残差平方和最小:

$$\min S(a_1,\dots,a_p) \triangleq \sum_{t=p+1}^{N}[y_t-(a_1y_{t-1}+\cdots+a_py_{t-p})]^2$$

最小值点 $\hat{a}_1,\dots,\hat{a}_p$ 称为参数的最小二乘估计。

用线性模型理论，写出最小二乘对应的正规方程为

$$Y=\begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \cdots \\ y_N \end{pmatrix}, \tag{19.5}$$

$$X=\begin{pmatrix} y_p & y_{p-1} & \cdots & y_1 \\ y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & & \vdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-p} \end{pmatrix} = \left(y_{p+t-j}\right)_{\substack{t=1,\dots,N-p;\\ j=1,\dots,p}} \tag{19.6}$$

$$a=\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad (X'X)a=X'Y \tag{19.7}$$

最小二乘估计 $\tilde{a}_1,\dots,\tilde{a}_p$ 是正规方程(19.7)的解。当 $p\times p$ 矩阵 X^TX 正定时方程有唯一解

$$\tilde{a}=(X^TX)^{-1}X^TY$$

19.3.2 LS 估计与 YW 估计的比较

样本自协方差函数估计为

$$\hat{\gamma}_k=\sum_{j=1}^{N-k}y_jy_{j+k}=\sum_{j=1}^{N-k}(x_j-\bar{x}_N)(x_{j+k}-\bar{x}_N)$$

注意对于 $1 \le k \le j \le p$,
矩阵 $(1/N)X^T X$ 的第 (j,k) 元素是

$$
\begin{aligned}
\tilde{\gamma}_{jk} =& \frac{1}{N}\sum_{t=1}^{N-p} y_{t+p-j}y_{t+p-k} = \frac{1}{N}\sum_{t=p-j+1}^{N-j} y_t y_{t+j-k} \\
=& \frac{1}{N}\sum_{t=1}^{N-(j-k)} y_t y_{t+j-k} - \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j-k} - \frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} \\
=& \hat{\gamma}_{j-k} - \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j-k} - \frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k}.
\end{aligned}
\tag{19.8}
$$

完全类似地得到 $(1/N)X^T Y$ 的第 j 个元素

$$
\begin{aligned}
\tilde{\gamma}_{j} =& \frac{1}{N}\sum_{t=1}^{N-p} y_{t+p-j}y_{t+p} = \frac{1}{N}\sum_{t=p-j+1}^{N-j} y_t y_{t+j} \\
=& \frac{1}{N}\sum_{t=1}^{N-j} y_t y_{t+j} - \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j} \\
=& \hat{\gamma}_{j} - \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j}.
\end{aligned}
\tag{19.9}
$$

由 $\{\varepsilon_t\}$ 独立同分布知 AR(p) 序列 $\{X_t\}$ 严平稳遍历，所以 $\tilde{\gamma}_{jk}$ 的分解式(19.8)中第二项

$$
\begin{aligned}
& \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j-k} \\
=& \frac{1}{N}\sum_{t=1}^{p-j} (x_t - \bar{x}_N)(x_{t+j-k} - \bar{x}_N) \\
=& \frac{1}{N}\sum_{t=1}^{p-j} x_t x_{t+j-k} - \bar{x}_N \frac{1}{N}\sum_{t=1}^{p-j} x_t - \bar{x}_N \frac{1}{N}\sum_{t=1}^{p-j} x_{t+j-k} + \bar{x}_N^2 \frac{p-j}{N}
\end{aligned}
\tag{19.10}
$$

因为 $\sum_{t=1}^{p-j} x_t x_{t+j-k}$ 是一个不随 N 而变的随机变量所以第一项 a.s. 趋于 0，同理 $\frac{1}{N}\sum_{t=1}^{p-j} x_t$ 和 $\frac{1}{N}\sum_{t=1}^{p-j} x_{t+j-k}$ 也 a.s. 趋于 0，由严平稳遍历性知 $\bar{x}_N$ a.s. 趋于 $\mu = EX_t$，所以 $\tilde{\gamma}_{jk}$ 的分解中的第二项 a.s. 趋于 0。

类似可知 $\tilde{\gamma}_j$ 的分解(19.9)中第二项 $\frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j}$a.s. 趋于 0。

考虑 $\tilde{\gamma}_{jk}$ 的分解(19.8)中的第三项。记 $Z_t = X_t - \mu$，则 $\{Z_t\}$ 为零均值严平稳遍历序列。作分解

$$
\begin{aligned}
&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} \\
=&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} (z_t - \bar{z}_N)(z_{t+j-k} - \bar{z}_N) \\
=&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} z_t z_{t+j-k} - \bar{z}_N \frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} z_t - \bar{z}_N \frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} z_{t+j-k} + \bar{z}_N^2 \frac{k}{N}
\end{aligned}
\tag{19.11}
$$

由 X_t 严平稳遍历可知 $\xi_t = Z_t Z_{t+j-k}$ 也是严平稳遍历序列，且 $E\xi_t = \gamma_{j-k}$。有

$$
\begin{aligned}
&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} z_t z_{t+j-k} \\
=&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} \xi_t \\
=&\frac{1}{N}\sum_{t=1}^{N-(j-k)} \xi_t - \frac{1}{N}\sum_{1}^{t=N-j} \xi_t \\
=&\frac{N-(j-k)}{N}\frac{1}{N-(j-k)}\sum_{t=1}^{N-(j-k)} \xi_t - \frac{N-j}{N}\frac{1}{N-j}\sum_{1}^{t=N-j} \xi_t \\
\to& 1 \cdot E\xi_t - 1 \cdot E\xi_t = 0, \quad \text{a.s. } (N \to \infty)
\end{aligned}
$$

类似有

$$
\begin{aligned}
&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} z_t \to 0 \quad \text{a.s.} \\
&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} z_{t+j-k} \to 0 \quad \text{a.s.} \\
&\bar{z}_n \to EZ_t = 0 \quad \text{a.s.}
\end{aligned}
$$

所以 $\tilde{\gamma}_{jk}$ 的分解(19.8)中的第三项 a.s. 趋于 0。

于是, 对 $1 \le k \le j \le p$ 有

$$
\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} \to 0,\ \tilde{\gamma}_j - \hat{\gamma}_j \to 0,\ \text{a.s.},
$$

由 $\hat{\gamma}_j$ 的强相合性知 $\tilde{\gamma}_{jk} \to \gamma_{j-k}$, $\tilde{\gamma}_j \to \gamma_j$, a.s.

这样, 当 $N \to \infty$

$$\begin{aligned}(X^TX)^{-1}X^TY =& \left(\frac{1}{N}X^TX\right)^{-1}\left(\frac{1}{N}X^TY\right) \\ \to& \Gamma_p^{-1}\gamma_p = a, \quad \text{a.s.}\end{aligned}$$

即最小二乘估计也是强相合的。

来计算 $\tilde{\gamma}_{jk} - \hat{\gamma}_{k-j}$ 和 $\tilde{\gamma}_j - \hat{\gamma}_j$ 的依概率上界。

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{k-j} = -\frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j-k} - \frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} \tag{19.12}$$

$$\tilde{\gamma}_j - \hat{\gamma}_j = \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j} \tag{19.13}$$

考虑

$$\begin{aligned}E\left|\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k}\right| \le& \sum_{t=N-j+1}^{N-(j-k)} E|y_t y_{t+j-k}| \\ \le& \sum_{t=N-j+1}^{N-(j-k)} \sqrt{Ey_t^2 \; Ey_{t+j-k}^2}\end{aligned}$$

对任意 t,

$$\begin{aligned}Ey_t^2 =& E(Z_t - \bar{Z}_N)^2 \\ \le& E(2Z_t^2 + 2\bar{Z}_N^2) \\ =& 2\gamma_0 + 2E\bar{Z}_N^2\end{aligned}$$

而

$$\begin{aligned}E\bar{Z}_N^2 =& E(\frac{1}{N}\sum_{t=1}^N Z_t)^2 \\ =& \frac{1}{N^2}E\left(\sum_{t=1}^N\sum_{s=1}^N Z_t Z_s\right) \\ =& \frac{1}{N^2}\sum_{t=1}^N\sum_{s=1}^N \gamma_{s-t} \\ \le& \gamma_0\end{aligned}$$

所以

$$Ey_t^2 \le 4\gamma_0,\ \forall t$$

所以

$$E\left|\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k}\right| \le 4\gamma_0 k,\ \forall N$$

引理：对随机变量序列 $\{\xi_n\}$，如果 $E|\xi_n|$ 有界，则 $\xi_n = O_p(1)$。可以用马尔可夫不等式证明。

于是(19.12)中的 $\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} = O_p(1)$，其它几项类似，所以有

$$\tilde{\gamma}_{kj} - \hat{\gamma}_{k-j} = O_p(\frac{1}{N}), \quad \tilde{\gamma}_j - \hat{\gamma}_j = O_p(\frac{1}{N}).$$

用 $\hat{a} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p)^T$ 和 $\tilde{a} = (\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_p)^T$ 分别表示 Yule-Walker 估计和最小二乘估计. 记 $\hat{\gamma}_p = (\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_p)^T$, 就有

$$\begin{aligned}
\tilde{a} - \hat{a} =& (X^T X)^{-1} X^T Y - \hat{\Gamma}_p^{-1} \hat{\gamma}_p && (19.14)\\
=& (\frac{1}{N} X^T X)^{-1} \Big(\frac{1}{N} X^T Y - \hat{\gamma}_p\Big) + \Big[(\frac{1}{N} X^T X)^{-1} - \hat{\Gamma}_p^{-1}\Big] \hat{\gamma}_p && (19.15)\\
=& (\frac{1}{N} X^T X)^{-1} O_p(1/N) + (\frac{1}{N} X^T X)^{-1} \Big[\hat{\Gamma}_p - \frac{1}{N} X^T X\Big] \hat{\Gamma}_p^{-1} \hat{\gamma}_p && (19.16)\\
=& O_p(1/N), \quad \text{当 } N \to \infty. && (19.17)
\end{aligned}$$

上式说明对 $1 \le j \le p$, $\hat{a}_j - \tilde{a}_j = O_p(1/N)$. 由于 $1/N$ 收敛到零的速度是很快的, 所以也可以说对较大的 N, 最小二乘估计和 Yule-Walker 估计的差别不大.

定理 19.2. *设 AR(p) 模型(19.1)中的白噪声 $\{\varepsilon_t\}$ 是独立同分布的, $E\varepsilon_t^4 < \infty$, $(\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_p)$ 是自回归系数 $(a_1, a_2, \dots, a_p)$ 的最小二乘估计. 则当 $N \to \infty$ 时*

$$\sqrt{N}(\tilde{a}_1 - a_1, \tilde{a}_2 - a_2, \dots, \tilde{a}_p - a_p)$$

依分布收敛到 p-维正态分布 $N(0, \sigma^2 \Gamma_p^{-1})$.

即最小二乘估计和 Y-W 估计有相同的渐近分布。定理19.2的证明由定理19.1和(19.17)直接得到.

例 19.1. Y-W 估计和 LS 估计的模拟比较。

取 $\{\varepsilon_t\}$ 为标准正态白噪声，模拟产生 AR(4) 序列

$$
\begin{aligned}
X_t =&1.16X_{t-1} - 0.37X_{t-2} - 0.11X_{t-3} \\
&+ 0.18X_{t-4} + \varepsilon_t, t = 1, 2, \dots, N
\end{aligned}
$$

分别取 $N = 25, 100, 400, 1600$，重复 $M = 10000$ 次产生模拟数据并分别用 Y-W 方法和最小二乘方法估计参数。

结果发现在样本量较小时，系数估计 Y-W 方法较好，预测方差 σ^2 估计最小二乘方法较好。样本量大时二者没有明显差别。

```
demo.ar.estb <- function(n=50, m=400){
  a <- c(1.16, -0.37, -0.11, 0.18)
  sig <- 1.0
  ests.yw <- matrix(0, nrow=m, ncol=5)
  ests.ls <- matrix(0, nrow=m, ncol=5)
  for(ii in seq(m)){
    x <- arima.sim(
      model=list(ar=a), n=n,
      rand.gen=function(n, ...) rnorm(n, 0, sig))
    ## do Levinson estimation and LS estimation
    fit.yw <- ar(x, aic=FALSE, order.max=4,
                 method="yule-walker")
    ests.yw[ii,] <- c(fit.yw$ar, fit.yw$var.pred)
    fit.ls <- ar(x, aic=F, order.max=4,
                 method="ols")
    ests.ls[ii,] <- c(fit.ls$ar, fit.ls$var.pred)
  }
  res <- matrix(0, nrow=7, ncol=5)
  rownames(res) <- c(" 真实参数", "Y-W 估计均值", "LS 估计均值",
                     "Y-W 估计标准差", "LS 估计标准差",
                     "Y-W 估计根均方误差", "LS 估计根均方误差"
                     )
  colnames(res) <- c("a1", "a2", "a3", "a4", "s2")
  res[1,] <- c(a, sig^2)
  res[2,] <- apply(ests.yw, 2, mean)
```

```r
  res[3,] <- apply(ests.ls, 2, mean)
  res[4,] <- apply(ests.yw, 2, sd)
  res[5,] <- apply(ests.ls, 2, sd)
  res[6,] <- sqrt(1/m*( (m-1)*res[4,]^2 + m*(res[2,]-res[1,])^2))
  res[7,] <- sqrt(1/m*( (m-1)*res[5,]^2 + m*(res[3,]-res[1,])^2))
  cat("\n\n==== AR 模型模拟: 样本量 =", n, " 重复", m, " 次估计比较\n")
  print(round(res, 4))

  invisible(res)
}

demo.ar.est <- function(n=c(25,50,100,400,1600), m=400){
  for(ni in n){
    demo.ar.estb(ni, m)
  }
}
set.seed(1)
demo.ar.est(m=10000)
```

```
##
##
## ==== AR模型模拟: 样本量= 25  重复 10000 次估计比较
##                     a1      a2      a3     a4     s2
## 真实参数          1.1600 -0.3700 -0.1100 0.1800 1.0000
## Y-W估计均值       0.9216 -0.3005 -0.0567 0.0196 1.3005
## LS估计均值        1.0049 -0.3760 -0.0610 0.0519 0.7489
## Y-W估计标准差     0.2094  0.2465  0.2254 0.1585 0.5076
## LS估计标准差      0.2540  0.3307  0.3170 0.2269 0.2649
## Y-W估计根均方误差 0.3173  0.2561  0.2316 0.2255 0.5899
## LS估计根均方误差  0.2976  0.3307  0.3208 0.2605 0.3650
##
##
## ==== AR模型模拟: 样本量= 50  重复 10000 次估计比较
```

```
##                      a1      a2      a3     a4     s2
## 真实参数           1.1600 -0.3700 -0.1100 0.1800 1.0000
## Y-W估计均值        1.0366 -0.3202 -0.0718 0.0906 1.1649
## LS估计均值         1.0951 -0.3718 -0.0783 0.1140 0.8915
## Y-W估计标准差      0.1454  0.1901  0.1833 0.1251 0.2911
## LS估计标准差       0.1575  0.2214  0.2201 0.1466 0.1958
## Y-W估计根均方误差  0.1907  0.1965  0.1872 0.1538 0.3345
## LS估计根均方误差   0.1704  0.2214  0.2223 0.1607 0.2238
##
##
## ==== AR模型模拟: 样本量= 100  重复 10000 次估计比较
##                      a1      a2      a3     a4     s2
## 真实参数           1.1600 -0.3700 -0.1100 0.1800 1.0000
## Y-W估计均值        1.1019 -0.3464 -0.0872 0.1336 1.0824
## LS估计均值         1.1354 -0.3765 -0.0922 0.1474 0.9468
## Y-W估计标准差      0.1006  0.1443  0.1393 0.0935 0.1760
## LS估计标准差       0.1038  0.1566  0.1544 0.1007 0.1402
## Y-W估计根均方误差  0.1162  0.1462  0.1412 0.1043 0.1943
## LS估计根均方误差   0.1067  0.1568  0.1554 0.1059 0.1499
##
##
## ==== AR模型模拟: 样本量= 400  重复 10000 次估计比较
##                      a1      a2      a3     a4     s2
## 真实参数           1.1600 -0.3700 -0.1100 0.1800 1.0000
## Y-W估计均值        1.1448 -0.3626 -0.1039 0.1682 1.0215
## LS估计均值         1.1539 -0.3710 -0.1055 0.1720 0.9877
## Y-W估计标准差      0.0496  0.0747  0.0744 0.0490 0.0757
## LS估计标准差       0.0498  0.0763  0.0762 0.0497 0.0706
## Y-W估计根均方误差  0.0518  0.0750  0.0746 0.0503 0.0787
## LS估计根均方误差   0.0502  0.0763  0.0763 0.0503 0.0716
##
##
## ==== AR模型模拟: 样本量= 1600  重复 10000 次估计比较
```

```
##                            a1      a2      a3     a4     s2
## 真实参数             1.1600 -0.3700 -0.1100 0.1800 1.0000
## Y-W估计均值          1.1561 -0.3680 -0.1084 0.1769 1.0051
## LS估计均值           1.1584 -0.3701 -0.1088 0.1779 0.9967
## Y-W估计标准差        0.0246  0.0376  0.0380 0.0247 0.0364
## LS估计标准差         0.0246  0.0378  0.0383 0.0248 0.0357
## Y-W估计根均方误差    0.0249  0.0376  0.0380 0.0249 0.0367
## LS估计根均方误差     0.0246  0.0378  0.0383 0.0249 0.0359
```

19.4 最大似然估计

Y-W 估计是矩估计，最小二乘估计与 Y-W 估计渐近相同。矩估计容易计算但可能精度不高。最大似然估计一般精度较高。

设 AR(p) 模型(19.1)的白噪声 $\{\varepsilon_t\}$ 是正态白噪声，则 $\{X_t\}$ 是正态平稳列，给定观测 $x_1, x_2, \dots, x_N$，似然函数是联合正态分布密度函数。

记参数 $\theta = (a_1, a_2, \dots, a_p, \sigma^2)$，则似然函数为

$$
\begin{aligned}
L(\theta) =& p(x_1, \dots, x_N | \theta) \\
=& p(x_1, \dots, x_p | \theta) \prod_{t=p+1}^{N} p(x_t | x_{t-1}, \dots, x_1, \theta)
\end{aligned}
$$

其中 $p(x_1, \dots, x_p|\theta)$ 是 $(X_1, \dots, X_p)$ 的联合密度，是一个联合正态密度；$p(x_t|x_{t-1}, \dots, x_1, \theta)$ 是 x_t 的条件密度，是一个一元正态条件密度。

一元正态条件密度完全由其条件期望和条件方差确定，而 $t > p$ 时

$$
\begin{aligned}
E(x_t|x_{t-1}, \dots, x_1, \theta) =& E(\varepsilon_t + a_1 x_{t-1} + \dots + a_p x_{t-p} | x_{t-1}, \dots, x_1, \theta) \\
=& 0 + a_1 x_{t-1} + \dots + a_p x_{t-p} \\
\text{Var}(x_t|x_{t-1}, \dots, x_1, \theta) =& E\left([x_t - E(x_t|x_{t-1}, \dots, x_1, \theta)]^2 | x_{t-1}, \dots, x_1, \theta \right) \\
=& E\left(\varepsilon_t^2 | x_{t-1}, \dots, x_1, \theta \right) \\
=& \sigma^2
\end{aligned}
$$

所以 $x_t|x_{t-1}, \dots, x_1, \theta$ 服从 $\text{N}(a_1 x_{t-1} + \dots + a_p x_{t-p}, \sigma^2)$。

$$
p(x_t|x_{t-1}, \dots, x_1, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{1}{2\sigma^2} (x_t - a_1 x_{t-1} - \dots - a_p x_{t-p})^2 \right\},\ t = p+1, \dots, n.
$$

但是，联合密度分解式中的 $p(x_1, \dots, x_p|\theta)$ 部分比较难以表示成 $a_1, \dots, a_p, \sigma^2$ 的简单函数。故此，当 N 较大时，仅考虑似然函数中的主要部分，忽略 $p(x_1, \dots, x_p|\theta)$ 这一部分，得到一个简化的似然函数 (profile likelihood)

$$\begin{aligned}
&\tilde{L}(a_1, \dots, a_p, \sigma^2) \\
=&\prod_{t=p+1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{1}{2\sigma^2}(x_t - a_1 x_{t-1} - \dots - a_p x_{t-p})^2\right\} \\
=&(2\pi\sigma^2)^{-(N-p)/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=p+1}^{N} (x_t - a_1 x_{t-1} - \dots - a_p x_{t-p})^2\right\}
\end{aligned}$$

这实际是给定 $(x_1, \dots, x_p)$ 条件下 $(x_{p+1}, \dots, x_N)$ 的条件密度。

实际中的数据不一定是零均值的，令 $y_t = x_t - \bar{X}_N$，将 $\{y_t\}$ 作为零均值的 AR(p) 序列的观测值。记 $a = (a_1, \dots, a_p)$，

$$S(a) = \sum_{t=p+1}^{N} (x_t - a_1 x_{t-1} - \dots - a_p x_{t-p})^2$$

则 $\tilde{L}$ 对应的对数似然函数为

$$\begin{aligned}
&\ell(a, \sigma^2) \\
=&-\frac{N-p}{2} \ln(2\pi) - \frac{N-p}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} S(a).
\end{aligned}$$

$\ell(a, \sigma^2)$ 关于 σ^2 求偏导并令偏导等于零，得

$$\sigma^2 = \frac{1}{N-p} S(a) \tag{19.18}$$

于是最大值点是 $S(a)$ 的最小值点，最大似然估计等于最小二乘估计。

19.5 AR 模型定阶

对于时间序列观测样本，假设它来自某个 AR(p) 模型，但是 p 一般是未知的。要想办法估计 p。

估计 p 可以用 AR(p) 的偏相关 p 步截尾性质，计算样本偏相关系数 $\{\hat{a}_{kk}\}$，看 $\{\hat{a}_{kk}\}$ 在何处截尾。估计 p 也可以在拟合优度和模型简单之间进行权衡，AIC 和 BIC 准则就是这样的办法。

19.5.1 样本偏相关系数

只要观测样本 $x_1, \ldots, x_N$ 不完全相同则样本自协方差阵 $\hat{\Gamma}_k$ 正定 $(k < N)$。样本偏相关系数 $\{\hat{a}_{k,k}\}$ 可以由样本自协方差唯一决定

$$\begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0 \end{pmatrix} \begin{pmatrix} \hat{a}_{k,1} \\ \hat{a}_{k,2} \\ \vdots \\ \hat{a}_{k,k} \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{pmatrix}$$

可以用 Levinson 递推公式计算样本偏相关系数。

定理 19.3. *如果 AR(p) 模型*(19.1)*中的白噪声时独立同分布的,则对任何* $k \geq p$*,有*

$$\lim_{N\to\infty} \hat{a}_{k,j} = \begin{cases} a_j, & \text{当} j \leq p, \\ 0, & \text{当} j > p. \end{cases} \quad a.s. \tag{19.19}$$

证明:

对 $k > p$ 证明，$k = p$ 情况类似。

由 §16.2的定理16.1知道样本自协方差函数 $\hat{\gamma}_k$ 是 γ_k 的强相合估计. 对任何矩阵 $(c_{j,k}(N))$ 定义极限符号

$$\lim_{N\to\infty} (c_{j,k}(N)) \triangleq (\lim_{N\to\infty} c_{j,k}(N)).$$

对任何 $k>p$, 利用 Γ_k 的正定性得到:

$$
\begin{aligned}
&\lim_{N\to\infty}\begin{pmatrix}\hat a_{k,1}\\ \hat a_{k,2}\\ \vdots\\ \hat a_{k,k}\end{pmatrix}\\
&=\lim_{N\to\infty}\begin{pmatrix}\hat\gamma_0 & \hat\gamma_1 & \cdots & \hat\gamma_{k-1}\\ \hat\gamma_1 & \hat\gamma_0 & \cdots & \hat\gamma_{k-2}\\ .. & .. & \cdots & ..\\ \hat\gamma_{k-1} & \hat\gamma_{k-2} & \cdots & \hat\gamma_0\end{pmatrix}^{-1}\times\lim_{N\to\infty}\begin{pmatrix}\hat\gamma_1\\ \hat\gamma_2\\ \vdots\\ \hat\gamma_k\end{pmatrix}\\
&=\begin{pmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{k-1}\\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2}\\ \vdots & \vdots & & \vdots\\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0\end{pmatrix}^{-1}\times\begin{pmatrix}\gamma_1\\ \gamma_2\\ \vdots\\ \gamma_k\end{pmatrix}\\
&=\begin{pmatrix}a_{k,1}\\ a_{k,2}\\ ..\\ a_{k,k}\end{pmatrix}.
\end{aligned}
$$

从 §9.2的定理9.2知道

$$(a_{k,1},a_{k,2},\dots,a_{k,k})=(a_1,a_2,..,a_p,0,\cdots,0). \tag{19.20}$$

所以(19.19)成立.

19.5.2 偏相关系数的检验

为检验 H_0： $a_{k,k}=0$，需要知道 $\hat a_{k,k}$ 的分布。但我们只能得到 $\hat a_{k,k}$ 的极限分布。

定理 19.4. 设 $(a_{k,1},a_{k,2},\dots,a_{k,k})$ 由(19.20)定义, 如果 *AR(p)* 模型(19.1)中的白噪声是独立同分布的, $E\varepsilon_t^4<\infty$, 则对确定的 $k>p$, 当 $N\to\infty$ 时,

$$\sqrt{N}(\hat a_{k,1}-a_{k,1},\ \hat a_{k,2}-a_{k,2},\ \cdots,\ \hat a_{k,k}-a_{k,k})$$

依分布收敛到 k-维正态分布 $N(0,\sigma^2\Gamma_k^{-1})$.

详见 (Brockwell & Davis, 1987)。

推论 19.1. *在定理19.4的条件下，对 $k > p$, $\sqrt{N}\hat{a}_{k,k}$ 依分布收敛到标准正态分布 $N(0,1)$.*

推论的证明：从定理19.4知道只需对 $k > p$ 证明 $\sigma^2\Gamma_k^{-1}$ 的第 (k,k) 元素是 1. 用 $A_{(j,j)}$ 表示矩阵 A 的第 (j,j) 元素. 利用 Γ_k^{-1} 的伴随矩阵表示得到

$$\Gamma_k^{-1} = \frac{1}{\det(\Gamma_k)} \begin{pmatrix} \det(\Gamma_{k-1}) & & * \\ & \ddots & \\ * & & \det(\Gamma_{k-1}) \end{pmatrix}.$$

于是

$$(\Gamma_k^{-1})_{(k,k)} = (\Gamma_k^{-1})_{(1,1)} = \frac{\det(\Gamma_{k-1})}{\det(\Gamma_k)},\ \forall k \geq 1.$$

对 $k \geq p$, 利用 Yule-Walker 方程得到

$$\Gamma_{k+1} \begin{pmatrix} 1 \\ -a_k \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{从而有} \quad \begin{pmatrix} 1 \\ -a_k \end{pmatrix} = \Gamma_{k+1}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

于是得到

$$\left(\sigma^2\Gamma_{k+1}^{-1}\right)_{(k+1,k+1)} = \left(\sigma^2\Gamma_{k+1}^{-1}\right)_{(1,1)} = 1,\ k \geq p.$$

推论证毕。

用样本偏相关系数可以为 AR 模型定阶。N 较大时若 $k > p$ 则 $\sqrt{N}\hat{a}_{k,k}$ 近似服从 N(0,1)，$\hat{a}_{k,k}$ 约以 95% 概率落入区间

$$(-1.96/\sqrt{N},\ 1.96/\sqrt{N}).$$

对某固定的 K，以

$$\hat{p} = \max\{j : |\hat{a}_{j,j}| > 1.96/\sqrt{N}, 1 \leq j \leq K\}$$

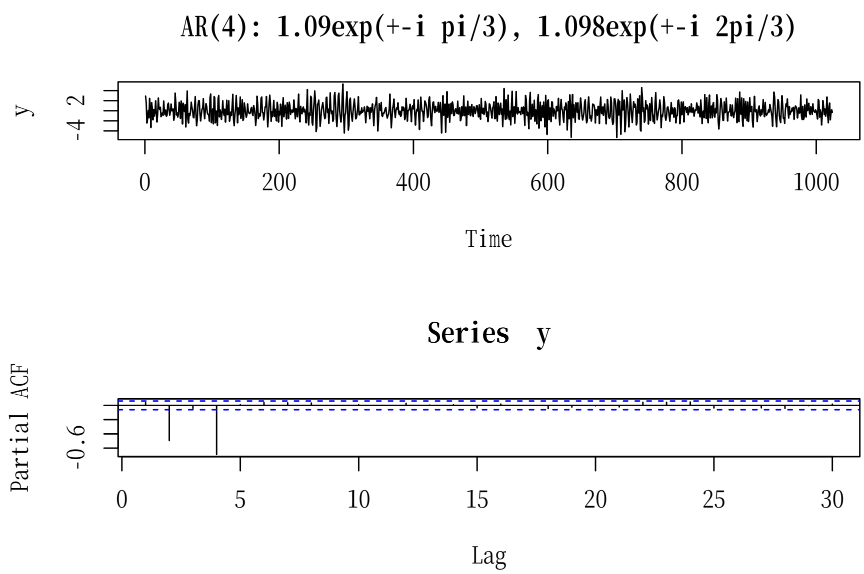
作为 p 的估计是合理的。实际问题中可以取 K 为允许的阶数的一个上界。

这种办法只要画出样本偏相关系数图就可以定阶。(R 中用 pacf(x) 作样本偏相关系数图)。

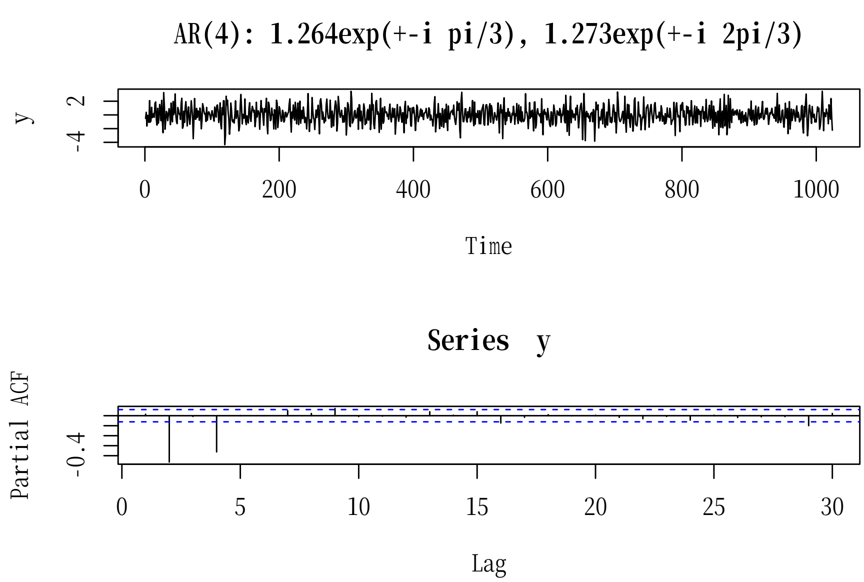
演示：AR order selection by PACF.

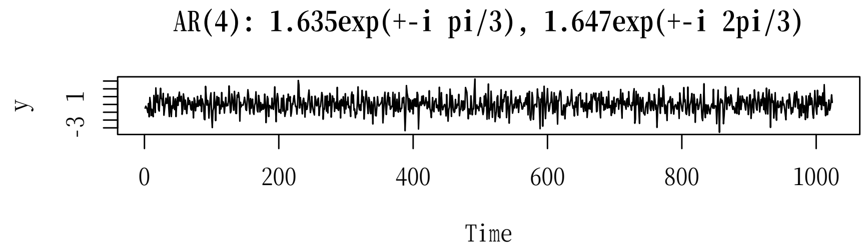
```
demo.pacf <- function(n=1024){
  rtlis <- list(c(complex(mod=1.09, arg=pi/3*c(1,-1)),
                  complex(mod=1.098, arg=pi*2/3*c(1,-1))),
                c(complex(mod=1.264, arg=pi/3*c(1,-1)),
                  complex(mod=1.273, arg=pi*2/3*c(1,-1))),
                c(complex(mod=1.635, arg=pi/3*c(1,-1)),
                  complex(mod=1.647, arg=pi*2/3*c(1,-1))),
                complex(mod=1.02, arg=pi/6*c(1,-1)),
                complex(mod=1.02, arg=pi/2*c(1,-1)),
                c(complex(mod=1.05, arg=pi/6*c(1,-1)),
                  complex(mod=1.05, arg=pi/2*c(1,-1))))
  tits <- c("AR(4): 1.09exp(+-i pi/3), 1.098exp(+-i 2pi/3)",
            "AR(4): 1.264exp(+-i pi/3), 1.273exp(+-i 2pi/3)",
            "AR(4): 1.635exp(+-i pi/3), 1.647exp(+-i 2pi/3)",
            "AR(2): mod=1.02 arg=+-pi/6",
            "AR(2): mod=1.02 arg=+-pi/2",
            "AR(4): mod=1.05 arg=+-pi/6,+-pi/2")
  oldpar <- par(mfrow=c(2,1)); on.exit(par(oldpar))
  for(ii in seq(along=rtlis)){
    rt = rtlis[[ii]]
    y <- ar.gen(n, rt, sigma=1.0, by.roots=TRUE,
                plot.it=FALSE)
    plot(y, main=tits[ii])
    pacf(y)
  }
}
set.seed(1)
demo.pacf(n=1024)
```

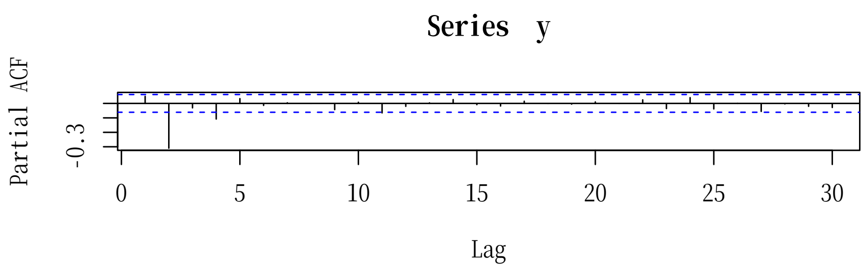
```
## Warning in polynomial(p): 强制改变时丢弃了虚数部分
```

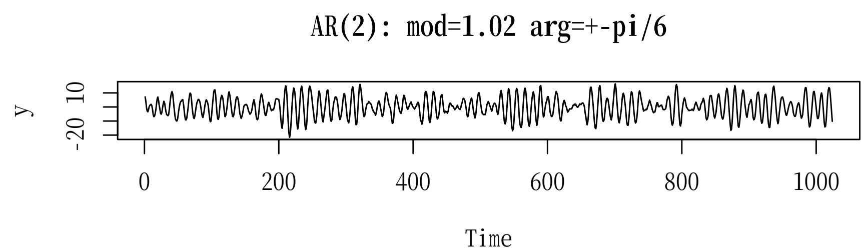


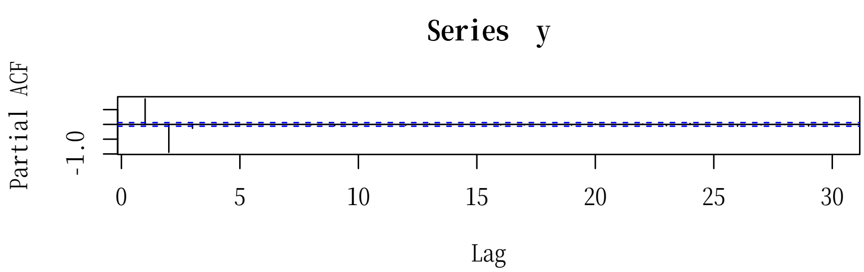

```
## Warning in polynomial(p): 强制改变时丢弃了虚数部分
```



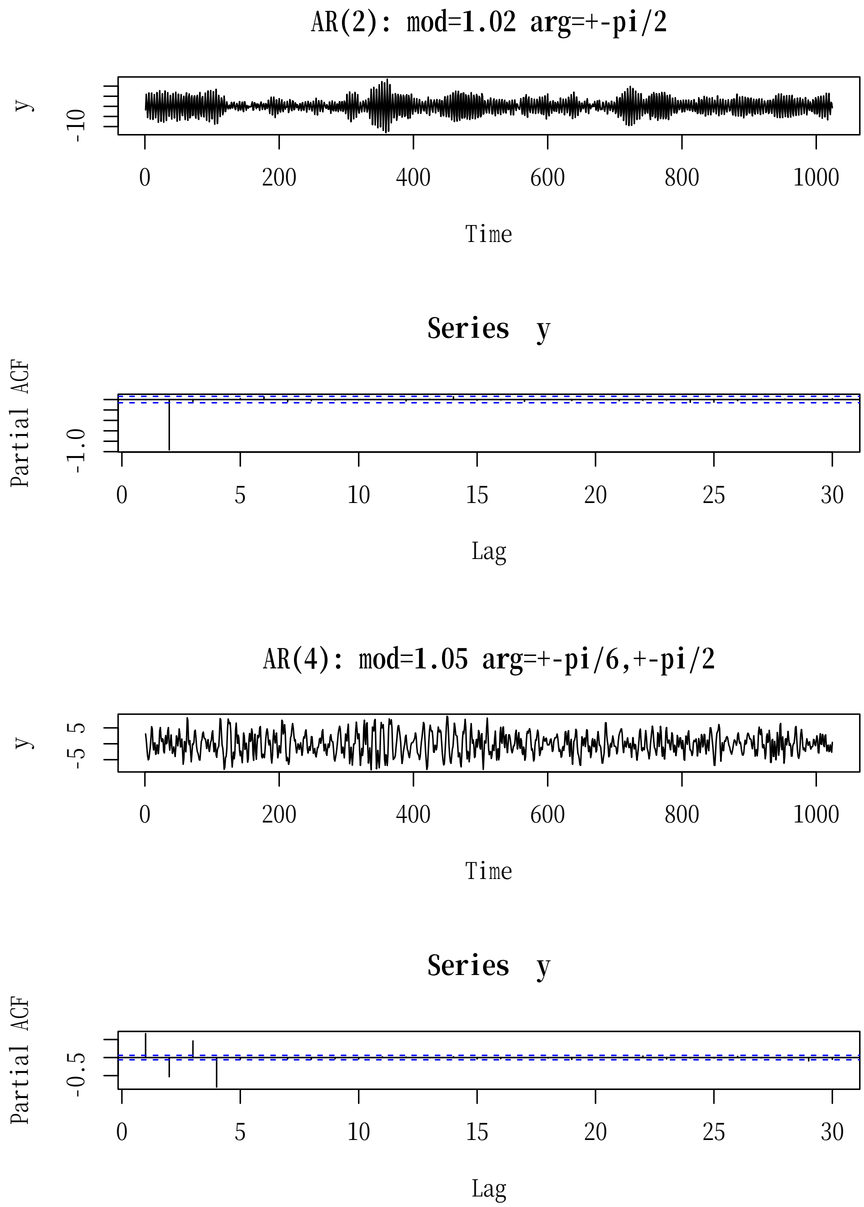









```
## Warning in polynomial(p): 强制改变时丢弃了虚数部分
```


AR(2): mod=1.02 arg=+-pi/2
y
-10
0
200
400
600
800
1000
Time
Series y
Partial ACF
-1.0
0
5
10
15
20
25
30
Lag
AR(4): mod=1.05 arg=+-pi/6,+-pi/2
y
-5 5
0
200
400
600
800
1000
Time
Series y
Partial ACF
-0.5
0
5
10
15
20
25
30
Lag

19.5.3 AIC 和 BIC 定阶

可以通过对拟合优度的要求加上对参数个数的惩罚指定一个准则来定阶。AIC 准则:

$$\text{AIC}(k) = \ln \hat\sigma_k^2 + \frac{2k}{N}, \quad k = 0, 1, \dots, P_0$$

(P_0 是可取的阶的上限)

BIC 定阶:

$$\text{BIC}(k) = \ln \hat\sigma_k^2 + \frac{k \ln N}{N}, \quad k = 0, 1, \dots, P_0$$

取 AIC 或 BIC 的最小值点作为 p 的估计。

AIC 定阶不相合, 倾向于高估；BIC 定阶在 $N \to \infty$ 时是强相合的，但对较小的 N 倾向于低估。

参考：

- HIROTUGU AKAIKE, Fitting Autoregressive Models for Prediction, Annals of the Institute of Statistical Mathematics AC-19 (1974), pp. 364 – 385
- GIDEON SCHWARZ, Estimating the Dimensions of a Model, Annals of Statistics 6(1978), pp. 461 – 464

演示：AR order selection by AIC and BIC. 模拟 AR(4) 模型：

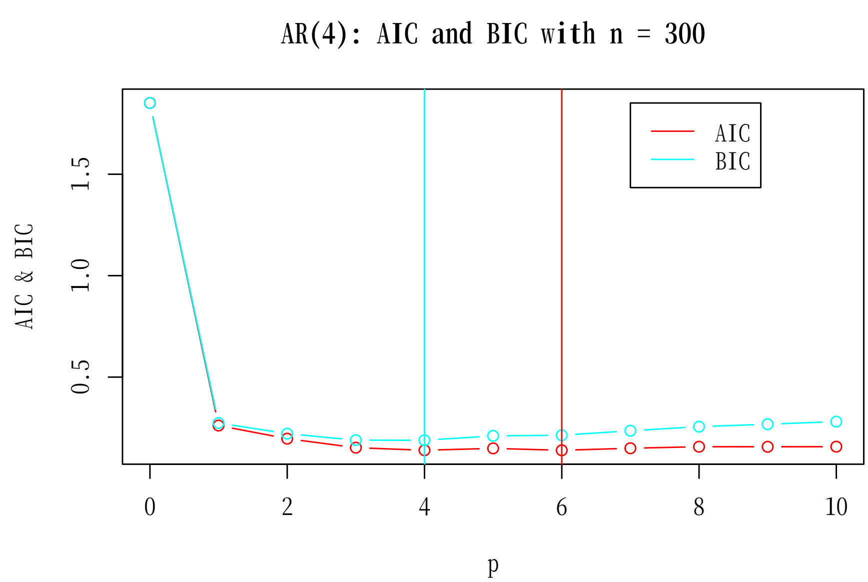
$$X_t = 1.16 X_{t-1} - 0.37 X_{t-2} - 0.11 X_{t-3} + 0.18 X_{t-4} + \varepsilon_t, \ \varepsilon_t \sim \text{N}(0, 1)$$

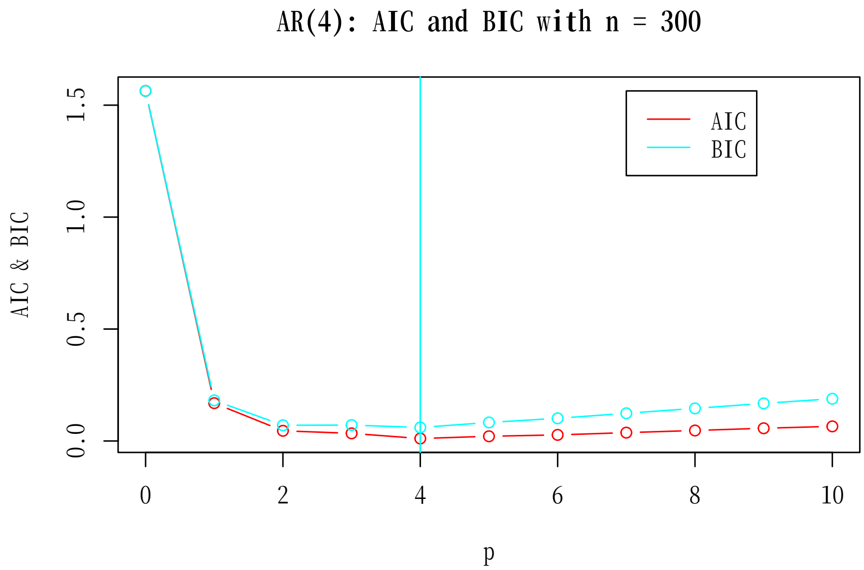
生成 400 组样本，用 AIC 或者 BIC 定阶，对前 5 组做了定阶选择图形。统计了选择的不同阶数的出现频次。

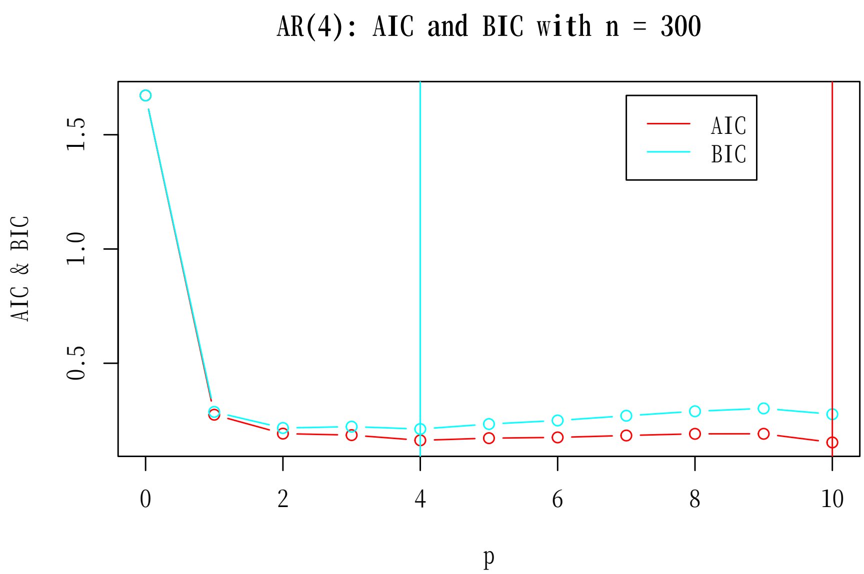
```
ar.bic <- function(x, order.max=10){
  aics <- numeric(order.max+1)
  bics <- numeric(order.max+1)
  aics[1] <- log(var(x))
  bics[1] <- log(var(x))
  n <- length(x)
  for(ii in 1:order.max){
    s2 <- ar(x, aic=F, order.max=ii,
             method="yule-walker")$var.pred
```

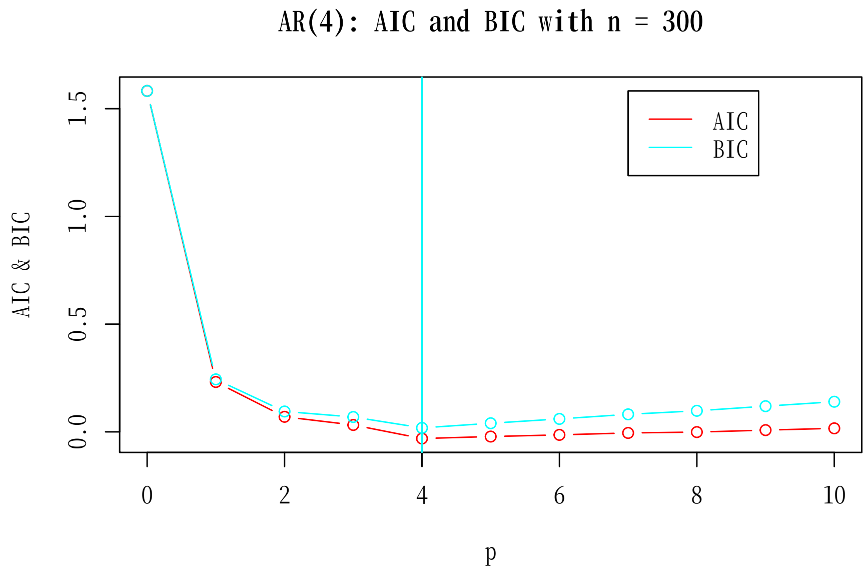
```
    aics[ii+1] <- log(s2) + ii*2/n
    bics[ii+1] <- log(s2) + ii*log(n)/n
  }
  p <- which.min(bics) - 1
  list(p=p, bics=bics, aics=aics)
}
demo.ar.order <- function(n=300, m=100){
  a <- c(1.16, -0.37, -0.11, 0.18)
  sig <- 1.0
  orders.aic <- numeric(m)
  orders.bic <- numeric(m)
  for(ii in seq(m)){
    x <- ar.gen(n, a, sigma=sig)
    ## do Levinson estimation and LS estimation
    fit.yw <- ar(x, aic=T, order.max=10,
                 method="yule-walker")
    p.aic <- fit.yw$order
    fit.bic <- ar.bic(x, order.max=10)
    aics <- fit.bic$aics
    bics <- fit.bic$bics
    p.bic <- fit.bic$p
    orders.aic[ii] <- p.aic
    orders.bic[ii] <- p.bic
    if(ii<=5){
      yr <- range(c(aics, bics))
      plot(0:10, aics, type="b", col="red",
           main=paste("AR(4): AIC and BIC with n =", n),
           xlab="p", ylab="AIC & BIC")
      lines(0:10, bics, type="b", col="cyan")
      abline(v=p.aic, col="red")
      abline(v=p.bic, col="cyan")
      legend(7,yr[2], lty=1,
             col=c("red", "cyan"),
```

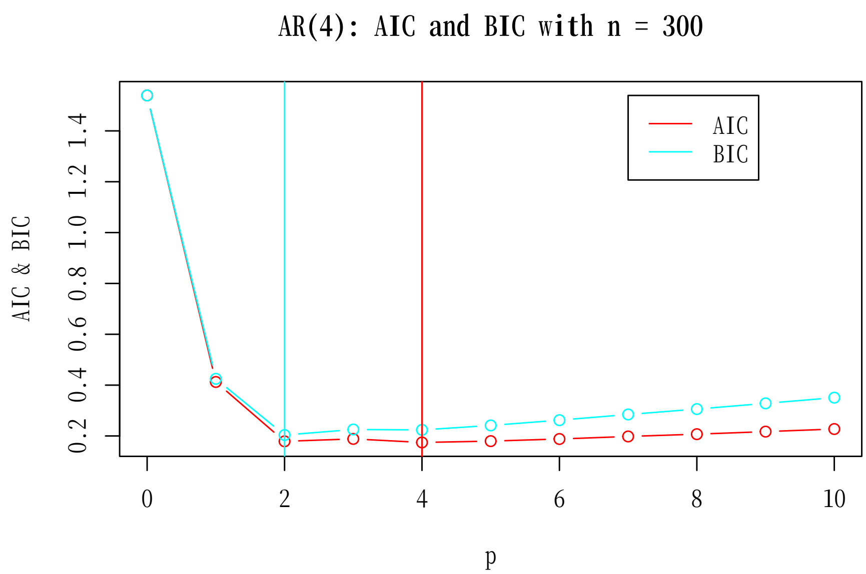
```
              legend=c("AIC", "BIC"))
    }
  }
  cat("\nAR 模型 AIC 和 BIC 定阶, 样本量 =", n, " 重复", m, " 次的结果:\n")
  cat("AIC order:\n")
  print(table(orders.aic))
  cat("BIC order:\n")
  print(table(orders.bic))
}
set.seed(1)
demo.ar.order(m=400)
```




AR(4): AIC and BIC with n = 300
AIC
BIC
AIC & BIC
1.5
1.0
0.5
0.0
0
2
4
6
8
10
p


AR(4): AIC and BIC with n = 300
AIC
BIC
AIC & BIC
1.5
1.0
0.5
0
2
4
6
8
10
p






```
##
## AR模型AIC和BIC定阶, 样本量= 300  重复 400 次的结果:
## AIC order:
```

```
## orders.aic
##   2   3   4   5   6   7   8   9  10
##  19  12 276  42  18  10  10   9   4
## BIC order:
## orders.bic
##   2   3   4   5
## 222  19 158   1
```

19.6 模型拟合检验

对观测数据 $x_1, x_2, \cdots, x_N$, 如果得到了 AR(p) 的阶 p 和自回归系数 $(a_1, a_2, \cdots, a_p)$ 的估计量 $\hat{p}$ 和, $(\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_{\hat{p}})$, 定义残差

$$\hat{\varepsilon}_t = y_t - \sum_{j=1}^{\hat{p}} \hat{a}_j y_{t-j}, \quad t = \hat{p}+1, \hat{p}+2, \dots, N. \tag{19.21}$$

这里的 $y_t = x_t - \bar{x}_N$ 是 x_t 零均值化.

对上述的残差序列进行白噪声的检验. 如果能够判定(19.21)是白噪声, 就认为建立的模型是合理的. 否则可以改动 $\hat{p}$ 的值后重新计算, 或改用 MA(q) 或 ARMA(p,q) 模型.

19.7 AR 谱密度估计

满足 AR(p) 模型(19.1)的 AR(p) 序列有谱密度函数

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 - \sum_{j=1}^{p} a_j e^{ij\lambda} \right|^{-2}. \tag{19.22}$$

把 σ^2, p 和 $(a_1, a_2, \dots, a_p)$ 的估计量 $\hat{\sigma}^2$, $\hat{p}$ 和 $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p)$ 代入(19.22)后, 得到 $f(\lambda)$ 的估计

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2}{2\pi} \left| 1 - \sum_{j=1}^{\hat{p}} \hat{a}_j e^{ij\lambda} \right|^{-2}. \tag{19.23}$$

通常称 $\hat{f}(\lambda)$ 为 AR 谱估计或极大熵谱估计.

对于 AIC 或 BIC 定阶 $\hat{p}$, 如果 $\{\varepsilon_t\}$ 是独立同分布的 WN(0,σ^2), 可以证明 $\hat{f}(\lambda)$ 一致收敛到 $f(\lambda)$.

19.8 附录：问题

问：AIC 和 BIC 是否受到 $\{X_t\}$ 的量纲的影响？

答不受影响。比如令 $Y_t = KX_t$, 则从 $\{Y_t\}$ 估计的新息方差 $\hat{\sigma}^2_{Y,p}$ 是从 $\{X_t\}$ 估计的新息方差 $\hat{\sigma}^2_{X,p}$ 的 K^2 倍，取 ln 后变成了 AIC 一致地增加 $\ln K^2$，最小值点不变。

用随机模拟方法验证：

```
demo.aic.sc <- function(fac=100, shift=20, n=300){
  a <- c(1.16, -0.37, -0.11, 0.18)
  sig <- 1
  x1 <- ar.gen(n, a, sigma=sig)
  x2 <- x1 * fac + shift
  ## do Levinson estimation and LS estimation
  fit1 <- ar(x1, aic=T, order.max=10,
             method="yule-walker")
  print(fit1$aic)
  fit2 <- ar(x2, aic=T, order.max=10,
             method="yule-walker")
  print(fit2$aic)
}
demo.aic.sc()
```

```
##           0           1           2           3           4           5
## 409.4603465  31.1461075   6.4023019   6.4645791   0.0000000   0.7616954
##           6           7           8           9          10
##   2.7610739   4.1566591   4.3684998   6.3193517   8.1556399
##           0           1           2           3           4           5
## 409.4603465  31.1461075   6.4023019   6.4645791   0.0000000   0.7616954
##           6           7           8           9          10
##   2.7610739   4.1566591   4.3684998   6.3193517   8.1556399
```

可见经过平移与伸缩的序列观测值给出的 AIC 值是完全相等的。

Chapter 20

MA 模型的参数估计

20.1 MA(1) 模型的参数估计

考虑 MA(1)

$$X_t = \varepsilon_t + b\varepsilon_{t-1}, \quad t = 1, 2, ...$$

其中 $|b| < 1$。$\gamma_0 = EX_t^2 = \sigma^2(1+b^2)$, $\gamma_1 = \sigma^2 b$. $\rho_1 = \frac{b}{1+b^2}$, 关于 b 的方程为

$$\rho_1 b^2 - b + \rho_1 = 0$$

当 $|\rho_1| \leq 0.5$ 时 b 才有实解:

$$b = \frac{1 - \sqrt{1 - 4\rho_1^2}}{2\rho_1}$$

(另一个解使 $|b| > 1$，抛弃) 用 $\hat{\rho}_1 = \hat{\gamma}_1 / \hat{\gamma}_0$ 代替理论值可得 b 的矩估计

$$\hat{b} = \frac{1 - \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}$$

如果 $\{\varepsilon_t\}$ 是独立同分布的白噪声, 由 §16.2定理16.1知道 $\hat{\rho}_1$ 是 ρ_1 的强相合估计: $\lim_{N\to\infty} \hat{\rho}_1 = \rho_1$, a.s., 于是

$$\lim_{N\to\infty} \hat{b} = \frac{1 - \sqrt{1 - 4\rho_1^2}}{2\rho_1} = b, \ a.s.,$$

所以 $\hat{b}$ 是 b 的强相合估计. 实际上利用 §16.2定理16.2还可以证明, 当 $N \to \infty$ 时, $\sqrt{N}(\hat{b} - b)$ 依分布收敛到正态分布 ([27]).

$$\mathrm{N}\left(0, \frac{1+b^2+b^4+b^6+b^8}{(1-b^2)^2}\right).$$

20.2 MA 模型的矩估计及其计算

考虑可逆 MA(q) 模型:

$$X_t = \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{20.1}$$

$\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$, 系数满足可逆条件：

$$B(z) = 1 + \sum_{j=1}^{q} b_j z^j \neq 0, \quad |z| \leq 1, \tag{20.2}$$

假定 q 已知，估计 $b = (b_1, \dots, b_q)^T$ 和 σ^2。

参数与自协方差函数关系为

$$\gamma_k = \sigma^2(b_0 b_k + b_1 b_{k+1} + \cdots + b_{q-k} b_q), \quad 0 \leq k \leq q. \tag{20.3}$$

(其中 $b_0 = 1$)

从样本估计出 $\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_q$ 后，可以用解非线性方程组的方法求解 b 和 σ^2，但不能保证解唯一，也不能保证可逆性条件。求解方法包括线性迭代方法和 Newton-Raphson 迭代方法。

还可以根据 §12.8计算矩估计。由 MA(q) 的 q 步截尾性，可定义

$$\tilde{\gamma}_k = \begin{cases} \hat{\gamma}_k, & 0 \leq k \leq q, \\ 0, & k > q, \end{cases}$$

用 $\{\tilde{\gamma}_k\}$ 作为 $\{\gamma_k\}$ 的估计代入 §12.8的计算公式。记

$$A=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{q\times q}, \quad C=\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{q\times 1} \tag{20.4}$$

$$\Omega_k=\begin{pmatrix} \tilde{\gamma}_1 & \tilde{\gamma}_2 & \cdots & \tilde{\gamma}_k \\ \tilde{\gamma}_2 & \tilde{\gamma}_3 & \cdots & \tilde{\gamma}_{k+1} \\ \cdots & \cdots & \cdots & \cdots \\ \tilde{\gamma}_q & \tilde{\gamma}_{q+1} & \cdots & \tilde{\gamma}_{q+k-1} \end{pmatrix}, \quad \gamma_q=\begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \\ \vdots \\ \tilde{\gamma}_q \end{pmatrix}$$

矩估计为

$$(\hat{b}_1,\dots,\hat{b}_q)^T=\frac{1}{\hat{\sigma}^2}(\gamma_q-A\hat{\Pi}C), \quad \hat{\sigma}^2=\hat{\gamma}_0-C^T\hat{\Pi}C \tag{20.5}$$

其中

$$\hat{\Pi}=\lim_{k\to\infty}\hat{\Omega}_k\tilde{\Gamma}_k^{-1}\hat{\Omega}_k$$

定理 20.1. 如果模型(20.1)中的 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0,\sigma^2)$, 则几乎必然地当 N 充分大后由(20.5)计算的 $\hat{b}_1,\hat{b}_2,\dots,\hat{b}_q$ 满足可逆条件(20.2).

证明由于当 $N\to\infty,\hat{\gamma}_k\to\gamma_k$, a.s., 所以

$$\begin{aligned}\hat{f}(\lambda)=&\frac{1}{2\pi}\sum_{k=-q}^{q}\hat{\gamma}_k e^{-ik\lambda}\\ &\to f(\lambda)=\frac{1}{2\pi}\sum_{k=-q}^{q}\gamma_k e^{-ik\lambda}, \text{ a.s., 当 } N\to\infty\end{aligned}$$

在 $[-\pi,\pi]$ 上一致成立. 于是当 N 充分大后 $\hat{f}(\lambda)$ 恒正. 利用 §12.6的引理12.1知道有惟一的 $(b_1',b_2',\dots,b_q')$ 满足可逆性条件(20.2), 并且使得

$$\hat{f}(\lambda)=\frac{\sigma_0^2}{2\pi}|1+\sum_{j=1}^{q}b_j'e^{-ij\lambda}|^2$$

是

$$Y_t=e_t+\sum_{j=1}^{q}b_j'e_{t-j}, \quad \{e_t\}\sim \mathrm{WN}(0,\sigma_0^2)$$

的谱密度. 这时, $\tilde{\gamma}_k = E(Y_t Y_{t+k})$. 再利用 §12.8知道

$$(\hat{b}_1, \hat{b}_2, \dots, \hat{b}_q) = (b_1', b_2', \cdots, b_q').$$

从定理20.1的证明知道这样得到的 $(\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_q)$ 满足

$$1 + \sum_{j=1}^{q} \hat{b}_j z^j \neq 0, \ |z| \leq 1$$

的充分条件是

$$\sum_{k=-q}^{q} \hat{\gamma}_k e^{-ik\lambda} > 0, \ \lambda \in [-\pi, \pi].$$

20.2.1 矩估计法的 R 程序

```r
# Given \gamma_0, \gamma_1, \dots, \gamma_q,
## Solve MA(q) coefficients b_1, \dots, b_q, \sigma^2
## Using Li Lei's algorithm.
## Input: gms -- \gamma_0, \gamma_1, \dots, \gamma_q
ma.solve <- function(gms, k=100){
  q <- length(gms)-1
  if(q==1){
    rho1 <- gms[2] / gms[1]
    b <- (1 - sqrt(1 - 4*rho1^2))/(2*rho1)
    s2 <- gms[1] / (1 + b^2)
    return(list(b=b, s2=s2))
  }

  A <- matrix(0, nrow=q, ncol=q)
  for(j in seq(2,q)){
    A[j-1,j] <- 1
  }
  cc <- numeric(q); cc[1] <- 1
  gamma0 <- gms[1]
```

```
  gammas <- numeric(q+k)
  gammas[1:(q+1)] <- gms
  gamq <- gms[-1]
  Gammak <- matrix(0, nrow=k, ncol=k)
  for(ii in seq(k)){
    for(jj in seq(k)){
      Gammak[ii,jj] <- gammas[abs(ii-jj)+1]
    }
  }

  Omk <- matrix(0, nrow=q, ncol=k)
  for(ii in seq(q)){
    for(jj in seq(k)){
      Omk[ii,jj] <- gammas[ii+jj-1+1]
    }
  }
  PI <- Omk %*% solve(Gammak, t(Omk))

  s2 <- gamma0 - c(t(cc) %*% PI %*% cc)
  b <- 1/s2 * c(gamq - A %*% PI %*% cc)
  return(list(b=b, s2=s2))
}
```

例 20.1. 考虑如下 MA(2) 模型：

$$X_t = \varepsilon_t - 0.36\varepsilon_{t-1} + 0.85\varepsilon_{t-2},\ \varepsilon_t \sim \text{WN}(0, 2^2)$$

对样本量 $N = 100, 300$ 重复产生样本 $M = 400$ 组，进行参数估计，计算估计参数的平均值、标准差、根均方误差。

```
demo.ma2.mom <- function(m=400){
  b <- c(-0.36, 0.85)
  sig <- 2.0
  ests.100 <- matrix(0, nrow=m, ncol=3)
```

```
  ests.300 <- matrix(0, nrow=m, ncol=3)
  for(ii in seq(m)){
    x <- arima.sim(
      model=list(ma=b), n=300,
      rand.gen=function(n, ...) rnorm(n, 0, sig))
    ## 样本量 100 的结果
    gms1 <- c(acf(x[1:100], type="covariance", lag.max=2, plot=FALSE)$acf)
    res1 <- ma.solve(gms1)
    ## 样本量 300 的结果
    gms2 <- c(acf(x[1:300], type="covariance", lag.max=2, plot=FALSE)$acf)
    res2 <- ma.solve(gms2)
    ests.100[ii,] <- c(res1$b, res1$s2)
    ests.300[ii,] <- c(res2$b, res2$s2)
  }
  res <- matrix(0, nrow=7, ncol=3)
  rownames(res) <- c(" 真实参数", " 样本量 100 估计均值", " 样本量 300 估计均值",
                     " 样本量 100 估计标准差", " 样本量 300 估计标准差",
                     " 样本量 100 估计根均方误差", " 样本量 300 估计根均方误差"
                     )
  colnames(res) <- c("b1", "b2", "s2")
  res[1,] <- c(b, sig^2)
  res[2,] <- apply(ests.100, 2, mean)
  res[3,] <- apply(ests.300, 2, mean)
  res[4,] <- apply(ests.100, 2, sd)
  res[5,] <- apply(ests.300, 2, sd)
  res[6,] <- sqrt(1/m*( (m-1)*res[4,]^2 + m*(res[2,]-res[1,])^2))
  res[7,] <- sqrt(1/m*( (m-1)*res[5,]^2 + m*(res[3,]-res[1,])^2))
  print(round(res, 4))

  invisible(res)
}
demo.ma2.mom(m=400)
```

```
##                               b1       b2      s2
## 真实参数                 -0.3600   0.8500  4.0000
## 样本量100估计均值         11.8755  -5.4916  3.8877
## 样本量300估计均值         -0.3270   0.6501  7.0421
## 样本量100估计标准差      247.6554 130.2699  7.1584
## 样本量300估计标准差        1.2432   2.8598 67.0377
## 样本量100估计根均方误差  247.6480 130.2614  7.1504
## 样本量300估计根均方误差    1.2421   2.8632 67.0229
```

可以看出矩估计法的误差太大，在中小样本情形无法接受。经检查，如果输入的是理论自协方差函数，求解的结果是完全正确的，所以并不是程序问题，而是矩估计法的问题。极大似然估计可以获得较满意的估计精度。

20.3 MA 模型参数估计的逆相关函数法

因为 AR 的 Yule-Warker 估计能保证最小相位性所以想到把 MA 变成一个 AR 再估计。

定义 20.1. 设平稳序列 $\{X_t\}$ 有恒正的谱密度 $f(\lambda)$. 通常称

$$f_y(\lambda) = \frac{1}{4\pi^2 f(\lambda)}. \tag{20.6}$$

为 $\{X_t\}$ 的**逆谱密度**. 称

$$\gamma_y(k) \triangleq \int_{-\pi}^{\pi} e^{ik\lambda} f_y(\lambda)\, d\lambda$$

为 $\{X_t\}$ 的**逆相关函数**或**逆自相关函数**. 严格来说应该叫做**逆自协方差函数**。

设 $\{X_t\}$ 为可逆 MA(q) 序列。其谱密度为

$$f(\lambda) = \frac{\sigma^2}{2\pi}|B(e^{i\lambda})|^2$$

逆谱密度为

$$f_y(\lambda) = \frac{\sigma^{-2}}{2\pi}|B(e^{i\lambda})|^{-2} \tag{20.7}$$

f_y 为如下 AR(q) 序列的谱密度

$$Y_t = -\sum_{j=1}^{q} b_j Y_{t-j} + e_t, \quad t \in \mathbb{Z} \tag{20.8}$$

其中 $\{e_t\} \sim \mathrm{WN}(0, \sigma^{-2})$。

$\{Y_t\}$ 的自协方差函数

$$\gamma_y(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f_y(\lambda)\, d\lambda, \quad k = 0, 1, 2, \ldots \tag{20.9}$$

是 $\{X_t\}$ 的逆相关函数。

只要能估计 $\{\gamma_y(k)\}$，设估计为 $\{\hat\gamma_y(k)\}$，就可以用 Y-W 方法估计 $\{Y_t\}$ 的参数 $b_1, \ldots, b_q$ 和 σ^{-2}：

$$\begin{pmatrix} \hat\gamma_y(0) & \hat\gamma_y(1) & \cdots & \hat\gamma_y(q-1) \\ \hat\gamma_y(1) & \hat\gamma_y(0) & \cdots & \hat\gamma_y(q-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat\gamma_y(q-1) & \hat\gamma_y(q-2) & \cdots & \hat\gamma_y(0) \end{pmatrix} \begin{pmatrix} -\hat b_1 \\ -\hat b_2 \\ \vdots \\ -\hat b_q \end{pmatrix} = \begin{pmatrix} \hat\gamma_y(1) \\ \hat\gamma_y(2) \\ \vdots \\ \hat\gamma_y(q) \end{pmatrix} \tag{20.10}$$

$$\hat\sigma^{-2} = \hat\gamma_y(0) + \hat b_1 \hat\gamma_y(1) + \cdots + \hat b_q \hat\gamma_y(q) \tag{20.11}$$

从而得到 MA(q) 序列 $\{X_t\}$ 的参数估计，且系数满足可逆性条件。

引理 20.1. 如果 $(a_1, a_2, \ldots, a_p)$ 和 σ^2 分别是 $AR(p)$ 模型

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z},$$

的自回归系数和白噪声 $\{\varepsilon_t\}$ 的方差，则 $\{X_t\}$ 的逆相关函数为

$$\gamma_y(k) = \begin{cases} \frac{1}{\sigma^2} \sum_{j=0}^{p-k} a_j a_{j+k}, & 0 \le k \le p, \quad a_0 \triangleq -1, \\ 0, & k > p. \end{cases}$$

证明 $\{X_t\}$ 有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} |A(e^{i\lambda})|^{-2}, \text{ 其中 } A(z) = 1 - \sum_{j=1}^{p} a_j z^j,$$

和逆谱密度

$$f_y(\lambda) = \frac{1}{4\pi^2 f(\lambda)} = \frac{1}{2\pi\sigma^2}|A(e^{i\lambda})|^2,$$

于是有逆相关函数

$$\gamma_y(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f_y(\lambda) d\lambda = \begin{cases} \frac{1}{\sigma^2}\sum_{j=0}^{p-k} a_j a_{j+k}, & 0 \le k \le p, \\ 0, & k > p. \end{cases}$$

满足(20.1)的可逆 MA(q) 序列 $\{X_t\}$ 可以写成无穷阶自回归的形式

$$X_t - \sum_{j=1}^{\infty} a_j X_{t-j} = \varepsilon_t, \quad t = 1, 2, .., \tag{20.12}$$

这里的回归系数 $\{a_j\}$ 由 $1/B(z)$ 在单位圆内的 Taylor 级数决定:

$$\frac{1}{B(z)} = 1 - \sum_{j=1}^{\infty} a_j z^j, \ |z| \le 1.$$

由于当 $j \to \infty$, a_j 是以负指数阶收敛到 0 的, 所以对较大的正整数 p 可以将无穷阶自回归模型(20.12)写成近似的长阶 (p 阶) 自回归的形式

$$X_t \approx \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t, \ t = 1, 2, .., \tag{20.13}$$

利用引理20.1知道 $\{X_t\}$ 的逆相关函数 $\gamma_y(k)$ 满足

$$\gamma_y(k) \approx \frac{1}{\sigma^2} \sum_{j=0}^{p-k} a_j a_{j+k}, \ 0 \le k \le q, \ a_0 \triangleq -1. \tag{20.14}$$

用样本 $x_1, x_2, \dots, x_N$ 计算逆相关函数 $\hat{\gamma}_y(k)$ 和 $\hat{b}_1, \dots, \hat{b}_q, \hat{\sigma}^2$ 的方法如下。

(1) 首先利用 $\{x_t\}$ 的样本自协方差函数 $\hat{\gamma}_k$ 建立一个 AR(p_N) 模型, 这里 p_N 可以是 AR 模型的 AIC 定阶, 也可以取作 $K\ln(N)$ 的整数部分, K 是一个正数.

(2) 对 $p \equiv p_N$ 解样本 Yule-Walker 方程(19.3), (19.4), 得到样本 Yule-Walker 系数

$$\hat{a} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p)^T \quad 和 \ \hat{\sigma}_p^2.$$

(3) 计算样本逆相关函数

$$\hat{\gamma}_y(k) = \frac{1}{\hat{\sigma}_p^2} \sum_{j=0}^{p-k} \hat{a}_j \hat{a}_{j+k}, \ k = 0, 1, 2, \dots, q, \ \hat{a}_0 \triangleq -1.$$

(4) 利用样本 Yule-Walker 方程(20.10)和(20.11)计算出 MA(q) 系数的估计 $\hat{b}_q = (\hat{b}_1, \hat{b}_2, .., \hat{b}_q)^T$ 和 $\hat{\sigma}^2$.

这种办法是用一个长阶自回归来近似 $\{X_t\}$，自回归序列的逆自相关函数很容易计算，这样得到逆相关函数。

逆相关函数法估计 MA 参数的 R 程序：

```
## 用输入的 $\gamma_0, \dots, \gamma_p$ 求解 AR(p) 模型参数
ar.yw <- function(gam){
  p <- length(gam) - 1
  G <- matrix(0, p, p)
  for(i in 1:p){
    for(j in 1:p){
      G[i,j] = gam[1 + abs(i-j)]
    }
  }
  gri <- gam[-1]
  a <- solve(G, gri)
  sig2 <- sum(gam * c(1, -a))
  list(ar = a, var.pred = sig2)
}

## 从 AR 模型参数计算逆相关函数
ar_racv <- function(a, sig2){
  p <- length(a)
  a <- c(-1, a)
  a2 <- c(a, numeric(p+1))
  res <- numeric(p+1)
  for(k in seq(0, p, by=1)){
    res[k+1] <- sum(a * a2[(k+1):(k+p+1)])
```

```r
  }
  res/sig2
}

## 用长阶自回归方法和逆相关函数方法估计 MA 模型
ma.solve.racv <- function(x, q=1){

  ## 拟合长阶自回归，用来估计逆自协方差函数
  n <- length(x)
  plar <- round(sqrt(n))
  if(plar < q) plar <- q

  mod1 <- ar(x, aic = FALSE, order.max = plar, method="yule-walker")
  a <- mod1$ar
  sigsq <- mod1$var.pred
  gamr <- ar_racv(a, sigsq)[1:(q+1)]
  res2 <- ar.yw(gamr)
  b <- -res2$ar
  sig2 <- 1/res2$var.pred

  list(ma = b, var.pred = sig2)
}
```

20.4 MA 模型参数估计的新息方法

用新息预报公式可以计算 $\{X_t\}$ 的样本新息。

$$
\begin{aligned}
\hat{\varepsilon}_{t+1} =& X_{t+1} - \hat{X}_{t+1} \triangleq X_{t+1} - L(X_{t+1}|X_1, \dots, X_t) \\
=& X_{t+1} - L(X_{t+1}|\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_t) \\
=& X_{t+1} - \sum_{j=1}^{t} \theta_{t,j} \hat{\varepsilon}_{t+1-j}, \quad t = 1, 2, \dots
\end{aligned}
$$

其中 $\hat{X}_1 \triangleq 0$，$\{\theta_{t,j}\}$ 可递推计算，预报误差 $\nu_t = E\hat{\varepsilon}_{t+1}^2$ 可递推计算。对 MA(q) 序列 $\{X_t\}$ 上述新息预报公式中当 $t \geq q, j > q$ 时 $\theta_{t,j} = 0$, 新息预报公式变成

$$\hat{X}_{t+1} = \sum_{j=1}^{q} \theta_{t,j} \hat{\varepsilon}_{t+1-j}, \quad t \geq q \tag{20.15}$$

对 MA(q) 序列 $\{X_t\}$，其白噪声项 $\{\varepsilon_t\}$ 是新息, $t \to \infty$ 时:

$$\begin{aligned} \varepsilon_{t+1} =& X_{t+1} - L(X_{t+1} | X_t, X_{t-1}, X_{t-2}, \dots) \\ \approx& X_{t+1} - L(X_{t+1} | X_t, X_{t-1}, X_{t-2}, \dots, X_1) \\ =& \hat{\varepsilon}_{t+1} \end{aligned}$$

这是因为 §23.2.2定理23.3，无穷长历史最佳线性预测是有限历史预测的极限。于是 t 较大时

$$\begin{aligned} X_t =& \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q} \\ \approx& \hat{\varepsilon}_t + b_1 \hat{\varepsilon}_{t-1} + \dots + b_q \hat{\varepsilon}_{t-q} \\ =& X_t - \hat{X}_t + b_1 \hat{\varepsilon}_{t-1} + \dots + b_q \hat{\varepsilon}_{t-q} \end{aligned}$$

说明

$$\hat{X}_t \approx b_1 \hat{\varepsilon}_{t-1} + \dots + b_q \hat{\varepsilon}_{t-q}$$

上式与(20.15)) 的新息预报公式比较可知当 t 较大时可以用 $\hat{b}_j = \theta_{t,j}$ 估计 b_j，用 ν_t 估计 σ^2。这种估计称为**新息估计**。

MA 参数的新息估计算法如下:

- 给定观测数据 $x_1, x_2, \dots, x_N$, 取 $m = o(N^{1/3})$.
- 计算样本自协方差函数 $\hat{\gamma}_0, \hat{\gamma}_1, \cdots, \hat{\gamma}_m$.
- b 和 σ^2 的新息估计

  $$(\hat{b}_1, \hat{b}_2, \dots, \hat{b}_q) = (\hat{\theta}_{m,1}, \hat{\theta}_{m,2}, \dots, \hat{\theta}_{m,q}), \quad \hat{\sigma}^2 = \hat{\nu}_m. \tag{20.16}$$

  由下面的新息预报递推公式得到.

  $$\begin{cases} \hat{\nu}_0 = \hat{\gamma}_0, \\ \hat{\theta}_{n,n-k} = \hat{\nu}_k^{-1} [\hat{\gamma}_{n-k} - \sum_{j=0}^{k-1} \hat{\theta}_{k,k-j} \hat{\theta}_{n,n-j} \hat{\nu}_j], & 0 \leq k \leq n-1, \\ \hat{\nu}_n = \hat{\gamma}_0 - \sum_{j=0}^{n-1} \hat{\theta}_{n,n-j}^2 \hat{\nu}_j, & 1 \leq n \leq m, \end{cases} \tag{20.17}$$

  其中 $\sum_{j=0}^{-1} (\cdot) \triangleq 0$.

- 递推次序是

$$\hat{\nu}_0;\ \hat{\theta}_{1,1}, \hat{\nu}_1;\ \hat{\theta}_{2,2}, \hat{\theta}_{2,1}, \hat{\nu}_2; \hat{\theta}_{3,3}, \hat{\theta}_{3,2}, \hat{\theta}_{3,1}, \hat{\nu}_3; \cdots.$$

下面的定理与推论给出了新息估计方法的相合性。

定理 20.2. 设 $\{X_t\}$ 是可逆 *ARMA(p, q)* 序列, 满足

$$A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z}.$$

$\{\psi_j\}$ 是 $B(z)/A(z)$ 的 *Taylor* 级数系数. 如果 $\{\varepsilon_t\}$ 是 *4* 阶矩有限的独立同分布 $WN(0, \sigma^2)$, 正整数列 $m = m(N) < N$ 满足当 $N \to \infty$ 时, $m \to \infty$ 和 $m = o(N^{1/3})$. 则对任何正整数 q, 当 $N \to \infty$

$$\sqrt{N}(\hat{\theta}_{m,1} - \psi_1, \hat{\theta}_{m,2} - \psi_2, \dots, \hat{\theta}_{m,q} - \psi_q)$$

依分布收敛到 q 维正态分布 $N(0, A)$, 其中 $q \times q$ 矩阵

$$A = (a_{i,j}), \quad a_{i,j} = \sum_{k=1}^{\min(i,j)} \psi_{i-k}\psi_{j-k}.$$

并且 $\hat{\nu}_m$ 依概率收敛到 σ^2.

参见 (Brockwell & Davis, 1987)。

推论 20.1. 对于 *MA(q)* 序列 $\{X_t\}$, 当模型(20.1)中的白噪声是 *4* 阶矩有限的独立同分布序列时, 新息估计(20.16)是相合估计, 当 N 趋于无穷时,

$$\sqrt{N}(\hat{b}_1 - b_1, \hat{b}_2 - b_2, \cdots, \hat{b}_q - b_q)$$

依分布收敛到 q 维正态分布 $N(0, A)$, 其中 $q \times q$ 矩阵

$$A = (a_{i,j}), \quad a_{i,j} = \sum_{k=1}^{\min(i,j)} b_{i-k}b_{j-k}, \quad b_0 \triangleq 1.$$

并且 $\hat{\nu}_m$ 依概率收敛到 σ^2.

注意需要 $m \to \infty$ 时新息方法得到的参数估计才是相合的，不能用 $\theta_{q,1}, \dots, \theta_{q,q}$ 作为 $b_1, \dots, b_q$ 的估计。实际中，m 也不能取得太大，因为新息预测的递推公式中用到 $\hat{\gamma}_{m-1}$。

20.5 MA 模型的定阶方法

由于 MA(q) 序列的特征是自相关系数 q 后截尾, 所以当样本自相关系数 $\hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$ 从某一点 $\hat{q}$ 后变得很小时, 可以 $\hat{q}$ 作为 q 的估计. 从 §16.3的定理16.2及§16.3.1知道, 对于 $m > q$, $\sqrt{N}\hat{\rho}_m$ 依分布收敛到期望为 0, 方差为

$$1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_q^2$$

的正态分布. 这样由样本自相关系数 $\hat{\rho}_k, k = 1, 2, \cdots, m$ 的图形可以大致得到 q 的估计, 同时也可以判断采用 MA(q) 模型的合理与否.

还可以用 AIC 定阶方法: 如果根据问题的背景或数据的特性能够判定 MA(q) 模型阶数 q 的上界是 Q_0. 对于 $m = 0, 1, 2, \ldots, Q_0$ 按前述的方法逐个拟合 MA(m) 模型. 白噪声方差 σ^2 的估计量记做 $\hat{\sigma}_m^2$. 定义 AIC 函数

$$\mathrm{AIC}(m) = \ln(\hat{\sigma}_m^2) + 2m/N, \quad m = 0, 1, \ldots, Q_0.$$

这里 N 是样本个数. AIC(m) 的最小值点 $\hat{q}$ (如不惟一, 应取小的) 称为 MA(q) 模型的 AIC 定阶.

20.6 MA 模型的拟合检验

从观测数据 $x_1, x_2, \cdots, x_N$ 得到模型的参数估计 $\hat{q}, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{\hat{q}}$ 和 $\hat{\sigma}^2$ 后, 取

$$\begin{aligned}
\hat{\varepsilon}_{1-\hat{q}} &= \hat{\varepsilon}_{2-\hat{q}} = \cdots = \hat{\varepsilon}_0 = 0, \\
y_t &= x_t - \bar{x}_N, \\
\hat{\varepsilon}_t &= y_t - \sum_{j=1}^{\hat{q}} \hat{b}_j \hat{\varepsilon}_{t-j}, \ t = 1, 2, \ldots, N.
\end{aligned}$$

对 $L = O(N^{1/3})$, 如果 $\{\hat{\varepsilon}_t : t = L, L+1, \cdots, N\}$ 能够通过白噪声检验, 就认为模型的选择合适. 否则改变 $\hat{q}$ 的取值, 拟合新的 MA 模型或改用其他的模型, 例如改用 ARMA 模型等.

20.7 MA 谱密度估计

如果从数据得到了 MA 模型的参数估计, 模型的检验也已经通过, 可以利用

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2}{2\pi} |1 + \sum_{j=1}^{\hat{q}} \hat{b}_j e^{ij\lambda}|^2$$

作为所关心的平稳序列的谱密度的估计. 这是因为如果观测数据确实是 MA(q) 序列(20.1)时, 它的谱密度是

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1 + \sum_{j=1}^{q} b_j e^{ij\lambda}|^2.$$

不难看出, 如果 $\hat{q}$, $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_q$ 和 $\hat{\sigma}^2$ 分别是 q, $b_1, b_2, .., b_q$ 和 σ^2 的相合估计, 则 $\hat{f}(\lambda)$ 是 $f(\lambda)$ 的相合估计.

Chapter 21

ARMA 模型的参数估计

对于零均值化后的平稳观测数据 $x_1, x_2, \cdots, x_N$, 如果拟合 AR(p) 和 MA(q) 模型的效果都不理想, 就要考虑 ARMA(p,q) 模型的拟合. 这时可以假设 $x_1, x_2, \cdots, x_N$ 满足如下的可逆 ARMA(p,q) 模型

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, t = 1, 2, ... \tag{21.1}$$

其中 $\{\varepsilon_t\}$ 是 WN(0,σ^2), 未知参数 $a = (a_1, a_2, \cdots, a_p)^T$ 和 $b = (b_1, b_2, \cdots, b_q)^T$ 使得多项式

$$A(z) = 1 - \sum_{j=1}^{p} a_j z^j, \quad B(z) = 1 + \sum_{j=1}^{q} b_j z^j \tag{21.2}$$

互素, 并且满足

$$A(z)B(z) \neq 0, \ |z| \leq 1. \tag{21.3}$$

以下仍然用 $\hat{\gamma}_k$ 表示由(19.2)定义的样本自协方差函数. 我们先设 p, q 是已知的.

21.1 ARMA(p,q) 模型的矩估计方法

利用 §13.4的(13.12)知道 ARMA(p,q) 序列的自协方差函数满足延伸的 Yule-Walker 方程

$$\begin{pmatrix} \gamma_{q+1} \\ \gamma_{q+2} \\ \vdots \\ \gamma_{q+p} \end{pmatrix} = \begin{pmatrix} \gamma_q & \gamma_{q-1} & \cdots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q-p+2} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \tag{21.4}$$

这是参数 a 的估计方程, 从它得到 a 的矩估计

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_q & \hat{\gamma}_{q-1} & \cdots & \hat{\gamma}_{q-p+1} \\ \hat{\gamma}_{q+1} & \hat{\gamma}_q & \cdots & \hat{\gamma}_{q-p+2} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}_{q+p-1} & \hat{\gamma}_{q+p-2} & \cdots & \hat{\gamma}_q \end{pmatrix}^{-1} \begin{pmatrix} \hat{\gamma}_{q+1} \\ \hat{\gamma}_{q+2} \\ \vdots \\ \hat{\gamma}_{q+p} \end{pmatrix}. \tag{21.5}$$

利用 §13.4的定理13.2知道(21.4)中的 $p\times p$ 矩阵 $\Gamma_{p,q}$ 是可逆的. 用 $\hat{\Gamma}_{p,q}$ 表示(21.5)中的 $p\times p$ 矩阵. 当 ARMA(p,q) 模型中的白噪声 $\{\varepsilon_t\}$ 独立同分布时, $\hat{\gamma}_k$ a.s. 收敛到 γ_k. 于是当 $N\to\infty$,

$$\det(\hat{\Gamma}_{p,q}) \to \det(\Gamma_{p,q}) \neq 0.$$

所以当 N 充分大后, (21.5)中的矩阵 $\hat{\Gamma}_{p,q}$ 也是可逆的. 这时矩估计是惟一的. 还可以看出, 在上述的条件下, 矩估计(21.5)是强相合的:

$$\lim_{N\to\infty} \hat{a}_j = a_j,\ a.s.,\ 1\le j\le p. \tag{21.6}$$

这种相合性要求大样本，在中小样本情形有很大可能误差较大，且不一定满足最小相位条件。在中小样本，矩估计不是一种实用的方法，即使作为最大似然估计的初值也不太合适。

下面估计 MA(q) 部分的参数. 由于

$$z_t \triangleq x_t - \sum_{j=1}^{p} a_j x_{t-j},\ t=p+1,p+2,\dots,N$$

满足 MA(q) 模型

$$z_t = B(\mathscr{B})\varepsilon_t,$$

所以得到 $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ 后,

$$y_t = x_t - \sum_{j=1}^{p} \hat{a}_j x_{t-j},\ t = p+1, p+2, \dots, N. \tag{21.7}$$

是一个 MA(q) 序列的近似观测数据. 它的样本自协方差函数由

$$\hat{\gamma}_y(k) = \sum_{j=0}^{p} \sum_{l=0}^{p} \hat{a}_j \hat{a}_l \hat{\gamma}_{k+j-l}, k = 0, 1, \cdots, q \tag{21.8}$$

定义, 其中 $\hat{a}_0 = -1$. 现在将(21.8)看成一个 MA(q) 序列的样本自协方差函数, 利用第20章的方法就可以估计出 MA(q) 部分的参数 b 和 σ^2.

矩估计法，其中 MA 部分用逆相关函数法的 R 程序：

```
## 用矩估计法估计 ARMA，其中 MA 部分用逆相关函数法
## gms 输入 $\gamma_k$, $k=0,1,\dots,p+q$。
arma.moment.racv <- function(x, p=1, q=1){
  n <- length(x)
  pmax <- round(sqrt(n))
  if(pmax < p+q) pmax <- p+q
  ## 计算样本自协方差函数
  gms <- c(acf(x, lag.max=pmax+p, type="covariance", plot=FALSE)$acf)

  ## 求解 AR 部分
  Gpq <- matrix(0, nrow=p, ncol=p)
  for(ii in seq(p)) for(jj in seq(p)){
    Gpq[ii,jj] <- gms[abs(q + ii - jj)+1]
  }
  gs <- gms[(q+1+1):(q+p+1)]
  a <- solve(Gpq, gs)
  aa <- c(-1, a)

  ## 从已有的自协方差估计与 AR 参数估计，计算 MA 部分的 $\gamma_j, j=0,1,\dots,pmax$。
  gys <- numeric(pmax+1)
  for(k in seq(0, pmax)){
    gys[k+1] <- sum(c(outer(0:p,0:p,
```

```
                            function(ii,jj) aa[ii+1] * aa[jj+1]
                            * gms[abs(k+jj-ii)+1])))
  }

  ## 从 MA 部分的自协方差估计，拟合长阶自回归模型
  res1 <- ar.yw(gys)

  ## 求 MA 部分的逆相关函数
  gamr <- ar_racv(res1$ar, res1$var.pred)[1:(q+1)]

  ## 利用逆相关函数为输入求解 YW 方程，得到 MA 参数估计
  res2 <- ar.yw(gamr)
  b <- -res2$ar
  sig2 <- 1/res2$var.pred
  ##sig2 <- gys[1] / sum(c(1, b)^2)

  list(ar=a, ma=b, var.pred=sig2)
}
## 从 AR 模型参数计算逆相关函数
ar_racv <- function(a, sig2){
  p <- length(a)
  a <- c(-1, a)
  a2 <- c(a, numeric(p+1))
  res <- numeric(p+1)
  for(k in seq(0, p, by=1)){
    res[k+1] <- sum(a * a2[(k+1):(k+p+1)])
  }
  res/sig2
}
```

21.2 ARMA(p,q) 模型的自回归逼近法

如果 ARMA 模型中已知 $\{\varepsilon_t\}$ 则 $a_1,\ldots,a_p$, $b_1,\ldots,b_q$ 可以看成是回归系数。ε_t 作为一步预报误差，可以用样本新息估计。但是样本新息直接计算困难，所以可以拟合长阶自回归模型，用自回归模型的残差作为一步预报误差的估计。

首先为数据建立 AR 模型. 取自回归阶数的上界 $P_0=[\sqrt{N}]$. 这里 $[a]$ 表示 a 的整数部分. 采用 AIC 定阶方法得到 AR 模型的阶数估计 $\hat{p}$ 和自回归系数的估计

$$(\hat{\phi}_1,\hat{\phi}_2,\ldots,\hat{\phi}_{\hat{p}}).$$

计算残差

$$\hat{\varepsilon}_t=x_t-\sum_{j=1}^{\hat{p}}\hat{\phi}_j x_{t-j},\quad t=\hat{p}+1,\hat{p}+2,\ldots,N.$$

然后写出近似的 ARMA(p,q) 模型

$$x_t=\sum_{j=1}^{p}a_j x_{t-j}+\hat{\varepsilon}_t+\sum_{j=1}^{q}b_j\hat{\varepsilon}_{t-j},\ t=L+1,L+2,\ldots,N.$$

$L=\max(\hat{p},p,q)$, a_j,b_k 是待定参数. 最后对目标函数

$$Q(a,b)=\sum_{t=L+1}^{N}\left(x_t-\sum_{j=1}^{p}a_j x_{t-j}-\sum_{j=1}^{q}b_j\hat{\varepsilon}_{t-j}\right)^2 \tag{21.9}$$

极小化, 得到最小二乘估计 $(\hat{a}_1,\ldots,\hat{a}_p,\hat{b}_1,\ldots\hat{b}_q)$. σ^2 的最小二乘估计由下式定义.

$$\hat{\sigma}^2=\frac{1}{N-L}Q(\hat{a}_1,\ldots,\hat{a}_p,\hat{b}_1,\ldots,\hat{b}_q).$$

21.3 正态时间序列的似然函数

设 $\{X_t\}$ 是零均值正态时间序列, 对 $n\ge 1$, $X_n=(X_1,X_2,..,X_n)^T$ 的协方差矩阵 Γ_n 正定. 采用 §24.3的记号可以得到最佳线性预测

$$\hat{X}_n\triangleq L(X_n|X_{n-1})=\sum_{j=1}^{n-1}\theta_{n-1,n-j}Z_j,$$

其中 $Z_j = X_j - \hat{X}_j$, $Z_1 = X_1$. 于是

$$\begin{aligned} X_n =& \hat{X}_n + Z_n = \sum_{j=1}^{n-1} \theta_{n-1,n-j} Z_j + Z_n \\ =& \sum_{j=1}^{n} \theta_{n-1,n-j} Z_j \qquad (\theta_{n-1,0} \triangleq 1) \\ =& (\theta_{n-1,n-1}, \theta_{n-1,n-2}, \cdots, \theta_{n-1,1}, 1) \cdot Z_n, \end{aligned}$$

其中 $Z_n = (Z_1, Z_2, \cdots, Z_n)^T$.

为了方便引入下三角矩阵

$$C = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 1 & 0 & \cdots & 0 \\ \theta_{2,2} & \theta_{2,1} & 1 & \cdots & 0 \\ \vdots v \cdots & \vdots & \cdots & \vdots & \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{pmatrix}$$

则有

$$X_n = C Z_n.$$

由于

$$r_{k-1} \triangleq E Z_k^2, \quad k = 1, 2, ..., n-1$$

是预测的均方误差 (原来新息预报中 r_{k-1} 用 ν_{k-1} 表示，但是下面 ν_{k-1} 要表示对 X_t 作变换后的序列 Y_t 的样本新息方差), 所以用 $Z_1, Z_2, \ldots, Z_n$ 的正交性得到

$$D \triangleq E(Z_n Z_n^T) = \text{diag}(r_0, r_1, \ldots, r_{n-1}).$$

由此得到 X_n 的协方差矩阵:

$$\begin{aligned} &\Gamma_n = E(C Z_n Z_n^T C^T) = C D C^T, \\ &\det(\Gamma_n) = \det(D) = r_0 r_1 \cdots r_{n-1}, \\ &X_n^T \Gamma_n^{-1} X_n = Z_n^T C^T (C D C^T)^{-1} C Z_n = \sum_{j=1}^{n} Z_j^2 / r_{j-1}. \end{aligned}$$

由于 X_n 的分布由 Γ_n 决定, 而 r_j, $\theta_{k,j}$ 都是 Γ_n 的函数, 所以可得基于 X_n 的

似然函数

$$
\begin{aligned}
L(\Gamma_n) =& \frac{1}{(2\pi)^{n/2}[\det(\Gamma_n)]^{1/2}} \exp\left(-\frac{1}{2}X_n^T\Gamma_n^{-1}X_n\right) \\
=& \frac{1}{(2\pi)^{n/2}(r_0 r_1 \cdots r_{n-1})^{1/2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n} Z_j^2/r_{j-1}\right).
\end{aligned}
\tag{21.10}
$$

21.4 ARMA 模型的最大似然估计

21.4.1 似然函数

设 $\{X_t\}$ 是满足 ARMA 模型(21.1)的平稳序列. 采用 §25.3中符号, 利用 §25.3的(25.13)得到逐步预报的递推公式

$$
\hat{X}_{k+1} = \begin{cases} \sum_{j=1}^{k} \theta_{k,j} Z_{k+1-j}, & 1 \le k < m, \\ \sum_{j=1}^{p} a_j X_{k+1-j} + \sum_{j=1}^{q} \theta_{k,j} Z_{k+1-j}, & k \ge m. \end{cases} \tag{21.11}
$$

这里 $m = \max(p, q)$,

$$
Z_k = X_k - \hat{X}_k = X_k - L(X_k | X_{k-1}, X_{k-2}, \dots, X_1), \tag{21.12}
$$

$$
r_{k-1} = EZ_k^2 = E(X_k - \hat{X}_k)^2 = \sigma^2 \nu_{k-1}. \tag{21.13}
$$

$\theta_{k,j}, \nu_k$ 可以通过(25.11)和(24.9)递推计算。而(25.11)中的 $\sigma^{-2}\gamma_k$ 由 §13.3的(13.9), (13.10)计算，即用 ARMA 自协方差的 Wold 系数表达式，Wold 系数可由模型参数递推计算。它们都是和 σ^2 无关的量, 由 ARMA 模型(21.1)的参数

$$
\beta = (a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_q)^T \tag{21.14}
$$

惟一决定. 从而 $\hat{X}_k$ 也是和 σ^2 无关的量, 仅由观测数据 $X_1, X_2, \dots, X_{k-1}$ 和 β 决定. 将(21.11)和(21.13)代入(21.10)就得到基于观测数据 $X_1, X_2, \cdots, X_N$ 的似然函数

$$
L(\beta, \sigma^2) = \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{N} Z_k^2/\nu_{k-1}\right)}{(2\pi)^{N/2}(\sigma^{2N}\nu_0\nu_1\cdots\nu_{N-1})^{1/2}} \tag{21.15}
$$

引入

$$S(\beta) = \sum_{k=1}^{N} Z_k^2/\nu_{k-1}. \tag{21.16}$$

忽略常数项后, 可以得到对数似然函数

$$\ln L(\beta, \sigma^2) = -\frac{N}{2}\ln\sigma^2 - \frac{1}{2}\ln(\nu_0\nu_1\cdots\nu_{N-1}) - \frac{1}{2\sigma^2}S(\beta). \tag{21.17}$$

利用

$$\frac{\partial}{\partial\sigma^2}\ln L(\beta, \sigma^2) = -\frac{N}{2\sigma^2} + \frac{S(\beta)}{2\sigma^4} = 0$$

得到

$$\sigma^2 = \frac{1}{N}S(\beta).$$

将上式代入(21.17)得到

$$l(\beta) \triangleq -\frac{2}{N}\ln L(\beta, S(\beta)/N) - 1 \tag{21.18}$$

$$= \frac{1}{N}\ln(\nu_0\nu_1\cdots\nu_{N-1}) + \ln[S(\beta)/N]. \tag{21.19}$$

通常称 $l(\beta)$ 是约化似然函数. 可以看出, $l(\beta)$ 的最小值点

$$\hat{\beta} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_q)^T \tag{21.20}$$

是 β 的最大似然估计, 而

$$\hat{\sigma}^2 = \frac{1}{N}S(\hat{\beta}) \tag{21.21}$$

是 σ^2 的最大似然估计.

21.4.2 最大似然估计的计算

在计算 $l(\beta)$ 的最小值点时, 可采用一般的最优化方法. 要计算 $l(\beta)$ 函数值，通过(25.11)和(24.9)递推计算出 $\theta_{k,j}$, ν_j 和 $Z_k = X_k - \hat{X}_k$, 然后计算出 $l(\beta)$.

为了加快搜索的速度和提高估计的精度, 初始值 $\beta^{(0)}$ 应当选择在 $l(\beta)$ 的最小值附近.

实际计算中, 初始值应当选成在 §21.1中定义的矩估计或21.2中定义的自回归逼近估计. 为了得的估计模型的合理性和可逆性, 还应当选择初始值 $\beta^{(0)}$ 使得 $A(z), B(z)$ 在闭单位圆内没有零点.

21.4.3 最大似然估计与最小二乘估计

由于当 $k \to \infty$, 用 §23.2定理23.2.2得到

$$\begin{aligned}\nu_k =&E(X_{k+1}-\hat{X}_{k+1})^2/\sigma^2\\ =&E[X_n - L(X_n|X_{n-1}, X_{n-2}, \dots, X_{n-k})]^2/\sigma^2\\ \to&E[X_n - L(X_n|X_{n-1}, X_{n-2}, \dots)]^2/\sigma^2\\ =&E\varepsilon_n^2/\sigma^2 = 1.\end{aligned}$$

所以 $N \to \infty$ 时,

$$\frac{1}{N}\ln(\nu_0\nu_1\cdots\nu_{N-1}) \to 0.$$

于是, 对较大的 N, $l(\beta)$ 和 $S(\beta)$ 的最小值点近似相等. 于是也可以用 $S(\beta)$ 的最小值点 $\tilde{\beta}$ 作为 β 的估计. 通常也称 $\tilde{\beta}$ 是 β 的最小二乘估计, 相应的白噪声方差 σ^2 的估计定义成

$$\tilde{\sigma}^2 = \frac{1}{N-p-q}S(\tilde{\beta}). \tag{21.22}$$

21.4.4 最大似然估计的极限分布

可以证明最小二乘估计 $\tilde{\beta}$ 和最大似然估计 $\hat{\beta}$ 有相同的极限分布, $\tilde{\sigma}^2$ 和 $\hat{\sigma}^2$ 有相同的极限分布.

定理 21.1. *如果 $\{X_t\}$ 是平稳可逆的 ARMA(p, q) 序列, 白噪声是独立同分布序列, $E\varepsilon_t^4 < \infty$. 则当 $N \to \infty$ 时,*

$$\sqrt{N}(\hat{\beta}-\beta)$$

依分布收敛到正态分布 $N(0, V(\beta))$, 其中

$$V(\beta) = \begin{cases} \sigma^2\begin{pmatrix} E(UU^T) & E(UV^T)\\ E(VU^T) & E(VV^T)\end{pmatrix}^{-1}, & p>0, q>0,\\ \sigma^2\Big(E(UU^T)\Big)^{-1}, & q=0, p>0,\\ \sigma^2\Big(E(VV^T)\Big)^{-1}, & p=0, q>0.\end{cases} \tag{21.23}$$

$U = (U_p, U_{p-1}, \dots, U_1)^T$, $V = (V_q, V_{q-1}, \dots, V_1)^T$,
$\{U_t\}$ 和 $\{V_t\}$ 分别是 AR(p) 和 AR(q) 序列, 分别满足 $A(\mathscr{B})U_t = \varepsilon_t$ 和 $B(\mathscr{B})V_t = \varepsilon_t$.

极限分布的利用:

用 v_{jj} 表示 $V(\beta)$ 的第 (j,j) 的元素, 则

$$\sqrt{N}(\hat\beta_j - \beta_j)$$

依分布收敛到正态分布 $N(0, v_{jj})$. 在实际问题中, 真值 $V(\beta)$ 是未知的, 通常用估计量 $V(\hat\beta)$ 代替. 于是可以用 $V(\hat\beta)$ 的第 (j,j) 的元 $\hat v_{jj}$ 作为 v_{jj} 的近似. 这样, 当 N 较大时, 利用

$$P[\sqrt{N}|\hat\beta_j - \beta_j|/\sqrt{v_{jj}} \leq 1.96] \approx 0.95$$

就可以在置信水平 0.95 下得到 β_j 的近似置信区间

$$[\ \hat\beta_j - 1.96\sqrt{\hat v_{jj}/N}, \quad \hat\beta_j + 1.96\sqrt{\hat v_{jj}/N}\ ].$$

21.5 例子：ARMA(4,2) 的估计与模拟分析

对于 ARMA(4,2) 模型

$$\begin{aligned} & X_t + 0.9X_{t-1} + 1.4X_{t-2} + 0.7X_{t-3} + 0.6X_{t-4} && (21.24)\\ =& \varepsilon_t + 0.5\varepsilon_{t-1} - 0.4\varepsilon_{t-2}, \quad t \in \mathbb{Z}, && (21.25) \end{aligned}$$

的模拟计算,
其中 $\{\varepsilon_t\}$ 是标准正态 WN$(0,1)$.

```
demo.arma.est <- function(n=300, m=100){
  a <- -c(0.9, 1.4, 0.7, 0.6)
  b <- c(0.5, -0.4)
  sig <- 1.0

  ests <- matrix(0, nrow=m, ncol=7)
  for(ii in seq(m)){
    x <- arima.sim(
      model=list(ar=a, ma=b), n=n,
      rand.gen=function(n, ...) rnorm(n, 0, sig))
    ## 用 R 的 arima 函数估计模型参数。最大似然估计法。
    fit.mle <- arima(x, order=c(4,0,2),
```

```
      include.mean=FALSE)
    ests[ii,] <- c(coef(fit.mle), fit.mle$sigma2)
  }
  res <- matrix(0, nrow=4, ncol=7)
  rownames(res) <- c(" 真实参数", " 估计均值",
                     " 估计标准差",
                     " 估计根均方误差"
                     )
  colnames(res) <- c("a1", "a2", "a3", "a4", "b1", "b2", "s2")
  res[1,] <- c(a, b, sig^2)
  res[2,] <- apply(ests, 2, mean)
  res[3,] <- apply(ests, 2, sd)
  res[4,] <- sqrt(1/m*( (m-1)*res[3,]^2 + m*(res[2,]-res[1,])^2))
  cat("\n\n==== ARMA(4,2) 模型模拟: 样本量 =", n, " 重复", m, " 次估计比较\n")
  print(round(res, 4))

  invisible(res)
}
set.seed(1)
demo.arma.est(m=400)
```

```
##
##
## ==== ARMA(4,2)模型模拟: 样本量= 300  重复 400 次估计比较
##                    a1      a2      a3      a4     b1      b2     s2
## 真实参数       -0.9000 -1.4000 -0.7000 -0.6000 0.5000 -0.4000 1.0000
## 估计均值       -0.8964 -1.3941 -0.6952 -0.5950 0.4995 -0.4043 0.9822
## 估计标准差      0.0639  0.0781  0.0748  0.0550 0.0770  0.0781 0.0839
## 估计根均方误差  0.0639  0.0782  0.0748  0.0551 0.0769  0.0781 0.0857
```

将这里的最大似然估计结果与教材 P.223 结果比较可以看出，用最大似然估计可以大大改善 MA 部分的参数估计精度。

21.6 ARMA 模型的检验

在得到了 ARMA(p,q) 模型的参数估计

$$(\hat{a}_1,\dots,\hat{a}_p),\ (\hat{b}_1,\cdots,\hat{b}_q)$$

后, 对模型进行检验是十分必要的.

首先要检验模型的平稳性和合理性, 即要检验估计的参数满足(21.3).

然后对取定的初值

$$x_0 = x_{-1} = \cdots = x_{-p+1} = \hat{\varepsilon}_0 = \cdots = \hat{\varepsilon}_{-q+1} = 0$$

递推计算模型的残差

$$\hat{\varepsilon}_t = x_t - \sum_{l=1}^{p} \hat{a}_l x_{t-l} + \sum_{j=1}^{q} \hat{b}_j \hat{\varepsilon}_{t-j},\ t = 1,2,\dots \tag{21.26}$$

取 $m = O(N^{1/3})$ 和 $m > \max(p,q)$. 如果残差

$$\hat{\varepsilon}_t, \quad t = m, m+1, .., N$$

可以通过白噪声的检验, 就认为模型合适. 否则寻找其他的模型.

在实际工作中, 参数 (p,q) 是未知的. 但是根据数据的性质有时可以知道阶数的大约范围. 可以在这个范围内对每一对 (p,q) 建立 ARMA(p,q) 模型, 如果一个模型可以通过检验, 就把这个模型留做备用.

如果不能确定阶数的范围, 可以采用从 $p+q=1, p+q=2, \dots$ 开始由低阶到高阶的依次搜寻的方法. 然后在所有备用的模型中选出 $p+q$ 最小的一个模型.

如果 $p+q$ 不能惟一决定 (p,q), 可以取 p 较大的一个. 也可以在所有的备用模型中, 采用下面的 AIC 定阶方法, 最后确定一个模型.

21.7 ARMA 模型的定阶方法

和 AR 模型的定阶方法相似, 给定 ARMA(p,q) 模型的阶数 (p,q) 的一个估计 $(k,j) = (\hat{p},\hat{q})$. 无论这个估计是怎样得到的, 按前面的方法都可以估计出 ARMA(k,j) 模型的参数. 用 $\hat{\sigma}^2 = \hat{\sigma}^2(k,j)$ 表示白噪声方差 σ^2 的估计. 一般来讲, 希望 $\hat{\sigma}^2$ 的取值越小越好. 因为 $\hat{\sigma}^2$ 越小表示模型拟合的越精确. 但是，较

小的残差方差 $\hat{\sigma}^2$ 通常对应于较大的阶数 k, j. 这样, 过多追求拟合的精度, 或说过分追求较小的残差方差 $\hat{\sigma}^2$ 会导致较大的 $\hat{p}$ 和 $\hat{q}$, 从而导致较多的待估参数. 其结果会使建立的模型关于数据过于敏感, 从而降低模型的稳健性.

另一方面，参数过多的模型拟合较好但是对新数据的解释能力差。AIC 定阶准则就是为了克服参数过多的弊病而提出的。

ARMA 模型的 AIC 定阶方法和 AR 模型的 AIC 定阶方法是相同的. 如果已知 p 的上界 P_0 和 q 的上界 Q_0, 对于每一对 (k, j), $0 \le k \le P_0$, $0 \le j \le Q_0$ 计算 AIC 函数

$$\mathrm{AIC}(k, j) = \ln(\hat{\sigma}^2(k, j)) + \frac{2(k + j)}{N}. \tag{21.27}$$

$\mathrm{AIC}(k, j)$ 的最小值点 $(\hat{p}, \hat{q})$ 称为 (p, q) 的 AIC 定阶. 如果最小值不惟一, 应先取 $k + j$ 最小的, 然后取 j 最小的. 一般 AIC 定阶并不是相合的. 也就是说, 如果平稳序列的观测 $\{x_t\}$ 确实是来自一个 $\mathrm{ARMA}(p, q)$ 模型时, AIC 定阶 $(\hat{p}, \hat{q})$ 并不依概率收敛到真正的 (p, q). 但是, 和 $AR(p)$ 模型的 AIC 定阶相似, 这种不相容性并不能否定 AIC 定阶的实用性. 因为一方面实际数据并不会真正满足某一个 $\mathrm{ARMA}(p, q)$ 模型, 所以真正的阶数并不存在, 采用 ARMA 模型只是对数据的一种处理方法. 另一方面, AIC 定阶有时会高估阶数, 但是并不会高估出很多. 这比低估阶数要好, 低估了阶数往往会造成更大的误差和模型的选择不当.

将(21.27)中的 $2(k + j)/N$ 改为 $(k + j)\ln N/N$ 就得到 $\mathrm{BIC}(k, j)$ 定阶.

21.8 ARMA 模型的谱密度估计

如果得到了 $\mathrm{ARMA}(p, q)$ 模型的参数估计, 模型的检验也已经通过. 可以利用

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2}{2\pi} \frac{|1 + \sum_{j=1}^{\hat{q}} \hat{b}_j e^{ij\lambda}|^2}{|1 - \sum_{j=1}^{\hat{p}} \hat{a}_j e^{ij\lambda}|^2} \tag{21.28}$$

作为谱密度的估计.

这是因为如果观测数据确实是 $\mathrm{ARMA}(p, q)$ 序列(21.1)时, 它的谱密度是

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|1 + \sum_{j=1}^{q} b_j e^{ij\lambda}|^2}{|1 - \sum_{j=1}^{p} a_j e^{ij\lambda}|^2}.$$

不难看出，如果 $\hat{p}, \hat{q}$, $\hat{a}_1, ..., \hat{a}_p$, $\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_q$ 和 $\hat{\sigma}^2$ 分别是 p, q, $a_1, ..., a_p$, $b_1, b_2, .., b_q$ 和 σ^2 的相合估计，则 $\hat{f}(\lambda)$ 是 $f(\lambda)$ 的相合估计.

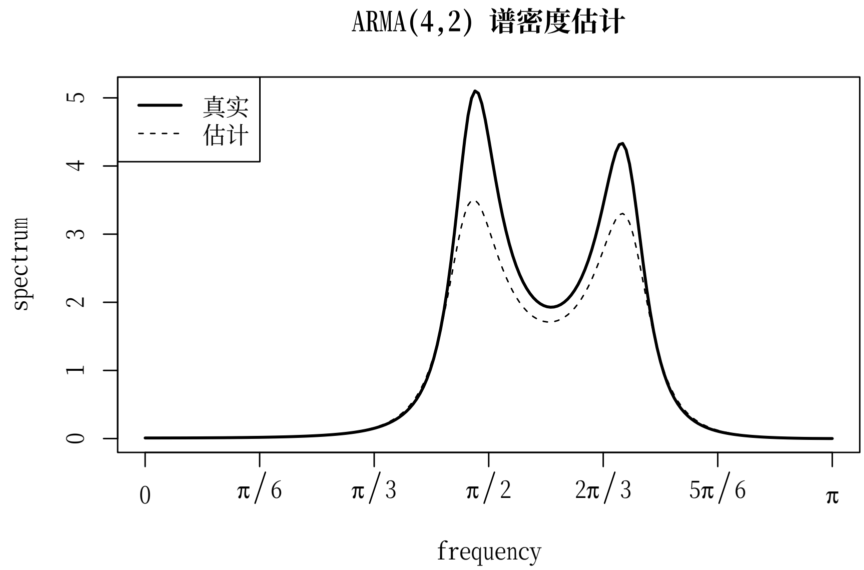
通常对谱密度进行估计只是为了解时间序列的频率特性. 后面有一章将主要介绍谱密度的估计方法.

21.8.1 谱密度估计示例

考虑 §21.5那里的 ARMA(4,2) 模型。模拟生成 300 个观测数据值，用最大似然法估计参数，然后做出真实谱密度与估计的谱密度的对比图形。

```
arma.spec <- function(a, b, sigma, ngrid=201){
  p <- length(a)
  q <- length(b)
  freqs <- seq(from=0, to=pi, length=ngrid)
  spec1 <- numeric(ngrid)
  spec2 <- numeric(ngrid)
  for(ii in seq(ngrid)){
    spec1[ii] <- 1 + sum(complex(mod=b, arg=freqs[ii]*seq(q)))
    spec2[ii] <- 1 - sum(complex(mod=a, arg=freqs[ii]*seq(p)))
  }
  spec = sigma^2 / (2*pi) * abs(spec1)^2 / abs(spec2)^2

  list(frequency=freqs, spec=spec)
}
demo.arma.spec.sim <- function(n=300, ngrids=201){
  a <- -c(0.9, 1.4, 0.7, 0.6)
  b <- c(0.5, -0.4)
  sig <- 1.0

  x <- arima.sim(
    model=list(ar=a, ma=b), n=n,
    rand.gen=function(n, ...) rnorm(n, 0, sig))
  ## 用 R 的 arima 函数估计模型参数。最大似然估计法。
  fit.mle <- arima(x, order=c(4,0,2),
```

```
    include.mean=FALSE)

  sres1 <- arma.spec(a=a, b=b, sigma=sig)
  freqs <- sres1$frequency
  spec <- sres1$spec
  plot(freqs, spec, type='l', lwd=2,
       main="ARMA(4,2) 谱密度估计",
       xlab="frequency", ylab="spectrum",
       axes=FALSE)
  axis(2)
  axis(1, at=(0:6)/6*pi,
       labels=c(0, expression(pi/6),
         expression(pi/3), expression(pi/2),
         expression(2*pi/3), expression(5*pi/6), expression(pi)))
  box()

  ## 添加估计值
  sres2 <- arma.spec(
    a=coef(fit.mle)[1:4],
    b = coef(fit.mle)[5:6],
    sigma=fit.mle$sigma2)
  lines(freqs, sres2$spec, lty=2)

  ## 图例
  legend("topleft", lty=c(1,2), lwd=c(2,1), legend=c(" 真实", " 估计"))
  invisible()
}
set.seed(1)
demo.arma.spec.sim()
```



21.9 ARIMA 模型的参数估计

对于不平稳的数据，常用的模型是 ARIMA(p,d,q)。其中 d 取为使得数据 $\{X_t\}$ 在差分后平稳的差分阶数。差分后是否平稳，可以使用各种单位根检验方法。一旦确定了 d 的值，令 $y_t=(1-\mathscr{B})^d X_t$，以 $\{y_t\}$ 为观测数据建立 ARMA(p,q) 模型即可。

21.10 季节 ARIMA 模型的参数估计

对于月度、季度频率的数据，在建模中还要考虑固定的季节变化的影响。如果数据本身平稳，可以采用如下的模型：

$$A(\mathscr{B})A_s(\mathscr{B}^s)X_t = B(\mathscr{B})B_s(\mathscr{B}^s)\varepsilon_t,$$

其中 s 是数据的频率，比如月度数据 $s=12$，季度数据 $s=4$。这样的模型可以产生系数系数的模型，能够用比较少的模型参数将季节性纳入模型。

如果数据不平稳，则数据中可能含有单位根，包括 $(1-\mathscr{B})^d$ 这样的根，和

$(1-\mathscr{B}^s)^D$ 这样的根，模型可以表示为

$$A(\mathscr{B})A_s(\mathscr{B}^s)(1-\mathscr{B})^d(1-\mathscr{B}^s)^D X_t = B(\mathscr{B})B_s(\mathscr{B}^s)\varepsilon_t.$$

将这样的模型记为 $\text{ARIMA}(p,d,q)(P,D,Q)_s$。为了建模，也只需要通过平稳性检验得到差分阶 d 和 D 的值，令

$$y_t = (1-\mathscr{B})^d(1-\mathscr{B}^s)^D X_t,$$

则 $\{y_t\}$ 服从平稳的季节 ARMA 模型，仍可以用最大似然方法估计参数。

例如，实际中常用的一个模型是 $\text{ARIMA}(0,1,1)(0,1,1)_s$ 模型，称为航空模型。其模型为

$$(1-\mathscr{B})(1-\mathscr{B}^s)X_t = (1-\theta\mathscr{B})(1-\Theta\mathscr{B}^s)\varepsilon_t.$$

Part IV

预测

Chapter 22

最佳线性预测的基本性质

对于时间序列进行统计分析的主要目的之一是解决时间序列的预测问题。任何时间序列 $\{X_t\}$ 都可以按

$$X_t = T_t + S_t + R_t$$

分解成趋势项 $\{T_t\}$、季节项 $\{S_t\}$ 和随机项 $\{R_t\}$ 的和。趋势项和季节项都可以被当做非随机的时间序列处理, 它们的预测问题往往是简单的。随机项 $\{R_t\}$ 一般是平稳序列。

于是, 时间序列预测问题的重点应当是平稳序列。本章主要讨论平稳序列的预测问题。平稳序列的方差有限, 所以我们总是假设本章中的随机变量的方差有限。由于平稳序列总是零均值平稳序列加上一个常数, 所以我们主要讨论零均值平稳序列的预测问题。

22.1 最佳线性预测

定义 22.1. 设 Y 和 $X = (X_1, \dots, X_n)^T$ 是均值为零, 方差有限的随机变量 (向量). 如果 $a \in \mathbb{R}^n$ 使得对任何的 $b \in \mathbb{R}^n$, 有

$$E(Y - a^T X)^2 \le E(Y - b^T X)^2.$$

则称 a^TX 是用 $X_1, X_2, \ldots, X_n$ 对 Y 进行预测的**最佳线性预测**, 记做 $L(Y|X)$ 或 $\hat{Y}$. 于是

$$\hat{Y} = L(Y|X) = a^TX. \tag{22.1}$$

定义 22.2. 如果 $EY = b, EX = \mu$, 定义

$$L(Y|X) = L(Y - b|X - \mu) + b, \tag{22.2}$$

并称 $L(Y|X)$ 是用 $X_1, X_2, \cdots, X_n$ 对 Y 进行预测时的最佳线性预测.

以下总设随机变量均值为零。用 $\Gamma = E(XX^T)$ 表示 X 的协方差阵。用 $\Sigma_{XY} = E(XY)$ 表示 X 和 Y 的协方差向量。

22.1.1 性质 1

如果 $a \in \mathbb{R}^n$, 使得

$$\Gamma a = \Sigma_{XY}, \tag{22.3}$$

则

$$L(Y|X) = a^TX,$$

并且有

$$E(Y - L(Y|X))^2 = EY^2 - E[L(Y|X)]^2 = EY^2 - a^T\Gamma a. \tag{22.4}$$

如果 Γ 和 Σ_{XY} 已知, 以 a 为未知数的线性方程组(22.3)被称为**预测方程**.

证明: 对任何 $b \in \mathbb{R}^n$,

$$\begin{aligned}
&E(Y - b^TX)^2 \\
=&E[Y - a^TX + (a^T - b^T)X]^2 \\
=&E(Y - a^TX)^2 + E[(a^T - b^T)X]^2 + 2E[(a^T - b^T)X(Y - a^TX)] \\
=&E(Y - a^TX)^2 + E[(a^T - b^T)X]^2 + 2(a^T - b^T)[E(XY) - E(XX^T)a] \\
=&E(Y - a^TX)^2 + E[(a^T - b^T)X]^2 \\
\geq&E(Y - a^TX)^2.
\end{aligned}$$

所以, a^TX 是 Y 的最佳线性预测. 利用(22.3)得到

$$
\begin{aligned}
&E[Y-L(Y|X)]^2=E(Y-a^TX)^2\\
=&EY^2+a^TE(XX^T)a-2a^TE(XY)\\
=&EY^2+a^T\Gamma a-2a^T\Gamma a\\
=&EY^2-a^T\Gamma a.
\end{aligned}
$$

注意：a 是预测方程的解等价于

$$E((Y-a^TX)X)=\Sigma_{XY}-\Gamma a=0$$

即 $Y-a^TX$ 与 X 正交。(注意 a^TX 是标量)

性质说明 $Y-a^TX$ 与 X 正交则 $a^TX=L(Y|X)$。

22.1.2 性质 2

- (1) 如果 $\Gamma=E(XX^T)$ 可逆, 则 $a=\Gamma^{-1}E(XY)$ 使得 $L(Y|X)=a^TX$。
- (2) 预测方程 $\Gamma a=E(XY)$ 总有解.
- (3) 如果 $\det(\Gamma)=0$，取正交矩阵 A 使得

  $$A\Gamma A^T=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_r,0,\dots,0),\quad \lambda_j>0,j=1,\dots,r.$$

  定义 $Z=AX=(Z_1,Z_2,\dots,Z_n)^T=(Z_1,Z_2,\dots,Z_r,0,\dots,0)^T$ 和 $\xi=(Z_1,Z_2,\cdots,Z_r)^T$, 则 $E(\xi\xi^T)$ 正定, 并且当取

  $$\alpha=[E(\xi\xi^T)]^{-1}E(\xi Y) \tag{22.5}$$

  时, $L(Y|X)=L(Y|\xi)=\alpha^T\xi$.

性质 2 的第二条说明最佳线性预测总存在，而且总可以由预测方程的解表示。

性质 2 的第三条说明当第一条不成立时，$L(Y|X)$ 可以通过 X 的基表示。

证明: 仅需证明 $\det(\Gamma)=0$ 时第三和第二条成立。

$$
\begin{aligned}
E(ZZ^T)=&E(AXX^TA^T)=A\Gamma A^T\\
=&\mathrm{diag}(\lambda_1,\lambda_2,\dots,\lambda_r,0,\dots,0)
\end{aligned}
$$

故 $Z_{r+1}=\cdots=Z_n=0$。且 $E(\xi\xi^T)=\mathrm{diag}(\lambda_1,\ldots,\lambda_r)$ 是正定阵。当 α 按(22.5)定义时, 有

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \lambda_r \end{pmatrix} \alpha = E(\xi Y).$$

注意 $A\Gamma A^T$ 与 $\lambda_1,\ldots,\lambda_r$ 的关系，可以导出

$$\begin{aligned} A\Gamma A^T \begin{pmatrix} \alpha \\ 0 \end{pmatrix} &= \mathrm{diag}(\lambda_1,\ldots,\lambda_r,0,\ldots,0) \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} E(\xi Y) \\ 0 \end{pmatrix} = E\left(\begin{pmatrix} \xi \\ 0 \end{pmatrix} Y \right) \\ &= E(ZY)=E(AXY)=AE(XY)=A\Sigma_{XY} \end{aligned}$$

两边同乘以 A^T, 记 $a=A^T\begin{pmatrix} \alpha \\ 0 \end{pmatrix}$ 有

$$\Gamma a=\Sigma_{XY}$$

由性质 1 知

$$L(Y|X)=a^TX=(\alpha^T,0,\ldots,0)AX=\alpha^T\xi$$

这在证明了第三条的同时也证明了第二条。

22.1.3 性质 3

尽管 a 由 $\Gamma a=E(XY)$ 决定时可以不惟一, 但 $L(Y|X)$ 总是 (a.s.) 惟一的.

证明: 预测方程总有解，设 a 为预测方程的一个解，则由性质 1 对 $\forall b\in\mathbb{R}^n$ 有

$$E(Y-b^TX)^2=E(Y-a^TX)^2+E((a-b)^TX)^2$$

若还有 b 使得 $L(Y|X)=b^TX$ 则 $E(Y-b^TX)^2=E(Y-a^TX)^2$，于是 $E((a-b)^TX)^2=0$, 即 $b^TX=a^TX$，a.s.

22.1.4 性质 4

- (1) 如果 $E(XY)=0$ 则 $L(Y|X)=0$.
- (2) 如果 $Y=\sum_{j=1}^n b_jX_j$ 则 $L(Y|X)=Y$.

这是线性预测的两个极端：

- 因变量和自变量不相关时线性预测无效；
- 因变量为自变量线性组合时可以完全线性预测。

证明：

(1) $\forall b$

$$\begin{aligned}E(Y-b^TX)^2 =& EY^2 + b^T\Gamma b - 2b^TE(XY)\\ =& EY^2 + b^T\Gamma b \geq EY^2 = E(Y-0)^2\end{aligned}$$

所以 $L(Y|X)=0$。

(2) 这时 $Y=b^TX$, $E(Y-bX)^2=0$，所以 $L(Y|X)=b^TX=Y$。

22.1.5 性质 5

设 $Y_1, Y_2, \cdots, Y_m$ 是随机变量, b_j 是常数. 如果 $Y=\sum_{i=1}^m b_iY_i$, 则

$$L(Y|X)=\sum_{j=1}^m b_jL(Y_j|X).$$

性质 5 说明求最佳线性预测的运算 $L(\cdot|X)$ 是一种线性运算.

证明: 设 a_i 为 $\Gamma a_i = E(XY_i)$ 的解 $(i=1,2,\dots,m)$，则 $L(Y_i|X)=a_i^TX$。取 $a=\sum_{i=1}^m b_ia_i$，则

$$\begin{aligned}\Gamma a =& \Gamma\left(\sum_{i=1}^m b_ia_i\right) = \sum_{i=1}^m b_i(\Gamma a_i) = \sum_{i=1}^m b_iE(XY_i)\\ =& E\left(X\sum_{i=1}^m b_iY_i\right) = E(XY)\end{aligned}$$

由性质 1 即知

$$L(Y|X) = a^TX = \sum_{i=1}^m b_ia_i^TX = \sum_{i=1}^m b_iL(Y_i|X)$$

22.1.6 性质 6

设 $X=(X_1, X_2, \cdots, X_n)^T$, $Z=(Z_1, Z_2, \cdots, Z_m)^T$. 如果 $E(XZ^T)=0$(X 与 Z 不相关), 则有

$$L(Y|X,Z)=L(Y|X)+L(Y|Z).$$

证明: 记 $\xi = \begin{pmatrix} X \\ Z \end{pmatrix}$，记 $\Sigma_{XX} = E(XX^T), \Sigma_{ZZ} = E(ZZ^T)$, 则

$$\Sigma \triangleq E(\xi\xi^T) = \begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{ZZ} \end{pmatrix}$$

设 a, b 使得 $\Sigma_{XX}a = E(XY)$, $\Sigma_{ZZ}b = E(ZY)$, 则 $L(Y|X) = a^T X$，$L(Y|Z) = b^T Z$, 取 $c = \begin{pmatrix} a \\ b \end{pmatrix}$，则

$$\begin{aligned} \Sigma c = & \begin{pmatrix} \Sigma_{XX}a \\ \Sigma_{ZZ}b \end{pmatrix} = \begin{pmatrix} E(XY) \\ E(ZY) \end{pmatrix} \\ = & E\left(\begin{pmatrix} X \\ Z \end{pmatrix} Y \right) = E(\xi Y) \end{aligned}$$

由性质 1

$$\begin{aligned} L(Y|X,Z) = & L(Y|\xi) = c^T \xi \\ = & a^T X + b^T Z = L(Y|X) + L(Y|Z) \end{aligned}$$

22.1.7 性质 7

设 $\tilde{Y} = b^T X$ 是 X 的线性组合, 则 $\tilde{Y} = L(Y|X)$ 的充分必要条件是

$$E(X_j(Y - \tilde{Y})) = 0, \ 1 \le j \le n. \tag{1.8}$$

即

$$E(X(Y - b^T X)) = 0$$

$L(Y|X)$ 可以看成 Y 在 X 张成的空间上的投影，此性质即投影应满足的性质。注意

$$E(X(Y - b^T X)) = E(XY) - \Gamma b.$$

即残差与自变量正交等价于系数 b 满足预测方程。

证明:

必要性: 设 $\tilde{Y}$ 为 $L(Y|X)$，由性质 2 知存在 a 满足预测方程，由性质 1 和性质 3 知 $L(Y|X) = a^T X = b^T X$。两边右乘以 X^T 取期望得

$$a^T \Gamma = b^T \Gamma$$

注意 $\Gamma a = E(XY)$ 所以由上式得 $\Gamma b = E(XY)$, 即条件成立。

充分性: 条件成立时 b 是预测方程的解，由性质 1 即知 $\tilde{Y} = b^T X$ 是最佳线性预测。

22.1.8 性质 8

如果

$$\begin{aligned}\hat{Y} =& L(Y|X_1, X_2, \cdots, X_n),\\ \tilde{Y} =& L(Y|X_1, X_2, \cdots, X_{n-1}),\end{aligned}$$

则有

$$L(\hat{Y}|X_1, X_2, \cdots, X_{n-1}) = \tilde{Y},$$

并且有

$$E(Y-\hat{Y})^2 \le E(Y-\tilde{Y})^2. \tag{22.6}$$

(22.6)表明在方差最小的意义下, $\hat{Y}$ 比 $\tilde{Y}$ 要好.
这是由于 $X_1, X_2, \cdots, X_n$ 中包含的信息比 $X_1, X_2, \cdots, X_{n-1}$ 中包含的信息多的原因.

证明 $Y_0 \triangleq L(\hat{Y}|X_1, X_2, \dots, X_{n-1})$ 是 $X_1, X_2, \dots, X_{n-1}$ 的线性组合, 利用 $Y - \hat{Y}$, $\hat{Y} - Y_0$ 都和 $X_1, \dots, X_{n-1}$ 正交, 得到

$$Y - Y_0 = (Y - \hat{Y}) + (\hat{Y} - Y_0)$$

和 $X_1, \dots, X_{n-1}$ 正交. 利用性质 7 即知 $Y_0 = L(Y|X_1, \dots, X_{n-1}) = \tilde{Y}$.

$\hat{Y}$ 是 $X_1, \dots, X_n$ 对 Y 的最佳线性预测而 $\tilde{Y}$ 是 $X_1, \dots, X_n$ 的一个线性组合所以有(22.6)成立。

这个性质实际是投影的性质。

22.1.9 性质 9(非零均值的最佳线性预测的意义)

如果 $EY = b$, $EX = \mu$, 按定义 $L(Y|X) = b + L(Y - b|X - \mu)$。事实上对任何 $c_0 \in \mathbb{R}$, $c \in \mathbb{R}^n$,

$$E[Y - L(Y|X)]^2 \le E[Y - (c_0 + c^T X)]^2. \tag{22.7}$$

证明设 $L(Y|X) = b + a^T(X-\mu)$, 则

$$
\begin{aligned}
& E\left\{Y - c_0 - c^T X\right\}^2 \\
=& E\left\{Y - b - a^T(X-\mu) + b + a^T(X-\mu) \right. \\
& \left. - [c_0 + c^T\mu + c^T(X-\mu)]\right\}^2 \\
=& E\left\{[Y - b - a^T(X-\mu)] + (b - c_0 - c^T\mu) + (a-c)^T(X-\mu)\right\}^2 \\
=& E[Y - b - a^T(X-\mu)]^2 + (b - c_0 - c^T\mu)^2 + (a-c)^T\Gamma(a-c) \\
\geq& E[Y - b - a^T(X-\mu)]^2 = E[Y - L(Y|X)]^2
\end{aligned}
$$

22.1.10 性质 10

设 X 和 Z 分别是 m 和 n 维向量, 如果有实矩阵 A, B 使得 $X = AZ$, $Z = BX$, 则 $L(Y|X) = L(Y|Z)$.

如果 X 和 Z 能互相线性表示则利用其预报 Y 能达到的下界是相同的，预报是一致的。

证明为习题。

22.1.11 性质总结

性质 1、2、3、7 说明 $L(Y|X)$ 存在唯一，且

$$
\begin{aligned}
& a^T X = L(Y|X) \\
\Longleftrightarrow & \Sigma a = \Sigma_{XY} \\
\Longleftrightarrow & E[X_j(Y - a^T X)] = 0, j = 1, 2, \dots, n.
\end{aligned}
$$

性质 1 还给出了勾股定理：

$$
E(Y^2) = E[L(Y|X)]^2 + E[Y - L(Y|X)]^2.
$$

性质 5 说明 $L(\cdot|X)$ 是定义在 L^2 空间的线性算子。

性质 4 说明如果 Y 与 X 不相关则预报为 EY，如果 Y 已经是 X 的线性组合则预报误差为 0。

性质 6 说明如果两组自变量不相关，则预报可以分别预报后求和；性质 10 说明如果两组自变量线性等价（可以互相线性表示），则预报相同。

性质 8 说明增加自变量可以使得预报均方误差减少。

22.1.12 例子

考虑 §13.7.1的 ARMA(4,2) 序列的预测。设已知 ARMA(4,2) 的参数。

$$\begin{aligned} X_t =& -0.9X_{t-1} - 1.4X_{t-2} - 0.7X_{t-3} - 0.6X_{t-4} \\ & + \varepsilon_t + 0.5\varepsilon_{t-1} - 0.4\varepsilon_{t-2},\ \varepsilon_t \sim \text{WN}(0,1), \end{aligned}$$

模拟生成观测到 $x_1, \dots, x_{21}$，用前 14 个观测值对最后的 7 个点作预测。

使用预测方程直接求解系数，Γ 使用理论值。预测方程中 Σ_{XY} 当 $Y = X_{n+k}$ 时为

$$g_k = E(X_n X_{n+k}) = (\gamma_{n+k-1}, \gamma_{n+k-2}, \dots, \gamma_k)^T$$

最佳线性预测为

$$\hat{X}_{n+k} \triangleq L(X_{n+k}|X_n) = (\Gamma_n^{-1} g_k)^T X_n = g_k^T \Gamma_n^{-1} X_n$$

预测方差为

$$\sigma^2(k) = \gamma_0 - (\Gamma_n^{-1} g_k)^T \Gamma_n (\Gamma_n^{-1} g_k) = \gamma_0 - g_k^T \Gamma_n^{-1} g_k$$

设 $\{X_t\}$ 为正态平稳列，则 $X_{n+k} - \hat{X}_{n+k}$ 作为有限线性组合也是正态分布的。$X_{n+k} - \hat{X}_{n+k} \sim \text{N}(0, \sigma^2(k))$, 可以构造 X_{n+k} 的置信区间 (预测区间)：

$$\Pr(|X_{n+k} - \hat{X}_{n+k}|/\sigma(k) \leq 1.96) = 0.95, \quad k = 1, 2, \dots, m$$

见演示。

对真实数据需要用 $x_1, x_2, \dots, x_N$ 估计 $\hat{\gamma}_k$, 然后用令 $x_n = (x_{N-n+1}, x_{N-n+2}, \dots, x_N)^T$, 用 x_n 预报 X_{N+k}, $k = 1, 2, \dots$。数据先减去均值再估计并预测，预测值要把均值加回去。

下面是从 ARMA 模型参数计算 Wold 系数的 R 函数，以及通过 Wold 系数计算理论自协方差函数的 R 函数。

```
## Wold coefficients for the ARMA model
arma.Wold <- function(n, a, b=numeric(0)){
  p <- length(a)
  q <- length(b)
  arev <- rev(a)
```

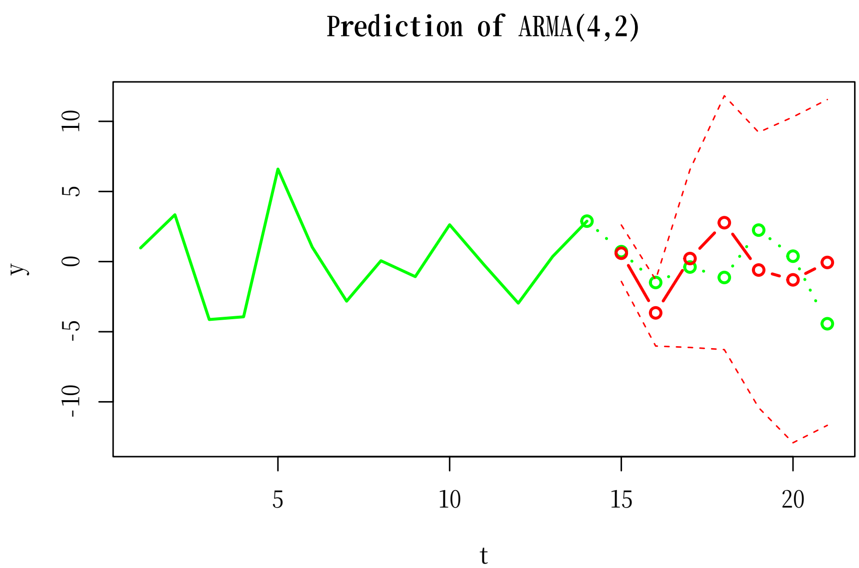
```
  psi <- numeric(n)
  psi[1] <- 1
  for(j in seq(n-1)){
    if(j <= q) bj=b[j]
    else bj=0
    psis <- psi[max(1, j+1-p):j]
    np <- length(psis)
    if(np < p) psis <- c(rep(0,p-np), psis)
    psi[j+1] <- bj + sum(arev * psis)
  }

  psi
}

## Calculate theoretical autocovariance function
## of ARMA model using Wold expansion
arma.gamma.by.Wold <- function(n, a, b=numeric(0), sigma=1){
  nn <- n + 100
  psi <- arma.Wold(nn, a, b)
  gam <- numeric(n)
  for(ii in seq(0, n-1)){
    gam[ii+1] <- sum(psi[1:(nn-ii)] * psi[(ii+1):nn])
  }
  gam <- (sigma^2) * gam
  gam
}
arma.gamma <- arma.gamma.by.Wold
```

下面的 R 程序模拟生成 21 个样本点，用开头的 14 个观测预测最后的 7 个观测（多步预测）。使用模型的理论自协方差函数求解预测系数。

在结果图形中，绿色线是预测用到的观测，红色点为预测值，绿色点为实际值，上下两条红色虚线是逐点的预测区间。

```r
## 对模拟 ARMA(4,2) 数据，
## 用理论协方差解预测方程进行 Y-W 预报
## 这应该和用 Levinson 公式得到的预测系数是一致的。
demo.pred.arma42 <- function(n=21){
  a <- c(-0.9, -1.4, -0.7, -0.6)
  b <- c(0.5, -0.4)
  ng <- 21

  x <- arima.sim(model=list(ar=a, ma=b), n=n)
  gams <- arma.gamma(ng, a, b, sigma=1)

  n.use <- 14  ## use n.use points in prediction
  m.pred <- 7  ## predit m.pred steps
  n.start <- n - m.pred - n.use + 1  ## which x are affected
  ## predict usging true ACV
  Ga <- matrix(0, nrow=n.use, ncol=n.use)
  for(ii in seq(n.use)) for(jj in seq(n.use)) {
    ind <- abs(ii-jj)+1
    Ga[ii,jj] <- gams[ind]
  }
  Gar <- solve(Ga)

  x.use <- x[(n.start+n.use-1):n.start] # reverse time order
  y.pred <- x
  y.pred[] <- NA
  errs <- numeric(m.pred)
  for(k in seq(m.pred)){
    g <- gams[(k+1):(k+n.use)]
    a <- Gar %*% g
    y.pred[n.start+n.use-1+k] <- sum(a * x.use)
    errs[k] <- gams[1] - g %*% Gar %*% g
  }
  lb <- y.pred[(n-m.pred+1):n]-1.96*errs
```

```
  ub <- y.pred[(n-m.pred+1):n]+1.96*errs
  yl <- range(c(lb,ub,x,y.pred), na.rm=T)
  plot(1:n, x, type="n",
       main="Prediction of ARMA(4,2)",
       xlab="t", ylab="y",
       ylim=yl)
  if(n.start > 1){
    lines(1:(n.start), x[1:n.start],
          lwd=2)  ## unused
  }
  lines(n.start:(n.start+n.use-1), x[n.start:(n.start+n.use-1)],
        col="green", lwd=2) ## used for predition
  lines((n-m.pred):n, x[(n-m.pred):n],
        type="b", col="green", lty=3, lwd=2)  ## true values
  lines((n-m.pred+1):n, y.pred[(n-m.pred+1):n],
        type="b", col="red", lwd=2)  ## predictons
  lines((n-m.pred+1):n, ub,
        type="l", col="red", lty=2)  ## upper bounds
  lines((n-m.pred+1):n, lb,
        type="l", col="red", lty=2)  ## lower bounds

  invisible()
}
set.seed(1)
demo.pred.arma42()
```



22.2 Hilbert 空间中的投影

下面说明最佳线性预测实际上是 Hilbert 空间中的投影.

用 L^2 表示全体方差有限的随机变量构成的 Hilbert 空间 (参见第5章). 设 H 是 L^2 的闭子空间, Y 属于 L^2. 可以证明 H 中存在惟一的 $\hat{Y}$ 使得

$$E(Y-\hat{Y})^2 = \inf_{\xi\in H} E(Y-\xi)^2 \tag{22.8}$$

定义 22.3. 如果 H 是 L^2 的闭子空间, $Y \in L^2$, $\hat{Y} \in H$ 使得(22.8)成立, 则称 $\hat{Y}$ 是 Y 在 H 上的投影. 记做 $P_H(Y)$, 并且称 P_H 是投影算子.

定义 22.4. 设 $Y \in L^2$, 如果对 H 中的任何 ξ, $E(Y\xi)=0$, 则称 Y 垂直于 H, 记作 $Y \perp H$.

22.2.1 投影存在唯一的证明

取 $Y_n \in H$ 使

$$d = \inf_{\xi\in H} E(Y-\xi)^2 = \lim_{n\to\infty} E(Y-Y_n)^2$$

则 $(Y_n+Y_m)/2 \in H$, 并且当 $n,m\to\infty$

$$
\begin{aligned}
&E(Y_n-Y_m)^2 && (22.9)\\
=&E[(Y_n-Y)-(Y_m-Y)]^2 && (22.10)\\
&+E[(Y_n-Y)+(Y_m-Y)]^2 && (22.11)\\
&-E[(Y_n+Y_m)-2Y]^2 && (22.12)\\
=&2E(Y_n-Y)^2+2E(Y_m-Y)^2-4E[(Y_n+Y_m)/2-Y]^2 && (22.13)\\
\le&2E(Y_n-Y)^2+2E(Y_m-Y)^2-4d && (22.14)\\
\to&2d+2d-4d=0. && (22.15)
\end{aligned}
$$

于是, $\{Y_n\}$ 是 H 中的基本列, 从而有 $\hat{Y}\in H$ 使得 Y_n 均方收敛到 $\hat{Y}$. 由内积的连续性知道

$$E(Y-\hat{Y})^2=\lim_{n\to\infty}E(Y-Y_n)^2=d.$$

于是, $\hat{Y}$ 满足(22.8).

如果又有 $\hat{\xi}\in H$ 也使得(22.8)成立, 仿照(22.15)的推导得到

$$
\begin{aligned}
&E(\hat{Y}-\hat{\xi})^2\\
=&E[(\hat{Y}-Y)-(\hat{\xi}-Y)]^2\\
&+E[(\hat{Y}-Y)+(\hat{\xi}-Y)]^2-E[(\hat{Y}+\hat{\xi})-2Y]^2\\
=&2E(\hat{Y}-Y)^2+2E(\hat{\xi}-Y)^2-4E[(\hat{Y}+\hat{\xi})/2-Y]^2\\
\le&2d+2d-4d=0.
\end{aligned}
$$

所以 $\hat{\xi}=\hat{Y}$, a.s.

22.2.2 投影的垂直性 (正交性)

定理 22.1. *设 $Y\in L^2$, $\hat{Y}\in H$, 则 $\hat{Y}=P_H(Y)$ 的充分必要条件是 $(Y-\hat{Y})\perp H$.*

最佳线性预测性质 7 是定理22.1的特例。

证明先证必要性.

设 $\hat{Y}=P_H(Y)$. 对 $\forall\xi\in H$, 我们证明

$$a \triangleq E[(Y-\hat{Y})\xi]=0.$$

无妨设 $E\xi^2 = 1$, 这时

$$
\begin{aligned}
d \triangleq & E(Y - \hat{Y})^2 \le E(Y - \hat{Y} - a\xi)^2 \\
= & E(Y - \hat{Y})^2 + E(a\xi)^2 - 2aE[(Y - \hat{Y})\xi] \\
= & d + a^2 - 2a^2 = d - a^2
\end{aligned}
$$

由此得到 $a = 0$.

来证明充分性。若 $\hat{Y} \in H$ 使 $Y - \hat{Y} \perp H$，则对 $\forall \xi \in H$ 有

$$
\begin{aligned}
& E(Y - \xi)^2 = E(Y - \hat{Y} + \hat{Y} - \xi)^2 \\
= & E(Y - \hat{Y})^2 + E(\hat{Y} - \xi)^2 + 2E[(Y - \hat{Y})(\hat{Y} - \xi)] \\
= & E(Y - \hat{Y})^2 + E(\hat{Y} - \xi)^2 \\
\ge & E(Y - \hat{Y})^2
\end{aligned}
$$

即 $\hat{Y} = P_H(Y)$。

22.2.3 最佳线性预报与投影的等价性

用 $L^2(X)$ 表示 $X = (X_1, X_2, \cdots, X_n)^T$ 的元素和常数 1 生成的 Hilbert 空间. 它是 $X_1, X_2, .., X_n$ 和常数 1 的线性组合的全体 (参见 (何书元, 2003) 第 1 章习题 6.5). 设 $\mu = (\mu_1, \mu_2, \cdots, \mu_n)^T = EX$. 对任何方差有限的随机变量 Y, 设 $EY = b$, $\hat{Y} = L(Y|X)$ 由(22.2)式定义. 则有

$$Y - \hat{Y} = (Y - b) - L(Y - b|X - \mu).$$

利用性质 7 知道

$$
\begin{aligned}
& E[1 \cdot (Y - \hat{Y})] = E(Y - b) - EL(Y - b|X - \mu) = 0, \\
& E[X_i(Y - \hat{Y})] = E[(X_i - \mu_i)(Y - \hat{Y})] + \mu_i E(Y - \hat{Y}) = 0.
\end{aligned}
$$

即得到 $(Y - \hat{Y})$ 垂直于 $H \triangleq L^2(X)$. 由定理22.1知道

$$L(Y|X) = P_H(Y).$$

基于上述原因, 当 H 是 $\{X_j : j \in T\}$ 和常数 1 生成的 Hilbert 空间, 我们也用

$$L(Y|1, X_j, j \in T) \quad 或 \quad L(Y|H)$$

表示 $P_H(Y)$, 这里 T 是一个可列的指标集.

下面记 $\|\xi\| = \sqrt{E\xi^2}$, $\forall \xi \in L^2$。$\|\xi\|$ 是 ξ 的长度。

22.2.4 投影算子的性质

定理 22.2. 设 H, M 是 L^2 的闭子空间, $X,Y \in L^2$, a,b 是常数.

- *(1)* $L(aX+bY|H) = aL(X|H) + bL(Y|H)$. *(*对应于最佳线性预测性质 *5)*
- *(2)* $\|Y\|^2 = \|L(Y|H)\|^2 + \|Y - L(Y|H)\|^2$. *(*对应于最佳线性预测性质 *1* 的 *(1.6)* 式*)*
- *(3)* $\|L(Y|H)\| \leq \|Y\|$.
- *(4)* $Y \in H$ 的充分必要条件是 $L(Y|H) = Y$. *(*对应于最佳线性预测性质 *4* 第 *(2)* 条*)*
- *(5)* Y 垂直于 H 的充分必要条件是 $L(Y|H) = 0$, *(*对应于最佳线性预测性质 *4* 第 *(1)* 条*)*
- *(6)* 如果 H 是 M 的子空间, 则 $P_H P_M = P_M P_H = P_H$, 并且对 $Y \in L^2$,

$$\|Y - L(Y|M)\| \leq \|Y - L(Y|H)\|.$$

*(*对应于最佳线性预测性质 *8)*

证明

(1) 设 $Z = aL(X|H) + bL(Y|H)$ 则 $Z \in H$。由

$$(aX + bY) - Z = a[X - L(X|H)] + b[Y - L(Y|H)]$$

看出 $(aX + bY) - Z \perp H$。所以

$$L(aX + bY|H) = Z = aL(X|H) + bL(Y|H)$$

这说明投影是线性算子。

(2) 由于 $L(Y|H) \in H$ 而 $Y - L(Y|H) \perp H$ 所以

$$\begin{aligned}
\|Y\|^2 =& E[(Y - L(Y|H)) + L(Y|H)]^2 \\
=& E[Y - L(Y|H)]^2 + E[L(Y|H)]^2 \\
& + 2E[(Y - L(Y|H))L(Y|H)] \\
=& E[Y - L(Y|H)]^2 + E[L(Y|H)]^2 \\
=& \|Y - L(Y|H)\|^2 + \|L(Y|H)\|^2
\end{aligned}$$

(3) 由 (2) 直接得到。

(4) 必要性：$Y \in H$ 时取 $L(Y|H) = Y$ 可得均方误差为 0。

充分性: 若 $L(Y|H) = Y$ 则由于投影必须属于 H 所以 $Y \in H$。

(5) 必要性: 若 $Y \perp H$ 则 $0 \in H, Y - 0 \perp H$ 所以 $L(Y|H) = 0$。

充分性: 若 $L(Y|H) = 0$ 则由 $Y - L(Y|H) \perp H$ 知 $Y \perp H$。

(6) $\forall Y \in L^2$，设 $\xi = P_M(Y), \eta = P_H(Y)$，来证 $P_H(\xi) = \eta$。事实上，$\eta \in H$ 且

$$\xi - \eta = (Y - \eta) - (Y - \xi)$$

其中 $Y - \eta$ 和 $Y - \xi$ 都与 H 垂直，所以 $P_H(\xi) = \eta$，即 $P_H P_M = P_H$。另外, $H \subseteq M$ 所以 $\eta \in M$，$P_M(\eta) = \eta$，即 $P_M P_H = P_H$。

由 $P_H(Y) \in M$ 和 $P_M(Y)$ 的定义马上可得

$$\|Y - P_H(Y)\|^2 \geq \|Y - P_M(Y)\|^2$$

22.3 最佳预测

最佳线性预测只用了自变量的线性函数而未考虑其他函数。设

$$M = \bar{\text{sp}}\{g(X):\ Eg^2(X) < \infty, g(\cdot)\text{是可测函数}\} \tag{22.16}$$

考虑用 M 中的元素逼近 Y。

定义 22.5. 设 M 由(22.16)定义. 用 $X = (X_1, X_2, \dots, X_n)^T$ 对 Y 进行预测时, 称

$$L(Y|M) \triangleq P_M(Y)$$

为 Y 的最佳预测.

最佳预测 $L(Y|M)$ 实际上是概率论中的条件数学期望 $E(Y|X)$.

$L^2(X)$ 是 M 的子空间，由定理22.2的 (6) 得

$$\|Y - L(Y|M)\| \leq \|Y - L(Y|X)\|$$

在预测均方误差最小的意义下最佳预测比最佳线性预测好。

但是由于 M 要比 $L^2(X)$ 复杂很多, 实际计算最佳预测往往比计算最佳线性预测困难得多.

对于正态序列来讲, 最佳预测和最佳线性预测是一致的.

定理 22.3. 如果 $(X_1, X_2, \dots, X_n, Y)^T$ 服从联合正态分布 $N(\mu, \Sigma)$, M 由(22.16)定义, 则

$$L(Y|M) = L(Y|X_1, X_2, \cdots, X_n). \tag{22.17}$$

证明设 $\hat{Y} = L(Y|X_1, X_2, \dots, X_n)$, 则 $(Y - \hat{Y})$ 与 X 正交. 由于 $E(Y - \hat{Y}) = 0$, 所以 $Y - \hat{Y}$ 与 X 不相关. 由正态分布的性质知道, $Y - \hat{Y}$ 与 X 独立, 从而和 M 中的任何随机变量独立. 对任何 $\xi \in M$, $E[\xi(Y - \hat{Y})] = (E\xi)E(Y - \hat{Y}) = 0$, 即 $Y - \hat{Y}$ 垂直于 M. 从 $\hat{Y} \in M$ 和定理22.1知道(22.17)成立.

22.3.1 例子

例 22.1. 设随机变量 ε, η 独立, 都服从标准正态分布 $N(0,1)$, 则 $E\eta^4 = 3$. 取 $X = \eta, Y = (3\varepsilon^2 - \eta^2)\eta$.

则 $EX = EY = 0$, $E(XY) = E(3\varepsilon^2\eta^2 - \eta^4) = 0$. 从而 $L(Y|X) = 0$. 计算

$$\begin{aligned} E(Y|X) =& E(3\varepsilon^2\eta - \eta^3|\eta) \\ =& 3\eta - \eta^3 = 3X - X^3 \end{aligned}$$

容易验证 $Y - (3\eta - \eta^3) = 3(\varepsilon^2 - 1)\eta$ 垂直于

$$M = \bar{sp}\left\{g(X) : Eg^2(X) < \infty, g(x) \text{ 是可测函数}\right\}.$$

于是, 从 $3\eta - \eta^3 \in M$ 知道: $L(Y|M) = 3X - X^3$.

22.4 附录：补充

22.4.1 正交直和投影

最佳线性预测的性质大都可以看成投影性质。其中性质 6 可以扩充为：设 M, N 是 L^2 的两个子 Hilbert 空间，对 $\forall \xi \in M$, $\forall \eta \in N$，有

$$\langle \xi, \eta \rangle = 0$$

称 M 与 N 正交。定义

$$M \oplus N = \{\xi + \eta : \xi \in M, \eta \in N\}$$

则

$$L(Y|M \oplus N) = L(Y|M) + L(Y|N)$$

用投影的残差正交性可以证明此性质。

Chapter 23

非决定性平稳序列及其 Wold 表示

23.1 非决定性平稳序列

对平稳序列, 考虑用所有的历史 $\{X_t,\ t \le n\}$ 对 X_{n+1} 进行最佳线性预测. 当预测误差是零时, X_{n+1} 的信息完全含在历史资料中. 这样的平稳序列被称为**决定性的**. §9.5中 Γ_{n+1} 不满秩造成 X_t 可以被 $X_{t-1}, \dots, X_{t-n}$ 完全线性预测，是决定性平稳列的特例。

最小序列: 用 $\{X_s : s \neq t\}$ 预报 X_t 误差不为零。决定性序列不是最小序列。

实际问题中, 决定性平稳序列描述事物的发展没有新的信息出现.

如果用 $\{X_t,\ t \le n\}$ 对 X_{n+1} 做线性预测的误差不是零, 说明 X_{n+1} 的信息不能由历史资料的线性组合及其极限完全确定, 我们称这种时间序列是**非决定性的**.

非决定性平稳序列描述事物的发展总伴随新的信息出现.

最小序列一定是非决定性的。

平稳序列的 Wold 定理表示告诉我们, 非决定性平稳序列总是可以分解成白噪声的单边滑动和加上一个决定性平稳序列.

从应用的角度讲, 非决定性平稳序列总是白噪声的单边滑动和加上一个离散谱序列.

23.1.1 最佳线性预测均方误差的极限

设 $\{X_n : n \in \mathbb{Z}\}$ 是零均值平稳序列. 记

$$X_{n,m} = (X_n, X_{n-1}, \cdots, X_{n-m+1})^T,$$

这里 n 表示向量的第一个脚标, m 表示向量的维数.

定义

$$\hat{X}_{n+1,m} = L(X_{n+1}|X_{n,m}).$$

从最佳线性预测的性质 8 知道 $\sigma_{1,m}^2 = E(X_{n+1} - \hat{X}_{n+1,m})^2$ 是 m 的单调减函数, 于是定义

$$\sigma_1^2 \triangleq \lim_{m\to\infty} \sigma_{1,m}^2 < \infty.$$

定理 23.1. $\sigma_1^2 \triangleq \lim_{m\to\infty} \sigma_{1,m}^2$ 与 n 无关.

证明: 设 $a = (a_1, a_2, \dots, a_m)^T$ 是预测方程(22.3)的解, 则 a 和 n 无关. 由于

$$Y_n \triangleq X_{n+1} - \sum_{j=1}^{m} a_j X_{n+1-j} = X_{n+1} - \hat{X}_{n+1,m}, \quad n \in \mathbb{Z}, \tag{23.1}$$

是平稳序列, 所以 $\sigma_{1,m}^2 = EY_n^2 = EY_0^2$ 与 n 无关. 最后 $\sigma_1^2 = \lim_{m\to\infty} \sigma_{1,m}^2$ 与 n 无关.

23.1.2 决定性与非决定性的严格定义

对充分大的 m, $L(X_{n+1}|X_{n,m})$ 表示用充分多的历史对未来 X_{n+1} 进行预测. $\sigma_{1,m}^2$ 表示的是预测的均方误差. 当 $m \to \infty$ 时, $\sigma_{1,m}^2 \to 0$ 说明 X_{n+1} 可以由所有历史 $X_n, X_{n-1}, \dots$ 进行完全预测. 当 $\sigma_1^2 > 0$ 说明 X_{n+1} 不可以由所有历史 $X_n, X_{n-1}, \dots$ 的线性组合以及极限进行完全预测.

定义 23.1. 设 $\{X_t\}$ 是零均值平稳序列.

- 如果 $\sigma_1^2 = 0$, 称 $\{X_t\}$ 是**决定性平稳序列**;

- 如果 $\sigma_1^2>0$, 称 $\{X_t\}$ 是**非决定性平稳序列**, 并且称 $\sigma_1^2=\lim_{m\to\infty}\sigma_{1,m}^2$ 为 $\{X_t\}$ 的**一步 (线性) 预测的均方误差**。

对于平稳序列 $\{X_t\}$, 如果 $EX_t=\mu$, 引入 $\{Z_t\}=\{X_t-\mu\}$ 和 m 维向量 $\mu_m=(\mu,\dots,\mu)^T$.

按照最佳线性预测的定义22.2,

$$\begin{aligned}\hat{X}_{n+1,m}=&\mu+L(X_{n+1}-\mu|X_{n,m}-\mu_m)\\ =&\mu+L(Z_{n+1}|Z_{n,m})=\mu+\hat{Z}_{n+1,m}.\end{aligned}$$

于是

$$E(Z_{n+1}-\hat{Z}_{n+1,m})^2=E(X_{n+1}-\hat{X}_{n+1,m})^2.$$

即对 X 预报的均方误差等于对中心化得到的 Z 预报的均方误差。因而, 当且仅当 $\{X_t-\mu\}$ 是决定性平稳序列时, 称 $\{X_t\}$ 是决定性平稳序列. 于是以后只需要讨论零均值的平稳序列.

23.1.3 可完全线性预测

设平稳列 $\{X_t\}$ 的 $n+1$ 阶自协方差阵 Γ_{n+1} 退化，$|\Gamma_n|>0$。则 $X_1,X_2,\cdots,X_{n+1}$ 线性相关, 所以 X_{n+1} 可以由 $X_n,X_{n-1},\dots,X_1$ 线性表示. 于是, $L(X_{n+1}|X_n,\dots,X_1)=X_{n+1}$. 当 $m\ge n$ 时,

$$L(X_{n+1}|X_n,\dots,X_{n-m+1})=X_{n+1},$$

即有 $\sigma_{1,m}^2=0$, $\{X_t\}$ 是决定性平稳列。

最简单的决定性平稳列是 $X_t\equiv\xi$, ξ 为随机变量。

23.1.4 离散谱序列

设零均值随机变量 $\xi_j,\eta_k(j,k=1,2,\dots,p)$ 两两正交, 满足

$$E(\xi_j^2)=E(\eta_j^2)=\sigma_j^2,\ j=1,2,\dots \tag{23.2}$$

对确定的 j, 定义简单离散谱序列

$$Z_j(t)=\xi_j\cos(t\lambda_j)+\eta_j\sin(t\lambda_j),\quad t\in\mathbb{Z}. \tag{23.3}$$

可以证明 $\{Z_j(t)\}$ 是平稳序列。事实上，易见 $EZ_j(t) \equiv 0$。而

$$\begin{aligned} E[Z_j(t)Z_j(s)] =& E\xi_j^2 \cos(\lambda_j t)\cos(\lambda_j s) + E\eta_j^2 \sin(\lambda_j t)\sin(\lambda_j s) \\ =& \sigma_j^2 \cos((t-s)\lambda_j) \end{aligned}$$

只依赖于 $t-s$。

$\{Z_j(t)\}$ 的每一次实现是周期函数, 由 §2.3 的定理 3.7 知道 $\{Z_j(t)\}$ 的 3 阶自协方差矩阵是退化的, 因此(23.3)是可完全线性预测的，$Z_j(n)$ 可以被 $Z_j(n-1), Z_j(n-2)$ 完全线性预测，是决定性序列。

事实上容易证明

$$Z_j(t) = (2\cos\lambda_j)Z_j(t-1) - Z_j(t-2)$$

定义离散谱序列

$$Z_t = \sum_{j=1}^{p} Z_j(t), \quad t \in \mathbb{Z}. \tag{23.4}$$

这是 p 个简单离散谱序列的叠加. 由 §9.4的定理9.4知道由(23.4)定义的离散谱序列也是决定性的. Z_n 可以被 $Z_{n-1}, Z_{n-2}, \dots, Z_{n-2p}$ 完全线性预测。

23.1.5 纯非决定性

决定性与非决定性取决于一步线性预报误差是否为零。

对非决定性序列，用 $\{X_s, s \le n\}$ 预报 X_{n+k} 的误差会随 k 增大而增大。记

$$\sigma_{k,m}^2 = E[X_{n+k} - L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})]^2$$

则 $\sigma_{k,m}$ 也是 m 的单调递减函数，与 n 无关。可定义

$$\sigma_k^2 = \lim_{m\to\infty} \sigma_{k,m}^2$$

在极限意义下可以证明 $\sigma_k^2 \ge \sigma_{k-1}^2$:

$$\begin{aligned} \sigma_k^2 &= \lim_{m\to\infty} E(X_{n+k} - L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m}))^2 & (23.5)\\ &= \lim_{m\to\infty} E[X_{n+k-1} - L(X_{n+k-1}|X_{n-1}, X_{n-2}, \dots, X_{n-1-m})]^2 & (23.6)\\ &\ge \lim_{m\to\infty} E[X_{n+k-1} - L(X_{n+k-1}|X_n, X_{n-1}, \dots, X_{n-m-1})]^2 & (23.7)\\ &= \lim_{m\to\infty} \sigma_{k-1,m+1}^2 = \sigma_{k-1}^2. & (23.8) \end{aligned}$$

注意上面证明中没有说明 $\sigma_{k,m}^2$ 是 k 的增函数。反例：AR(2) 序列

$$X_t = \frac{1}{2}X_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$$

平稳解为

$$X_t = \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j \varepsilon_{t-2j}$$

$$\begin{aligned}
\gamma_0 =& \frac{4}{3}\sigma^2 \quad \gamma_1 = 0 \\
\gamma_2 =& \frac{2}{3}\sigma^2 \\
L(X_t|X_{t-1}) =& 0 \quad \sigma_{1,1}^2 = \frac{4}{3}\sigma^2 \\
L(X_t|X_{t-2}) =& \frac{1}{2}X_{t-2} \quad \sigma_{2,1}^2 = \sigma^2 < \sigma_{1,1}^2
\end{aligned}$$

由最佳线性预测定义知

$$\begin{aligned}
\sigma_{k,m}^2 =& E[X_{n+k} - L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})]^2 \\
\le& E[X_{n+k} - 0]^2 = \gamma_0
\end{aligned}$$

所以 $\sigma_k^2 \le \gamma_0$。$k \to \infty$ 时如果 $\sigma_k^2 \to \gamma_0$ 则最佳线性预测与用平均值 0 预测效果相同，没有作用。

定义 23.2. 设 $\{X_t\}$ 是非决定性的平稳序列. 如果 $\lim_{k\to\infty} \sigma_k^2 = \gamma_0$, 则称 $\{X_t\}$ 是**纯非决定性的**.

纯非决定性的平稳列不能作长期预报。非决定性但不是纯非决定性的平稳列作长期预报是有意义的; 当然，决定性序列可以精确地长期预报。

对纯非决定性的平稳序列, 有如下的结果:

$$\lim_{k\to\infty} \lim_{m\to\infty} E[L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})]^2 = 0. \tag{23.9}$$

实际上, 记 $\hat{X}_{n+k,m} = L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})$. 由投影的正交性得

$$\sigma_{k,m}^2 = E(X_{n+k} - \hat{X}_{n+k,m})^2 = EX_{n+k}^2 - E\hat{X}_{n+k,m}^2.$$

于是得到

$$\lim_{k\to\infty} \lim_{m\to\infty} E\hat{X}_{n+k,m}^2 = \lim_{k\to\infty} \lim_{m\to\infty} (\gamma_0 - \sigma_{k,m}^2) = \gamma_0 - \gamma_0 = 0. \tag{23.10}$$

从(23.9)也可看出, 对于纯非决定性的平稳序列做长期或超长期预测是不合适的.

23.2 Wold 表示定理

23.2.1 线性闭包

设 A 为 Hilbert 空间 H 的子集，记 $\mathrm{sp}(A)$ 或 L_A 为 A 的所有有限线性组合构成的集合，记 $\overline{\mathrm{sp}}(A)$ 或 $\bar{L}_A$ 为 $\mathrm{sp}(A)$ 的元素及其元素极限组成的集合，记 H_A 为包含 A 的最小的闭子空间。

引理 23.1. *设 A 为 Hilbert 空间 H 的子集则*

$$H_A = \overline{sp}(A)$$

于是 $\forall \xi \in H_A$，必存在 $\xi_n \in sp(A), n = 1, 2, ...$ 使得

$$\|\xi_n - \xi\| \to 0, \quad n \to \infty.$$

称 $\overline{\mathrm{sp}}(A)$ 为 A 的**线性闭包**，或由 A 生成的子希尔伯特空间，或由 A 张成的子希尔伯特空间。

证明: 易见 $\mathrm{sp}(A) \subset \overline{\mathrm{sp}}(A) \subset H$。

首先，H_A 存在而且是 H 的闭子空间。事实上，令

$$H_A = \bigcap_{B\text{是}H\text{的闭子空间且}B\supset A} B$$

因为 $H \supset A$ 所以 H_A 非空。易见 H_A 也是线性空间，且也是闭集，所以 H_A 是包含 A 的最小闭子空间。

易见 $\mathrm{sp}(A)$ 为 H 的子线性空间，且由 $A \subset H_A$ 和 H_A 是线性空间知 $\mathrm{sp}(A) \subset H_A$。因为 H_A 是闭集所以 $\overline{\mathrm{sp}}(A) \subset H_A$。

另一方面，可以证明 $\overline{\mathrm{sp}}(A)$ 是闭子空间，由 H_A 的定义及 $A \subset \overline{\mathrm{sp}}(A)$ 得 $H_A \subset \overline{\mathrm{sp}}(A)$。

易见 $\overline{\mathrm{sp}}(A)$ 是 H 的线性子空间。下面证明 $\overline{\mathrm{sp}}(A)$ 是闭集。

设 $\xi_n \in \overline{\mathrm{sp}}(A)$, $\xi \in H$ 使得 $\lim_{n\to\infty} \|\xi_n - \xi\| = 0$, 只要证明 $\xi \in \overline{\mathrm{sp}}(A)$。对 ξ_n，存在 $\eta_n \in \mathrm{sp}(A)$ 使得

$$\|\xi_n - \eta_n\| < \frac{1}{n}$$

所以

$$\begin{aligned}\|\eta_n - \xi\| \le& \|\xi_n - \eta_n\| + \|\xi_n - \xi\| \\ \le& \frac{1}{n} + \|\xi_n - \xi\| \\ \to& 0,\ (n \to \infty)\end{aligned}$$

即 $\xi \in \overline{\mathrm{sp}}(A)$，所以 $\overline{\mathrm{sp}}(A)$ 是闭子空间，于是 $\overline{\mathrm{sp}}(A) \supset H_A$，从而 $H_A = \overline{\mathrm{sp}}(A)$。证毕。

23.2.2 无穷历史的线性预测

记 H_n 为 $X_n, X_{n-1}, \dots$ 生成的闭子空间(线性闭包)。$L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})$ 当 $m \to \infty$ 时为 $L(X_{n+k}|H_n)$(见后面的定理23.3)。

定理 23.2. 设 $Y \in L^2$, $\xi \in H_n$, 则 $\xi = L(Y|H_n)$ 的充分必要条件是

$$Y - \xi \perp X_j, \quad j = n, n-1, n-2, \dots \tag{23.11}$$

证明:

必要性: 由定理22.1得到 $Y - \xi \perp H_n$ 所以有(23.11)。

充分性: 记 $A = \{X_n, X_{n-1}, \dots\}$，则由(23.11)可知 $Y - \xi \perp L_A$。由引理23.1, 对 $\eta \in H_n$ 有 $\eta_m \in L_A$ 使 $\eta_m \to \eta$，由内积的连续性可得

$$E((Y-\xi)\eta) = \lim_{m\to\infty} E((Y-\xi)\eta_m) = 0$$

即 $Y - \xi \perp H_n$，由定理22.1即得 $\xi = L(X_{n+k}|H_n)$。

H_n 为 $\{X_s, s \le n\}$ 所张成的子 Hilbert 空间，$L(X_{n+k}|H_n)$ 是一个投影。下面的定理说明这个投影是有穷维最佳线性预测的极限。

定理 23.3. 设 $X_{n,m} = (X_n, X_{n-1}, \dots, X_{n-m+1})^T$, 当 $m \to \infty$ 时

$$L(Y|X_{n,m}) \xrightarrow{m.s.} \hat{Y} \triangleq L(Y|H_n) \tag{23.12}$$

证明: 记 $\hat{Y}_m = L(Y|X_{n,m})$。先证明 $\{\hat{Y}_m\}$ 是 H_n 中基本列。显然 $\hat{Y}_m \in H_n$，设当 $m \to \infty$ 时

$$\eta_m^2 \triangleq E(Y - \hat{Y}_m)^2 \to \eta^2 \qquad \text{(注意单调性)}$$

对 $m, k \to \infty$，注意 $\hat{Y}_m, \hat{Y}_{m+k}$ 都和 $Y - \hat{Y}_{m+k}$ 正交，得

$$\begin{aligned}
&\|\hat{Y}_m - \hat{Y}_{m+k}\|^2 = \|\hat{Y}_m - Y + Y - \hat{Y}_{m+k}\|^2 \\
=&\|\hat{Y}_m - Y\|^2 + \|Y - \hat{Y}_{m+k}\|^2 + 2\langle \hat{Y}_m - Y, Y - \hat{Y}_{m+k}\rangle \\
=&\eta_m^2 + \eta_{m+k}^2 - 2\langle Y, Y - \hat{Y}_{m+k}\rangle \\
=&\eta_m^2 + \eta_{m+k}^2 - 2\langle Y - \hat{Y}_{m+k}, Y - \hat{Y}_{m+k}\rangle \\
=&\eta_m^2 + \eta_{m+k}^2 - 2\eta_{m+k}^2 \to 0
\end{aligned}$$

因此 $\{\hat{Y}_m\}$ 是 H_n 的基本列，在 H_n 中存在唯一极限 ξ。

由内积的连续性, 对任何 $X_s, s \le n$ 有

$$\langle X_s, Y - \xi\rangle = \lim_{m\to\infty} \langle X_s, Y - L(Y|X_{n,m})\rangle = 0$$

由定理23.2得到 $\xi = L(Y|H_n)$。

23.2.3 无穷历史最优线性预测方差

由内积连续性，

$$\begin{aligned}
\sigma_1^2 \triangleq& \lim_{m\to\infty} \|X_{n+1} - L(X_{n+1}|X_{n,m})\|^2 \\
=&\|X_{n+1} - L(X_{n+1}|H_n)\|^2 = \|X_1 - L(X_1|H_0)\|^2
\end{aligned}$$

最后一个等号是因为等号右边也可以写成有限自变量预报均方误差极限。

$\sigma_1^2 = 0 \Leftrightarrow X_1 = L(X_1|H_0)$，所以 $\sigma_1^2 = 0 \Rightarrow X_1 \in H_0$。反之，如果 $X_1 \in H_0$，则 $E(X_1 - X_1)^2 = 0$ 最小所以 $X_1 = L(X_1|H_0)$。即 $\sigma_1^2 = 0 \Leftrightarrow X_1 \in H_0$。这

时 $\{X_t\}$ 是决定性序列。类似地，

$$\begin{aligned}\sigma_k^2 \triangleq & \lim_{m\to\infty} \|X_{n+k} - L(X_{n+k}|X_{n,m})\|^2 \\ = & \|X_{n+k} - L(X_{n+k}|H_n)\|^2 = \|X_k - L(X_k|H_0)\|^2\end{aligned}$$

定理 23.4. 设 $\{X_t\}$ 是零均值平稳列，

(1) $\{X_t\}$ 是决定性序列当且仅当对某个 n 有

$$X_{n+1} \in H_n; \tag{23.13}$$

并且如果(23.13)对某个 n 成立则对所有 n 成立，这时 $H_n = H_{n-1}, \forall n \in \mathbb{Z}$。

(2) $\{X_t\}$ 是纯非决定性的当且仅当对某个 n，有

$$\sigma_k^2 = \|X_{n+k} - L(X_{n+k}|H_n)\|^2 \to \gamma_0, \quad k \to \infty \tag{23.14}$$

并且如果(23.14)对某个 n 成立则对所有 n 成立。

23.2.4 Wold 表示定理

定理 23.5 (Wold 表示定理)**.** 任一非决定性的零均值平稳列可以表示成

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} + V_t, \quad t \in \mathbb{Z} \tag{23.15}$$

其中

(1) $\varepsilon_t = X_t - L(X_t|X_{t-1}, X_{t-2}, \dots)$ 是零均值白噪声，满足

$$\begin{aligned}& E\varepsilon_t^2 = \sigma^2 > 0, \quad a_0 = 1 \\ & a_j = E(X_t \varepsilon_{t-j})/\sigma^2, \\ & \sum_{j=0}^{\infty} a_j^2 < \infty\end{aligned}$$

(2) $\{U_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},\ t \in \mathbb{Z}\}$ 和 $\{V_t\}$ 都是平稳列且两者互相正交；

(3) 定义 $H_\varepsilon(t) = \bar{sp}\{\varepsilon_s : s \le t\}$, $H_U(t) = \bar{sp}\{U_s : s \le t\}$, 则 $\forall t$

$$H_U(t) = H_\varepsilon(t)$$

(4) $\{U_t\}$ 是纯非决定性的平稳序列，有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} a_j e^{ij\lambda} \right|^2$$

(5) $\{V_t\}$ 是决定性的平稳序列。对任何 $t, k \in \mathbb{Z}$，$V_t \in H_{t-k}$。

定义 23.3. 在 Wold 表示定理中

- (1) 称(23.15)是 $\{X_t\}$ 的 **Wold 表示**；
- (2) 称 $\{U_t\}$ 是 $\{X_t\}$ 的**纯非决定性部分**，称 $\{V_t\}$ 是 $\{X_t\}$ 的**决定性部分**；
- (3) 称 $\{a_j\}$ 是 $\{X_t\}$ 的 **Wold 系数**；
- (4) 称一步预测误差 $\varepsilon_t = X_t - L(X_t | X_{t-1}, X_{t-2}, \dots)$ 为 $\{X_t\}$ 的**新息序列**;
- (5) 称 $\sigma^2 = E\varepsilon_t^2$ 为**一步 (线性) 预测的均方误差**。

这里新息的意思是不能被历史线性预测的部分。由 Wold 定理可知任何纯非决定性平稳序列可以表达为新息的单边滑动和。事实上，任何白噪声的单边滑动和 (系数平方可和) 一定是纯非决定性的，但其中的白噪声不一定恰好是新息。

23.2.4.1 ARMA 序列的 Wold 表示

设 $\{X_t\}$ 是 ARMA(p,q) 序列，模型方程为

$$A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z},$$

设 $A^{-1}(z)B(z)$ 有 Taylor 展开式

$$\Psi(z) = A^{-1}(z)B(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \le 1$$

则

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \tag{23.16}$$

下面证明 $\{X_t\}$ 是纯非决定性的平稳序列，(23.16)是 $\{X_t\}$ 的 Wold 表示，$\{\varepsilon_t\}$ 是 $\{X_t\}$ 的新息序列，$\{\psi_j\}$ 是 $\{X_t\}$ 的 Wold 系数。

我们只对比较容易的可逆 ARMA 的情况证明。由(23.16)看出 $X_t \in H_\varepsilon(t)$, 利用可逆性，$\varepsilon_t = B^{-1}(\mathscr{B})A(\mathscr{B})X_t \in H_t$, 所以 $H_t = H_\varepsilon(t)$. 于是

$$X_t - \varepsilon_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \in H_\varepsilon(t-1) = H_{t-1}$$

来证 $X_t - \varepsilon_t = L(X_t|H_{t-1})$。只要证明 $X_t - (X_t - \varepsilon_t) \perp H_{t-1}$.

事实上，由于 ε_t 与 $\varepsilon_{t-j}, j \geq 1$ 正交可知 $\varepsilon_t \perp H_\varepsilon(t-1) = H_{t-1}$。故

$$\begin{aligned} X_t - \varepsilon_t =& L(X_t|H_{t-1}) \\ \varepsilon_t =& X_t - L(X_t|H_{t-1}) \end{aligned}$$

即 $\{\varepsilon_t\}$ 是 $\{X_t\}$ 的新息列，$\sigma^2 = E\varepsilon_t^2$ 是一步预测均方误差，在(23.16)两边同乘以 ε_{t-j} 后取期望，利用内积的连续性可得

$$\psi_j = \langle X_t, \varepsilon_{t-j} \rangle / \sigma^2$$

即 $\{\psi_j\}$ 是 $\{X_t\}$ 的 Wold 系数列。

23.2.4.2 Wold 表示定理证明

取 $\varepsilon_t = X_t - L(X_t|H_{t-1})$, $H_\varepsilon(t) = \overline{\text{sp}}\{\varepsilon_s : s \leq t\}$, 易见 $\varepsilon_t \in H_t$，由 H_t 的单调性可知 $\varepsilon_t \in H_s, \forall t < s$，因此 $H_\varepsilon(t) \subset H_t$。

来证明 $\{\varepsilon_t\}$ 是白噪声。

由定理23.3

$$\begin{aligned} L(X_t|H_{t-1}) &= \lim_{m\to\infty} L(X_t|X_{t-1}, \dots, X_{t-m}) \quad (L^2) \\ &= \lim_{m\to\infty} a_m^T X_{t-1,m} \end{aligned}$$

其中 a_m 是 m 阶 Y-W 系数 (预测方程的解)，不依赖于 t，由内积的连续性

$$\begin{aligned} E\varepsilon_t^2 &= \lim_{m\to\infty} \|X_t - L(X_t|X_{t-1}, \dots, X_{t-m})\|^2 \\ &= \lim_{m\to\infty} (\gamma_0 - a_m^T \Gamma_m a_m) \quad (\text{与}t\text{无关}) \\ &= \lim_{m\to\infty} \sigma_{1,m}^2 = \sigma^2 > 0 \quad (\text{由非决定性定义}) \end{aligned}$$

对 $s>t$，$\varepsilon_s \perp H_{s-1} \supset H_t$ 所以 $\varepsilon_s \perp \varepsilon_t$, ($s>t$ 时)。即

$$\{\varepsilon_t\} \sim \mathrm{WN}(0,\sigma^2), \quad \sigma^2 > 0$$

定义 $V_t = X_t - L(X_t|H_\varepsilon(t))$, 则 $V_t \in H_t$。来证明 $\{\varepsilon_t\}$ 和 $\{V_t\}$ 正交。

由投影性质,$V_t \perp H_\varepsilon(t)$,即 $V_t \perp \varepsilon_s, \forall s \leq t$。当 $s>t$ 时,注意 $\varepsilon_s \perp H_{s-1} \supset H_t$，而 $V_t \in H_t$ 所以 $\varepsilon_s \perp V_t, \forall s > t$。于是 $\{\varepsilon_t\}$ 和 $\{V_t\}$ 正交。

来证明 $\{\varepsilon_t\}$ 和 $\{X_t\}$ 平稳相关。

当 $s>t$ 时 $\varepsilon_s \perp H_t$ 所以 $\langle \varepsilon_s, X_t\rangle = 0, \forall s > t$。

当 $s \leq t$ 时，由定理23.3，

$$\begin{aligned} L(X_t|H_{t-1}) &= \lim_{m\to\infty} L(X_t|X_{t-1,m}) \quad (L^2) \\ &= \lim_{m\to\infty} a_m^T X_{t-1,m} \end{aligned}$$

由内积的连续性, $s \leq t$ 时

$$\begin{aligned} \langle X_t, \varepsilon_s\rangle &= \lim_{m\to\infty} \langle X_t, X_s - a_m^T X_{s-1,m}\rangle \\ &= \gamma_{t-s} - \lim_{m\to\infty}(a_{m1}\gamma_{t-s+1} + \cdots + a_{mm}\gamma_{t-s+m}) \end{aligned}$$

只依赖于 $t-s$。所以 $\{\varepsilon_t\}$ 和 $\{X_t\}$ 平稳相关。

由 $\{\varepsilon_t\}$ 和 $\{X_t\}$ 平稳相关，若定义

$$a_j = \langle X_t, \varepsilon_{t-j}\rangle/\sigma^2 \quad (j \geq 0)$$

则 a_j 与 t 无关。且

$$\begin{aligned} a_0 &= \langle X_t, \varepsilon_t\rangle/\sigma^2 \\ &= \langle \varepsilon_t + L(X_t|H_{t-1}), \varepsilon_t\rangle/\sigma^2 \\ &= \langle \varepsilon_t, \varepsilon_t\rangle/\sigma^2 = 1 \end{aligned}$$

令 $U_t = L(X_t|H_\varepsilon(t))$, 则 $V_t = X_t - L(X_t|H_\varepsilon(t)) = X_t - U_t$, $X_t = U_t + V_t$。来证明

$$U_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \tag{23.17}$$

定义 $U_{t,n} = L(X_t|\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-n})$. 设 $U_{t,n} = \sum_{j=0}^n b_j \varepsilon_{t-j}$，由 $\{\varepsilon_t\}$ 与 $\{X_t\}$ 平稳相关可知 $\{b_j\}$ 与 t 无关。对 $j = 0, 1, \dots, n$

$$\begin{aligned}\sigma^2 a_j =& \langle X_t, \varepsilon_{t-j}\rangle = \langle U_{t,n} + (X_t - U_{t,n}), \varepsilon_{t-j}\rangle \\ =& \langle U_{t,n}, \varepsilon_{t-j}\rangle = \sigma^2 b_j\end{aligned}$$

即 $b_j = a_j$，

$$U_{t,n} = L(X_t|\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-n}) = \sum_{j=0}^{n} a_j \varepsilon_{t-j}$$

注意 $U_{t,n}$ 是 X_t 的投影所以 $\|U_{t,n}\|^2 \leq \|X_t\|^2 = \gamma_0$，所以

$$\|U_{t,n}\|^2 = \sigma^2 \sum_{j=0}^{n} a_j^2 \leq \gamma_0 < \infty$$

故 $\sum_{j=0}^\infty a_j^2 < \infty$，$\sum_{j=0}^n a_j \varepsilon_{t-j}$ 均方收敛到 $\sum_{j=0}^\infty a_j \varepsilon_{t-j}$。由定理23.3知 $U_{t,n}$ 均方收敛到 $U_t = L(X_t|H_\varepsilon(t))$，所以(23.17)成立。

$\{U_t\}$ 是线性平稳列，其谱密度立即可得 (结论 (4) 的第二部分)。

由于 $\{\varepsilon_t\}$ 与 $\{V_t\}$ 正交所以 $V_s \perp H_\varepsilon(t), \forall s, t \in \mathbb{Z}$，而 $U_t \in H_\varepsilon(t)$ 所以 $V_s \perp U_t$, $\{V_t\}$ 与 $\{U_t\}$ 正交。由 $\{U_t\}$ 和 $\{V_t\}$ 正交，$X_t = U_t + V_t$, $\{X_t\}$ 和 $\{U_t\}$ 平稳可知 $V_t = X_t - U_t$ 也是平稳列。

至此定理的 (1)(2) 已证明。

来证明第 (3) 条结论。定义 $H_U(t) = \bar{\text{sp}}\{U_s: s \leq t\}$, 来证明 $H_U(t) = H_\varepsilon(t)$。

显然 $U_t \in H_\varepsilon(t)$ 所以 $H_U(t) \subset H_\varepsilon(t)$。只要证明 $H_\varepsilon(t) \subset H_U(t)$。

注意 $\varepsilon_t \in H_t \subset \bar{\text{sp}}\{U_s, V_s : s \leq t\}$，由引理23.1，存在 $\xi_m \in L\{V_t, V_{t-1}, \dots, V_{t-m}\}$, $\eta_m \in L\{U_t, U_{t-1}, \dots, U_{t-m}\}$, 使

$$\|\xi_m + \eta_m - \varepsilon_t\|^2 \to 0 \ (m \to \infty)$$

但前面已证明 $\{V_t\}$ 与 $\{\varepsilon_t\}$ 正交，也与 $\{U_t\}$ 正交，所以

$$\begin{aligned}&\|\xi_m+\eta_m-\varepsilon_t\|^2\\=&\|\xi_m\|^2+\|\eta_m-\varepsilon_t\|^2\\\geq&\|\eta_m-\varepsilon_t\|^2\end{aligned}$$

令 $m\to\infty$ 得 $\|\eta_m-\varepsilon_t\|^2\to 0$，由引理23.1知 $\varepsilon_t\in H_U(t)$。所以 $H_\varepsilon(t)\subset H_U(t)$，$H_\varepsilon(t)=H_U(t)$。结论 (3) 证毕。

来证明 $\{U_t\}$ 是纯非决定性的 (结论 (4))。利用定理22.2(4)(5)

$$\begin{aligned}&L(U_{t+k}|H_U(t))=L(U_{t+k}|H_\varepsilon(t))\\=&L\left[\sum_{j=0}^{k-1}a_j\varepsilon_{t+k-j}+\sum_{j=k}^{\infty}a_j\varepsilon_{t+k-j}|H_\varepsilon(t)\right]\\=&\sum_{j=k}^{\infty}a_j\varepsilon_{t+k-j}\end{aligned}$$

于是

$$\|U_{t+k}-L(U_{t+k}|H_U(t))\|^2=\|\sum_{j=0}^{k-1}a_j\varepsilon_{t+k-j}\|^2$$
$$=\sigma^2\sum_{j=0}^{k-1}a_j^2\to\sigma^2\sum_{j=0}^{\infty}a_j^2=EU_t^2$$

按定义可知 $\{U_t\}$ 为纯非决定性的。

注意：这个证明对一般单边线性序列不适用。

已证明 $\{V_t\}$ 平稳，来证明 $\{V_t\}$ 是决定性的。用定理23.4。注意 $\varepsilon_{t-j}\in H_{t-j}$，

$$\begin{aligned}V_t=&X_t-U_t=X_t-\varepsilon_t-\sum_{j=1}^{\infty}a_j\varepsilon_{t-j}\\=&L(X_t|H_{t-1})-\sum_{j=1}^{\infty}a_j\varepsilon_{t-j}\end{aligned}$$

其中 $L(X_t|H_{t-1})\in H_{t-1}$，$\sum_{j=1}^{\infty}a_j\varepsilon_{t-j}\in H_\varepsilon(t-1)\subset H_{t-1}$，所以 $V_t\in H_{t-1}$。

注意 $H_{t-1} \subset \bar{\text{sp}}\{U_s, V_s: \ s \leq t-1\}$, 由引理23.1，存在 $\xi_m \in L_{\{V_{t-1},\ldots,V_{t-m}\}}$, $\eta_m \in L_{\{U_{t-1},\ldots,U_{t-m}\}}$, 使

$$\|\xi_m + \eta_m - V_t\|^2 \to 0$$

已证明 $\{V_t\}$ 与 $\{U_t\}$ 正交所以 η_m 与 $\xi_m - V_t$ 正交，于是

$$\begin{aligned} &\|\xi_m + \eta_m - V_t\|^2 = \|\xi_m - V_t\|^2 + \|\eta_m\|^2 \\ \geq &\|\xi_m - V_t\|^2 \end{aligned}$$

故 $\|\xi_m - V_t\|^2 \to 0 (m \to \infty)$，由引理23.1知 $V_t \in H_V(t-1) = \bar{\text{sp}}\{V_s: \ s \leq t-1\}$。由定理23.4知 $\{V_t\}$ 为决定性的，且 $H_V(t) = H_V(t-j), j \in \mathbb{Z}$，所以 $V_t \in H_V(t-j), \ t, j \in \mathbb{Z}$。

23.2.4.3 关于新息的讨论

$\varepsilon_t = X_t - L(X_t|H_{t-1})$ 是 X_t 提供的比 $X_{t-1}, X_{t-2}, \ldots$ 多的信息 (线性意义下)。

可以证明

$$H_t = \text{sp}(\varepsilon_t) \oplus H_{t-1} \tag{23.18}$$

这样

$$H_t = \bigoplus_{j=0}^{\infty} \text{sp}(\varepsilon_{t-j}) \oplus H_{-\infty},$$

其中 $H_{-\infty} = \bigcap_t H_t$。$V_t \in H_{-\infty}$。

事实上，(23.18)右侧两项正交，都是闭子空间，且都包含于 H_t，则 (参见23.5.2)

$$\text{sp}(\varepsilon_t) \oplus H_{t-1} = \{\alpha\varepsilon_t + \xi : \alpha \in \mathbb{R}, \xi \in H_{t-1}\}$$

是 H_t 内的闭子空间, $\text{sp}(\varepsilon_t) \oplus H_{t-1} \subset H_t$。

来证 $H_t \subset \text{sp}(\varepsilon_t) \oplus H_{t-1}$。由引理23.1只要证明 $X_s \in \text{sp}(\varepsilon_t) \oplus H_{t-1}, s \leq t$。

当 $s < t$ 时显然，对 X_t，因为

$$X_t = \varepsilon_t + L(X_t|H_{t-1})$$

所以 $X_t \in \text{sp}(\varepsilon_t) \oplus H_{t-1}$, 于是有

$$H_t = \text{sp}(\varepsilon_t) \oplus H_{t-1}.$$

23.3 Kolmogorov 公式

考虑多步预报的均方误差。设 $\{X_t\}$ 是非决定性的平稳列，由 Wold 表示定理(5)，$V_{t+n} \in H_t$，所以用无穷长历史进行的最佳线性预测为

$$
\begin{aligned}
L(X_{t+n}|H_t) =& L(U_{t+n}|H_t) + L(V_{t+n}|H_t) \qquad (23.19)\\
=& L(\sum_{j=0}^{\infty} a_j \varepsilon_{t+n-j}|H_t) + V_{t+n} \qquad (23.20)\\
=& \sum_{j=n}^{\infty} a_j \varepsilon_{t+n-j} + V_{t+n} \qquad (23.21)
\end{aligned}
$$

称 $L(X_{t+n}|H_t)$ 是 X_{t+n} 的 n 步预报，由 Wold 分解公式知预报误差为

$$
X_{t+n} - L(X_{t+n}|H_t) = \sum_{j=0}^{n-1} a_j \varepsilon_{t+n-j} \qquad (23.22)
$$

预报的均方误差为

$$
\sigma^2(n) = \sigma^2 \sum_{j=0}^{n-1} a_j^2 \qquad (23.23)
$$

$n \to \infty$ 时 $\sigma^2(n) \to EU_t^2$.

定理 23.6 (Kolmogorov 公式)**.** 设 $\{U_t\}$ 是非决定性平稳序列 $\{X_t\}$ 的纯非决定性部分，$f(\lambda)$ 是 $\{U_t\}$ 的谱密度. 则有

$$
\sigma^2 = E[X_t - L(X_t|H_{t-1})]^2 = 2\pi \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda\right). \qquad (23.24)
$$

公式(23.24)的证明需要较多解析函数的知识. 当 $\{U_t\}$ 是白噪声时，公式(23.24)明显是成立的.

从 Kolmogorov 公式(23.24)看到, 如果 $\{X_t\}$ 是非决定性的, 则它的纯非决定性部分的谱密度 $f(\lambda)$ 必是 ln 可积的, 即

$$
\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda > -\infty. \qquad (23.25)
$$

23.4 最佳预测和最佳线性预测相等的条件

设 $\{X_t\}$ 是平稳序列, 用 $\mathscr{F}_t = \sigma\{X_t, X_{t-1}, \ldots\}$ 表示由 $X_t, X_{t-1}, \ldots$ 生成的 σ-代数. 称条件数学期望

$$E(X_{t+k}|\mathscr{F}_t)$$

是用全体历史 $\{X_j : j \le t\}$ 对 X_{t+k} 进行预测时的**最佳预测**.

最佳预测是均方误差最小的，这是因为条件数学期望 $E(X_{t+k}|\mathscr{F}_t)$ 是 $X_{t-1}, X_{t-2}, \ldots$ 的函数, 二阶矩有限:

$$E[E(X_{t+k}|\mathscr{F}_t)]^2 \le E[E(X_{t+k}^2|\mathscr{F}_t)] = EX_{t+k}^2 < \infty.$$

由概率论中数学期望性质可以证明对任意二阶矩有限的 $X_t, X_{t-1}, \ldots$ 的函数 ξ 有

$$E(X_{t+k} - \xi)^2 \ge E[X_{t+k} - E(X_{t+k}|\mathscr{F}_t)]^2 \tag{23.26}$$

事实上，对 $\xi \in \mathscr{F}_t$，

$$\begin{aligned}
&E\left[(X_{t+k} - \xi)^2\right] \\
=&E\left\{\left[(X_{t+k} - E(X_{t+k}|\mathscr{F}_t)) + (E(X_{t+k}|\mathscr{F}_t) - \xi)\right]^2\right\} \\
=&E\left\{[X_{t+k} - E(X_{t+k}|\mathscr{F}_t)]^2\right\} + E\left\{[E(X_{t+k}|\mathscr{F}_t) - \xi)]^2\right\} \\
&+ 2E\left\{(X_{t+k} - E(X_{t+k}|\mathscr{F}_t))(E(X_{t+k}|\mathscr{F}_t) - \xi)\right\}
\end{aligned}$$

而交叉项

$$\begin{aligned}
&E\left\{(X_{t+k} - E(X_{t+k}|\mathscr{F}_t))(E(X_{t+k}|\mathscr{F}_t) - \xi)\right\} \\
=&E\left\{E\left[(X_{t+k} - E(X_{t+k}|\mathscr{F}_t))(E(X_{t+k}|\mathscr{F}_t) - \xi) \,|\, \mathscr{F}_t\right]\right\} \\
=&E\left\{(E(X_{t+k}|\mathscr{F}_t) - \xi)\; E\left[X_{t+k} - E(X_{t+k}|\mathscr{F}_t) \,|\, \mathscr{F}_t\right]\right\} \\
=&\, 0
\end{aligned}$$

所以

$$\begin{aligned}
&E\left[(X_{t+k} - \xi)^2\right] \\
=&E\left\{[X_{t+k} - E(X_{t+k}|\mathscr{F}_t)]^2\right\} + E\left\{[E(X_{t+k}|\mathscr{F}_t) - \xi)]^2\right\} \\
\ge&E\left\{[X_{t+k} - E(X_{t+k}|\mathscr{F}_t)]^2\right\}
\end{aligned}$$

最佳预测一般比最佳线性预测好，但是对纯非决定性序列如果其新息是独立序列则二者等价。

定理 23.7. 设平稳序列 $\{X_t\}$ 有 *Wold* 表示

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \tag{23.27}$$

则

$$L(X_{t+n}|H_t) = E(X_{t+n}|\mathscr{F}_t), \ n \geq 1, \quad t \in \mathbb{Z}, \tag{23.28}$$

成立的充分必要条件是

$$E(\varepsilon_{t+1}|\varepsilon_t, \varepsilon_{t-1}, \dots) = 0, \quad t \in \mathbb{Z}. \tag{23.29}$$

(23.29)的条件称为鞅差。零均值独立白噪声列是鞅差的特例。

推论 23.1. 设 *ARMA(p,q)* 序列 $\{X_t\}$ 中的新息 $\{\varepsilon_t\}$ 是独立白噪声，则用全体历史 $\{X_t, X_{t-1}, \dots\}$ 对 X_{t+n} 进行预测时，最佳预测和最佳线性预测相等.

23.5 附录：补充

非决定性也称为非奇异，决定性序列称为奇异序列。见谢衷洁《时间序列分析》P.82。纯非决定性序列叫做正则序列，见谢衷洁《时间序列分析》P.118 第 13 题。

23.5.1 关于预测的分类

```
knitr::include_graphics("figs/forecast-class.png")
```

23.5.2 正交直和分解

定理：设 H 为 Hilbert 空间，H_1 和 H_2 是 H 的闭子空间，H_1 与 H_2 正交，$M = H_1 \oplus H_2 = \{x + y : x \in H_1, y \in H_2\}$，则 M 是 H 的闭子空间，且 $\forall \xi \in H$,

$$L(\xi|M) = L(\xi|H_1) \oplus L(\xi|H_2),$$

其中 $\oplus$ 表示求和，且求和的两项正交。

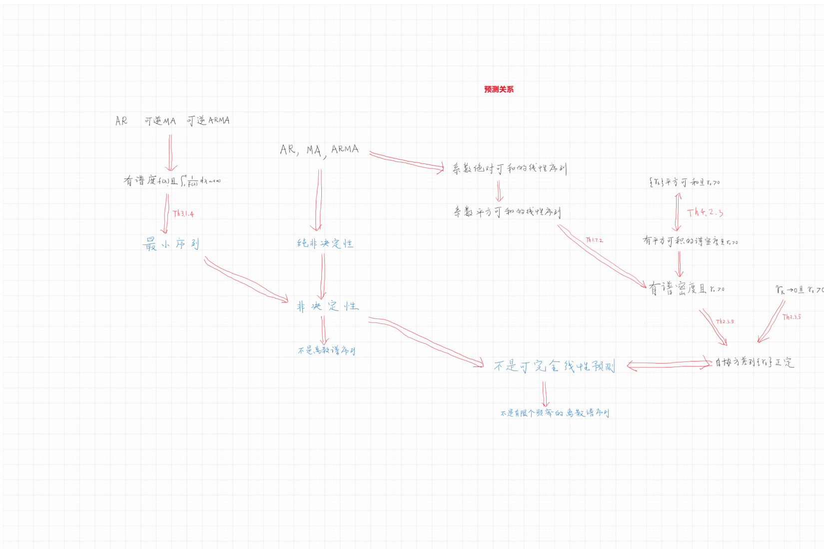


图 23.1: 时间序列预测分类

证明: 对 $\xi, \eta \in M$，应有 $\xi_1, \eta_1 \in H_1, \xi_2, \eta_2 \in H_2$ 使得 $\xi = \xi_1 + \xi_2, \eta = \eta_1 + \eta_2$，于是 $\xi + \eta = (\xi_1 + \eta_1) + (\xi_2 + \eta_2) \in M$；对 $\alpha \in \mathbb{R}$，$\alpha\xi = (\alpha\xi_1) + (\alpha\xi_2) \in M$，所以 M 是 H 的子线性空间。

对 M 中的基本列 $\{\xi_n\}$，有分解 $\xi_n = \xi_{1,n} + \xi_{2,n}$，$\xi_{1,n} \in H_1, \xi_{2,n} \in H_2$,

$$0 = \lim_{n,m\to\infty} \|\xi_n - \xi_m\|^2 = \lim_{n,m\to\infty} \left(\|\xi_{1,n} - \xi_{1,m}\|^2 + \|\xi_{2,n} - \xi_{2,m}\|^2\right)$$

从而 $\{\xi_{1,n}\}$ 是 H_1 的基本列，$\{\xi_{2,n}\}$ 是 H_2 的基本列，存在 $\xi_1 \in H_1, \xi_2 \in H_2$，使得 $\lim \|\xi_{1,n} - \xi_1\| = 0, \lim \|\xi_{2,n} - \xi_2\| = 0$, 于是

$$\lim_{n\to\infty} \|\xi_n - (\xi_1 + \xi_2)\|^2 = \lim_{n\to\infty} \left(\|\xi_{1,n} - \xi_1\|^2 + \|\xi_{2,n} - \xi_2\|^2\right) = 0$$

因 $\xi_1 + \xi_2 \in M$，所以 M 是 H 的闭子空间。

$\forall \xi \in H$, 令 $\xi_1 = L(\xi|H_1), \xi_2 = L(\xi|H_2)$, 则 $\xi_1 + \xi_2 \in M$，且

$$\begin{aligned}
\xi - (\xi_1 + \xi_2) =&(\xi - \xi_1) + (-\xi_2) \perp H_1, \\
\xi - (\xi_1 + \xi_2) =&-\xi_1 + (\xi - \xi_2) \perp H_2,
\end{aligned}$$

因此 $\xi - (\xi_1 + \xi_2)$ 与 $M = H_1 + H_2$ 正交，从而 $\xi_1 + \xi_2 = L(\xi|M)$，且 $\xi_1 \perp \xi_2$。结论得证。

23.5.3 单边线性序列与 Wold 表示

单边线性序列一定是纯非决定性的，但其中的白噪声不一定是新息，所以表达式本身不一定是 Wold 表示。

如

$$X_t = \varepsilon_t + 2\varepsilon_{t-1}$$

是纯非决定性序列，其谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1 + 2e^{i\lambda}|^2 = \frac{4\sigma^2}{2\pi}|1 + \frac{1}{2}e^{i\lambda}|^2$$

但 $\varepsilon_t \neq X_t - L(X_t|H_{t-1})$。

令

$$\eta_t = (1 - \frac{1}{2}\mathscr{B})^{-1} X_t$$

则 $\{\eta_t\}$ 是 $\{X_t\}$ 的系数绝对可和的线性滤波，故平稳，且 $\{\eta_t\}$ 的谱密度为

$$f_\eta(\lambda) = |1 - \frac{1}{2}e^{-i\lambda}|^{-2} f(\lambda) = \frac{4\sigma^2}{2\pi}$$

所以 $\{\eta_t\}$ 是 WN(0, $4\sigma^2$)，

$$X_t = \eta_t + \frac{1}{2}\eta_{t-1}$$

是可逆 MA(1) 模型。由上面关于可逆 ARMA 模型的一般结论可知 η_t 是 $\{X_t\}$ 的新息，而 ε_t 与 η_t 方差不同，不会 a.s. 相等，由新息的唯一性可知 $\{\varepsilon_t\}$ 不是 $\{X_t\}$ 的新息。所以单边线性序列中的白噪声列不一定是新息。

23.5.4 离散谱序列可完全线性预测的直接证明

对

$$Z_j(t) = \xi_j \cos(t\lambda_j) + \eta_j \sin(t\lambda_j),\ t \in \mathbb{Z},$$

有

$$\begin{aligned}
&2\cos\lambda_j Z_j(t-1) - Z_j(t-2) \\
=&\xi_j\{2\cos\lambda_j \cos[(t-1)\lambda_j] - \cos[(t-2)\lambda_j]\} \\
&+ \eta_j\{2\cos\lambda_j \sin[(t-1)\lambda_j] - \sin[(t-2)\lambda_j]\} \\
=&\xi_j\{\cos(t\lambda_j) + \cos[(t-2)\lambda_j] - \cos[(t-2)\lambda_j]\} \\
&+ \eta_j\{\sin(t\lambda_j) + \sin[(t-2)\lambda_j] - \sin[(t-2)\lambda_j]\} \\
=&Z_j(t)
\end{aligned}$$

所以

$$L(Z_j(t)|Z_j(t-1), Z_j(t-2)) = 2\cos\lambda_j Z_j(t-1) - Z_j(t-2).$$

但是混合多个频率的离散谱序列的预测公式就没有这么容易。

Chapter 24

时间序列的递推预测

§10.1已经用 Y-W 方程给出了非决定性平稳序列的预测公式，这里我们进行更深入的讨论。

- 非决定性平稳序列可以根据自协方差函数用 Levinson 递推得到预测系数和预测方差；
- 对 AR(p) 模型只要使用自回归系数预测；
- 对 MA 模型和 ARMA 模型则只能递推得到预测系数。

这里研究可以化简 MA 模型和 ARMA 模型预测的公式, 以及非平稳时也可用的递推公式。假设自协方差函数已知，实际中可以用样本自协方差函数代替。

24.1 一般时间序列的递推预测

24.1.1 递推预测的正交分解

设 $\{Y_t\}$ 是方差有限的零均值时间序列, 对任何正整数 n, 用

$$L_n = \overline{\text{span}}\{Y_1, Y_2, \dots, Y_n\}$$

表示 $Y_1, \dots, Y_n$ 的线性组合的全体. 定义 $\boldsymbol{Y}_n = (Y_1, Y_2, \dots, Y_n)^T$,

$$\hat{Y}_1 = 0, \ \hat{Y}_n = L(Y_n | \boldsymbol{Y}_{n-1}), \ n = 2, \dots . \tag{24.1}$$

引入预测误差 W_n 及其方差 ν_{n-1} 如下:

$$W_n = Y_n - \hat{Y}_n, \quad \nu_{n-1} = EW_n^2. \quad n = 1, 2, \cdots. \tag{24.2}$$

由最佳线性预测的性质 7 知道 W_n 和 L_{n-1} 中的任何随机变量正交, 并且 $W_n \in L_n$. 于是 $\{W_n\}$ 是一个正交序列, 满足

$$E(W_n W_k) = \nu_{n-1}\delta_{n-k}.$$

这里 δ_t 是 Kronecker 函数.

用

$$M_n \triangleq \overline{\text{span}}\{W_1, W_2, \dots, W_n\}$$

表示 $W_1, W_2, \cdots, W_n$ 的线性组合全体. 则 $M_n \subset L_n$.

对 $n \in \mathbb{N}$, 我们用归纳法证明 $Y_n \in M_n$.

首先 $Y_1 = W_1 \in M_1$. 如果对 $k \leq n$ 已经证明 $Y_k \in M_k$, 注意 $\hat{Y}_{n+1} = L(Y_{n+1}|\boldsymbol{Y}_n) \in M_n$, 于是

$$Y_{n+1} = (Y_{n+1} - \hat{Y}_{n+1}) + \hat{Y}_{n+1} = W_{n+1} + \hat{Y}_{n+1} \in M_{n+1}.$$

这就证明了对 $n \in \mathbb{N}$, $Y_n \in M_n$ 成立.

于是得到

$$\begin{aligned} L_n =& \overline{\text{span}}\{Y_1, Y_2, \dots, Y_n\} && (24.3)\\ =& \overline{\text{span}}\{W_1, W_2, \dots, W_n\} = M_n, && (24.4)\\ & n = 1, 2, \dots && (24.5)\end{aligned}$$

§22.1性质 10 和(24.5)式告诉我们用 $\boldsymbol{W}_n = (W_1, W_2, \dots, W_n)^T$ 对 Y_{n+1} 进行预测和用 $\boldsymbol{Y}_n = (Y_1, Y_2, \dots, Y_n)^T$ 对 Y_{n+1} 进行预测是等价的. 由于 $\{W_t\}$ 是正交序列, 所以用 $\boldsymbol{W}_n$ 对 Y_{n+1} 进行预测有很多的方便. 类似于在正交基上的投影，可以直接计算坐标。

24.1.2 递推预测定理

定理 24.1. *设 $\{Y_t\}$ 是零均值时间序列 (不要求平稳!). 如果 $(Y_1, Y_2, \dots, Y_{m+1})^T$ 的协方差矩阵*

$$\Big(E(Y_s Y_t)\Big)_{1 \leq s, t \leq m+1} \tag{24.6}$$

正定，则最佳线性预测

$$
\begin{aligned}
\hat{Y}_{n+1} \triangleq& L(Y_{n+1}|\boldsymbol{Y}_n),\ n=1,2,\dots,m \\
=& \sum_{j=1}^{n} \theta_{n,j} W_{n+1-j} = \sum_{k=0}^{n-1} \theta_{n,n-k} W_{k+1} \qquad (24.7) \\
=& \theta_{n,1} W_n + \theta_{n,2} W_{n-1} + \dots + \theta_{n,n} W_1 \qquad (24.8) \\
=& \theta_{n,n} W_1 + \theta_{n,n-1} W_2 + \dots + \theta_{n,1} W_n
\end{aligned}
$$

其中的系数 $\{\theta_{n,j}\}$ 和预测的均方误差 $\nu_n = EW_{n+1}^2$ 满足如下的递推公式.

$$
\begin{cases}
\nu_0 = EY_1^2, \\
\theta_{n,n-k} = \left[E(Y_{n+1}Y_{k+1}) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \nu_j \right] / \nu_k, \\
\qquad 0 \leq k \leq n-1, \\
\nu_n = EY_{n+1}^2 - \sum_{k=0}^{n-1} \theta_{n,n-k}^2 \nu_k,
\end{cases} \qquad (24.9)
$$

其中约定 $\sum_{j=0}^{-1} (\cdot) \triangleq 0$.

递推的顺序是

$$
\begin{array}{lllll}
\nu_0 & & & & \\
\theta_{1,1} & \nu_1 & & & \\
\theta_{2,2} & \theta_{2,1} & \nu_2 & & \\
\theta_{3,3} & \theta_{3,2} & \theta_{3,1} & \nu_3 & \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{array}
$$

从协方差可以递推计算系数 $\{\theta_{n,k}\}$ 和 $\{\nu_n\}$，并递推计算

$$
\begin{aligned}
\hat{Y}_1 =& 0, & W_1 =& Y_1 - \hat{Y}_1 \\
\hat{Y}_2 =& \theta_{1,1} W_1, & W_2 =& Y_2 - \hat{Y}_2 \\
\hat{Y}_3 =& \theta_{2,2} W_1 + \theta_{2,1} W_2, & W_3 =& Y_3 - \hat{Y}_3 \\
\hat{Y}_4 =& \theta_{3,3} W_1 + \theta_{3,2} W_2 + \theta_{3,1} W_3, & W_4 =& Y_4 - \hat{Y}_4 \\
& \dots\dots\dots & & \dots\dots\dots
\end{aligned}
$$

证明

从自协方差矩阵(24.6)的正定性知道 $\nu_n = EW_{n+1}^2 > 0$. 以下设 $0 \le k \le n-1$. 在(24.7)两边同乘 W_{k+1} 后求数学期望, 由 $\{W_k\}$ 的正交性得

$$E(\hat{Y}_{n+1}W_{k+1}) = \theta_{n,n-k}\nu_k. \tag{24.10}$$

利用 $W_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$ 和 W_{k+1} 垂直, 得

$$E(Y_{n+1}W_{k+1}) = E(\hat{Y}_{n+1}W_{k+1}) = \theta_{n,n-k}\nu_k \tag{24.11}$$

注意到

$$\hat{Y}_{k+1} = \sum_{j=1}^{k} \theta_{k,j} W_{k+1-j} = \sum_{j=0}^{k-1} \theta_{k,k-j} W_{j+1},$$

于是利用(24.11)可得

$$\begin{aligned}
\theta_{n,n-k} =& E(Y_{n+1}W_{k+1})/\nu_k \\
=& E\left[Y_{n+1}(Y_{k+1} - \sum_{j=0}^{k-1} \theta_{k,k-j} W_{j+1})\right]/\nu_k \\
=& \left[E(Y_{n+1}Y_{k+1}) - \sum_{j=0}^{k-1} \theta_{k,k-j} E(Y_{n+1}W_{j+1})\right]/\nu_k \\
=& [E(Y_{n+1}Y_{k+1}) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j]/\nu_k.
\end{aligned}$$

最后, 利用 $\nu_n = EW_{n+1}^2 = EY_{n+1}^2 - E\hat{Y}_{n+1}^2$ 和(24.7)得到预测的均方误差公式:

$$\nu_n = EY_{n+1}^2 - \sum_{j=1}^{n} \theta_{n,j}^2 \nu_{n-j} = EY_{n+1}^2 - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j.$$

24.1.3 多步预报问题

下面考虑用 $\{Y_1, Y_2, \cdots, Y_n\}$ 预测 Y_{n+k+1} 的问题. 设 $(Y_1, Y_2, \ldots, Y_{n+k+1})^T$ 的自协方差矩阵正定. 仍记 $\hat{Y}_{n+k+1} = L(Y_{n+k+1}|Y_{n+k})$, 用 W_j 表示预测误差 $Y_j - L(Y_j|Y_{j-1})$, 则有

$$\hat{Y}_{n+k+1} = \sum_{j=1}^{n+k} \theta_{n+k,j} W_{n+k+1-j}. \tag{24.12}$$

注意, 对 $j \geq 0$, W_{n+j+1} 垂直于 L_n, $W_{n-j} \in L_n$. 根据 §22.1中最佳线性预测的性质 4、5、8 或定理22.2得到

$$\begin{aligned}
&L(Y_{n+k+1}|Y_n) = L(\hat{Y}_{n+k+1}|Y_n) \\
=&L[\sum_{j=1}^{n+k}\theta_{n+k,j}W_{n+k+1-j} \Bigm| W_n] \\
=&L[\sum_{j=k+1}^{n+k}\theta_{n+k,j}W_{n+k+1-j} \mid W_n] \\
=&\sum_{j=k+1}^{n+k}\theta_{n+k,j}W_{n+k+1-j}. \qquad (24.13)\\
=&\sum_{j=0}^{n-1}\theta_{n+k,n+k-j}W_{j+1}
\end{aligned}$$

由投影的正交性, 得到预测的均方误差

$$\begin{aligned}
&E[Y_{n+k+1} - L(Y_{n+k+1}|Y_n)]^2 \\
=&EY_{n+k+1}^2 - E[L(Y_{n+k+1}|Y_n)]^2 \\
=&EY_{n+k+1}^2 - \sum_{j=k+1}^{n+k}\theta_{n+k,j}^2\nu_{n+k-j}, \qquad (24.14)\\
=&EY_{n+k+1}^2 - \sum_{j=0}^{n-1}\theta_{n+k,n+k-j}^2\nu_j^2
\end{aligned}$$

其中的系数 $\theta_{n+k,j}$ 和预测的均方误差 ν_{n+k-j} 可用递推公式(24.9)计算, 只不过因为 $Y_{n+1},\dots,Y_{n+k}$ 未知所以 $W_{n+1},\dots,W_{n+k}$ 不能计算。

24.2 正态时间序列的区间预测

如果 $\{Y_t\}$ 是正态时间序列, 则 $\hat{Y}_{n+1}$ 也是最佳预测. $W_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$ 作为 $Y_1, Y_2, \cdots, Y_{n+1}$ 的线性组合服从正态分布 $N(0, \nu_n)$. 利用

$$P\left(|Y_{n+1} - \hat{Y}_{n+1}|/\sqrt{\nu_n} \leq 1.96\right) = 0.95$$

可以得到 Y_{n+1} 的置信度为 0.95 的置信区间 (预测区间)

$$[\hat{Y}_{n+1} - 1.96\sqrt{\nu_n},\ \hat{Y}_{n+1} + 1.96\sqrt{\nu_n}]$$

24.3 平稳序列的递推预测

设 $\gamma_k = E(X_{t+k}X_t)$ 是零均值平稳序列 $\{X_t\}$ 的自协方差函数, Γ_n 是 $\{X_t\}$ 的 n 阶自协方差矩阵. 设 $X_n = (X_1, X_2, ..., X_n)^T$, $Z_n = X_n - L(X_n|X_{n-1})$, 可以把定理24.1改述如下.

推论 24.1. 设 $\{X_t\}$ 是零均值平稳序列, 对任何 $n \in \mathbb{N}$, 自协方差矩阵 Γ_n 正定. 则最佳线性预测

$$\begin{aligned}\hat{X}_{n+1} &\triangleq L(X_{n+1}|X_n)\\ &= \sum_{j=1}^{n} \theta_{n,j} Z_{n+1-j}, \qquad (24.15)\\ &= \sum_{j=0}^{n-1} \theta_{n,n-j} Z_{j+1} \;\; n = 1, 2, ...\end{aligned}$$

其中的系数 $\{\theta_{n,j}\}$ 和预测的均方误差 $\nu_n = EZ_{n+1}^2$ 满足如下的递推公式:

$$\begin{cases} \nu_0 = \gamma_0 \\ \theta_{n,n-k} = [\gamma_{n-k} - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j]/\nu_k, \\ \qquad\qquad 0 \le k \le n-1, \\ \nu_n = \gamma_0 - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j, \end{cases} \qquad (24.16)$$

其中 $\sum_{j=0}^{-1}(\cdot) \triangleq 0$, 递推的顺序是

$$\begin{array}{lllll} \nu_0 & & & & \\ \theta_{1,1} & \nu_1 & & & \\ \theta_{2,2} & \theta_{2,1} & \nu_2 & & \\ \theta_{3,3} & \theta_{3,2} & \theta_{3,1} & \nu_3 & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

由于预测误差 $Z_n = X_n - L(X_n|X_{n-1})$ 和 X_{n-1} 正交, 所以是不被 X_{n-1} 的线性组合包含的信息. 基于这个原因, 人们又称 Z_n 是**样本新息**. 从 §23.2的讨论知道,

$$\nu_n = E[X_1 - L(X_1|X_0, X_{-1}, \cdots, X_{-n+1})]^2 \to \sigma^2, \; (n \to \infty)$$

这里 σ^2 是用全体历史 $X_t, X_{t-1}, ...$ 预测 X_{t+1} 时的均方误差. $\sigma^2 > 0$ 表示 $\{X_t\}$ 是非决定性的.

Chapter 25

ARMA 序列的递推预测

新息预测方法从理论上很完美，对平稳列只需要知道自协方差函数列 $\{\gamma_k\}$，还可以预测非平稳列。但是对样本数据协方差需要估计。注意新息预测系数中需要用到 γ_n，但是如果样本量不够的话估计 γ_n 会产生很大误差，导致预测误差很大。如果我们知道序列服从 ARMA 模型，就可以用比较少的自协方差估计得到模型参数估计，用模型参数来得到预报公式，这样的预报结果受随机误差影响比较小。另外，ARMA 模型中的白噪声如果是独立白噪声，最佳线性预报还是最佳预报。这个结果虽然是对无穷长历史得到的但是样本量足够大时也近似成立。

25.1 AR 序列的预测

25.1.1 AR 序列的一步预测

设 $\{X_t\}$ 满足 AR(p) 模型

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{25.1}$$

其中 $\{\varepsilon_t\}$ 是零均值白噪声, 特征多项式

$$A(z) = 1 - \sum_{j=1}^{p} a_j z^j \neq 0, \quad |z| \leq 1$$

考虑用 $X_n = (X_1, X_2, \dots, X_n)^T$ 预测 X_{n+1} 的问题. 设 $\{\gamma_n\}$ 是 $\{X_t\}$ 的自协方差函数.

对于 $1 \le n \le p-1$, 由最佳线性预测的性质 1 知道

$$\begin{aligned}\hat{X}_{n+1} =& L(X_{n+1}|X_n) = \gamma_n^T \Gamma_n^{-1} X_n \\ =& a_{n,1}X_n + a_{n,2}X_{n-1} + \cdots + a_{n,n}X_1\end{aligned}$$

其中 Γ_n 是 $\{X_t\}$ 的 n 阶自协方差矩阵, $\gamma_n = (\gamma_n, \gamma_{n-1}, \dots, \gamma_1)^T$, $(a_{n,1}, a_{n,2}, \dots, a_{n,n})$ 是 n 阶 Yule-Walker 系数。预测的均方误差是

$$E(X_{n+1} - \hat{X}_{n+1})^2 = \gamma_0 - \gamma_n^T \Gamma_n^{-1} \gamma_n.$$

可以用 Levinson 递推公式计算线性组合系数系数和均方误差。(见 §24.3)

对于 $n \ge p$, 由于当 $k \ge 1$, $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ 与 ε_{t+k} 正交, 所以对 $n \ge p$

$$\begin{aligned}\hat{X}_{n+1} =& L(\varepsilon_{n+1} + \sum_{j=1}^{p} a_j X_{n+1-j}|X_n) \\ =& \sum_{j=1}^{p} a_j X_{n+1-j}\end{aligned}$$

由于 $X_{n+1} - \hat{X}_{n+1}$ 与 $X_n, \dots, X_1$ 正交，可见 $n \ge p$ 时

$$\hat{X}_{n+1} = L(X_{n+1}|X_n, X_{n-1}, \dots, X_{n-p+1}).$$

25.1.2 AR 序列的多步预测

下面考虑用 X_n 预测 $X_{n+k}(k \ge 1)$ 的问题. 对于 $n \ge p$, 对 k 用归纳法容易证明 (习题 4.1)

$$L(X_{n+k}|X_n) = L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-p+1}).$$

当 $n \ge p$ 时，记

$$\hat{X}_{n,m} = \begin{cases} L(X_m|X_n), & m > n \\ X_m, & m \le n \end{cases}$$

则

$$
\begin{aligned}
&L(X_{n+k}|X_n) = L(\sum_{j=1}^{p} a_j X_{n+k-j} + \varepsilon_{n+k}|X_n) \\
=&L(\sum_{j=1}^{p} a_j X_{n+k-j}|X_n) = \sum_{j=1}^{p} a_j \hat{X}_{n,n+k-j}, \\
&k = 1, 2, ...
\end{aligned}
$$

可以递推计算 $\hat{X}_{n,m}, m > n$。

25.1.3 例：AR(1) 预测

考虑 AR(1) 模型

$$X_n = a_1 X_{n-1} + \varepsilon_n$$

则对任何 $m \in \mathbb{N}$,

$$
\begin{aligned}
L(X_{n+1}|X_n, X_{n-1}, ..., X_{n-m+1}) =& a_1 X_n, \\
L(X_{n+2}|X_n, X_{n-1}, ..., X_{n-m+1}) =& a_1 L(X_{n+1}|X_n) \\
=& a_1^2 X_n, \\
& \cdots\cdots \\
L(X_{n+k}|X_n, X_{n-1}, ..., X_{n-m+1}) =& a_1^k X_n.
\end{aligned}
$$

当 $k \to \infty$, 利用 $|a_1| < 1$ 和控制收敛定理得到

$$
\begin{aligned}
&E[X_{n+k} - L(X_{n+k}|X_n, X_{n-1}, \cdots, X_{n-m})]^2 \\
=&E(X_{n+k} - a_1^k X_n)^2 = E(X_{n+k}^2) - a_1^{2k} E(X_n^2) \\
=&\gamma_0(1 - a_1^{2k}) \to \gamma_0 = EX_0^2,\ \text{当}\ k \to \infty.
\end{aligned}
$$

这与 AR(1) 序列是纯非决定性的平稳序列有关. 实际上任何 ARMA(p, q) 序列都是纯非决定性的 (见 §23.2.4.1).

25.1.4 例：降雨量预测

平均降雨量为 $\bar{X} = 540$mm. 用 $X_t, t = 1, 2, ...$ 表示该地区的逐年降雨量. 如果

$$Y_t = X_t - \bar{X}$$

满足 AR(2) 模型

$$Y_t = -0.54Y_{t-1} + 0.3Y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$$

给定观测 $X_1 = 560$, $X_2 = 470$, $X_3 = 580$, $X_4 = 496$, $X_5 = 576$, 求 X_8 的最佳线性预测.

首先中心化得,

$$\begin{aligned}
&Y_1 = 560 - 540 = 20,\ Y_2 = 470 - 540 = -70,\\
&Y_3 = 580 - 540 = 40,\ Y_4 = 496 - 540 = -44,\\
&Y_5 = 576 - 540 = 36.
\end{aligned}$$

用递推公式计算。记 $\hat{Y}_j = L(Y_j|Y_1, Y_2, ..., Y_5)$, $6 \leq j \leq 8$, 则有

$$\begin{aligned}
\hat{Y}_6 =& -0.54 \times 36 + 0.3 \times (-44) = -32.64,\\
\hat{Y}_7 =& -0.54\hat{Y}_6 + 0.3 \times 36 = 28.43\\
\hat{Y}_8 =& -0.54\hat{Y}_7 + 0.3\hat{Y}_6 = -25.14
\end{aligned}$$

最后加上平均值得

$$\begin{aligned}
\hat{X}_6 &= 540 - 32.64 = 507.36\\
\hat{X}_7 &= 540 + 28.43 = 568.43\\
\hat{X}_8 &= 540 - 25.14 = 514.86.
\end{aligned}$$

在本例中, 特征多项式 $A(z) = 1 + 0.54z - 0.3z^2$ 有两个实根 $z_1 = -1.1355$, $z_2 = 2.9355$. 最靠近单位圆的根的辐角是 π, 所以序列有周期 $T = 2$ 的特性. 预测数据也体现了围绕均值 540 上下交替变化的特性.

25.2 MA 序列的预测

设 $\{\varepsilon_t\}$ 是 $\mathrm{WN}(0, \sigma^2)$, 实系数多项式 $B(z) = 1 + b_1 z + \cdots + b_q z^q$ 在单位圆内无根:

$$B(z) \neq 0, |z| < 1$$

满足 MA(q) 模型

$$X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z}, \tag{25.2}$$

的 MA(q) 序列 $\{X_t\}$ 的自协方差函数 $\{\gamma_k\}$ 是 q 后截尾的. 假设 σ^2, $b_1, b_2, \ldots, b_q$ 已知, 我们考虑用 $X_1, X_2, \ldots, X_n$ 预测 X_{n+k} 的问题.

从(24.5)(新息预报) 可以看出, 对 $n \geq 1$

$$L_n = \overline{\text{span}}\{X_1, X_2, \cdots, X_n\} = \overline{\text{span}}\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\}, \tag{25.3}$$

这里 $\hat{\varepsilon}_n = X_n - L(X_n|X_{n-1})$ 是 $\{X_t\}$ 的逐步预测误差, $X_n = (X_1, \ldots, X_n)^T$.

以下假定 $n \geq q$. 由于 $\{\hat{\varepsilon}_t\}$ 是正交序列, 并由 q 步截尾性可知 X_{n+1} 与 L_{n-q} 正交，所以 X_{n+1} 与 $\hat{\varepsilon}_{n-q}, \hat{\varepsilon}_{n-q-1}, \ldots, \hat{\varepsilon}_1$ 正交，根据最佳线性预测的性质 6、4、10 得到

$$L(X_{n+1}|X_n) = L(X_{n+1}|\hat{\varepsilon}_n, \ldots, \hat{\varepsilon}_1) \quad (\text{性质 10 及 (4.3)}) \tag{25.4}$$

$$= L(X_{n+1}|\hat{\varepsilon}_n, \ldots, \hat{\varepsilon}_{n-q+1}) + L(X_{n+1}|\hat{\varepsilon}_{n-q}, \ldots, \hat{\varepsilon}_1) \quad (\text{性质 6}) \tag{25.5}$$

$$= L(X_{n+1}|\hat{\varepsilon}_n, \ldots, \hat{\varepsilon}_{n-q+1}) \quad (\text{性质 4}) \tag{25.6}$$

$$= \sum_{j=1}^{q} \theta_{n,j} \hat{\varepsilon}_{n+1-j}. \quad (\text{新息预测公式}) \tag{25.7}$$

预测的均方误差

$$\nu_n = E\hat{\varepsilon}_{n+1}^2 = \gamma_0 - \sum_{j=1}^{q} \theta_{n,j}^2 \nu_{n-j}. \tag{25.8}$$

这里的 $\{\theta_{n,j}\}$, ν_n 可以利用(24.16)进行递推计算, 但注意 $\{\gamma_k\}$ 是 q 步截尾的。

25.3 ARMA 序列的预测

对于满足 ARMA(p, q) 模型

$$A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z}, \tag{25.9}$$

的 ARMA 序列 $\{X_t\}$, 定义 $m = \max(p, q)$ 和

$$Y_t = \begin{cases} X_t/\sigma, & t = 1, 2, \ldots, m, \\ A(\mathscr{B})X_t/\sigma, & t = m+1, \ldots. \end{cases} \tag{25.10}$$

则 $\{Y_t\}$ 由 ARMA(p, q) 模型(25.9)的参数 $\beta = (a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q)^T$ 和标准白噪声 $\{\varepsilon_t/\sigma\}$ 决定, 从而不依赖 σ. 假设模型(25.9)中的参数已知, 我们考虑用 $X_1, X_2, \cdots, X_n$ 对 X_{n+k} 进行逐步预测的问题.

从 Y_t 的定义知道, 对 $t \geq 1$, $Y_t \in L_t = \overline{\text{span}}\{X_1, X_2, \dots, X_t\}$, 并且 $X_1, X_2, \dots, X_m \in \overline{\text{span}}\{Y_1, Y_2, \dots, Y_m\}$. 容易看出, 对 $t > m$,

$$X_t = \sigma Y_t + \sum_{j=1}^{p} a_j X_{t-j} \ \in \ \overline{\text{span}}\{Y_1, Y_2, \dots, Y_t\}.$$

于是再利用(24.5)(新息与原序列互相线性表示) 得到

$$\begin{aligned} L_n =& \overline{\text{span}}\{X_1, X_2, \dots, X_n\} \\ =& \overline{\text{span}}\{Y_1, Y_2, \dots, Y_n\} = \overline{\text{span}}\{W_1, W_2, ..., W_n\}, \end{aligned}$$

其中 $W_t = Y_t - L(Y_t|Y_{t-1})$, $W_1 = Y_1$ 是 $\{Y_t\}$ 的样本新息.

用 γ_k 表示 $\{X_t\}$ 的自协方差函数, 取 $b_0 = 1$, $b_j = 0$, 当 $j > q$. 可以计算出

$$E(Y_s Y_t) = \begin{cases} \sigma^{-2}\gamma_{t-s}, & 1 \leq s \leq t \leq m, \\ \sigma^{-2}[\gamma_{t-s} - \sum_{j=1}^{p} a_j \gamma_{t-s-j}], & 1 \leq s \leq m < t, \\ \sum_{j=0}^{q} b_j b_{j+t-s}, & t \geq s > m, \end{cases} \tag{25.11}$$

其中的 $\{\gamma_k\}$ 可用 §13.3的(13.9)和(13.10)计算.

定义 $\{X_t\}$ 的逐步预测误差

$$Z_t = X_t - L(X_t|X_{t-1}),\ Z_1 = X_1,$$

则对 $1 \leq t \leq m$,

$$\begin{aligned} W_t =& X_t/\sigma - L(X_t/\sigma|X_{t-1}) \\ =& \sigma^{-1}[X_t - L(X_t|X_{t-1})] = \sigma^{-1} Z_t \end{aligned}$$

对 $t \geq m+1$,

$$\begin{aligned} W_t =& \sigma^{-1}[X_t - \sum_{j=1}^{p} a_j X_{t-j} - L(X_t - \sum_{j=1}^{p} a_j X_{t-j}|X_{t-1})] \\ =& \sigma^{-1}[X_t - L(X_t|X_{t-1})] = \sigma^{-1} Z_t \end{aligned}$$

所以 $\{X_t\}$ 和 $\{Y_t\}$ 的预测误差之间总有以下的关系

$$Z_t = \sigma W_t,\ EZ_t^2 = \sigma^2 EW_t^2,\ t = 1, 2, \dots \tag{25.12}$$

以下仍用 ν_{t-1} 表示 EW_t^2, 就有 $EZ_t^2 = \sigma^2\nu_{t-1}$. 对于 $1 \le n < m = \max(p,q)$, 从逐步预测公式(24.7)得到

$$\hat{Y}_{n+1} = L(Y_{n+1}|\boldsymbol{Y}_n) = \sum_{j=1}^{n} \theta_{n,j} W_{n+1-j}.$$

于是对 $1 \le n < m$,

$$\begin{aligned}\hat{X}_{n+1} =& L(X_{n+1}|\boldsymbol{X}_n) = L(\sigma Y_{n+1}|\boldsymbol{Y}_n) \\ =& \sigma\hat{Y}_{n+1} = \sum_{j=1}^{n} \theta_{n,j}\sigma W_{n+1-j} = \sum_{j=1}^{n} \theta_{n,j} Z_{n+1-j}.\end{aligned}$$

对于 $n \ge m$, 利用 ARMA 序列的因果性即 $E(X_t \varepsilon_{t+k}) = 0$, $k = 1, 2, \dots$, 得到

$$Y_{n+1} = \sigma^{-1} B(\mathscr{B})\varepsilon_{n+1} = \sigma^{-1} \sum_{j=0}^{q} b_j \varepsilon_{n+1-j}$$

与 $\{X_j : 1 \le j \le n-q\}$ 正交, 从而与

$$\overline{\text{span}}\{Y_j : 1 \le j \le n-q\} = \overline{\text{span}}\{W_j : 1 \le j \le n-q\}$$

中的任何随机变量正交.

利用最佳线性预测的性质 6、4 得到

$$L(Y_{n+1}|\boldsymbol{Y}_n) = \sum_{j=1}^{q} \theta_{n,j} W_{n+1-j}, \ n \ge m = \max(p,q).$$

于是对 $n \ge m$, 由 Y 和 X 的关系得到

$$\begin{aligned}\hat{X}_{n+1} =& L(\sum_{j=1}^{p} a_j X_{n+1-j} + \sigma Y_{n+1}|\boldsymbol{X}_n) \\ =& \sum_{j=1}^{p} a_j X_{n+1-j} + \sigma \sum_{j=1}^{q} \theta_{n,j} W_{n+1-j} \\ =& \sum_{j=1}^{p} a_j X_{n+1-j} + \sum_{j=1}^{q} \theta_{n,j} Z_{n+1-j}.\end{aligned}$$

总结上述推导得到:

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{n,j} Z_{n+1-j}, & 1 \le n < m, \\ \sum_{j=1}^{p} a_j X_{n+1-j} + \sum_{j=1}^{q} \theta_{n,j} Z_{n+1-j}, & n \ge m. \end{cases} \tag{25.13}$$

预测的均方误差仍然是 $EZ_{n+1}^2 = \sigma^2\nu_n$. 这里的 $\{\theta_{n,j}\}$, $\nu_n = EW_{n+1}^2$ 可以利用(25.11)和(24.9)进行递推计算. $Z_{n+1} = X_{n+1} - \hat{X}_{n+1}$ 也可以递推计算。

因为 $\{Y_t\}$ 和 σ 无关, 所以 $\theta_{n,k}$, $\{W_n\}$ 以及 ν_{n-1} 都是和 σ^2 无关的量. 它们只依赖于参数 $a = (a_1, a_2, \dots, a_p)^T$ 和 $b = (b_1, b_2, \dots, b_q)^T$. 这个性质在研究 ARMA 模型的最大似然估计时将得到应用.

25.3.1 预报例子

考虑 §13.7.1和 §22.1.12的 ARMA(4,2) 模型。

$$\begin{aligned} X_t =& -0.9X_{t-1} - 1.4X_{t-2} - 0.7X_{t-3} - 0.6X_{t-4} \\ &+ \varepsilon_t + 0.5\varepsilon_{t-1} - 0.4\varepsilon_{t-2},\ \varepsilon_t \sim \text{WN}(0,1), \end{aligned}$$

已计算自协方差函数 $\gamma_0, \gamma_1, \dots, \gamma_{20}$, 由此用(25.11)计算出 $E(Y_t Y_S), (1 \le s, t \le 21)$. 用 §24.1的公式(24.9)计算出新息预报系数 $\theta_{n,j}$ 如下：

n	1	2	3	4	$\cdots$	19	20
$\theta_{n,1}$	−0.226	−0.4017	−0.5705	0.1807	$\cdots$	0.4875	0.489
$\theta_{n,2}$	0	−0.6865	−0.6353	−0.1597	$\cdots$	−0.3937	−0.394
$\theta_{n,3}$	0	0	0.3699	−0.0000	$\cdots$	−0.0001	0.000
$\theta_{n,4}$	0	0	0	−0.0000	$\cdots$	0.0000	−0.000

该模型的观测数据 $x_1, \cdots, x_{21}$ 为:

−0.4587	0.7125	1.9948	−4.5285	−0.7514	5.8782	−0.1273
−2.9223	−0.7581	1.1422	2.1107	−0.5640	−2.4452	−0.5105

利用公式(25.13)和(24.9)可以计算出逐步预测 $\hat{X}_{k+1} = L(X_{k+1}|X_k)$ 和逐步预测的均方误差 $\nu_{k-1} = EW_k^2$ 如下:

j	$\hat{X}_j$	ν_{j-1}
1	0	6.670
2	0.104	6.330
3	0.070	2.505
4	-1.654	2.387
5	0.232	1.268
6	5.385	1.233
7	-1.788	1.142
8	-4.398	1.114
9	-0.837	1.086
10	0.839	1.069
11	2.259	1.056
12	-1.395	1.046
13	-2.354	1.038
14	0.467	1.031
15	2.585	1.026
16	1.069	1.022
17	-1.668	1.018
18	0.161	1.016
19	5.725	1.013
20	-0.229	1.011
21	-3.468	1.010

从以上数据看出 ν_k 收敛到 $\sigma^2 = 1$ 的速度是较理想的 (参见习题 4.4). 由于 $Z_k = X_k - \hat{X}_k$ 服从正态分布 $(0, \nu_{k-1})$, 所以真值 x_t 的置信度为 0.95 的置信下、上限分别是

$$\hat{X}_k - 1.96\sqrt{\nu_{k-1}}, \quad \hat{X}_k + 1.96\sqrt{\nu_{k-1}}.$$

下面的 R 函数 `ts.ipred.coef()` 利用协方差阵 Γ_n（对非平稳列）或者自协方差函数列 $\{\gamma_k\}$（对平稳列）求递推预测系数 $\{\theta_{nj}\}$ 和方差 ν_j，`ts.ipred()` 做递推预测。这里还收录了计算 ARMA 模型 Wold 系数和理论自协方差函数的函数，用 Levinson 递推求解预测系数的函数，Levinson 递推方法方法仅适用于平稳列。

```r
## Wold coefficients for the ARMA model
arma.Wold <- function(n, a, b=numeric(0)){
  p <- length(a)
  q <- length(b)
  arev <- rev(a)
  psi <- numeric(n)
  psi[1] <- 1
  for(j in seq(n-1)){
    if(j <= q) bj=b[j]
    else bj=0
    psis <- psi[max(1, j+1-p):j]
    np <- length(psis)
    if(np < p) psis <- c(rep(0,p-np), psis)
    psi[j+1] <- bj + sum(arev * psis)
  }

  psi
}

## Calculate theoretical autocovariance function
## of ARMA model using Wold expansion
arma.gamma.by.Wold <- function(n, a, b=numeric(0), sigma=1){
  nn <- n + 100
  psi <- arma.Wold(nn, a, b)
  gam <- numeric(n)
  for(ii in seq(0, n-1)){
    gam[ii+1] <- sum(psi[1:(nn-ii)] * psi[(ii+1):nn])
  }
  gam <- (sigma^2) * gam
  gam
}
arma.gamma <- arma.gamma.by.Wold
```

```
### 非平稳时间序列新息递推预测的系数和一步预测误差方差计算。
### 输入:
###   Gam --- 如果序列非平稳，为自协方差矩阵 n*n
###           如果序列平稳，为自协方差列\gamma_0, \gamma_1, \dots, \gamma_{n-1}
### 输出:
###  用新息预报方法对 Y_1, Y_2, \dots, Y_n 进行一步最佳线性预报的系数和均方误差。
###  theta -- (n-1)*(n-1) 矩阵。第 k 行为预报 Y_{k+1} 的样本新息系数:
###     \hat Y_{k+1} = theta[k,1]*W_k + theta[k,2]*W_{k-1} + \dots + theta[k,k]*W_1
###     W_1=Y_1, W_k = Y_k - \hat Y_k, k=2,3,\dots,n
###  nu --- n 个元素，nu[k] 为预报 Y_k 的均方误差。
ts.ipred.coef <- function(Gam){
  if(!is.matrix(Gam)){
    n <- length(Gam)
    Gam <- outer(1:n, 1:n,
                 function(i,j) Gam[abs(i-j)+1])
  } else {
    n <- nrow(Gam)
  }
  ## X_1, X_2, \dots, X_{n-1} 的自协方差矩阵为 Gam

  ## 新息预报递推
  theta <- matrix(0, n-1, n-1)
  nu <- numeric(n)
  nu[1] <- Gam[1,1]
  theta[1,1] <- Gam[2,1] / Gam[1,1]
  nu[2] <- Gam[2,2] - theta[1,1]^2 * nu[1]
  for(k in 2:(n-1)){
    theta[k, k] <- Gam[k+1,0+1] / nu[1]
    for(j in 1:(k-1)){
      theta[k,k-j] <- (Gam[k+1,j+1]
                      - sum(theta[j, j:1]*theta[k,k:(k-j+1)]*nu[1:j])) /
                        nu[j+1]
    }
```

```
    nu[k+1] <- Gam[k+1,k+1] - sum(theta[k,k:1]^2 * nu[1:k])
  }
  list(theta=theta, nu=nu)
}

## 时间序列递推预报 (新息预报)，假定已知自协方差列（平稳时）或自协方差阵（非平稳时）
## 输入:
##  x -- 时间序列。可以包含要预报的部分作为对比。
##  gams -- 向量或矩阵。向量时为自协方差列，矩阵时为自协方差阵
##  demean -- 是否减去均值再预报
##  end -- 递推预报是对每个时间点都进行预报，end 是预报到那个时间点，
##         这受到 gams 大小的限制，只能预报到与自协方差阵阶数相同的时间点
## conf -- 计算预报区间的置信度
ts.ipred <- function(x, gams, demean=FALSE, end=length(x), conf=0.95){
  stationary <- !is.matrix(gams)
  res1 <- ts.ipred.coef(gams)
  theta <- res1$theta
  nu <- res1$nu

  if(demean) {xmean <- mean(x); x <- x - xmean}

  if(!stationary) end <- length(nu)
  W <- numeric(end)
  pred <- numeric(end)
  lb <- numeric(end)
  ub <- numeric(end)
  lam <- qnorm(1 - (1-conf)/2)

  pred[1] <- 0
  W[1] <- x[1]
  for(n in 1:(end-1)){
    pred[n+1] <- sum(theta[n, n:1]*W[1:n])
    W[n+1] <- x[n+1] - pred[n+1]
```

```
  }
  lb <- pred - lam*sqrt(nu[1:end])
  ub <- pred + lam*sqrt(nu[1:end])

  if(demean){
    pred <- xmean + pred
    lb <- xmean + lb
    ub <- xmean + ub
  }
  list(x=x, pred=pred, lb=lb, ub=ub, conf=conf)
}

### 对平稳列，已知自协方差列时用 Levinson 递推计算
### 逐个一步预报系数 (Y-W 系数) 和一步预测误差方差
### 输入: gams[1:n] 为\gamma_k, k=0,1,\dots,n-1
### 输出: 预报 Y_1, Y_2, \dots, Y_{n} 所需的系数和均方误差。
###   coef.YW -- (n-1)*(n-1) 矩阵，第 k 行的 1:k 元素为
###         a_{k,1}, a_{k,2}, \dots, a_{k,k},
###         用来预报 Y_{k+1}:
###         L(Y_{k+1} | Y_1, Y_2, \dots, Y_k)
###       = a_{k,1} Y_{k} + a_{k,2} Y_{k-1} + \dots + a_{k,k} Y_1
###       k=1,2,\dots,n-1
###       L(Y_1| ) = 0
###   sigmasq -- n 向量，预报 Y_1, Y_2, \dots, Y_n 的一步预报均方误差
###       其 1 号元素是预报 Y_1 的均方误差，2 号元素是预报 Y_2 的均方误差，
###       ..., n 号元素是预报 Y_n 的均方误差。
Levinson.coef <- function(gams){
  n <- length(gams)

  ## ayw 保存 Y-W 系数 a[i,j], i=1,2,\dots,n-1, j=1,2,\dots,i
  ayw <- matrix(0, n-1, n-1)
  ## ss 保存一步预报误差方差
  ss <- numeric(n)
```

```
  ss[1] <- gams[0+1] ## 预报 Y_1 的均方误差，等于\gamma_0
  ayw[1,1] <- gams[1+1] / gams[0+1] ## 用 Y_1 预报 Y_2 的系数
  ss[2] <- ss[1] * (1 - ayw[1,1]^2) ## 用 Y_1 预报 Y_2 的均方误差
  if(n>2) for(k in 1:(n-2)){
    ## 用 Y_1, \dots, Y_{k+1} 预报 Y_{k+2} 的系数
    ayw[k+1,k+1] <- (gams[(k+1)+1] - sum(ayw[k,1:k] * gams[(k:1)+1])) / ss[k+1]
    ayw[k+1,1:k] <- ayw[k, 1:k] - ayw[k+1,k+1]*ayw[k, k:1]
    ## 用 Y_1, \dots, Y_{k+1} 预报 Y_{k+2} 的均方误差
    ss[k+2] <- ss[k+1] * (1 - ayw[k+1,k+1]^2)
  }

  list(coef.YW=ayw, sigmasq=ss)
}

### 平稳时间序列已知自协方差列，用 Levinson 递推计算 Y-W 系数并逐步预报。
### 输入：
###  x --- 序列 X_1, \dots, X_n
###  gams --- 理论自协方差列\gamma_0, \gamma_1, \dots, \gamma_{m-1}, m <= n
###  max.lag --- 最多使用多少个历史值预报。缺省使用所有历史值，或理论自协方差列个数减一
###  end --- 从时刻 1 预测到时刻 end，缺省对每个时间点做一步预报。
###  demean --- 预报时是否中心化
###  conf --- 计算预测区间时的置信度。
### 输出：
###  x --- n 个元素，原始序列
###  pred --- n 个元素，每一步的预报值
###  lb, ub --- 对应的预测下限和上限。
sts.pred.levinson <- function(x, gams, max.lag=length(gams)-1,
                              end=length(x),
                              demean=FALSE, conf=0.95){
  n <- length(x)
  m <- length(gams)
  if(max.lag > m-1){
```

```
    msg <- paste('可用最大历史长度', max.lag,
                 '大于可用理论自协方差个数减一', m-1)
    stop(msg)
  }
  if(demean) {
    xmean <- mean(x)
    x[] <- x - xmean
  }

  pred <- numeric(end)
  lb <- numeric(end)
  ub <- numeric(end)
  lam <- qnorm(1 - (1-conf)/2)

  res1 <- Levinson.coef(gams)
  ayw <- res1$coef.YW ## (m-1)*(m-1) 矩阵，第 k 行用于 k 个历史进行一步预报。
  ss <- res1$sigmasq  ## 长度为 m 的向量，第 k 个为用 k-1 个历史预报下一步的均方误差。

  pred[1] <- 0
  for(k in seq(end-1)){
    if(k <= max.lag) {
      pred[k+1] <- sum(ayw[k, 1:k] * x[k:1])
    } else {
      pred[k+1] <- sum(ayw[max.lag, 1:max.lag] * x[k:(k-max.lag+1)])
    }
  }
  if(demean) pred[] <- pred + xmean

  ss2 <- c(ss[1:min(end,max.lag+1)],
           rep(ss[max.lag+1], end-min(end,max.lag+1)))

  lb[] <- pred - lam*sqrt(ss2)
  ub[] <- pred + lam*sqrt(ss2)
```

```
  list(x=x, pred=pred, lb=lb, ub=ub, conf=conf)
}
```

下面的 R 函数生成 ARMA(4,2) 的长度为 21 的模拟数据，但是用理论自协方差函数计算递推预测系数，对最后的 7 个点作一步最佳线性预测，计算预测区间，并与直接用 Levinson 递推得到 YW 方程解的方法进行比较，结果一致。注意这与 §22.1.12的预测不太一样，那里是基于前 14 个点对后 7 个点作多步预测。一步预测误差一般低于多步预测误差。

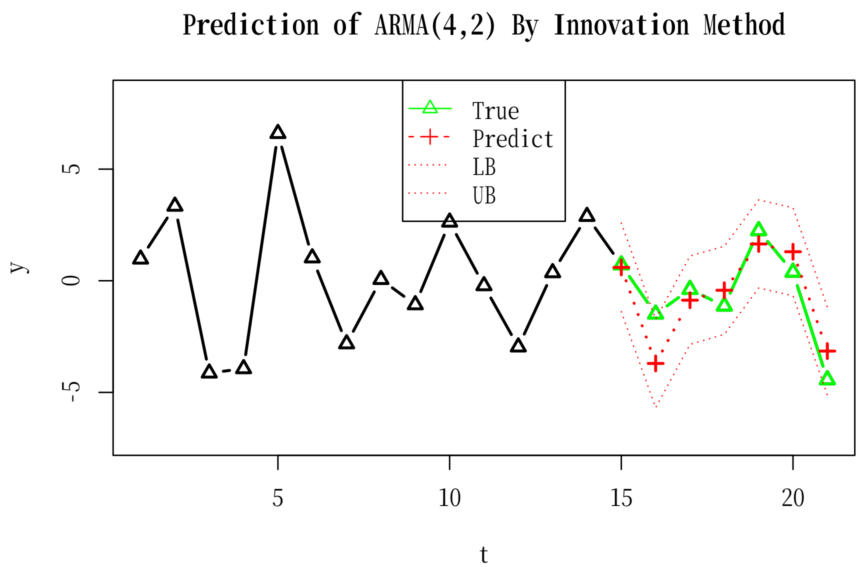
```
## 从 ARMA(4,2) 的模拟样本用理论自协方差
## 和一般新息预报公式进行递推预报,
## 并与用 Levinson 递推公式解 Y-W 方程的一步预报对比。
demo.ipred.arma42 <- function(){
  a <- c(-0.9, -1.4, -0.7, -0.6)
  b <- c(0.5, -0.4)
  ng <- 21
  n <- 21
  gams <- arma.gamma(ng, a, b, sigma=1)

  x <- arima.sim(model=list(ar=a, ma=b), n=n)

  n.old <- 14  ## 不预报这些项
  m.pred <- 7  ## 对 15--21 进行一步预报

  ## 用理论自协方差函数计算递推预测系数
  res0 <- ts.ipred.coef(gams)
  ##cat("==== Iterative predition coefficients:\n")
  ##print(round(res0[["theta"]], 2))
  ##cat("==== Iterative predition MSEs:\n")
  ##print(round(res0[["nu"]], 2))
  ## 用理论自协方差函数计算 YW 方程解
  res0b <- Levinson.coef(gams)
  ##cat("==== Levinson YW solution predition coefficients:\n")
```

```
##print(round(resOb[["coef.YW"]], 2))
##cat("==== Levinson YW solution predition MSEs:\n")
##print(round(resOb[["sigmasq"]], 2))

res1 <- ts.ipred(x, gams, demean=FALSE, end=n, conf=0.95)
pred <- res1$pred
lb <- res1$lb
ub <- res1$ub

res2 <- sts.pred.levinson(x=x, gams=gams, demean=FALSE, conf=0.95)

cat("==== Iterative predition and Levinson YW predition:\n")
print(round(cbind(x, pred, res2$pred, lb, res2$lb, ub, res2$ub), 4))

yl <- range(c(x, pred, lb, ub))
plot(1:n, x, type="n",
     main="Prediction of ARMA(4,2) By Innovation Method",
     xlab="t", ylab="y",
     ylim=yl)
lines(1:n, x[1:n],
      type='b',
      col="black",
      lty=1, pch=2,
      lwd=2)  ## 不预报的真实值
lines((n.old+1):(n.old+m.pred), x[(n.old+1):(n.old+m.pred)],
      type='b',
      col="green",
      lty=1, pch=2,
      lwd=2) ## 预报部分的真实值
lines((n.old+1):(n.old+m.pred), pred[(n.old+1):(n.old+m.pred)],
      type="b",
      col="red",
      lty=3, pch=3,
```

```r
        lwd=2)  ## 预报值
  lines((n.old+1):(n.old+m.pred), lb[(n.old+1):(n.old+m.pred)],
        col="red", lty=3, lwd=1)  ## 预报下界
  lines((n.old+1):(n.old+m.pred), ub[(n.old+1):(n.old+m.pred)],
        col="red", lty=3, lwd=1)  ## 预报上界
  legend("top",
         col=c('green', 'red', 'red', 'red'),
         pch=c(2,3,NA,NA),
         lty=c(1,2,3,3),
         legend=c('True', 'Predict', 'LB', 'UB'))
  invisible()
}
set.seed(1)
demo.ipred.arma42()
```

```
## ==== Iterative predition and Levinson YW predition:
## Time Series:
## Start = 1
## End = 21
## Frequency = 1
##           x    pred res2$pred      lb res2$lb      ub res2$ub
##  1  0.9736  0.0000    0.0000 -5.0622 -5.0622  5.0622  5.0622
##  2  3.3414 -0.2200   -0.2200 -5.1512 -5.1512  4.7111  4.7111
##  3 -4.1315 -2.0990   -2.0990 -5.2016 -5.2016  1.0037  1.0037
##  4 -3.9365 -0.7431   -0.7431 -3.7718 -3.7718  2.2855  2.2855
##  5  6.5983  6.1507    6.1507  3.9436  3.9436  8.3578  8.3578
##  6  1.0273  1.1262    1.1262 -1.0507 -1.0507  3.3032  3.3032
##  7 -2.8216 -5.1024   -5.1024 -7.1975 -7.1975 -3.0074 -3.0074
##  8  0.0599 -0.2545   -0.2545 -2.3236 -2.3236  1.8145  1.8145
##  9 -1.0723 -1.4527   -1.4527 -3.4955 -3.4955  0.5901  0.5901
## 10  2.6286  2.2890    2.2890  0.2620  0.2620  4.3160  4.3160
## 11 -0.2186  0.7958    0.7958 -1.2182 -1.2182  2.8099  2.8099
## 12 -2.9626 -3.3527   -3.3527 -5.3572 -5.3572 -1.3482 -1.3482
```

```
## 13  0.3625  2.3391   2.3391  0.3423  0.3423  4.3359  4.3359
## 14  2.8848  1.3270   1.3270 -0.6637 -0.6637  3.3178  3.3178
## 15  0.7091  0.5973   0.5973 -1.3885 -1.3885  2.5831  2.5831
## 16 -1.4901 -3.7039  -3.7039 -5.6857 -5.6857 -1.7221 -1.7221
## 17 -0.3880 -0.8703  -0.8703 -2.8487 -2.8487  1.1081  1.1081
## 18 -1.1331 -0.4252  -0.4252 -2.4008 -2.4008  1.5505  1.5505
## 19  2.2462  1.6478   1.6478 -0.3256 -0.3256  3.6211  3.6211
## 20  0.3857  1.3009   1.3009 -0.6705 -0.6705  3.2722  3.2722
## 21 -4.4308 -3.1497  -3.1497 -5.1194 -5.1194 -1.1800 -1.1800
```



25.4 ARMA 序列多步预测

设 $n > m$,

$$
\begin{aligned}
&L(X_{n+k+1}|X_n)\\
=&\begin{cases}
\sum_{j=1}^{p} a_j L(X_{n+k+1-j}|X_n) + \sum_{j=k+1}^{q} \theta_{n+k,j} Z_{n+k+1-j}, & 0 \le k < q, n > m\\
\sum_{j=1}^{p} a_j L(X_{n+k+1-j}|X_n), & k \ge q, n > m.
\end{cases}
\end{aligned}
$$

25.5 ARIMA 序列预测

模型

$$A(\mathscr{B})(1-\mathscr{B})^d X_t = B(\mathscr{B})\varepsilon_t, \quad \varepsilon_t \sim \mathrm{WN}(0,\sigma^2), \quad t \in N.$$

则 $Y_t = (1-\mathscr{B})^d X_t,\ t = d+1, d+2, \cdots, n$ 满足 ARMA(p,q) 模型

$$A(\mathscr{B})Y_t = B(\mathscr{B})\varepsilon_t, \quad t \in N.$$

考虑用 $X_1, X_2, \cdots, X_n$ 预测 X_{n+k} 的问题.

首先利用 ARMA 序列的预测方法得到用 $Y_{d+1}, Y_{d+2}, \cdots, Y_n$ 预测

$$Y_{n+1}, Y_{n+2}, \cdots, Y_{n+k}$$

时的最佳线性预测

$$\tilde{Y}_{n+j} = L(Y_{n+j}|Y_{d+1}, Y_{d+2}, \cdots, Y_n),\ j = 1, 2, \cdots, k.$$

再由公式

$$(1-\mathscr{B})^d \hat{X}_t = \tilde{Y}_t,\ t = n+1, n+2, \cdots, n+k$$

得到近似的递推公式:

$$\hat{X}_{n+k} = \tilde{Y}_{n+k} - \sum_{j=1}^{d} C_d^j (-1)^j \hat{X}_{n+k-j},\ k \geq 1,$$

其中 $\hat{X}_{n-j} = X_{n-j}$, 当 $j \geq 0$.

25.6 附录：ARMA 预报的补充

AR(p) 观测值为 $X_1, X_2, \dots, X_n$，假设已知理论参数，进行逐步一步拟合和预报:

$$\begin{aligned}\hat{X}_t =& L(X_t|X_{t-1}, \dots, X_1),\ t = 1, 2, \dots, n \\ \hat{X}_{n+k} =& L(X_{t+k}|X_1, X_2, \dots, X_n),\ k = 1, 2, \dots\end{aligned}$$

- 前 p 个用一般 Levinson 递推得到 Y-W 系数和方差；
- 后 $n-p$ 个用 AR 系数预报;
- 从第 $n+1$ 开始用递推预报。

25.6.1 思考：AR 序列有限历史最佳线性预测多步预测的均方误差

多步预报的均方误差？

$$
\begin{aligned}
&L(X_{t+k}|X_1, \dots, X_n) \\
=&L(X_{t+k}|X_n, X_{n-1}, \dots, X_{n-p+1}
\end{aligned}
$$

解方程

$$
\Gamma_p b = (\gamma_k, \gamma_{k+1}, \dots, \gamma_{k+p-1})^T
$$

得 k 步预报均方误差

$$
\sigma_k^2 = \gamma_0 - b_1\gamma_k - b_2\gamma_{k+1} - \cdots - b_p\gamma_{k+p-1}
$$

有没有简化公式？

25.6.2 MA 模型的新息问题

MA(q) 序列满足自协方差函数 q 后截尾。在零均值情况下，X_{n+1} 与 $X_1, X_2, \dots, X_{n-q}$ 不相关，即正交。那么，$L(X_{n+1}|X_n, \dots, X_1)$ 能否表示为 X_n, X_{n-1}, X_{n-q+1} 的线性组合？如果这样，MA 的新息估计法就是不必要的。

以 MA(1) 为例。模型为

$$
X_t = \varepsilon_t + b\varepsilon_{t-1}
$$

其中 $|b| < 1$, $\{\varepsilon_t\}$ 是 WN(0, σ^2)。

令

$$
\hat\varepsilon_n = X_n - L(X_n|X_1, \dots, X_{n-1}),\ n = 1, 2, \dots
$$

则 $\{\hat\varepsilon_t\}$ 是正交随机变量列，$L(X_1, \dots, X_n) = L(\hat\varepsilon_1, \dots, \hat\varepsilon_n)$。于是由投影性质，

$$
\begin{aligned}
\hat X_{n+1} =& L(X_{n+1}|X_1, \dots, X_n) = L(X_{n+1}|\hat\varepsilon_1, \dots, \hat\varepsilon_n) \\
=& \sum_{t=1}^n L(X_{n+1}|\hat\varepsilon_t)
\end{aligned}
$$

但是，因为对 $t \leq n-1$，X_{n+1} 与 X_t 正交（不相关），所以 X_{n+1} 与 $L(X_1, \dots, X_{n-1})$ 正交，X_{n+1} 与 $L(\hat\varepsilon_1, \dots, \hat\varepsilon_{n-1})$ 正交，由投影性质可知 $t \leq n-1$ 时 $L(X_{n+1}|\hat\varepsilon_t) = 0$，于是

$$
\hat X_{n+1} = L(X_{n+1}|\hat\varepsilon_n) = \theta_{n,n}\hat\varepsilon_n
$$

其中 $\theta_{n,j}$ 是新息递推预报系数。

对 MA(1) 序列，$\gamma_0 = \sigma^2(1+b^2)$，$\gamma_1 = \sigma^2(2b)$。如果直接计算 $L(X_{n+1}|X_1,\dots,X_n)$，需要求解线性方程组

$$\begin{pmatrix} 1+b^2 & 2b & 0 & \cdots & 0 & 0 & 0 \\ 2b & 1+b^2 & 2b & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2b & 1+b^2 & 2b \\ 0 & 0 & 0 & \cdots & 0 & 2b & 1+b^2 \end{pmatrix} a_n = \begin{pmatrix} 2b \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

其中 $a_n = (a_{n1},\dots,a_{nn})^T$，$L(X_{n+1}|X_1,\dots,X_n) = a_{n1}X_n + \cdots + a_{nn}X_1$。$a_n$ 并不是只有 a_{n1} 一个非零元素。

思考题：利用 Levinson 递推公式求出 a_n 的表达式。

下面用 R 程序对一些特殊 b 值求解 a_n。

```
Levinson.coef <- function(gams){
  n <- length(gams)

  ## ayw 保存 Y-W 系数 a[i,j], i=1,2,\dots,n-1, j=1,2,\dots,i
  ayw <- matrix(0, n-1, n-1)
  ## ss 保存一步预报误差方差
  ss <- numeric(n)

  ss[1] <- gams[0+1] ## 预报 Y_1 的均方误差，等于\gamma_0
  ayw[1,1] <- gams[1+1] / gams[0+1] ## 用 Y_1 预报 Y_2 的系数
  ss[2] <- ss[1] * (1 - ayw[1,1]^2) ## 用 Y_1 预报 Y_2 的均方误差
  if(n>2) for(k in 1:(n-2)){
    ## 用 Y_1, \dots, Y_{k+1} 预报 Y_{k+2} 的系数
    ayw[k+1,k+1] <- (gams[(k+1)+1] - sum(ayw[k,1:k] * gams[(k:1)+1])) / ss[k+1]
    ayw[k+1,1:k] <- ayw[k, 1:k] - ayw[k+1,k+1]*ayw[k, k:1]
    ## 用 Y_1, \dots, Y_{k+1} 预报 Y_{k+2} 的均方误差
    ss[k+2] <- ss[k+1] * (1 - ayw[k+1,k+1]^2)
  }
```

```
  list(coef.YW=ayw, sigmasq=ss)
}

demo.ma1.yw <- function(b=0.5, maxn=10){
  gams <- c(1+b^2, 2*b, rep(0.0, maxn-1))
  res <- Levinson.coef(gams)
  print(round(res$coef.YW, 2))
}
demo.ma1.yw(b=0.5, maxn=10)
```

```
##        [,1]   [,2]  [,3]  [,4]   [,5]  [,6] [,7]  [,8]  [,9] [,10]
##  [1,]  0.80   0.00  0.00  0.00   0.00  0.00 0.00  0.00  0.00  0.00
##  [2,]  2.22  -1.78  0.00  0.00   0.00  0.00 0.00  0.00  0.00  0.00
##  [3,] -1.03   2.29 -1.83  0.00   0.00  0.00 0.00  0.00  0.00  0.00
##  [4,]  0.44   0.45 -1.00  0.80   0.00  0.00 0.00  0.00  0.00  0.00
##  [5,]  1.23  -0.54 -0.56  1.24  -0.99  0.00 0.00  0.00  0.00  0.00
##  [6,] 58.31 -71.89 31.55 32.45 -72.11 57.69 0.00  0.00  0.00  0.00
##  [7,] -0.02   1.02 -1.26  0.55   0.57 -1.26 1.01  0.00  0.00  0.00
##  [8,]  0.79   0.01 -0.81  0.99  -0.44 -0.45 1.00 -0.80  0.00  0.00
##  [9,]  2.17  -1.71 -0.03  1.75  -2.16  0.95 0.97 -2.16  1.73  0.00
## [10,] -1.09   2.36 -1.86 -0.03   1.90 -2.35 1.03  1.06 -2.35  1.88
```

输出结果是一个矩阵，第 n 行是 $a_{n,j}, j=1,2,\dots,n$ 以及补充的 0。可以看出并不是只有 a_{n1} 非零。

Part V

谱密度估计

Chapter 26

潜周期模型

26.1 介绍

实际问题中有很多的数据表现出明显的周期特性. 例如考虑北京地区的气温变化, 以每个小时的平均气温为一个记录数据, 则观测数据 $x_1, x_2, \cdots, x_N$ 以每年的 365×24 小时为一个大周期, 以每天的 24 小时为一个小周期. 对于这种具有明显周期的数据可以考虑用潜周期模型(26.1)描述.

在信号处理领域, 余弦波信号是一种常见信号. 在随机干扰背景下能否成功地检测出各信号的角频率成分及其振幅是一个有实际意义的问题. 通常的余弦波信号也是用潜周期模型(26.1)描述的.

$$x_t = \sum_{j=1}^{k} A_j \cos(\omega_j t + \phi_j) + \xi_t, \quad t \in \mathbb{N}_+, \tag{26.1}$$

其中 $0 < \omega_1 < \omega_2 < \cdots < \omega_k \leq \pi$. 正数 A_j 是相应于第 j 个角频率 ω_j 的振幅. 对 $\omega_j > 0$, 从

$$\cos(\omega_j(t + 2\pi/\omega_j) + \phi_j) = \cos(\omega_j t + \phi_j), \quad t \in \mathbb{Z},$$

知道相应于角频率 ω_j 的周期是 $T_j = 2\pi/\omega_j$. $\phi_j \in [0, 2\pi)$ 是相应于角频率 λ_j 的初始相位. $\{\xi_t\}$ 是一个零均值的线性平稳序列, 被称为有色噪声. 它是对周期

叠加项

$$y_t = \sum_{j=1}^{k} A_j \cos(\omega_j t + \phi_j) \tag{26.2}$$

的随机干扰. 当 $k = 0$ 定义 $\sum_{j=1}^{0} \equiv 0$. 这时 $\{x_t\} = \{\xi_t\}$ 是平稳序列.

满足模型(26.1)的时间序列被称为潜频率或潜周期序列. 在模型(26.1)中, 可以不要求 $\omega_1 > 0$. 但是 $\omega_1 = 0$ 就相当于(26.1)的右边有一个常数项. 由于在实际问题的计算中, 需要先对原始数据进行零均值化, 所以常数项实际上是不起作用的. 模型(26.1)中还可以要求振幅 $A = (A_1, A_2, \cdots, A_q)$ 和初相位 $\phi = (\phi_1, \phi_2, \cdots, \phi_q)$ 是随机的. 但是由于在时间序列分析中, 我们只能得到时间序列的一次实现, 因而在数据中 A 和 ϕ 的取值都是常数. 这样就没有必要把 A 和 ϕ 当作随机向量来考虑了. 所以, 我们认为 A_j, ϕ_j 是常数.

在模型(26.1)中, 用 $\sigma = \sqrt{\mathrm{Var}(\xi_t)}$ 表示噪声项的标准差. 称

$$\eta_j = A_j / \sigma \tag{26.3}$$

为相应于角频率 λ_j 的信噪比 (信号和噪声的比率). 明显, 信噪比 η_j 越大, 说明频率项 $A_j \cos(\omega_j t + \phi_j)$ 的作用越大, 在给定的数据中越容易发现或估计出相应的角频率 ω_j 和振幅 A_j. 以下的分析说明, 无论是怎样的信噪比, 只要有较大的数据量都可以十分准确地估计出每个角频率、振幅和初始相位.

模型(26.1)是三角函数项的叠加, 所以又被称为调和模型. 它在气象, 天文, 机械振动, 共振研究和调和信号处理方面有广泛的应用.

如果用三角公式把模型(26.1)进行展开, 可以得到等价的模型

$$x_t = \sum_{j=1}^{k} [a_j \cos(\omega_j t) + b_j \sin(\omega_j t)] + \xi_t, \quad t \in \mathbb{N}_+, \tag{26.4}$$

其中

$$a_j = A_j \cos(\phi_j), \quad b_j = -A_j \sin(\phi_j).$$

如果 $\{\omega_j\}$ 已知，就变成了一个线性回归问题。

模型(26.1)还可以写成复的形式. 在模型(26.1)中, 定义

$$q = 2k \tag{26.5}$$

$$\tag{26.6}$$

$$\lambda_j = \begin{cases} \omega_j, & j = 1, 2, \cdots, k \\ -\omega_{j-k}, & j = k+1, k+2, \cdots, q. \end{cases} \tag{26.7}$$

$$\tag{26.8}$$

$$\alpha_j = \begin{cases} \frac{1}{2} A_j \exp(i\phi_j), & j = 1, 2, \cdots, k \\ \frac{1}{2} A_{j-k} \exp(-i\phi_{j-k}), & j = k+1, k+2, \cdots, q. \end{cases} \tag{26.9}$$

就得到

$$x_t = \sum_{j=1}^{q} \alpha_j \exp(it\lambda_j) + \xi_t, \quad t \in \mathbb{N}_+. \tag{26.10}$$

模型(26.10)是复值潜周期模型.
这个模型更方便进行理论推导。

为了说话的方便, 对潜周期模型(26.10)通常还要求如下的条件:

(1) $\{\xi_t\}$ 是一个零均值线性平稳序列, 它是对周期叠加项

$$y_t = \sum_{j=1}^{q} \alpha_j \exp(it\lambda_j) \tag{26.11}$$

的随机干扰.

(2) 实向量 $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_q)^T$ 满足

$$-\pi < \lambda_1 < \lambda_2 < \cdots < \lambda_q \le \pi. \tag{26.12}$$

(3) 复值向量 $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_q)^\tau$ 中的每一个分量不等于零:

$$\prod_{j=1}^{q} \alpha_j \neq 0.$$

当 $q \ge 1$, λ 是潜频率序列 $\{x_t\}$ 的角频率, α 是 $\{x_t\}$ 的振幅. 对于 $\lambda_j \neq 0$, 由于

$$\exp\left(i\lambda_j(t + 2\pi/\lambda_j)\right) = \exp(i\lambda_j t), \quad t \in \mathbb{Z},$$

所以相应于角频率 λ_j 的周期是 $T_j = 2\pi/\lambda_j$, 振幅是 α_j.

满足模型(26.10)的时间序列也被称为潜频率序列. 在模型(26.10)中, 只有潜频率序列 $\{x_t\}$ 是可以观测的. 如果观测序列 $\{x_t\}$ 和随机干扰项 $\{\xi_t\}$ 都是实值的, 由(26.11)定义的 $\{y_t\}$ 就是实值的. 这时如果设实数 a_j, b_j 使得 $\alpha_j = a_j + ib_j$, 利用三角公式就可以得到

$$y_t = \sum_{j=1}^{q} \alpha_j \exp(i\lambda_j t) = \sum_{j=1}^{q} (a_j + ib_j)[\cos(\lambda_j t) + i\sin(\lambda_j t)] \tag{26.13}$$

$$= \sum_{j=1}^{q} [a_j \cos(\lambda_j t) - b_j \sin(\lambda_j t)] + i\sum_{j=1}^{q} [a_j \sin(\lambda_j t) + b_j \cos(\lambda_j t)] \tag{26.14}$$

$$= \sum_{j=1}^{q} [a_j \cos(\lambda_j t) - b_j \sin(\lambda_j t)]. \tag{26.15}$$

于是, 利用

$$\sum_{j=1}^{q} [a_j \sin(\lambda_j t) + b_j \cos(\lambda_j t)] = 0, \quad t \in \mathbb{N}_+, \tag{26.16}$$

对(26.15)进行适当的整理合并, 就得到(26.2). 参数之间的关系仍然由(26.9)式决定.

对实值的潜周期模型(26.1)进行统计分析不如对复值潜周期模型(26.10)的统计分析来的方便. 如果对复值模型 (26.10)给出了参数 q, λ 和 α 的估计, 通过参数之间的变换(26.9)就可以得到模型(26.1)的参数估计.

26.2 潜周期模型的参数估计

为了分析潜周期模型中有色噪声项对周期信号 $\{y_t\}$ 影响的大小, 我们需要了解噪声 $\{\xi_t\}$ 的加权部分和的收敛性质.

定理 26.1. 设 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$, 线性滤波器 $\{c_j\}$ 满足

$$\sum_{j=0}^{\infty} j|c_j| < \infty. \tag{26.17}$$

平稳噪声 $\{\xi_t\}$ 由

$$\xi_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \tag{26.18}$$

定义, 则有如下的结果

$$\limsup_{N\to\infty} \frac{1}{\sqrt{N\ln N}} \sup_{\lambda} \left|\sum_{t=1}^{N} \xi_t e^{-i\lambda t}\right| \le \sqrt{2\pi \sup_{\lambda} f(\lambda)},\ a.s.,$$

其中

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left|\sum_{j=0}^{\infty} c_j e^{ij\lambda}\right|^2$$

是 $\{\xi_t\}$ 的谱密度.

定理26.1是对模型(26.10)进行统计分析的基本定理. 除了它, 我们还需要了解 Dirichlet 核

$$D_N(\lambda) = \frac{\sin(N\lambda/2)}{\sin(\lambda/2)},\ \lambda \in (-\pi, \pi]$$

的基本性质. $D_N(\lambda)$ 是偶函数, 在 $\lambda = 0$ 取得最大值 $D_N(0) = N$, 在区间 $[0, \pi/2N]$ 上单调减少, 利用不等式 $x \ge \sin x > 2x/\pi$, $x \in (0, \pi/2)$, 得到

$$D_N(x) \ge D_N(\pi/2N) = \frac{\sin(\pi/4)}{\sin(\pi/4N)} \ge \frac{2\sqrt{2}N}{\pi},\ x \in [0, \frac{\pi}{2N}]. \tag{26.19}$$

另一方面, 对于 $\lambda \ge 1/(2\sqrt{N})$,

$$|D_N(\lambda)| \le \left(\sin(\frac{1}{4\sqrt{N}})\right)^{-1} \le \frac{4\pi\sqrt{N}}{2} = 2\pi\sqrt{N}. \tag{26.20}$$

26.2.1 复值潜周期模型的初估计

在利用数据进行参数估计前应当先对数据进行零均值化处理, 以下设数据已经零均值化. 对于来自复值潜周期模型(26.10)的观测数据 $x_1, x_2, \cdots, x_N$, 引入函数

$$S_N(\lambda) = \sum_{t=1}^{N} x_t e^{-i\lambda t}, \tag{26.21}$$

这称为 $\{x_t, t = 1, 2, \dots, N\}$ 序列的离散傅里叶变换。类似地定义

$$C_N(\xi, \lambda) = \sum_{t=1}^{N} \xi_t e^{-i\lambda t},\quad C_N(\lambda) = \sum_{t=1}^{N} e^{-i\lambda t}.$$

利用求和公式

$$C_N(-\lambda) = D_N(\lambda) \exp(i\lambda(N+1)/2),$$

得到

$$|C_N(\lambda)| = |D_N(\lambda)|. \tag{26.22}$$

于是有

$$\begin{aligned} S_N(\lambda) &= \sum_{t=1}^{N} y_t e^{-i\lambda t} + \sum_{t=1}^{N} \xi_t e^{-i\lambda t} & (26.23)\\ &= \sum_{j=1}^{q} \alpha_j \sum_{t=1}^{N} \exp(i(\lambda_j - \lambda)t) + C_N(\xi, \lambda) & (26.24)\\ &= \sum_{j=1}^{q} \alpha_j C_N(\lambda - \lambda_j) + C_N(\xi, \lambda). & (26.25) \end{aligned}$$

由于 $S_N(\lambda)$, $C_N(\lambda)$ 和 $C_N(\xi, \lambda)$ 都是 λ 的周期为 2π 的函数, 为说话方便引入

$$\lambda_{q+1} = \lambda_1 + 2\pi, \quad \lambda_0 = \lambda_q - 2\pi. \tag{26.26}$$

对于角频率 λ_j, $(1 \le j \le q)$, 定义

$$\delta = \min_{1\le j\le q}(\delta_j), \quad \text{其中} \quad \delta_j = \min\{|\lambda_j - \lambda_{j+1}|, |\lambda_j - \lambda_{j-1}|\}. \tag{26.27}$$

如果样本量 N 充分大, 且使得

$$\delta_j \le \frac{1}{\sqrt{N}},$$

我们就可以检测出 λ_j. 实际上, 在 λ_j 的 $\pi/2N$ 邻域内, 利用(26.19)和(26.20)得到

$$\begin{aligned} |S_N(\lambda)| &\ge |\alpha_j C_N(\lambda - \lambda_j)| - \sum_{l\ne j} |\alpha_l C_N(\lambda - \lambda_l)| - |C_N(\xi, \lambda)| & (26.28)\\ &\ge |\alpha_j D_N(\lambda - \lambda_j)| - \sum_{l\ne j} |\alpha_l| 2\pi\sqrt{N} - |C_N(\xi, \lambda)| & (26.29)\\ &\ge |\alpha_j| 2\sqrt{2}N/\pi - O(\sqrt{N \ln N}). & (26.30)\\ &\ge 0.9|\alpha_j|N, \quad \text{a.s.}, & (26.31) \end{aligned}$$

而在所有的 λ_j, $0 \le j \le q+1$ 的 $1/(2\sqrt{N})$ 邻域外, 也就是在集合

$$\mathcal{A} = \{\lambda \in [-\pi, \pi] : \ |\lambda - \lambda_j| \ge 1/(2\sqrt{N}), \ j = 0, 1, 2, \cdots, q+1\} \tag{26.32}$$

上, 利用(26.20)和定理26.1 得到

$$|S_N(\lambda)| \le \sum_{l=1}^{q} |\alpha_l C_N(\lambda_l - \lambda)| + |C_N(\xi, \lambda)| \tag{26.33}$$

$$\le \sum_{l=1} |\alpha_l| 2\pi\sqrt{N} + |C_N(\xi, \lambda)| \tag{26.34}$$

$$= O(\sqrt{N \ln N}). \tag{26.35}$$

于是看出当 N 充分大后, 实值连续函数 $|S_N(\lambda)|$ 在区间 $[-\pi, \pi]$ 上的图形具有如下的形状:

1. $|S_N(\lambda)|$ 在每个 λ_j 的 $1/(2\sqrt{N})$ 邻域内有一峰群, 其最高峰的高度大于 $0.9|\alpha_j|N$. 最高峰的下面隐藏着角频率 λ_j.
2. 在所有的 λ_j 的 $1/(2\sqrt{N})$ 邻域外, 也就是在集合 $\mathcal{A}$ 上, $|S_N(\lambda)| = O(\sqrt{N \ln N})$.
3. 峰群的个数就是潜周期模型中的周期 (或角频率) 个数的估计.

根据 $|S_N(\lambda)|$ 的图形形状, 可以给出潜周期模型(26.10)中的角频率个数 q, 角频率向量 λ 和振幅向量 α 的估计方法如下.

由于很难把 $|S_N(\lambda)|$ 的所有取值计算出来, 为了节约计算时间, 同时也能够大致的保持 $|S_N(\lambda)|$ 的图形, 可以采用如下的离散化计算方法.

26.2.1.1　方法一

当潜周期模型中的各振幅 α_j 的绝对值差别不大, 并且平稳噪声 $\{\xi_t\}$ 的谱密度 $f(\lambda)$ 没有明显的峰值时可以采用本方法.

取 $A_N > 0$ 满足: 当 $N \to \infty$ 时, $A_N = O(N^{0.75})$.

第一步: 定义 $\mu(j) = j\pi/(2N) - \pi$, 计算

$$d(\mu(j)) = |S_N(\mu(j))|, \quad j = 1, 2, \cdots, 4N. \tag{26.36}$$

在下面计算中, 如果最大值不惟一, 可任取其中一个.

第二步: 计算 $d(\mu(j))$ 的最大值 $d(\mu(j_1))$. 当 $d(\mu(j_1)) \le A_N$, 定义 $\hat{q} = 0$, 停止计算. 否则, 定义

$$\mathcal{A}(1) = \left\{\lambda \in (-\pi, \pi] : \ 1/\sqrt{N} \le |\lambda - \mu(j_1)| \le 2\pi - 1/\sqrt{N}\right\}.$$

在 $\mathcal{A}(1)$ 中, 计算 $d(\mu(j))$ 的最大值 $d(\mu(j_2))$. 当 $d(\mu(j_2))\le A_N$, 定义 $\hat{q}=1$, $\hat{\lambda}_1=\mu(j_1)$, 停止计算. 依次类推. 当 $d(\mu(j_l))$ 在

$$\begin{aligned}&\mathcal{A}(p-1)\\&=\Big\{\lambda\in(-\pi,\pi]:\ 1/\sqrt{N}\le|\lambda-\mu(j_l)|\le 2\pi-1/\sqrt{N}, l=1,2,\cdots,p-1\Big\}\end{aligned}$$

中的个最大值 $d(\mu(j_p))>A_N$, 在

$$\begin{aligned}&\mathcal{A}(p)\\&=\Big\{\lambda\in(-\pi,\pi]:\ 1/\sqrt{N}\le|\lambda-\mu(j_l)|\le 2\pi-1/\sqrt{N}, l=1,2,\cdots,p\Big\}\end{aligned}$$

中的最大值 $d(\mu(j_{p+1}))\le A_N$ 时,
定义 q 的估计为 $\hat{q}=p$. 并将最大值点 $\mu(j_1),\mu(j_2),\cdots,\mu(j_{\hat{q}})$ 从小到大重排后得到角频率 λ 的初估计

$$\hat{\lambda}=(\hat{\lambda}_1,\hat{\lambda}_2,\cdots,\hat{\lambda}_{\hat{q}(N)}). \tag{26.37}$$

26.2.1.2 方法二

当潜周期模型中的各振幅 α_j 的绝对值有较大的差别, 但是平稳噪声 $\{\xi_t\}$ 的谱密度 $f(\lambda)$ 没有明显的峰值时可以采用本方法.

取正整数 $A_N>0$ 满足: 当 $N\to\infty$, $A_N=o(N)$ 和 $\sqrt{N\ln N}=o(A_N)$.

第一步: 定义 $\mu(j)=j\pi/(2N)-\pi$. 计算

$$d(\mu(j))=|S_N(\mu(j))-S_N(\mu(j-6))|,\quad j=1,2,\cdots,4N. \tag{26.38}$$

第二步和方法一中的第二步相同. 最后得到角频率向量 λ 的初估计(26.37).

26.2.1.3 方法三

当潜周期模型中的各振幅 α_j 的绝对值差别较大, 并且平稳噪声 $\{\xi_t\}$ 的谱密度 $f(\lambda)$ 有陡峭的峰值时可以采用本方法.

第一步: 取正数 $B_0\in(1,3\sqrt{2})$, 定义 $\mu(j)=j\pi/(2N)-\pi$. 计算

$$d(\mu(j))=\frac{S_N(\mu(j))+\sqrt{N}\ln N}{S_N(\mu(j-6))+\sqrt{N}\ln N},\ j=1,2,\cdots,4N \tag{26.39}$$

第二步和方法一中的第二步相同. 最后得到角频率向量 λ 的初估计(26.37).

应当说明, 上面的三种计算方法只是为了计算机编写程序的方便. 实际问题中根据 $|S_N(\lambda)|$ 的具体形状确定潜周期的个数是最行之有效的. 按上面的方法得到的估计量实际上已经是很准确的估计了. 它的估计精度可以由以下的定理进行描述.

定理 26.2. *设模型(26.10)中的平稳噪声 $\{\xi_t\}$ 满足定理26.1中的条件, 则几乎必然地当 N 充分大后, 由上面的三种方法定义的 $\hat{q}$ 和初估计 $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_q$ 满足*

$$\hat{q} = q, \quad |\hat{\lambda}_j - \lambda_j| \le \pi/N.$$

有了角频率的估计量(26.37), 就可以定义振幅 α_j 的估计如下:

$$\hat{\alpha}_j = \frac{1}{N}\sum_{t=1}^{N} x_t e^{-i\hat{\lambda}_j t}, \ 1 \le j \le \hat{q}. \tag{26.40}$$

由角频率的初估计 $\hat{\lambda}_j$ 定义的 $\hat{\alpha}_j$ 在理论上还不能保证有很好的估计精度. 也就是说, 当 $N \to \infty$ 时, $\hat{\alpha}_j$ 收敛到 α_j 速度还不够好. 为了得到精度更高的估计量, 还需要对角频率的初估计 $\hat{\lambda}_j$ 进行改造. 得到更精确的估计, 称为精估计.

26.2.2 角频率的精估计

得到了角频率的初估计(26.37)后, 可以用以下的方法改进估计的精度, 得到精度更高的估计量.

26.2.2.1 方法一

对每一个初估计 $\hat{\lambda}_j$, 在它的 $8/N$ 邻域 $[\hat{\lambda}_j - 4/N, \hat{\lambda}_j + 4/N]$ 中加密计算函数 $|S_N(\lambda)|$, 得到加密计算后 $|S_N(\lambda)|$ 的最大值点 $\tilde{\lambda}_j$. 如果计算的密度可以到达 $o(N^{-1.6})$, 则称 $\tilde{\lambda}_j$ 为 λ_j 的周期图最大估计. 用周期图最大估计 $\tilde{\lambda}_j$ 代替(26.40)中的初估计 $\hat{\lambda}_j$ 得到振幅 α_j 的精估计:

$$\tilde{\alpha}_j = \frac{1}{N}\sum_{t=1}^{N} x_t e^{-i\tilde{\lambda}_j t}, \ 1 \le j \le \hat{q}. \tag{26.41}$$

对于潜周期模型(26.10)的精估计, 有以下的强相合性结果.

定理 26.3. 设模型(26.10)中的平稳噪声 $\{\xi_t\}$ 满足定理*26.1*中的条件, 则有如下的结果*:*

$$\limsup_{N\to\infty}\sqrt{\frac{N^3}{\ln N}}|\tilde{\lambda}_j-\lambda_j|=0,\ \ a.s.,$$
$$\limsup_{N\to\infty}\sqrt{\frac{N}{\ln N}}|\tilde{\alpha}_j-\alpha_j|=0,\ \ a.s.$$

在参数估计问题中, 估计量的 a.s. 收敛速度一般只达到 $O((N/\ln\ln N)^{-1/2})$, 这里角频率 λ_j 的周期图最大估计 $\tilde{\lambda}$ 的 a.s. 收敛速度可以达到 $o((N^3/\ln N)^{-1/2})$.

26.2.2.2 方法二

改进角频率的初估计 $\hat{\lambda}_j$ 还可以采用二次分析法以减少计算量. 传统的二次分析方法在 Fourier 分析和时间序列分析中有很多的应用. 利用角频率的初估计和二次分析方法可以给出改进角频率的初估计的方法如下.

用 $\arg(c)$ 表示复数 c 的辐角. 取正整数 $M>1$, 则存在正整数 m 满足

$$N=mM+M',0\le M'<M. \tag{26.42}$$

对于 $s=1,2,\cdots,M$, 定义

$$\begin{aligned}
&T_j(s)=\sum_{t=(s-1)m+1}^{sm}x_t\exp(-it\hat{\lambda}_j),\\
&\theta_j=\arg\Big[\sum_{s=1}^{M}T_j(s)/|T_j(s)|\Big],\\
&z_j(s)=\arg[T_j(s)\exp(-i\theta_j)],\\
&\beta_j=\sum_{s=1}^{M}z_j(s)\Big(s-\frac{M+1}{2}\Big)\Big/\sum_{s=1}^{M}\Big(s-\frac{M+1}{2}\Big)^2.
\end{aligned}$$

最后, 改进的初估计由

$$\tilde{\lambda}_j=\hat{\lambda}_j+\beta_j/m \tag{26.43}$$

定义. 将改进的初估计 $\tilde{\lambda}_j$ 代入(26.41), 就得到 α_j 的估计 $\tilde{\alpha}_j$. 对于用二次分析法改进的估计量, 也可以证明定理26.3中的收敛速度.

26.2.3 实值模型的参数估计

最后回到实值潜周期模型(26.1)的参数估计问题. 由于观测数据是实值的, 所以 $|S_N(\lambda)|$ 是偶函数, 因而只要按照前述的方法在 $[0,\pi]$ 上找出峰群的个数 $\hat{k}$ 作为角频率个数 k 的估计. 每个峰群中的最高峰下对应于一个角频率的估计 $\hat{\lambda}_j$. 设 $\hat{\alpha}_j$ 由(26.40)或(26.41)定义. 定义 A_k 的估计 $\hat{A}_k$ 如下:

如果 $\hat{\lambda}_k=\pi$, 取 $\hat{A}_k=\hat{\alpha}_k$, $\hat{\phi}_1=0$, $\hat{\omega}_1=\pi$;

如果 $\hat{\lambda}_j\in(0,\pi)$, 取 $\hat{\omega}_j=\hat{\lambda}_j$, $\hat{A}_j=2|\hat{\alpha}_j|$, 初始相位 ϕ_j 的估计取作 $\hat{\phi}_j=\arg(\hat{\alpha}_j)$.

不难看出, 这样定义的估计量都有很好的强相合性质.

26.2.4 模型的检测

对于实值模型(26.1)得到周期个数的估计 $\hat{k}$, 角频率的估计 $\hat{\omega}_j$, $1\le j\le\hat{k}$, 振幅的估计 $\hat{A}_j$, $1\le j\le\hat{k}$ 和初相位的估计 $\hat{\phi}_j$, $1\le j\le\hat{k}$ 后, 为了检测拟合的模型

$$x_t=\sum_{j=1}^{\hat{k}}\hat{A}_j\cos(\hat{\omega}_j t+\hat{\phi}_j)+\xi_t,\quad t\in\mathcal{N}_+$$

是否合理, 需要计算残差

$$\hat{\xi}_t=x_t-\sum_{j=1}^{\hat{k}}\hat{A}_j\cos(\hat{\omega}_j t+\hat{\phi}_j),\ t=1,2,\cdots,N \tag{26.44}$$

和它的样本自协方差函数

$$\hat{\gamma}_k=\frac{1}{N}\sum_{j=1}^{N-k}(\hat{\xi}_j-\hat{\mu})(\hat{\xi}_{j+k}-\hat{\mu}),\ k=1,2,\cdots,[\sqrt{N}],$$

这里 $\hat{\mu}$ 是 $\hat{\xi}_t$ 的样本均值.

如果 $\hat{\gamma}_k$ 有收敛到零的性质, 就可以认为模型合适. 需要的话, 对于实值的潜周期模型(26.1), 还可以进一步为残差(26.44)建立 AR 或 ARMA 模型. 如果建立的 AR 或 ARMA 模型可以通过模型检测, 就应当肯定拟合模型的合理性.

例 26.1. 观测序列 $\{x_t\}$ 来自潜周期模型(26.1),

其中 $k=5$,

$$\omega=(0.23,0.98,1.54,1.98,2.67)^T,$$
$$A=(1.44,2.89,1.98,4.98,1.78)^T,$$
$$\phi=(0.2,2.9,0.8,2.3,1.6)^T.$$

有色噪声 $\{\xi_t\}$ 是由

$$\xi_t=1.16\xi_{t-1}-0.37\xi_{t-2}-0.11\xi_{t-3}+0.18\xi_{t-4}+\varepsilon_t \tag{26.45}$$

产生的 $AR(4)$ 序列. 这里 $\{\varepsilon_t\}$ 是正态 WN(0,1), $\mathrm{Var}(\xi_t)=4.422$. 信噪比

$$A/\sqrt{\mathrm{Var}(\xi_t)}=(0.6847\quad 1.3741\quad 0.9414\quad 2.3678\quad 0.8463)^T. \tag{26.46}$$

观测数据的 80 个样本由图26.1给出.

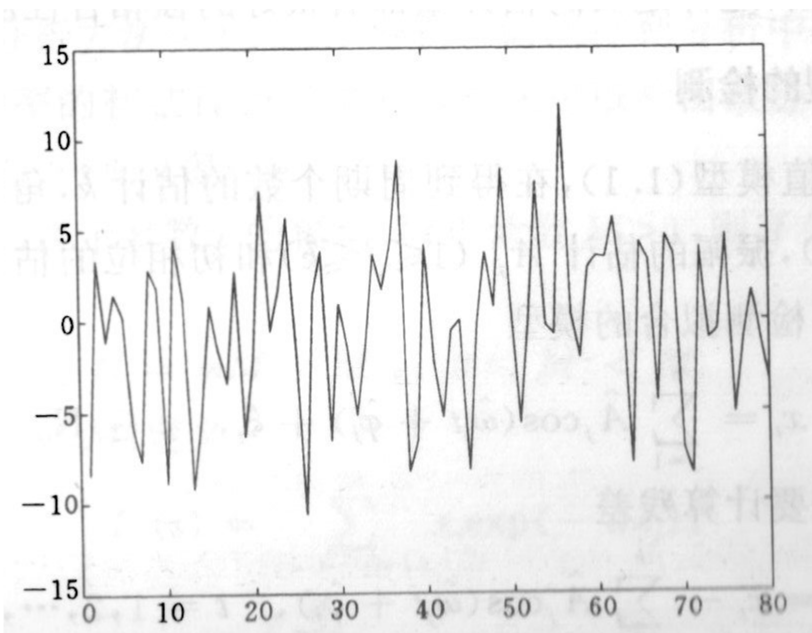


图 26.1: 原始数据

解答: 仅从数据图很难判断数据的潜周期个数. 图26.2和图26.3 分别是 $N=30,200$ 时函数 $|S_N(\lambda)|$ 的图形. 由于观测数据是实值的, 只需绘出 $[0,\pi]$ 部分. 计算密度是 $\pi/(2N)$.

实际计算 $S_N(\lambda)$ 时, 先对数据进行了零均值化处理.

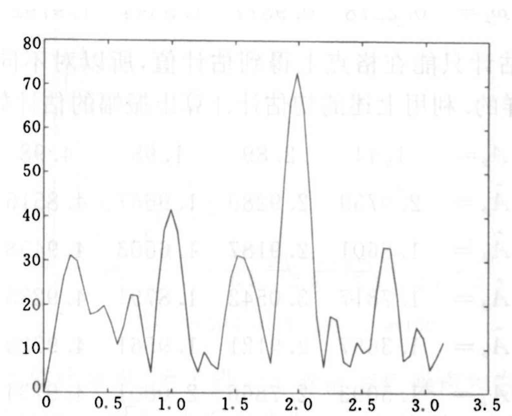


图 26.2: $|S_{30}(\lambda)|$

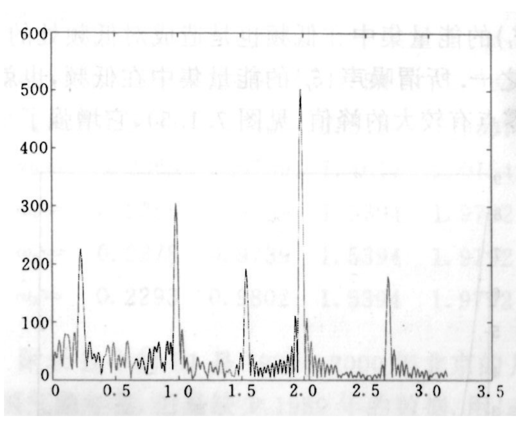


图 26.3: $|S_{200}(\lambda)|$

从上述图形看出, 信噪比高的角频率更容易被识别出来. 从方法一可以得到角频率的估计如下.

真值	$\omega_k =$	0.23	0.98	1.54	1.98	2.67
$N = 30$	$\hat{\omega}_k =$	0.2094	0.9948	1.5708	1.9897	2.6704
$N = 50$	$\hat{\omega}_k =$	0.2513	0.9739	1.5394	1.9792	2.6704
$N = 100$	$\hat{\omega}_k =$	0.2513	0.9739	1.5394	1.9792	2.6861
$N = 200$	$\hat{\omega}_k =$	0.2356	0.9739	1.5394	1.9792	2.6861
$N = 500$	$\hat{\omega}_k =$	0.2278	0.9817	1.5394	1.9792	2.6704

由于初估计只能在格点上得到估计值, 所以对不同的 N, 估计值有时是一样的. 利用上述的初估计计算出振幅的估计如下:

真值	$A_k =$	1.44	2.89	1.98	4.98	1.78
$N = 30$	$\hat{A}_k =$	2.0759	2.9289	1.9657	4.8515	2.2403
$N = 50$	$\hat{A}_k =$	1.8601	2.9187	1.6503	4.9458	2.1115
$N = 100$	$\hat{A}_k =$	1.7817	3.0542	1.8791	4.9235	1.8209
$N = 200$	$\hat{A}_k =$	1.3657	2.9121	1.9561	4.9729	1.8273
$N = 500$	$\hat{A}_k =$	1.5093	2.7866	2.0801	4.9734	1.8009

可以看出, 当 $N \leq 100$, 对角频率 $\omega_1 = 0.23$ 和振幅 $A_1 = 1.44$ 的估计精度较差.
这是因为角频率 ω_1 的信噪比最小的原因. 另外, 在本例中, 噪声 $\{\xi_t\}$ 的能量集中在低频也是造成对低频处的角频率估计不准的原因之一. 所谓噪声 $\{\xi_t\}$ 的能量集中在低频也就是说 $\{\xi_t\}$ 的谱密度在零点有较大的峰值 (见图 7.1.5), 它增强了 $|C_N(\xi, \lambda)|$ 在零点的振荡, 以致影响到对潜频率 λ_1 和振幅 A_1 的估计精度. 参见图 7.1.6. 本例中, $|C_N(\xi, \lambda)|$ 在低频比其他地方振荡要大很多.

对于参数估计来讲, 为了进一步弄清有色噪声和白噪声的不同影响, 我们把有色噪声 $\{\xi_t\}$ 改为方差为 4.422 的正态白噪声, 以保证信噪比(26.46)不变. 经过计算, 得到的角频率估计在低频确是有所改进.

计算的结果如下.

真值	$\omega_k =$	0.23	0.98	1.54	1.98	2.67
$N = 30$	$\hat{\omega}_k =$	0.2094	0.9425	1.5184	1.9897	2.6704
$N = 50$	$\hat{\omega}_k =$	0.2369	0.9739	1.4017	2.0154	2.8636
$N = 100$	$\hat{\omega}_k =$	0.2356	0.9739	1.5394	1.9792	2.6861
$N = 200$	$\hat{\omega}_k =$	0.2278	0.9739	1.5394	1.9792	2.6704
$N = 500$	$\hat{\omega}_k =$	0.2293	0.9802	1.5394	1.9792	2.6672

Chapter 27

平稳序列谱表示

本章内容：

- 随机积分;
- 平稳序列谱表示；
- 线性平稳列的谱表示；
- 离散谱序列的特征及其谱表示；
- 平稳序列的频率性质；
- 平稳序列的分解。

内容待补充。

借助于 L^2 空间中的随机积分，可以将平稳序列表示成随机积分形势，能够帮助我们理解平稳时间序列的时域和谱域的关系。

复值随机变量：如果 ξ, η 是随机变量，$Z = \xi + i\eta$ 称为复值随机变量。

复值时间序列：如果 $\{\xi_t\}, \{\eta_t\}$ 是时间序列，$Z_t = \xi_t + \eta_t$ 称为复值时间序列。

学习本章需要学生学过实变函数论中的勒贝格测度和勒贝格积分概念。

27.1 随机积分定义

27.1.1 随机测度

定义 27.1 (正交增量过程)**.** 称复值随机过程 $\{Z(\lambda), \lambda \in [-\pi, \pi]$ 是**正交增量过程**，如果

(1) 对一切 $\lambda \in [-\pi, \pi]$，$EZ(\lambda) = 0$, $E|Z(\lambda)|^2 < \infty$;

(2) 对任何 $-\pi \leq \lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4 \leq \pi$，有

$$E\left\{[Z(\lambda_2) - Z(\lambda_1)]\overline{[Z(\lambda_4) - Z(\lambda_3)]}\right\} = 0,$$

其中 $\bar{\xi}$ 表示 ξ 的复共轭。

定义 27.2 (右连续的正交增量过程)**.** 称正交增量过程 $\{Z(\lambda)\}$ 是**右连续**的，如果当 $\delta \downarrow 0$ 时，对任何 $\lambda \in [-\pi, \pi]$ 都有

$$E|Z(\lambda + \delta) - Z(\lambda)|^2 \to 0.$$

本章中无特殊声明情况下，正交增量过程总假定是右连续的，并且满足 $Z(-\pi) = 0$。

称 $[-\pi, \pi]$ 上任何单调不减、右连续的非负有界函数为**分布函数**。这是取值于 $[-\pi, \pi]$ 的随机变量的分布函数的推广，不要求在 π 处等于 1。

任何正交增量过程都唯一地对应一个在 $-\pi$ 处等于零的分布函数。

定理 27.1. *设 $\{Z(\lambda)\}$ 是正交增量过程，右连续，且 $Z(-\pi) = 0$，则存在唯一的分布函数 $F(\lambda)$，满足对任何 $-\pi \leq \lambda < \mu \leq \pi$，有*

$$F(\mu) - F(\lambda) = E|Z(\mu) - Z(\lambda)|^2, \ F(-\pi) = 0. \tag{27.1}$$

实际上，$F(\cdot)$ 在 $[-\pi, \pi]$ 定义了一个有限测度，定理说明 $Z(\cdot)$ 像是一个测度，但结果是随机变量。称 $Z(\cdot)$ 为一个随机测度。

称(27.1)确定的分布函数 $F(\cdot)$ 为正交增量过程 $\{Z(\lambda)\}$ 的分布分数。

证明：取 $F(\lambda) = E|Z(\lambda)|^2$。则

$$F(-\pi) = E|0|^2 = 0.$$

对 $\lambda < \mu$，利用正交增量性得

$$\begin{aligned}F(\mu) =&E|Z(\mu) - Z(\lambda) + Z(\lambda) - Z(-\pi)|^2 \\ =&E|Z(\mu) - Z(\lambda)|^2 + E|Z(\lambda) - Z(-\pi)|^2 \\ =&E|Z(\mu) - Z(\lambda)|^2 + E|Z(\lambda)|^2 \\ =&E|Z(\mu) - Z(\lambda)|^2 - F(\lambda)\end{aligned}$$

所以(27.1)成立，且 $F(\cdot)$ 单调不减，由(27.1)和 $Z(\cdot)$ 的右连续性可知 $\mu \downarrow \lambda$ 时 $F(\mu) \to F(\lambda)$，即 $F(\cdot)$ 右连续。

对任何满足(27.1)条件的分布函数 F，在(27.1)中令 $\lambda = -\pi$ 就有 $F(\mu) = E|Z(\mu)|^2$，是唯一确定的。

例 27.1 (布朗运动)**.** 设 $\{B(t), t \in [-\pi, \pi]\}$ 为独立增量过程，对 $\lambda < \mu$，满足 $B(-\pi) = 0$，

$$B(\mu) - B(\lambda) \sim \mathrm{N}(0, \sigma^2(\mu - \lambda)).$$

则 $\{B(t)\}$ 是 $[-\pi, \pi]$ 上的布朗运动。

$$E|B(\mu) - B(\lambda)|^2 = \sigma^2(\mu - \lambda)$$

所以是右连续过程。按定理27.1的构造方法，令

$$F(\lambda) = E|B(\lambda)|^2 = \sigma^2(\lambda + \pi),\ \lambda \in [-\pi, \pi],$$

则 $F(-\pi) = 0$, $F(\cdot)$ 是 $[-\pi, \pi]$ 上的分布函数，是 $\{B(t)\}$ 对应的分布函数。

27.1.2 平方可积函数空间

设 $\Omega = [-\pi, \pi]$, $\mathscr{B}$ 是 Ω 上的 Borel σ 代数，则分布函数 F 是可测空间 $(\Omega, \mathscr{B})$ 上的有限测度，$(\Omega, \mathscr{B}, F)$ 是一个测度空间。

用 $L^2(F)$ 表示 $(\Omega, \mathscr{B}, F)$ 上复值平方可积函数的全体：

$$L^2(F) = \left\{ g(x) : \int_{-\pi}^{\pi} |g(x)|^2 dF(x) < \infty \right\} \tag{27.2}$$

在内积

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(s)\overline{g(s)}\, dF(s)$$

下，$L^2(F)$ 是复数域上的 Hilbert 空间。参见 (何书元, 2003) 第八章习题 1.2。

27.1.3 随机积分定义的步骤

设 $\{Z(\lambda)\}$ 是正交增量过程，F 是其分布函数，要定义随机积分：

$$I(g) = \int_{-\pi}^{\pi} g(s) dZ(s)$$

结果 $I(g)$ 是一个复值随机变量。定义类似于勒贝格积分定义步骤：

- 先对阶梯函数 g 定义；
- 讨论这样的随机积分的性质；
- 将阶梯函数的随机积分推广到一般平方可积函数。

27.1.4 对阶梯函数定义随机积分

设 g 是 $L^2(F)$ 中的阶梯函数，表示为

$$g(s) = \sum_{j=0}^{n} a_j I_{(\lambda_j, \lambda_{j+1}]}(s),\ -\pi = \lambda_0 < \lambda_1 < \cdots < \lambda_{n+1} = \pi, \tag{27.3}$$

其中 a_j 是复常数，$I_A(s)$ 是集合 A 的示性函数。

定义

$$I(g) = \sum_{j=0}^{n} a_j [Z(\lambda_{j+1}) - Z(\lambda_j)] \tag{27.4}$$

则

$$EI(g) = 0.$$

阶梯函数 g 的表达式不必唯一，但是结果 $I(g)$ 是唯一的。

用 $\mathscr{D}$ 表示形如(27.3)的阶梯函数全体。对于复常数 a, b 和 $f, g \in \mathscr{D}$, $af + bg$ 仍是阶梯函数，且有 n, $\{\lambda_j\}$, $\{a_j\}$, $\{b_j\}$ 使得

$$f(s) = \sum_{j=0}^{n} a_j I_{(\lambda_j, \lambda_{j+1}]}(s),\ g(s) = \sum_{j=0}^{n} b_j I_{(\lambda_j, \lambda_{j+1}]}(s),$$

于是

$$af(s)+bg(s)=\sum_{j=0}^{n}[aa_j+bb_j]I_{(\lambda_j,\lambda_{j+1}]}(s)\in\mathscr{D}, \tag{27.5}$$

阶梯函数的随机积分 $I(g)$ 满足如下性质：

$$I(af+bg)=aI(f)+bI(g), \tag{27.6}$$

$$E[I(f)\overline{I(g)}]=\int_{-\pi}^{\pi}f(s)\overline{g(s)}\,dF(\lambda), \tag{27.7}$$

$$E|I(f)-I(g)|^2=\int_{-\pi}^{\pi}|f(s)-g(s)|^2\,dF(s). \tag{27.8}$$

性质(27.6)证明：

由(27.5)和阶梯函数的随机积分定义有

$$\begin{aligned}
I(af+bg)=&\sum_{j=0}^{n}[aa_j+bb_j][Z(\lambda_{j+1})-Z(\lambda_j)]\\
=&a\sum_{j=0}^{n}a_j[Z(\lambda_{j+1})-Z(\lambda_j)]+b\sum_{j=0}^{n}b_j[Z(\lambda_{j+1})-Z(\lambda_j)]\\
=&aI(f)+bI(g).
\end{aligned}$$

性质(27.7)证明：

利用正交增量性可得

$$\begin{aligned}
&E[I(f)\overline{I(g)}]\\
=&\sum_{j=0}^{n}\sum_{k=0}^{n}a_j\bar{b}_kE\left\{[Z(\lambda_{j+1})-Z(\lambda_j)][\overline{Z(\lambda_{k+1})}-\overline{Z(\lambda_k)}]\right\}\\
=&\sum_{j=0}^{n}a_j\bar{b}_jE|Z(\lambda_{j+1})-Z(\lambda_j)|^2=\sum_{j=0}^{n}a_j\bar{b}_j[F(\lambda_{j+1}-F(\lambda_j)]\\
=&\sum_{j=0}^{n}\int_{-\pi}^{\pi}a_j\bar{b}_jI_{(\lambda_j,\lambda_{j+1}]}(s)\,dF(s)=\int_{-\pi}^{\pi}\sum_{j=0}^{n}a_j\bar{b}_jI_{(\lambda_j,\lambda_{j+1}]}(s)\,dF(s)\\
=&\int_{-\pi}^{\pi}f(s)\overline{g(s)}\,dF(s).
\end{aligned}$$

性质(27.8)证明：

在(27.7)中将 f, g 都替换成 $f-g$，利用(27.6)式可得

$$\begin{aligned}
E|I(f)-I(g)|^2 =& E|I(f-g)|^2 \\
=& E\left\{I(f-g)\overline{I(f-g)}\right\} \\
=& \int_{-\pi}^{\pi}[f(s)-g(s)]\overline{[f(s)-g(s)]}\,dF(s) \\
=& \int_{-\pi}^{\pi}|f(s)-g(s)|^2\,dF(s).
\end{aligned}$$

27.1.5 随机积分定义推广到一般函数

用 L^2 表示方差有限的复值随机变量的全体，L^2 在内积

$$\langle X, Y\rangle = E(X\bar{Y})$$

下是复数域上的 Hilbert 空间。对 $g \in \mathscr{D}$，$I(g) \in L^2$。考虑 g 为一般 $L^2(F)$ 函数的随机积分。

由实变函数论，对 $g \in L^2(F)$，存在函数序列 $g_n \in \mathscr{D}$ 使得

$$\lim_{n\to\infty}\int_{-\pi}^{\pi}|g_n(s)-g(s)|^2\,dF(s) = 0.$$

来证明 $I(g_n)$ 是 L^2 中的基本列。

利用(27.8)和不等式

$$|a+b|^2 \leq 2|a|^2 + 2|b|^2$$

可得

$$\begin{aligned}
E|I(g_n)-I(g_m)|^2 =& \int_{-\pi}^{\pi}|g_n(s)-g_m(s)|^2\,dF(s) \\
=& \int_{-\pi}^{\pi}\left|[g_n(s)-g(s)]-[g_m(s)-g(s)]\right|^2\,dF(s) \\
\leq& 2\int_{-\pi}^{\pi}|g_n(s)-g(s)|^2\,dF(s) + 2\int_{-\pi}^{\pi}|g_m(s)-g(s)|^2\,dF(s) \\
\to& 0, \quad n, m \to \infty.
\end{aligned}$$

由 L^2 的完备性可知 $I(g_n)$ 在 L^2 中有极限，记为 $I(g)$，使得

$$E|I(g_n)-I(g)|^2 \to 0, \quad n \to \infty. \tag{27.9}$$

还要说明 $I(g)$ 的值不依赖于 g_n 的选取。假设又有 $f_n \in \mathscr{D}$ 使得

$$\int_{-\pi}^{\pi} |f_n(s) - g(s)|^2 \, dF(s) \to 0, \quad n \to \infty.$$

则有

$$\begin{aligned}
&E|I(f_n) - I(g)|^2 = E\left|[I(f_n) - I(g_n)] + [I(g_n) - I(g)]\right|^2 \\
\le& 2E|I(f_n) - I(g_n)|^2 + 2E|I(g_n) - I(g)|^2 \\
=& 2\int_{-\pi}^{\pi} |f_n(s) - g_n(s)|^2 \, dF(s) + 2E|I(g_n) - I(g)|^2 \\
=& 2\int_{-\pi}^{\pi} \left|[f_n(s) - g(s)] - [g_n(s) - g(s)]\right|^2 \, dF(s) + 2E|I(g_n) - I(g)|^2 \\
\le& 4\int_{-\pi}^{\pi} |f_n(s) - g(s)|^2 \, dF(s) + 4\int_{-\pi}^{\pi} |g_n(s) - g(s)|^2 \, dF(s) + 2E|I(g_n) - I(g)|^2 \\
\to& 0, \quad n \to \infty.
\end{aligned}$$

即必有相同的 L^2 极限，两个极限 a.s. 相等。

将上述方法定义的 $I(g)$ 称为函数 $g \in L^2(F)$ 关于随机测度 $\{Z(\lambda)\}$ 的**随机积分**，记作

$$I(g) = \int_{-\pi}^{\pi} g(\lambda) \, dZ(\lambda). \tag{27.10}$$

右连续正交增量过程 $\{Z(\lambda), \lambda \in [-\pi, \pi]\}$ 且 $Z(-\pi) = 0$ 则称为一个**随机测度**。

随机积分还有其它的不同的定义，比如，按每条轨道的积分，用类似达布和极限定义的积分，等等。

27.2 随机积分的性质

定理 27.2. 设 $\{Z(\lambda), \lambda \in [-\pi, \pi]\}$ 是右连续正交增量过程，$Z(-\pi) = 0$, 有分布函数 $F(\lambda) = E|Z(\lambda)|^2$，对于 $f, g \in L^2(F)$，随机积分 $I(\cdot)$ 有如下性质：

- *(1)* $EI(f) = 0$.
- *(2)* $E(af + bg) = aI(f) + bI(g)$, 其中 a, b 为复常数。
- *(3)*
$$E[I(f)\overline{I(g)}] = \int_{-\pi}^{\pi} f(s)\overline{g(s)} \, dF(s).$$

- *(4)*
$$E|I(f)-I(g)|^2 = \int_{-\pi}^{\pi} |f(s)-g(s)|^2 \, dF(s).$$

证明：设阶梯函数 $f_n, g_n \in \mathscr{D}$ 使得 $n \to \infty$ 时

$$E|I(f_n)-I(f)|^2 \to 0, \quad E|I(g_n)-I(g)|^2 \to 0.$$

(1) 利用内积的连续性，

$$EI(f) = \lim_{n\to\infty} E[I(f_n)\cdot 1] = 0.$$

(2) 阶梯函数 $af_n + bg_n$ 在 $L^2(F)$ 中收敛到 $af+bg$，由(27.6)，

$$I(af_n + bg_n) = aI(f_n) + bI(g_n)$$

易见 $aI(f_n)+bI(g_n)$ 在 L^2 中收敛到 $aI(f)+bI(g)$，由随机积分定义可知

$$I(af+bg) = aI(f) + bI(g).$$

(3) 利用内积的连续性和(27.7)得到

$$\begin{aligned}
E[I(f)\overline{I(g)}] &= \lim_{n\to\infty} E[I(f_n)\overline{I(g_n)}] \\
&= \lim_{n\to\infty} \int_{-\pi}^{\pi} f_n(s)\overline{g_n(s)} \, dF(s) \\
&= \int_{-\pi}^{\pi} f(s)\overline{g(s)} \, dF(s),
\end{aligned}$$

最后一个等式利用了 $L^2(F)$ 内积连续性。

(4) 在性质 (3) 中将 f,g 都替换成 $f-g$，再利用性质 (2) 可得

$$\begin{aligned}
E|I(f)-I(g)|^2 =& E|I(f-g)|^2 = E[I(f-g)\overline{I(f-g)}] \\
=& \int_{-\pi}^{\pi} (f(s)-g(s))\overline{(f(s)-g(s))} \, dF(s) \\
=& \int_{-\pi}^{\pi} |f(s)-g(s)|^2 \, dF(s).
\end{aligned}$$

27.3 平稳序列的谱表示

设 $\{Z(\lambda)\}$ 为随机测度，有相应的分布函数 $F(\lambda)$，由于 $e^{it\lambda} \in L^2(F)$，所以用随机积分定义时间序列

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dZ(\lambda). \tag{27.11}$$

由定理27.2，

$$\begin{aligned} EX_t =& 0, \\ E(X_{t+k}\overline{X_t}) =& \int_{-\pi}^{\pi} e^{i(t+k)\lambda} e^{it\lambda}\, dF(\lambda) = \int_{-\pi}^{\pi} e^{ik\lambda}\, dF(\lambda) \end{aligned}$$

所以 $\{X_t\}$ 是复值零均值平稳列，自协方差函数

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda}\, dF(\lambda),$$

F 是 $\{X_t\}$ 的谱函数。

从随机测度可以定义对应的复值零均值平稳列；反之，每一个复值零均值平稳列，都有一个对应的随机测度，使其表示成随机积分。

定理 27.3 (谱表示定理)**.** 对复值零均值平稳列 $\{X_t\}$，有 $[-\pi, \pi]$ 上的随机测度 $\{Z_X(\lambda)\}$ 使得

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dZ_X(\lambda), \tag{27.12}$$

并且 $\{Z_X(\lambda)\}$ 对应的分布函数 F 为 $\{X_t\}$ 的谱函数。如果另有随机测度 $\{\xi(\lambda)\}$ 也满足上述条件，则

$$P(\xi(\lambda) = Z_X(\lambda)) = 1, \quad \forall \lambda \in [-\pi, \pi].$$

称 $\{Z_X(\lambda)\}$ 为 $\{X_t\}$ 的**随机测度**。定理证明见 (Brockwell & Davis, 1987) P.138 节 4.8。

例 27.2 (布朗运动与白噪声)**.** 对例27.1中的布朗运动 $\{B(t), t \in [-\pi, \pi]\}$, 因为 $\{B(t)\}$ 是随机测度，定义

$$\varepsilon_t = \int_{-\pi}^{\pi} e^{it\lambda} dB(\lambda), \quad t \in \mathbb{Z},$$

$\{\varepsilon_t\}$ 有谱函数

$$F(\lambda) = \sigma^2(\lambda + \pi)$$

和谱密度

$$f(\lambda) = \sigma^2.$$

所以 $\{\varepsilon_t\}$ 是 WN(0, $2\pi\sigma^2$)。由联合正态分布性质和随机积分定义可以证明 $\{\varepsilon_t\}$ 是正态时间序列，所以是独立同分布零均值白噪声列。

27.4 线性平稳列的谱表示

设 $\{\varepsilon_t\}$ 为 WN(0, σ^2)，则其谱密度为 $f_\varepsilon(\lambda) = \frac{\sigma^2}{2\pi}$，谱函数为

$$F_\varepsilon(\lambda) = \int_{-\pi}^{\lambda} f(s)\,ds = \frac{\sigma^2}{2\pi}(\lambda + \pi).$$

由谱表示定理，存在随机测度 $\{Z_\varepsilon(\lambda)\}$，使得

$$\begin{aligned} \varepsilon_t =& \int_{-\pi}^{\pi} e^{it\lambda}\,dZ_\varepsilon(\lambda), \\ F_\varepsilon(\lambda) =& E|Z_\varepsilon(\lambda)|^2. \end{aligned}$$

设实数列 $\{a_j\}$ 平方可和，考虑线性平稳列

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}.$$

由线性平稳列性质，$\{X_t\}$ 有谱密度和谱函数如下：

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|\sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda}\right|^2, \quad F(\lambda) = \int_{-\pi}^{\lambda} f(s)\,ds. \tag{27.13}$$

级数 $\sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda}$ 在 $L^2[-\pi,\pi]$ 中均方收敛，所以在 L^2 中

$$\begin{aligned} X_t =& \sum_{j=-\infty}^{\infty} a_j \int_{-\pi}^{\pi} e^{i(t-j)\lambda}\,dZ_\varepsilon(\lambda) \\ =& \int_{-\pi}^{\pi} e^{it\lambda} \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda}\,dZ_\varepsilon(\lambda) \\ =& \int_{-\pi}^{\pi} e^{it\lambda}\,dZ_X(\lambda), \end{aligned}$$

其中 $\{Z_X(\lambda)\}$ 是 $\{X_t\}$ 的随机测度。

$\{Z_X(\lambda)\}$ 与 $\{Z_\varepsilon(\lambda)\}$ 之间的关系是

$$Z_X(\lambda) = \int_{-\pi}^{\lambda} \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda}\, dZ_\varepsilon(\lambda).$$

事实上，如果用上式定义一个随机过程 $Z(\lambda)$，可以证明其为右连续独立增量过程，$Z(-\pi)=0$，且对任意 $L^2[-\pi,\pi]$ 中的函数 g 都有

$$\int_{-\pi}^{\pi} g(s) dZ(s) = \int_{-\pi}^{\pi} g(s) \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda}\, dZ_\varepsilon(\lambda),$$

因此 $Z(\lambda)$ 也是 $\{X_t\}$ 的随机测度。

27.5 离散谱序列的特征

设平稳序列 $\{X_t\}$ 的谱函数 $F(\lambda)$ 是阶梯函数，来证明 $\{X_t\}$ 是离散谱序列

设 $\{X_t\}$ 有谱表示

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dZ_X(\lambda), \quad t \in \mathbb{Z}. \tag{27.14}$$

则

$$F(\lambda) = E|Z_X(\lambda)|^2.$$

不妨设 $F(\lambda)$ 仅在 $\lambda_j, j=1,2,\dots$ 处有跳跃且 $-\pi < \lambda_1 < \lambda_2 < \cdots \le \pi$。设 $F(\cdot)$ 在 λ_j 处跳跃高度为 $\sigma_j^2 > 0$。这时

$$F(\lambda) = \sum_{j=1}^{\infty} \sigma_j^2 I_{[\lambda_j,\pi]}(\lambda) = \sum_{j=1}^{\infty} F_j(\lambda),$$

其中

$$F_j(\lambda) = \sigma_j^2 I_{[\lambda_j,\pi]}(\lambda), \quad j=1,2,\dots$$

$F(\pi) = \sum_{j=1}^{\infty} \sigma_j^2 < \infty$。

自协方差函数用谱函数表示为

$$\begin{aligned}
E(X_{t+k}\bar{X}_t) &= \int_{-\pi}^{\pi} e^{ik\lambda}\, dF(\lambda) \\
&= \sum_{j=1}^{\infty} \sigma_j^2 e^{ik\lambda_j} \\
&= \sum_{j=1}^{\infty} \int_{-\pi}^{\pi} e^{ik\lambda}\, dF_j(\lambda).
\end{aligned}$$

定义

$$g_j(s) = \begin{cases} 1, & s = \lambda_j, \\ 0, & s \neq \lambda_j, \end{cases}$$

则 $g_j \in L^2(F)$，于是可以定义随机积分

$$\xi_j = \int_{-\pi}^{\pi} g_j(\lambda)\, dZ_X(\lambda), \quad j = 1, 2, ...$$

利用定理27.2的 (3) 可得

$$E(\xi_j\bar{\xi}_k) = \int_{-\pi}^{\pi} g_j(s)\overline{g_k(s)}\, dF(s) = \delta_{j-k}\sigma_j^2. \tag{27.15}$$

定义

$$Y_t = \int_{-\pi}^{\pi} \left[1 - \sum_{j=1}^{\infty} g_j(\lambda)\right] e^{it\lambda}\, dZ_X(\lambda),$$

则 $EY_t = 0$，

$$E|Y_t|^2 = \int_{-\pi}^{\pi} \left|1 - \sum_{j=1}^{\infty} g_j(\lambda)\right|^2 dF(\lambda) = 0$$

（被积函数在每个跳跃点都等于零，）所以 $Y_t = 0$, a.s.。

所以 a.s. 意义下

$$\begin{aligned}X_t &= \int_{-\pi}^{\pi} e^{it\lambda}\, dZ_X(\lambda) \\ &= \int_{-\pi}^{\pi} e^{it\lambda} \sum_{j=1}^{\infty} g_j(\lambda)\, dZ_X(\lambda) + Y_t \\ &= \sum_{j=1}^{\infty} \int_{-\pi}^{\pi} e^{it\lambda_j} g_j(\lambda)\, dZ_X(\lambda) \\ &= \sum_{j=1}^{\infty} \xi_j e^{it\lambda_j}.\end{aligned}$$

这就是复值离散谱序列的表达式。

定理 27.4. 如果平稳序列 $\{X_t\}$ 的谱函数 $F(\lambda)$ 是阶梯函数，且只在 λ_j 有跳跃高度 σ_j^2 $(j=1,2,...)$，则当 $-\pi < \lambda_1 < \lambda_2 < \cdots \le \pi$ 时，存在复值零均值随机变量序列，使得 $E(\xi_j \bar{\xi}_k) = \delta_{j-k}\sigma_j^2$，且

$$X_t = \sum_{j=1}^{\infty} \xi_j e^{it\lambda_j}, \quad a.s., \quad t \in \mathbb{Z}.$$

特别地，当 $F(\lambda)$ 只有 p 个跳跃点时，

$$X_t = \sum_{j=1}^{p} \xi_j e^{it\lambda_j}, \quad \text{a.s.}, \quad t \in \mathbb{Z}.$$

27.6 离散谱序列的随机测度

设 $-\pi < \lambda_1 < \lambda_2 < \cdots \le \pi$,

$$F(\lambda) = \sum_{j=1}^{\infty} \sigma_j^2 I_{[\lambda_j,\pi]}(\lambda),\ \lambda \in [-\pi,\pi],$$

$F(\pi) = \sum \sigma_j^2 < \infty$，复值随机变量序列 $\{\xi_j\}$ 满足 $E(\xi_j \bar{\xi}_k) = \sigma_j^2 \delta_{j-k}$。

$$X_t = \sum_{j=1}^{\infty} \xi_j e^{it\lambda_j},\ t \in \mathbb{Z}.$$

令

$$Z(\lambda) = \sum_{j=1}^{\infty} \xi_j I_{[\lambda_j,\pi]}(\lambda), \tag{27.16}$$

来证明 $\{Z(\lambda)\}$ 是 $\{X_t\}$ 的随机测度。

定义

$$Z_j(\lambda) = \xi_j I[\lambda_j, \pi](\lambda),\ j = 1, 2, ...$$

则 $j \neq k$ 时 $\{Z_j(\lambda)\}$ 与 $\{Z_k(\lambda)\}$ 为相互正交的随机过程。

对每个给定的 j，对任意 $-\pi \leq s_1 < s_2 \leq s_3 < s_4 \leq \pi$，都有

$$[Z(s_2) - Z(s_1)]\overline{[Z(s_4) - Z(s_3)]} = |\xi_j|^2 I_{(s_1,s_2]}(\lambda_j) I_{(s_3,s_4]}(\lambda_j) \equiv 0,$$

所以 $\{Z_j(\lambda)\}$ 是正交增量过程，由定义 $Z_j(-\pi) = 0$。对 $\lambda < \pi$，总可以取 $\epsilon > 0$, $\lambda + \epsilon < \pi$，使得 $Z(\lambda + \epsilon) = Z(\lambda)$，于是右连续。在 $\lambda = \pi$ 处不用讨论是否右连续。于是 $\{Z_j(\lambda)\}$ 是随机测度，相互之间正交。

由随机积分的定义可以证明关于 $Z_j(\lambda)$ 的随机积分 $\int_{-\pi}^{\pi} g(s) dZ_j(s) = g(\lambda_j)\xi_j$。

相应于 $\{Z_j(\lambda)\}$ 的分布函数是

$$F_j(\lambda) = E|Z_j(\lambda)|^2 = \sigma_j^2 I_{[\lambda_j,\pi]}(\lambda).$$

且

$$\int_{-\pi}^{\pi} e^{it\lambda}\, dZ_j(\lambda) = \xi_j e^{it\lambda_j}.$$

定义

$$Z(\lambda) = \sum_{j=1}^{\infty} Z_j(\lambda),$$

则 $\{Z(\lambda)\}$ 也是随机测度，对应的分布函数为

$$\begin{aligned}
F(\lambda) =& E|Z(\lambda)|^2 = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} E[Z_j(\lambda)\overline{Z_k(\lambda)}] \\
=& \sum_{j=1}^{\infty} E|Z_j(\lambda)|^2 = \sum_{j=1}^{\infty} \sigma_j^2 I_{[\lambda_j,\pi]}(\lambda).
\end{aligned}$$

定义平稳序列

$$Y_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dZ(\lambda),$$

利用内积的连续型和随机积分定义，得

$$
\begin{aligned}
Y_t =& \int_{-\pi}^{\pi} \sum_{j=1}^{\infty} e^{it\lambda} \, dZ_j(\lambda) \\
=& \sum_{j=1}^{\infty} \int_{-\pi}^{\pi} e^{it\lambda} \, dZ_j(\lambda) \\
=& \sum_{j=1}^{\infty} \xi_j e^{it\lambda_j}, \ t \in \mathbb{Z}.
\end{aligned}
$$

可见 $\{Z(\lambda)\}$ 是 $\{X_t\}$ 的随机测度。

在离散谱序列的这两节中可以不要求 λ_1 最小。

27.7 谱密度的频率特征

设平稳序列 $\{X_t\}$ 有谱表示

$$
X_t = \int_{-\pi}^{\pi} e^{it\lambda} \, dZ_X(\lambda).
$$

谱函数与随机测度之间满足 $F(\lambda) = E|Z_X(\lambda)|^2$。当谱函数绝对连续时，$\{X_t\}$ 有谱密度 $f(\lambda) = F'(\lambda)$。

如果 $f(\lambda)$ 在 λ_0 处有一个峰值，则 $F(\lambda)$ 在 λ_0 左右的增量为

$$
\begin{aligned}
& F(\lambda_0 + \delta) - F(\lambda_0 - \delta) \\
=& E|Z(\lambda_0 + \delta) - Z(\lambda_0 - \delta)|^2 \\
=& \int_{\lambda_0 - \delta}^{\lambda_0 + \delta} f(\lambda) \, d\lambda,
\end{aligned}
$$

可以看成是随机测度在 λ_0 附近集中的能量。

例 27.3. 考虑 AR(2) 模型

$$
X_t = -0.276 X_{t-1} - 0.756 X_{t-2} + \varepsilon_t, \ \varepsilon_t \sim \text{WN}(0, 4)
$$

谱密度 $f(\lambda)$ 在 $\lambda_0 = 1.73$ 处有唯一的峰，见图27.1。

```
ar.true.spectrum(c(-0.276, - 0.756), title = "AR(2) 序列谱密度")
```

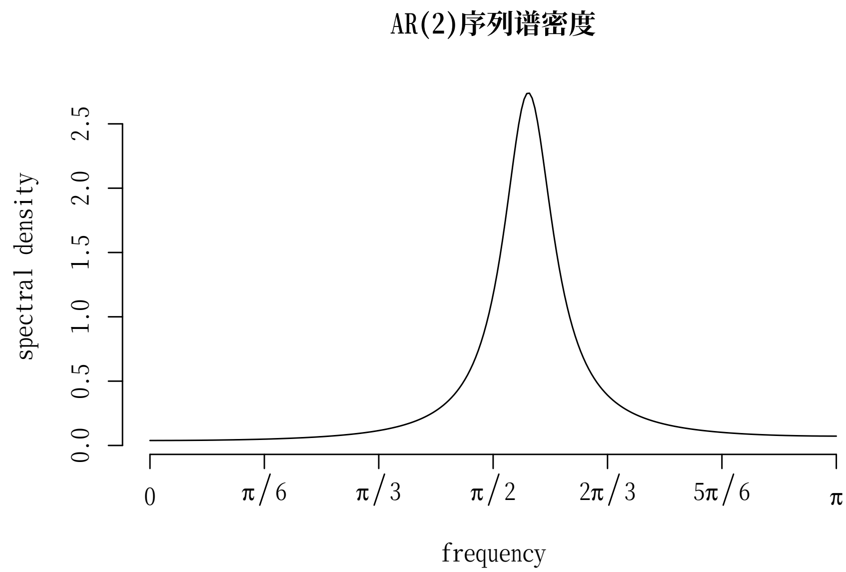


图 27.1: AR(2) 谱密度

用 $g(s)$ 表示集合 $(\lambda_0 - \delta, \lambda_0 + \delta]$ 的示性函数。δ 的取值与 λ_0 处 $f(\lambda)$ 的峰的宽度有关。将 X_t 分解为

$$\begin{aligned} X_t =& \int_{-\pi}^{\pi} g(\lambda) e^{it\lambda} \, dZ_X(\lambda) + \int_{-\pi}^{\pi} [1 - g(\lambda)] e^{it\lambda} \, dZ_X(\lambda) \\ =& \xi_t + \eta_t. \end{aligned} \tag{27.17}$$

其中 $\{\xi(t)\}$ 是平稳序列，有自协方差函数

$$\gamma_\xi(k) = \int_{-\pi}^{\pi} e^{ik\lambda} g(\lambda) f(\lambda) \, d\lambda,$$

和谱密度 $g(\lambda)f(\lambda)$。$\{\eta_t\}$ 也是平稳序列，有自协方差函数

$$\gamma_\eta(k) = \int_{-\pi}^{\pi} e^{ik\lambda} [1 - g(\lambda)] f(\lambda) \, d\lambda,$$

和谱密度 $[1 - g(\lambda)]f(\lambda)$。

$\{\xi_t\}$ 的随机测度为 $\int_{-\pi}^{\lambda} g(s)dZ_X(s)$，$\{\eta_t\}$ 的随机测度为 $\int_{-\pi}^{\lambda}[1-g(s)]dZ_X(s)$，因为 $g(s)[1-g(s)] \equiv 0$ 所以两个随机测度正交，故 $\{\xi_t\}$ 与 $\{\eta_t\}$ 也相互正交。

所以 $\{X_t\}$ 的自协方差函数 γ_k 可分解为

$$
\begin{aligned}
\gamma_k =& \gamma_\xi(k) + \gamma_\eta(k) \\
=& \int_{\lambda_0-\delta}^{\lambda_0+\delta} e^{ik\lambda} f(\lambda)\, d\lambda + \int_{-\pi}^{\pi} [1-g(\lambda)] e^{ik\lambda} f(\lambda)\, d\lambda.
\end{aligned}
$$

在角频率 λ_0 附近，谱密度 $f(\lambda)$ 为自协方差函数 $\{\gamma_k\}$ 提供的能量是

$$
\int_{\lambda_0-\delta}^{\lambda_0+\delta} e^{ik\lambda} f(\lambda)\, d\lambda \approx e^{ik\lambda_0} \int_{\lambda_0-\delta}^{\lambda_0+\delta} f(\lambda)\, d\lambda
$$

又有

$$
\xi_t \approx e^{it\lambda_0}[Z_X(\lambda_0+\delta) - Z_X(\lambda_0-\delta)]
$$

是 $\{X_t\}$ 受到角频率 λ_0 影响的波动，周期对应于 $\frac{2\pi}{\lambda_0}$，$[Z_X(\lambda_0+\delta)-Z_X(\lambda_0-\delta)]$ 为随机振幅，平方期望和方差约为 $\int_{\lambda_0-\delta}^{\lambda_0+\delta} f(\lambda)\, d\lambda$。

当谱密度 $f(\lambda)$ 在 λ_0 处有尖锐的峰时，$\{\gamma_k\}$ 的变化集中在 $e^{ik\lambda_0}$ 附近，$\{X_t\}$ 的变化集中在 $e^{it\lambda_0}$ 周期波动附近，γ_k 和 $\{X_t\}$ 都明显表现出周期为 $\frac{2\pi}{\lambda_0}$ 的波动。

同理，当 $\{X_t\}$ 的谱密度在 $\lambda_1, \dots, \lambda_p$ 处有峰值时，$\{\gamma_k\}$ 和 $\{X_t\}$ 的轨道会体现出角频率 $\lambda_1, \dots, \lambda_p$ 的波动。λ_j 处谱密度峰值越大、形状越陡峭，λ_j 的影响越明显。

对实平稳列 $\{X_t\}$，谱密度存在时必为偶函数。这时谱密度在 λ_0 处为 γ_k 提供的贡献约为 $e^{ik\lambda_0} \int_{\lambda_0-\delta}^{\lambda_0+\delta} f(\lambda)\, d\lambda$，在 $-\lambda_0$ 处为 γ_k 提供的贡献约为 $e^{-ik\lambda_0} \int_{\lambda_0-\delta}^{\lambda_0+\delta} f(\lambda)\, d\lambda$，在 $\pm\lambda_0$ 处为 γ_k 提供的总能量约为

$$
2\cos(k\lambda_0) \int_{\lambda_0-\delta}^{\lambda_0+\delta} f(\lambda)\, d\lambda.
$$

27.8 平稳序列的分解

实变函数论给出了如下结论：每个分布函数 $F(\lambda)$ 都可以唯一分解成如下三个部分：

$$
F(\lambda) = F_1(\lambda) + F_2(\lambda) + F_3(\lambda) \tag{27.18}
$$

其中 F_1 绝对连续，F_2 为跳跃函数，F_3 是奇异部分，F_3 相对于勒贝格测度都为 0。分解称为勒贝格分解。

对平稳列从谱域也有类似分解。

定理 27.5. *设零均值复值平稳列 $\{X_t\}$ 有谱函数 $F(\lambda)$，相应于 $F(\lambda)$ 的勒贝格分解 (1.20)，$\{X_t\}$ 可以唯一分解成三个相互正交的零均值平稳序列的和：*

$$X_t = X_1(t) + X_2(t) + X_3(t),\ t \in \mathbb{Z}, \tag{27.19}$$

其中 $\{X_i(t)\}$ 有谱函数 $F_i(\lambda)$。

27.5说明 $X_2(t)$ 是离散谱序列。可以证明，$X_1(t)$ 是线性平稳列。$\{X_3(t)\}$ 是奇异部分。如果不考虑奇异部分，平稳列都可以分解成线性平稳列与离散谱序列的叠加。

Chapter 28

谱估计

28.1 平稳序列的周期图

周期图是估计谱密度的基础。对零均值平稳序列的观测 $x_1, x_2, \dots, x_N$，记

$$
\begin{aligned}
\bar{x} =& \frac{1}{N} \sum x_k, \\
\hat{\gamma}_k =& \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}), \\
\hat{\gamma}_{-k} =& \hat{\gamma}_k, k = 0, 1, \dots, N-1, \\
\tilde{f}(\lambda) =& \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \hat{\gamma}_k e^{-ik\lambda}.
\end{aligned}
$$

记

$$
J_N(\lambda) = \frac{1}{2\pi N} \left| \sum_{k=1}^{N} (x_k - \bar{x}) e^{-ik\lambda} \right|^2, \quad \lambda \in [-\pi, \pi]
$$

引理 28.1.

$$
\tilde{f}(\lambda) = J_N(\lambda).
$$

证明： 记 $y_t = (x_t - \bar{x}) e^{-it\lambda}$, $t = 1, 2, \dots, N$, 用 1 表示元素都是 1 的 N 维向

量。则

$$\begin{aligned}
J_N(\lambda) =& \frac{1}{2\pi N}\sum_{j=1}^{N}\sum_{k=1}^{N}(x_k-\bar{x})(x_j-\bar{x})e^{-i(k-j)\lambda} \\
=& \frac{1}{2\pi N}\sum_{j=1}^{N}\sum_{k=1}^{N} y_k\bar{y}_j \\
=& \frac{1}{2\pi N}1^T\begin{pmatrix} y_1\bar{y}_1 & y_1\bar{y}_2 & \cdots & y_1\bar{y}_N \\ y_2\bar{y}_1 & y_2\bar{y}_2 & \cdots & y_2\bar{y}_N \\ \vdots & \vdots & & \vdots \\ y_N\bar{y}_1 & y_N\bar{y}_2 & \cdots & y_N\bar{y}_N \end{pmatrix} 1 \\
=& \frac{1}{2\pi}(\hat{\gamma}_0 + \hat{\gamma}_1 e^{i\lambda} + \cdots + \hat{\gamma}_{N-1}e^{i(N-1)\lambda} \\
& + \hat{\gamma}_1 e^{-i\lambda} + \cdots + \hat{\gamma}_{N-1}e^{-i(N-1)\lambda}) \\
=& \tilde{f}(\lambda)
\end{aligned}$$

定义 **Fourier 频率点**为

$$\lambda_j = \frac{2\pi j}{N}, j = 1, 2, \dots, N-1.$$

记

$$I_N(\lambda) = \frac{1}{2\pi N}\left|\sum_{k=1}^{N} x_k e^{-ik\lambda}\right|^2,$$

由于

$$\sum_{k=1}^{N} e^{-ik\lambda} = \frac{1-e^{-iN\lambda}}{1-e^{-i\lambda}}e^{-i\lambda}$$

和 $e^{-iN\lambda_j} = e^{-i2\pi j} = 1$ 得 $\sum_{k=1}^{N} e^{-ik\lambda_j} = 0$。所以在 Fourier 频率点上

$$I_N(\lambda_j) = J_N(\lambda_j), j = 1, 2, \dots, N-1.$$

定义 28.1 (周期图)**.** $I_N(\lambda)$ 称为观测数据的**周期图**。

定理 28.1. 对于零均值时间序列的观测值 $x_1, x_2, \dots, x_N$，若定义

$$\hat{\gamma}_{\pm k} = \frac{1}{N}\sum_{t=1}^{N-k} x_t x_{t+k}, \quad k = 0, 1, \dots, N-1 \tag{28.1}$$

则

$$I_N(\lambda) = \frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\hat{\gamma}_k e^{-ik\lambda}, \quad \lambda \in [-\pi, \pi]. \tag{28.2}$$

定理 28.2 (周期图的渐近无偏性). 如果零均值平稳序列 $\{X_t\}$ 的自协方差函数绝对可和:

$$\sum_{k=1}^{\infty} |\gamma_k| < \infty$$

则 $I_N(\lambda)$ 是谱密度 $f(\lambda)$ 的渐近无偏估计:

$$EI_N(\lambda) \to f(\lambda), \quad N \to \infty.$$

证明:

$$\begin{aligned}
EI_N(\lambda) =& \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} E\hat{\gamma}_k e^{-ik\lambda} \quad (\text{这里}\hat{\gamma}_k\text{按不减去均值的公式定义}) \\
=& \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \frac{N-|k|}{N} \gamma_k e^{-ik\lambda} \\
=& \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \gamma_k e^{-ik\lambda} - \frac{1}{2\pi N} \sum_{k=-N+1}^{N-1} |k| \gamma_k e^{-ik\lambda} \\
\to& \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-ik\lambda} \\
& (\text{由 Kronecker 引理及}\{\gamma_k\}\text{绝对可和知后一项趋于零}) \\
=& f(\lambda)
\end{aligned}$$

最后一个等式用到了定理9.1。

下面的定理说明，$I_N(\lambda)$ 不是 $f(\lambda)$ 的相合估计。

定理 28.3. 设 $\{\varepsilon_t\}$ 是独立同分布的 *WN*$(0, \sigma^2)$, 实数列 $\{c_j\}$ 满足

$$\sum_{j=0}^{\infty} j|c_j| < \infty$$

线性平稳列

$$X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \tag{28.3}$$

$I_N(\lambda)$ 为 $X_1, X_2, \ldots, X_N$ 的周期图，$f(\lambda)$ 为 $\{X_t\}$ 的谱密度，则

$$\limsup_{N\to\infty} \frac{1}{\ln\ln N} I_N(\lambda) = \begin{cases} f(\lambda), & \lambda \neq 0, \pi \\ 2f(\lambda), & \lambda = 0, \pi, \end{cases} \quad a.s. \tag{28.4}$$

定理说明 $|I_N(\lambda) - f(\lambda)|$ 不趋于零。

例 28.1. 设 $\{X_t\}$ 为标准正态白噪声。其谱密度 $f(\lambda) = \frac{1}{2\pi}$。$\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j \sim \text{N}(0, 1)$.

$$I_N(0) = \frac{1}{2\pi} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j \right)^2$$

为 χ^2_1 分布的 $\frac{1}{2\pi}$ 倍, 与 N 无关。

周期图一般不是谱密度的相合估计, 因为使用了所有的 $\hat\gamma_k$, 其中 k 接近于 $N-1$ 时 $\hat\gamma_k$ 估计偏差很大:

$$E\hat\gamma_k - \gamma_k = \frac{N-k}{N}\gamma_k - \gamma_k = -\frac{k}{N}\gamma_k.$$

减少使用的 $\hat\gamma_k$ 可以改善估计，称为**截断周期图**:

$$\hat f(\lambda) = \frac{1}{2\pi} \sum_{k=-N_1}^{N_1} \hat\gamma_k e^{-ik\lambda} \tag{28.5}$$

$(N_1 < N - 1)$

例 28.2. AR(2) 序列的周期图和截断周期图。

$$X_t = 0.1132 X_{t-1} - 0.64 X_{t-2} + \varepsilon_t,\ \varepsilon_t \sim \text{WN}(0, 4)$$

理论谱密度为

$$f(\lambda) = \frac{4}{2\pi} \frac{1}{|1 - 0.1132 e^{-i\lambda} + 0.64 e^{-i2\lambda}|^2}$$

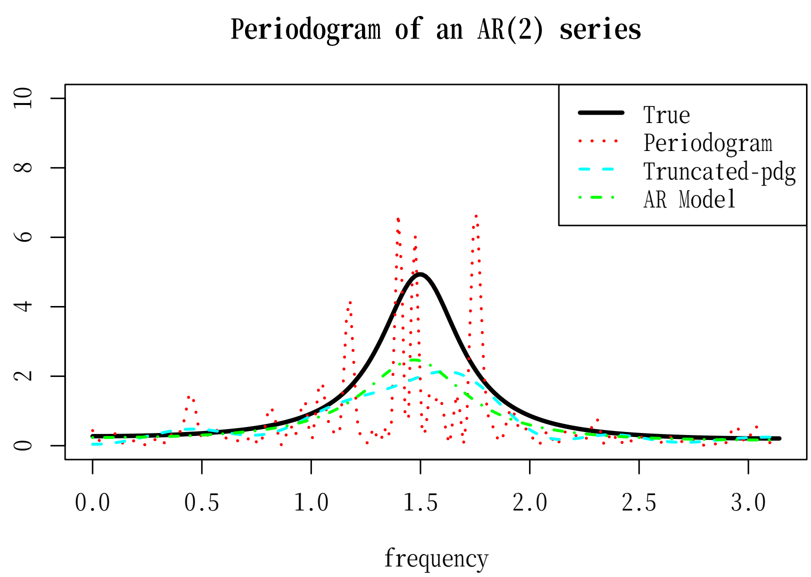
产生模拟数据（长度 $N = 123$），将理论谱密度、周期图、截断周期图与极大熵谱估计对比作图。取 $N_1 = \sqrt{N}$。

```
direct.periodogram <- function(y, freqs=seq(0,pi,length=101),
                               demean=FALSE){
  N <- length(y)
  if(demean) y <- y - mean(y)
  II <- numeric(length(freqs))
  for(j in seq(along=freqs)){
```

```
    II[j] <- 1/(2*pi*N)*Mod(sum(y * complex(mod=1, arg=freqs[j]*seq(1,N))))^2
  }

  list(frequencies=freqs, spectrum=II)
}
## AR theoretical spectrum given AR coefficients
ar.true.spectrum <- function(a, ngrid=256, sigma=1, plot.it=TRUE){
  p <- length(a)
  freqs <- seq(from=0, to=pi, length=ngrid)
  spec <- numeric(ngrid)
  for(ii in seq(ngrid)){
    spec[ii] <- 1 - sum(complex(mod=a, arg=freqs[ii]*seq(p)))
  }
  spec = sigma^2 / (2*pi) / abs(spec)^2
  if(plot.it){
    plot(freqs, spec, type='l',
         main="AR True Spectral Density",
         xlab="frequency", ylab="spectrum",
         axes=FALSE)
    axis(2)
    axis(1, at=(0:6)/6*pi,
         labels=c(0, expression(pi/6),
           expression(pi/3), expression(pi/2),
           expression(2*pi/3), expression(5*pi/6), expression(pi)))
  }
  list(frequencies=freqs, spectrum=spec,
       ar.coefficients=a, sigma=sigma)
}

demo.ar.pdg <- function(seed=7){
  set.seed(seed)
  N <- 120
  a <- c(0.1132, -0.64)
```

```
y <- arima.sim(list(ar=a), n=N)*2

## 理论谱密度
ft <- function(lambda){
  4 / (2*pi) / Mod(1 - complex(mod=0.113, arg=-lambda)
                 + complex(mod=0.64, arg=-2*lambda))^2
}
lams <- seq(0, pi, length=201)
spec0 <- ft(lams)

## 周期图
spec1 <- direct.periodogram(y, freqs=lams, demean=FALSE)$spectrum

## 截断周期图
N1 <- round(sqrt(N))
gams <- c(acf(y, lag.max=N1,
              type="covariance", plot=FALSE)$acf)
spec2 <- numeric(length(lams))
for(j in seq(along=lams))
  spec2[j] <- 1/(2*pi)*(gams[1] +
                        2*sum(gams[2:(N1+1)] *
                              cos(-seq(N1)*lams[j])))

## AR 模型
arr <- ar(y)
spec3 <- ar.true.spectrum(arr$ar, ngrid=length(lams),
                          sigma=sqrt(arr$var.pred),
                          plot.it=FALSE)$spectrum

plot(lams, spec0, lwd=3, type="l",
     ylim=c(0,10),
     main="Periodogram of an AR(2) series",
     xlab="frequency", ylab="")
```

```
  lines(lams, spec1, col="red", lwd=2, lty=3)
  lines(lams, spec2, col="cyan", lwd=2, lty=2)
  lines(lams, spec3, col="green", lwd=2, lty=4)
  legend("topright", lwd=c(3,2,2,2), lty=c(1,3,2,4),
         col=c("black", "red", "cyan", "green"),
         legend=c("True", "Periodogram", "Truncated-pdg", "AR Model"))

}
demo.ar.pdg()
```



28.2 加窗谱估计

28.2.1 加时窗的谱估计

克服周期图不相合性的经典方法是加窗谱估计。令

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \lambda_N(k) \hat{\gamma}_k e^{-ik\lambda} \tag{28.6}$$

其中 $\lambda_N(k)$ 是 $|k|$ 的减函数，称为**时窗**, 称 $\hat f(\lambda)$ 为**加时窗窗谱估计**，简称**加窗谱估计**。

例 28.3. 取时窗为

$$\lambda_N(k) = \begin{cases} 1, & |k| \leq \sqrt{N} \\ 0, & |k| > \sqrt{N} \end{cases}$$

即为截断周期图估计。

例 28.4. MA(q) 序列。则谱密度为

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-q}^{q} \gamma_k e^{-ik\lambda}.$$

取 $M \geq q$，时窗

$$\lambda_N(k) = \begin{cases} 1, & |k| \leq M \\ 0, & |k| > M \end{cases}$$

加窗谱估计为

$$\hat f(\lambda) = \frac{1}{2\pi} \sum_{k=-M}^{M} \hat\gamma_k e^{-ik\lambda}.$$

如果 $\{\varepsilon_t\}$ 是独立白噪声，则 $N \to \infty$ 时 $\hat\gamma_k \to \gamma_k$, a.s., 有

$$\begin{aligned}
\lim_{N\to\infty} \hat f(\lambda) =& \frac{1}{2\pi} \sum_{k=-M}^{M} \gamma_k e^{-ik\lambda} \\
=& \frac{1}{2\pi} \sum_{k=-q}^{q} \gamma_k e^{-ik\lambda} = f(\lambda), \quad \text{a.s.}
\end{aligned}$$

加窗后 $\hat f(\lambda)$ 为 $f(\lambda)$ 的强相合估计。

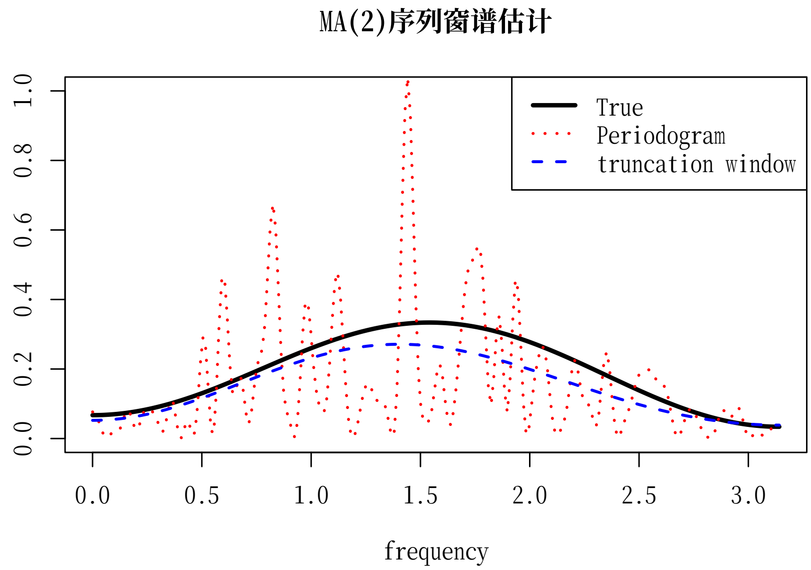
见演示：例 3.2（MA(2) 截断窗谱估计）。

```
## MA theoretical spectrum given MA coefficients
ma.true.spectrum <- function(a, ngrid=256, sigma=1,
                             tit="True MA Spectral Density",
                             plot.it=TRUE){
  p <- length(a)
  freqs <- seq(from=0, to=pi, length=ngrid)
```

```
  spec <- numeric(ngrid)
  for(ii in seq(ngrid)){
    spec[ii] <- 1 + sum(complex(mod=a, arg=freqs[ii]*seq(p)))
  }
  spec = sigma^2 / (2*pi) * abs(spec)^2
  if(plot.it){
    plot(freqs, spec, type='l',
         main=tit,
         xlab="frequency", ylab="spectrum",
         axes=FALSE)
    axis(2)
    axis(1, at=(0:6)/6*pi,
         labels=c(0, expression(pi/6),
           expression(pi/3), expression(pi/2),
           expression(2*pi/3), expression(5*pi/6), expression(pi)))
    box()
  }
  list(frequencies=freqs, spectrum=spec,
       ma.coefficients=a, sigma=sigma)
}
### 截断时窗谱估计
spec.rect <- function(y, freqs=seq(0,pi,length=101),
                      M = round(2*sqrt(length(y))),
                      demean=FALSE){
  N <- length(y)
  if(demean) y <- y - mean(y)
  spec <- numeric(length(freqs))
  gams <- c(acf(y, lag.max=M,
                type="covariance", plot=FALSE)$acf)
  for(j in seq(along=freqs))
    spec[j] <- 1/(2*pi)*(gams[1] +
                         2*sum(gams[2:(M+1)] *
                                 cos(seq(M)*freqs[j])))
```

```r
  list(frequencies=freqs, spectrum=spec)
}

demo.ma.spec <- function(seed=1,
                         window=c('truncation', 'Bartlett',
                           'Daniell', 'Turkey-Hamming',
                           'Turkey-Hanning',
                           'Parzen',
                           'Bartlett-Priestley')){
  set.seed(seed)
  N <- 100
  b <- c(0.0943, -0.4444)
  sigma <- 1
  y <- arima.sim(list(ma=b), n=N)*sigma

  ## 理论谱密度
  lams <- seq(0, pi, length=201)
  res1 <- ma.true.spectrum(b, sigma=sigma, plot.it=FALSE)
  lams <- res1$frequencies
  spec0 <- res1$spectrum

  ## 周期图
  spec1 <- direct.periodogram(y, freqs=lams, demean=FALSE)$spectrum

  ## 加窗谱估计
  if(window[1]=='truncation'){
    spec2 <- spec.rect(y, freqs=lams, M=3, demean=FALSE)$spectrum
  } else if(window[1]=='Bartlett'){
    spec2 <- spec.Bartlett(y, freqs=lams, M=10, demean=FALSE)$spectrum
  } else if(window[1]=='Daniell'){
    spec2 <- spec.Daniell(y, freqs=lams, M=6, demean=FALSE)$spectrum
  } else if(window[1]=='Turkey-Hamming'){
```

```
    spec2 <- spec.Turkey(y, freqs=lams, demean=FALSE, M=8, a=0.23)$spectrum
  } else if(window[1]=='Turkey-Hanning'){
    spec2 <- spec.Turkey(y, freqs=lams, demean=FALSE, M=8, a=0.25)$spectrum
  } else if(window[1]=='Parzen'){
    spec2 <- spec.Parzen(y, freqs=lams, demean=FALSE, M=10)$spectrum
  } else if(window[1]=='Bartlett-Priestley'){
    spec2 <- spec.Priestley(y, freqs=lams, demean=FALSE, M=10)$spectrum
  }

  ylim <- c(0,1.0)
  plot(lams, spec0, lwd=3, type="l",
       ylim=ylim,
       main="MA(2) 序列窗谱估计",
       xlab="frequency", ylab="")
  lines(lams, spec1, col="red", lwd=2, lty=3)
  lines(lams, spec2, col="blue", lwd=2, lty=2)
  legend("topright", lwd=c(3,2,2), lty=c(1,3,2),
         col=c("black", "red", "blue"),
         legend=c("True", "Periodogram", paste(window,'window')))
}
demo.ma.spec(window="truncation")
```



28.2.2 加谱窗的谱估计

$$
\begin{aligned}
I_N(\lambda) =& \frac{1}{2\pi N} \left| \sum_{k=1}^{N} x_k e^{-ik\lambda} \right|^2 \\
=& \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \hat{\gamma}_k e^{-ik\lambda}
\end{aligned}
$$

所以

$$
\hat{\gamma}_k = \int_{-\pi}^{\pi} I_N(s) e^{iks} ds, \quad |k| \leq N-1. \tag{28.7}
$$

(其中 $\hat{\gamma}_k$ 使用不减去均值的计算公式)

于是

$$\begin{aligned}\hat{f}(\lambda) =&\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\lambda_N(k)\hat{\gamma}_k e^{-ik\lambda}\\ =&\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\lambda_N(k)e^{-ik\lambda}\int_{-\pi}^{\pi}I_N(s)e^{iks}ds\\ =&\int_{-\pi}^{\pi}I_N(s)\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\lambda_N(k)e^{-ik(\lambda-s)}ds\\ \triangleq&\int_{-\pi}^{\pi}I_N(s)W_N(\lambda-s)ds,\end{aligned}$$

其中

$$W_N(\lambda)\triangleq\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\lambda_N(k)e^{-ik\lambda}$$

定义 28.2. 设 $W_N(\lambda)$ 为满足一定要求的权函数，

$$\hat{f}(\lambda)=\int_{-\pi}^{\pi}I_N(s)W_N(\lambda-s)ds \tag{28.8}$$

称为 $f(\lambda)$ 的**加谱窗谱估计**，简称为**加窗谱估计**。权函数 $W_N(\lambda)$ 称为**谱窗**。

加谱窗的谱估计是对周期图用一个权函数做卷积，取 $W_N(\cdot)$ 为偶函数，在 0 点最高的钟形曲线，则加谱窗谱估计是对周期图的局部光滑，可以克服周期图的不相合性。直观想法是取小的 $\delta>0$ 令 $W_N(\lambda)=\frac{1}{2\delta}I_{[\lambda-\delta,\lambda+\delta]}$，$\hat{f}(\lambda)=\frac{1}{2\delta}\int_{\lambda-\delta}^{\lambda+\delta}I_N(s)ds$。更一般的 $W_N(\lambda)$ 是其它的局部加权平均方法。

时窗 $\lambda_N(k)$ 和谱窗 $W_N(\lambda)$ 的关系为

$$\begin{aligned}W_N(\lambda)=&\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\lambda_N(k)e^{-ik\lambda}\\ \lambda_N(k)=&\int_{-\pi}^{\pi}W_N(s)e^{iks}ds.\end{aligned}$$

谱窗应有的性质：

- $\int_{-\pi}^{\pi}W_N(\lambda)d\lambda=1$ (等价于 $\lambda_N(0)=1$)。
- $\int_{-\pi}^{\pi}W_N^2(\lambda)d\lambda<\infty$;
- $\forall\varepsilon>0$, $N\to\infty$ 时 $\sup_{|\lambda|>\varepsilon}W_N(\lambda)\to 0$; (即 $N\to\infty$ 时谱窗向着 $\lambda=0$ 处集中)

- 对称性: $W_N(-\lambda) = W_N(\lambda)$;
- 对任何正数 A，当 $N \to \infty$ 时，

$$\max_{|\mu| \leq A/N} \left| \frac{\int_{-\pi}^{\pi} W_N(\lambda) W_N(\mu + \lambda) d\lambda}{\int_{-\pi}^{\pi} W_N^2(\lambda) d\lambda} - 1 \right| \to 0.$$

(要求谱窗随 $N \to \infty$ 不要变窄太快)

28.2.3 常用谱窗和时窗

28.2.3.1 截断窗

取 $M_N = o(N)$, $M_N \to \infty$, 比如 $M_N = A[\sqrt{N}]$, A 取值 1 到 3 之间。截断窗函数:

$$\lambda_N(k) = \begin{cases} 1, & |k| \leq M_N \\ 0, & |k| > M_N. \end{cases}$$

谱估计：

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{|k| \leq M_N} \hat{\gamma}_k e^{-ik\lambda}$$

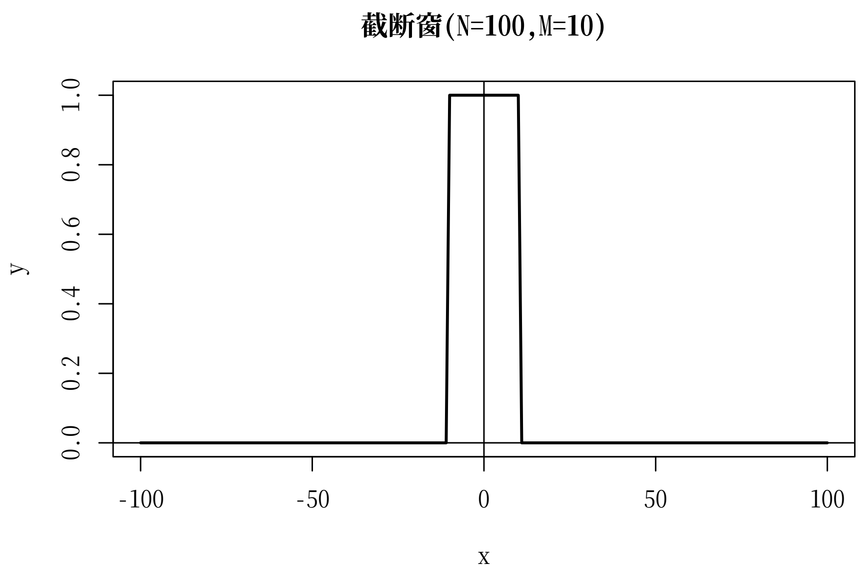
对应谱窗

$$\begin{aligned} W_N(\lambda) =& \frac{1}{2\pi} \sum_{|k| \leq M_N} e^{-ik\lambda} \\ =& \frac{1}{2\pi} \frac{\sin[(2M_n + 1)\lambda/2]}{\sin(\lambda/2)} \triangleq \frac{1}{2\pi} D_{2M_N+1}(\lambda) \end{aligned}$$

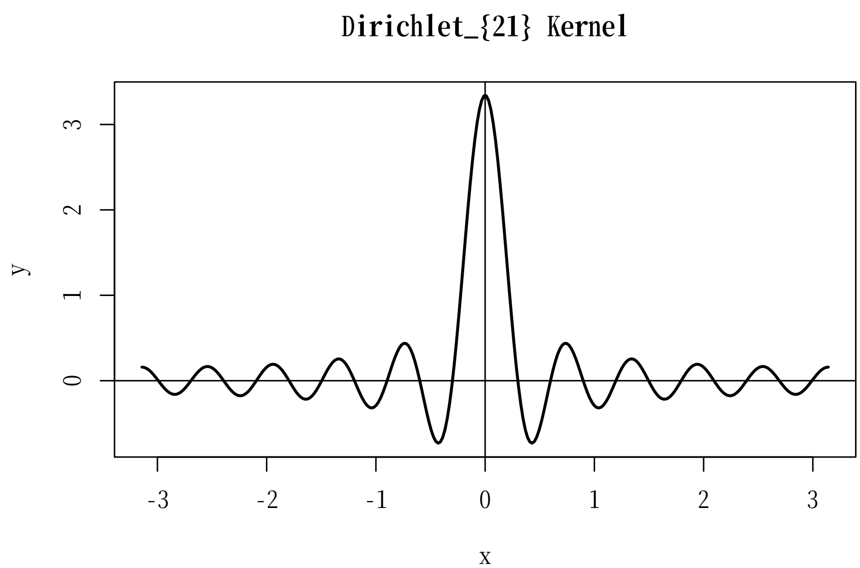
$D_K(\lambda)$ 称为 Dirichlet 核。它有正有负。

```
demo.rect <- function(){
  N <- 100
  M <- 10
  x <- seq(-N, N)
  y <- numeric(length(x))
  y[abs(x) <= M] <- 1
  plot(x, y, type="l", lwd=2,
       main=paste(" 截断窗 (N=100,M=10)", sep=""))
  abline(h=0)
  abline(v=0)
```

```
}
demo.rect()
```




```
demo.dirichlet <- function(K=11){
  x <- seq(-pi, pi, length=300)
  y <- 1/(2*pi) * sin(K*x/2) / sin(x/2)
  plot(x, y, type="l", lwd=2,
       main=paste("Dirichlet_{", K, "} Kernel", sep=""))
  abline(h=0)
  abline(v=0)
}
demo.dirichlet(21)
```



28.2.3.2 Bartlett 窗

时窗函数

$$\lambda_N(k) = \begin{cases} 1 - \frac{|k|}{M_N}, & |k| \leq M_N \\ 0, & |k| > M_N \end{cases}$$

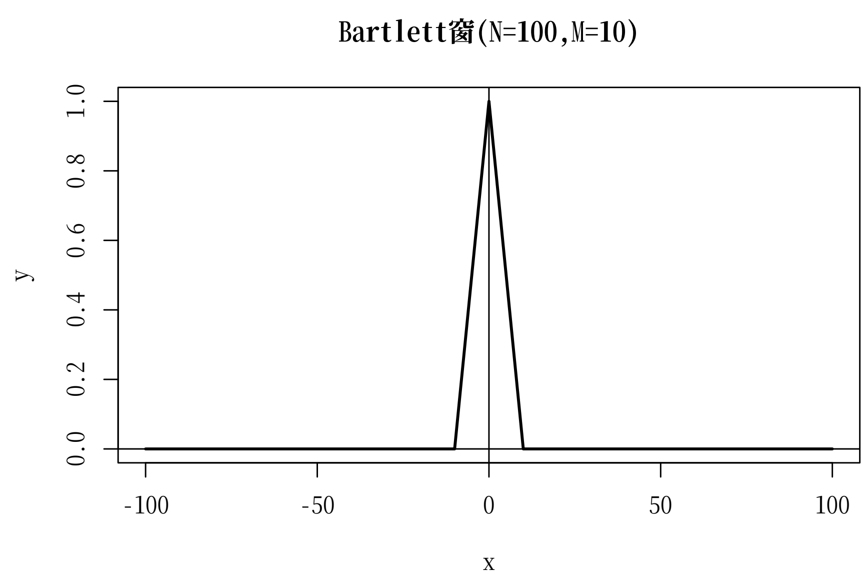
谱窗函数

$$\begin{aligned} W_N(\lambda) =& \frac{1}{2\pi} \sum_{|k| \leq M_N} \left(1 - \frac{|k|}{M_N} \right) e^{-ik\lambda} \\ =& \frac{1}{2\pi M_N} \left(\frac{\sin(M_N \lambda/2)}{\sin(\lambda/2)} \right)^2 \triangleq F_{M_N}(\lambda) \end{aligned}$$

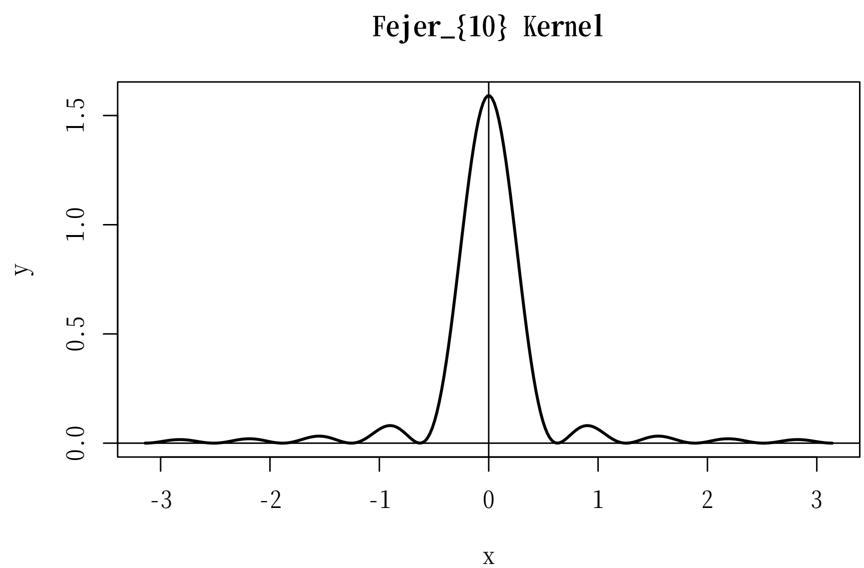
$F_{M_N}(\lambda)$ 称为 Fejer 核。是全非负的。

```
demo.tri <- function(){
  N <- 100
  M <- 10
  x <- seq(-N, N)
  y <- numeric(length(x))
  y[abs(x) <= M] <- 1 - abs(x[abs(x) <= M])/M
  plot(x, y, type="l", lwd=2,
```

```
      main=paste("Bartlett 窗 (N=100,M=10)", sep=""))
  abline(h=0)
  abline(v=0)
}
demo.tri()
```




```
demo.fejer <- function(K=10){
  x <- seq(-pi, pi, length=300)
  y <- 1/(2*pi *K) * (sin(K*x/2) / sin(x/2))^2
  plot(x, y, type="l", lwd=2,
      main=paste("Fejer_{", K, "} Kernel", sep=""))
  abline(h=0)
  abline(v=0)
}
demo.fejer()
```



28.2.3.3 Daniell 窗

谱窗函数：

$$W_N(\lambda) = \begin{cases} \frac{M_N}{2\pi}, & |\lambda| \le \frac{\pi}{M_N}, \\ 0, & |\lambda| > \frac{\pi}{M_N}. \end{cases}$$

谱估计

$$\hat{f}(\lambda) = \frac{M_N}{2\pi} \int_{\lambda - \frac{\pi}{M_N}}^{\lambda + \frac{\pi}{M_N}} I_N(s) ds$$

是 $I_N(s)$ 在 $[\lambda - \frac{\pi}{M_N}, \lambda + \frac{\pi}{M_N}]$ 的平均。对应的时窗函数为

$$\lambda_N(k) = \sin \frac{k\pi}{M_N} / \frac{k\pi}{M_N}.$$

28.2.3.4 Turkey 窗

时窗函数为

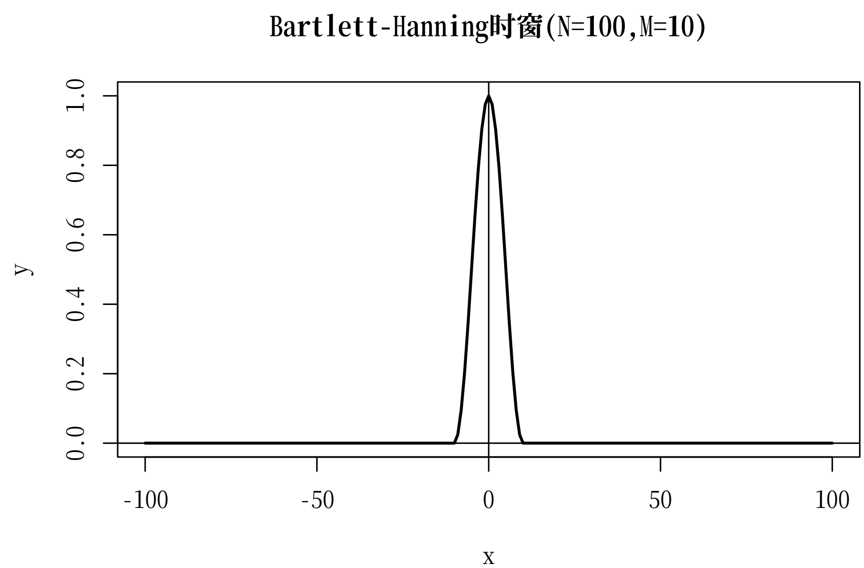
$$\lambda_N(k) = \begin{cases} 1 - 2a + 2a \cos \frac{k\pi}{M_N}, & |k| \le M_N, \\ 0, & |k| > M_N. \end{cases}$$

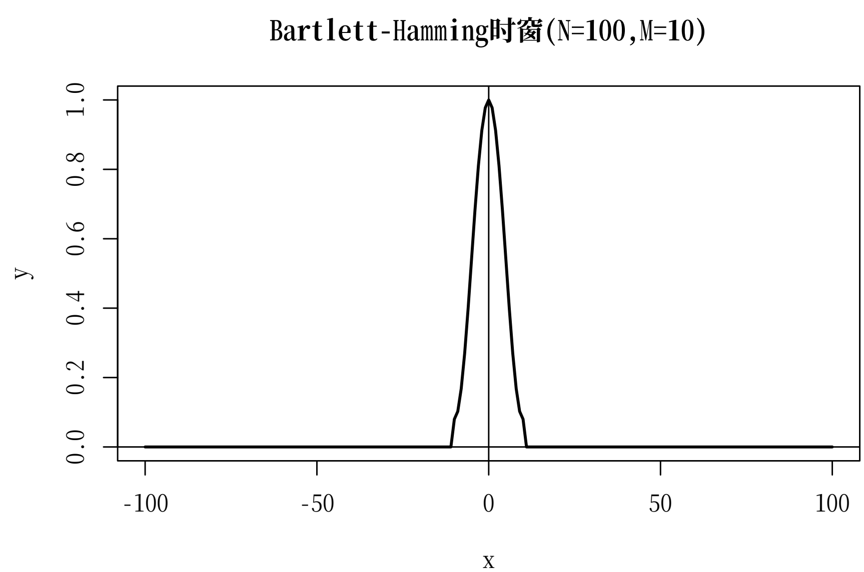
$a \in (0, 0.25]$. $a = 0.25$ 称为 Turkey-Hanning 窗，$a = 0.23$ 称为 Turkey-Hamming 窗。对应谱窗函数为

$$
\begin{aligned}
W_N(\lambda) =& aD(\lambda - \frac{\pi}{M_N}) + (1-2a)D(\lambda) + aD(\lambda + \frac{\pi}{M_N}), \\
D(\lambda) \triangleq& \frac{\sin[(2M_N+1)\lambda/2]}{2\pi\sin(\lambda/2)}
\end{aligned}
$$

谱窗有负值。

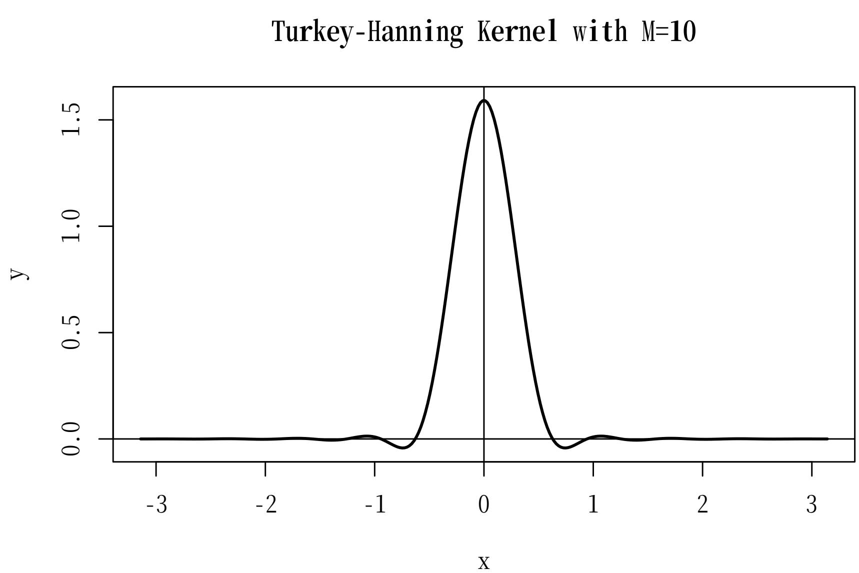
```
demo.turkey.t <- function(){
  N <- 100
  M <- 10
  x <- seq(-N, N)
  y <- numeric(length(x))
  a <- 0.25 # Hanning
  y[abs(x) <= M] <- 1 - 2*a + 2*a*cos(x[abs(x) <= M]*pi/M)
  plot(x, y, type="l", lwd=2,
       main=paste("Bartlett-Hanning 时窗 (N=100,M=10)", sep=""))
  abline(h=0)
  abline(v=0)

  a <- 0.23 # Hamming
  y[abs(x) <= M] <- 1 - 2*a + 2*a*cos(x[abs(x) <= M]*pi/M)
  plot(x, y, type="l", lwd=2,
       main=paste("Bartlett-Hamming 时窗 (N=100,M=10)", sep=""))
  abline(h=0)
  abline(v=0)
}
demo.turkey.t()
```



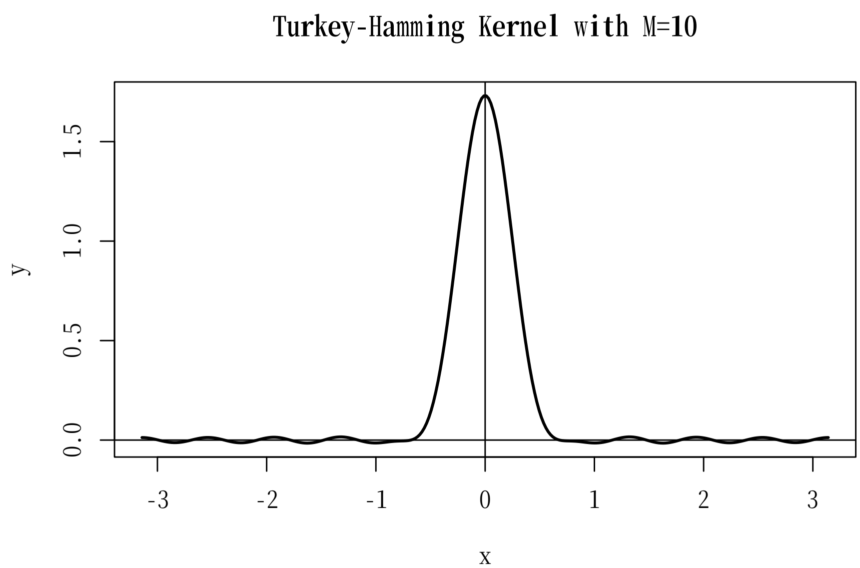



```
demo.turkey <- function(a = 0.25, M=10){
  x <- seq(-pi, pi, length=300)
  D <- function(x) 1/(2*pi) * sin((2*M+1)*x/2) / sin(x/2)
```

```
  y = a*D(x - pi/M) + (1-2*a)*D(x) + a*D(x + pi/M)
  if(a==0.25){
    tit <- paste("Turkey-Hanning Kernel with M=", M, sep="")
  } else if(a==0.23){
    tit <- paste("Turkey-Hamming Kernel with M=", M, sep="")
  } else {
    tit <- paste("Turkey Kernel with M=", M, " a=", a, sep="")
  }
  plot(x, y, type="l", lwd=2,  main=tit)
  abline(h=0)
  abline(v=0)
}
demo.turkey(a=0.25)
```




```
demo.turkey(a=0.23)
```



28.2.3.5 Parzen 窗

时窗函数为

$$\lambda_N(k) = \begin{cases} 1 - 6(k/M_N)^2 + 6(|k|/M_N)^3, & |k| \le M_N/2, \\ 2(1 - |k|/M_N)^3, & M_N/2 < k \le M_N, \\ 0, & |k| > M_N. \end{cases}$$

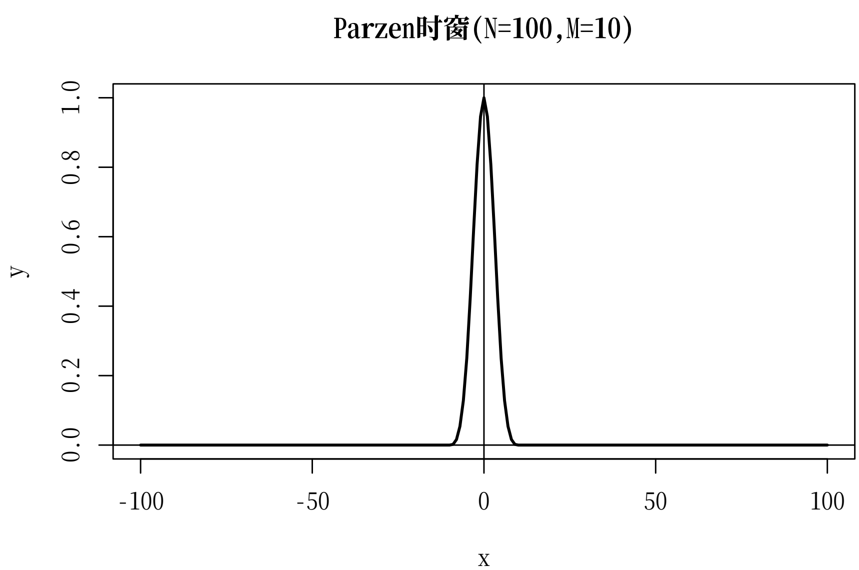
对应谱窗函数为

$$W_N(\lambda) = \frac{3}{8\pi M_N^3} \left(\frac{\sin(M_N \lambda/4)}{\sin(\lambda/2)/2} \right)^4 \left(1 - \frac{2}{3} \sin^2 \frac{\lambda}{2} \right).$$

谱窗非负。Parzen 窗在这几个加窗谱估计中估计方差最小。

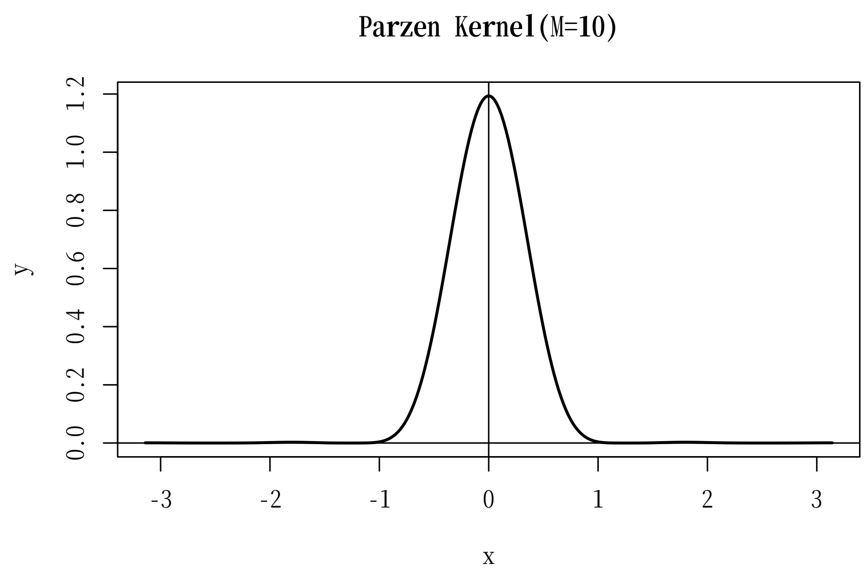
```
demo.parzen.t <- function(){
  N <- 100
  M <- 10
  M1 <- M/2
  x <- seq(-N, N)
  y <- numeric(length(x))
```

```
  a <- 0.25 # Hanning
  sele <- abs(x)<=M1
  y[sele] <- 1 - 6*(x[sele]/M)^2 + 6*(abs(x[sele]/M)^3)
  sele <- abs(x)>M1 & abs(x) <= M
  y[sele] <- 2*(1 - abs(x[sele])/M)^3
  plot(x, y, type="l", lwd=2,
       main=paste("Parzen 时窗 (N=100,M=10)", sep=""))
  abline(h=0)
  abline(v=0)
}
demo.parzen.t()
```




```
demo.parzen <- function(M=10){
  x <- seq(-pi, pi, length=300)
  y <- ((3/(8*pi*M^3)) * (sin(M*x/4) / (sin(x/2)/2))^4 *
        (1 - 2/3*sin(x/2)^2))
  plot(x, y, type="l", lwd=2,
       main=paste("Parzen Kernel(M=", M, ")", sep=""))
```

```
  abline(h=0)
  abline(v=0)
}
demo.parzen()
```



28.2.3.6 Bartlett-Priestley 窗

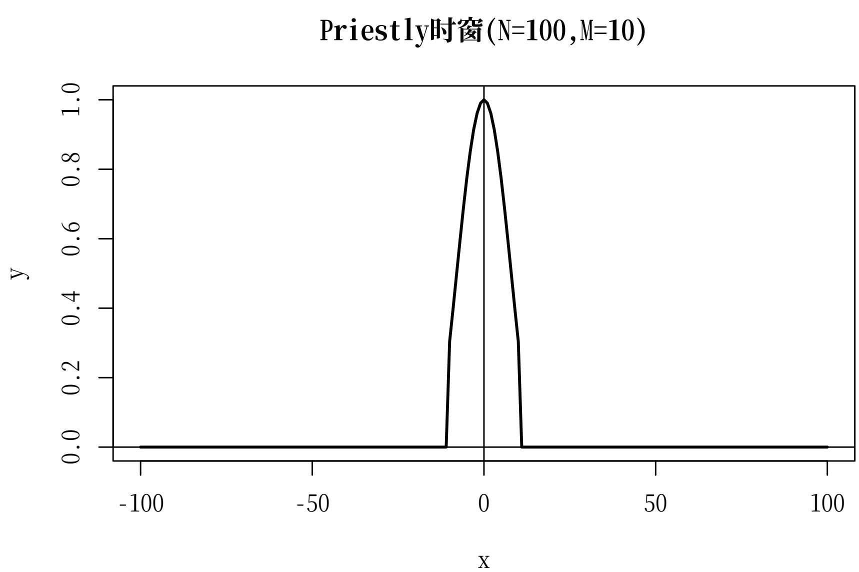
谱窗函数为

$$W_N(\lambda) = \begin{cases} \frac{3M_N}{4\pi}[1-(M_N\lambda/\pi)^2], & |\lambda| \le \pi/M_N \\ 0, & |\lambda| > \pi/M_N \end{cases}$$

是二次函数。对应的时窗函数为

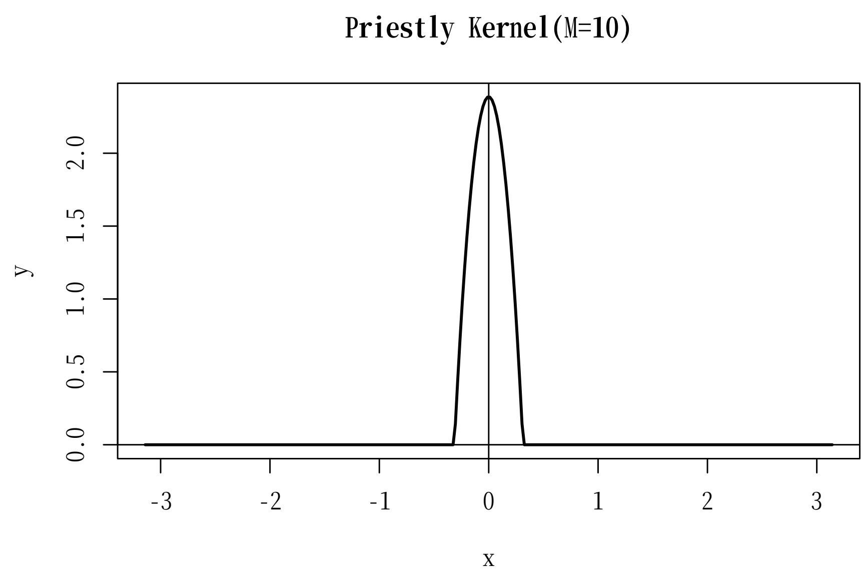
$$\lambda_N(k) = \begin{cases} \frac{3M_N^2}{(\pi k)^2}\left[\frac{\sin(\pi k/M_N)}{\pi k/M_N} - \cos(\pi k/M_N)\right], & 0 < |k| \le M_N, \\ 1, & k = 0, \\ 0, & |k| > M_N \end{cases}$$

```
demo.priestly.t <- function(){
  N <- 100
  M <- 10
  x <- seq(-N, N)
  y <- numeric(length(x))
  sele <- abs(x)<=M & x != 0
  y[sele] <- (3*M^2/(pi*x[sele])^2 *
               (sin(pi*x[sele]/M)/(pi*x[sele]/M) - cos(pi*x[sele]/M)))
  y[abs(x)<0.5] <- 1
  plot(x, y, type="l", lwd=2,
       main=paste("Priestly 时窗 (N=100,M=10)", sep=""))
  abline(h=0)
  abline(v=0)
}
demo.priestly.t()
```




```
demo.priestly <- function(M=10){
  x <- seq(-pi, pi, length=300)
```

```
  y <- numeric(length(x))
  sele <- abs(x) <= pi/M
  y[sele] <- 3*M/4/pi*(1 - (M*x[sele]/pi)^2)
  plot(x, y, type="l", lwd=2,
       main=paste("Priestly Kernel(M=", M, ")", sep=""))
  abline(h=0)
  abline(v=0)
}
demo.priestly()
```



28.3 加窗谱估计的比较

28.3.1 方差的比较

定理 28.4. *设 $\{X_t\}$ 是平稳正态序列，有连续可微的谱密度 $f(\lambda)$，谱窗函数 $W_N(\lambda)$ 满足 §28.2.2中的五个条件。加窗谱估计 $\hat{f}(\lambda)$ 由(28.8)定义。则有如下结果:*

(1) $\hat{f}(\lambda)$ 是 $f(\lambda)$ 的渐近无偏估计。

(2) $\hat{f}(\lambda)$ 是 $f(\lambda)$ 的均方相合估计。

(3) 当 N 充分大后，有

$$
\begin{aligned}
Var(\hat{f}(\lambda)) \approx& \begin{cases} \frac{M_N}{N} f(\lambda)^2 K^2, & \lambda \neq 0, \pm\pi, \\ 2\frac{M_N}{N} f(\lambda)^2 K^2, & \lambda = 0, \pm\pi, \end{cases} \\
K^2 =& \frac{2\pi}{M_N} \int_{-\pi}^{\pi} W_N^2(\lambda) d\lambda.
\end{aligned}
$$

虽然减小 M_N 会减小方差，但是可能因为减小 M_N 使得谱窗变宽，增加估计的光滑程度，所以可能会增加偏差，降低分辨率。

各谱窗方差比较:

窗名	K^2
截断窗	2.0000
Bartlett	0.6667
Daniell	1.0000
Turkey-Hamming	0.7948
Parzen	**0.5392**
Bartlett-Priestley	1.2000

28.3.2 分辨率的比较

谱估计对峰的个数、位置很感兴趣。两个相邻的峰需要能分开，不被混在一起。只有谱窗宽度比两个峰距离更短时才有可能分开。

例 28.5. 考虑 AR(4) 模型

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + a_3 X_{t-3} + a_4 X_{t-4} + \varepsilon,\ t \in \mathbb{Z}$$

其中

$$
\begin{aligned}
\{\varepsilon_t\} \sim& \text{WN}(0, 4), \\
a =& (-0.9337, -1.4599, -0.7528, -0.6355)^T
\end{aligned}
$$

特征多项式有两对共轭复根

$$1.115 e^{\pm 1.5i},\ 1.125 e^{\pm 2.2i}$$

谱密度 $f(\lambda)$ 在 $\lambda = 1.5, 2.2$ 处有明显峰值，距离为 $2.2 - 1.5 = 0.7$。

考虑用 Daniell 谱窗做谱估计：

$$W_N(\lambda) = \begin{cases} M_N/(2\pi) & |\lambda| \leq \pi/M_N \\ 0 & |\lambda| > \pi/M_N \end{cases}$$

M_N 越大，谱窗越窄。加窗谱估计为

$$\hat{f}(\lambda) = \frac{M_N}{2\pi} \int_{\lambda-\pi/M_N}^{\lambda+\pi/M_N} I_N(s) ds$$

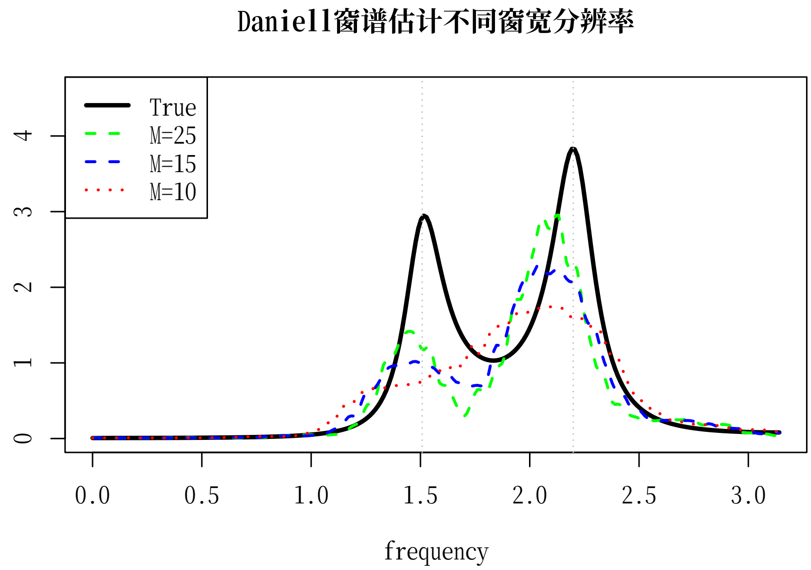
由周期图的渐近无偏性

$$E\hat{f}(\lambda) \approx \frac{M_N}{2\pi} \int_{\lambda-\pi/M_N}^{\lambda+\pi/M_N} f(s) ds$$

所以 $E\hat{f}(\lambda)$ 是 $f(s)$ 在 λ 的小邻域 $[\lambda - \pi/M_N, \ \lambda + \pi/M_N]$ 上的近似平均。如果这个小邻域太宽 (M_N 太小)，就会把两个峰混在一起。

```
### Daniel 谱窗谱估计
spec.Daniell <- function(y, freqs=seq(0,pi,length=101),
                         M = round(2*sqrt(length(y))),
                         demean=FALSE){
  N <- length(y)
  if(demean) y <- y - mean(y)
  spec <- numeric(length(freqs))
  gams <- c(acf(y, lag.max=N-1,
                type="covariance", plot=FALSE)$acf)
  jjs <- seq(N-1)
  for(j in seq(along=freqs))
    spec[j] <- 1/(2*pi)*(gams[1] +
                         2*sum(gams[jjs+1] *
                              ( sin(jjs*pi/M) / (jjs*pi/M) ) *
                              cos(jjs*freqs[j])))

  list(frequencies=freqs, spectrum=spec)
}
```

```r
demo.resolution <- function(seed=1){
  set.seed(seed)
  N <- 120
  a <- c(-0.9337, -1.4599, -0.7528, -0.6355)
  sigma <- 1
  y <- arima.sim(list(ar=a), n=N)*sigma

  ## 理论谱密度
  lams <- seq(0, pi, length=201)
  res1 <- ar.true.spectrum(a, sigma=sigma, plot.it=FALSE)
  lams <- res1$frequencies
  spec0 <- res1$spectrum

  spec1 <- spec.Daniell(y, freqs=lams, demean=FALSE, M=25)$spectrum
  spec2 <- spec.Daniell(y, freqs=lams, demean=FALSE, M=15)$spectrum
  spec3 <- spec.Daniell(y, freqs=lams, demean=FALSE, M=10)$spectrum


  ylim <- c(0, max(c(spec0, spec1, spec2, spec3)*1.2))
  plot(lams, spec0, lwd=3, type="l",
       ylim=ylim,
       main="Daniell 窗谱估计不同窗宽分辨率",
       xlab="frequency", ylab="")
  lines(lams, spec1, col="green", lwd=2, lty=2)
  lines(lams, spec2, col="blue", lwd=2, lty=2)
  lines(lams, spec3, col="red", lwd=2, lty=3)
  abline(v=c(1.508,2.199), lty=3, col="gray")
  legend("topleft", lwd=c(3,2,2,2), lty=c(1,2,2,3),
         col=c("black", "green", "blue", "red"),
         legend=c("True", "M=25", "M=15", "M=10"))
}
demo.resolution()
```



谱窗宽度的选择:

只有谱窗宽度比两个峰距离小很多时才有可能分开。但是，谱窗太窄 (M_N 太大)，估计方差会变大，估计的结果变得很不光滑，出现许多虚假的峰。谱密度和谱窗的宽度用带宽来描述。

设谱密度 $f(\lambda)$ 连续，在 λ_0 处有一个明显的峰值，如果有 $\omega_1 < \lambda_0 < \omega_2$ 使得

$$f(\omega_1) = f(\omega_2) = \frac{1}{2} f(\lambda_0), \tag{28.9}$$

$$f(\lambda) > \frac{1}{2} f(\lambda_0), \quad 当 \lambda \in (\omega_1, \omega_2), \tag{28.10}$$

则称 $\omega_2 - \omega_1$ 为 $f(\lambda)$ 在 λ_0 处的带宽。

若 $f(\lambda)$ 在 λ_0 处有一个明显的低谷，如果有 $\omega_1 < \lambda_0 < \omega_2$ 使得

$$f(\omega_1) = f(\omega_2) = 2 f(\lambda_0),$$
$$f(\lambda) < 2 f(\lambda_0), \quad 当 \lambda \in (\omega_1, \omega_2),$$

则称 $\omega_2 - \omega_1$ 为 $f(\lambda)$ 在 λ_0 处的带宽。

加窗谱估计要求谱窗的带宽比谱密度的带宽小很多，才能分辨出谱密度的峰值情况。

半功率带宽: $B_{\mathrm{HP}} = 2\theta_1$，$\theta_1$ 由

$$W_N(\theta_1) = \frac{1}{2} W_N(0).$$

这种谱窗带宽定义与(28.9)的谱密度带宽定义是一致的。

Parzen 带宽: $B_{\mathrm{P}} = 1/W_N(0)$.

Jenkin 带宽:

$$B_{\mathrm{J}} = \frac{2\pi}{\sum_{|k| \leq N-1} \lambda_N^2(k)}.$$

增加 M_N 可以减小谱窗的窗宽从而提高分辨率，但是会增大方差。不同 M_N 取值对分辨率的影响参见例28.5。

Part VI

波动率模型

Chapter 29

条件异方差模型

29.1 资产收益率

设 t 为某个固定时间单位的个数（比如天数、月数、年数），以天为例，用 P_t 表示某金融资产在第 t 天的价格。令

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}},$$

称为第 t 天的**简单收益率**。$1 + R_t$ 称为毛收益率。令

$$r_t = \log(1 + R_t) = \log \frac{P_t}{P_{t-1}} = \log P_t - \log P_{t-1},$$

称为第 t 天的**对数收益率**。

如果已知 $r_1, \dots, r_t$，则易见

$$\frac{P_t}{P_0} = \prod_{j=1}^{t} \frac{P_j}{P_{j-1}} = \prod_{j=1}^{t} (1 + R_j),$$

而

$$\log \frac{P_t}{P_0} = \sum_{j=1}^{t} \log \frac{P_j}{P_{j-1}} = \sum_{j=1}^{t} r_j,$$

所以对数收益率更容易进行数学推导。

29.2 ARCH 模型

对于资产收益率序列 $\{r_t\}$，如果它是高斯过程，则最优线性预测就是最优预测，通常只要建立一个 ARMA 这样的线性模型就足够了。但是，在金融市场中，资产收益率常常不是正态分布的，而且其条件分布 $X_t|\mathscr{F}_{t-1}$ 的条件方差不是恒定的，会随时间 t 变化，金融资产的收益率会有“波动率聚集”现象，即某一段时间的 r_t 波动较大，而另一端时间的 r_t 波动较小，所以有必要在对条件期望 $E(r_t|\mathscr{F}_{t-1})$ 建模的同时对条件方差 $\text{Var}(r_t|\mathscr{F}_{t-1})$ 建模。

设 $\{\varepsilon_t\}$ 是对条件均值建模后的残差，理论上，这相当于 Wold 分解中的新息，设其满足

$$E(\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\dots)=0.$$

$\{\varepsilon_t\}$ 是二阶平稳的宽白噪声，但这并不要求 $\{\varepsilon_t\}$ 独立，我们可以考虑能够表现波动率聚集的模型。因为 ε_t^2 代表了波动大小，我们尝试对其建立如下的 AR(1) 模型：

$$\varepsilon_t^2=\alpha_0+\alpha_1\varepsilon_{t-1}^2+\eta_t,$$

其中 $\alpha_0>0$, $\alpha_1\geq 0$，$\{\eta_t\}$ 是独立同分布零均值白噪声列。这样的模型中 ε_t^2 和 ε_{t-1}^2 是正相关的，所以能够体现波动率聚集性质。设 $E\varepsilon_t^2=\sigma^2$，则

$$\sigma^2=\alpha_0+\alpha_1\sigma^2+0,$$

将 ε_t^2 中心化，得

$$(\varepsilon_t^2-\sigma^2)=\alpha_1(\varepsilon_{t-1}^2-\sigma^2)+\eta_t,$$

可见 $\varepsilon_t^2-\sigma^2$ 满足一个 AR(1) 模型，其中 $0\leq\alpha_1<1$。记 $A(z)=1-\alpha_1 z$，则

$$A(\mathscr{B})(\varepsilon_t^2-\sigma^2)=\eta_t,$$

所以 ε_t^2 有平稳解

$$\varepsilon_t^2=\sigma^2+A^{-1}(\mathscr{B})\eta_t=\sigma^2+\sum_{j=0}^{\infty}\alpha_1^j\eta_{t-j}.$$

由此平稳解可见 η_t 与 $\{\varepsilon_{t-1},\varepsilon_{t-2},\dots\}$ 独立，从而

$$\sigma_t^2=\text{Var}(\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\dots)=E(\varepsilon_t^2|\varepsilon_{t-1},\varepsilon_{t-2},\dots)=\alpha_0+\alpha_1\varepsilon_{t-1}^2.$$

这就给出了条件方差的一个模型。进一步地可以设

$$\sigma_t^2=\alpha_0+\alpha_1\varepsilon_{t-1}^2+\cdots+\alpha_p\varepsilon_{t-p}^2.$$

注意上式中已经没有 η_t。

显然，这个模型中的系数应使得 $\alpha_0 > 0$, $\alpha_j \ge 0$, $j = 1, 2, \dots, p$，且多项式 $A(z) = 1 - \alpha_1 z - \dots - \alpha_p z^p$ 满足最小相位性。注意对应的 AR(p) 模型不能保证 ε_t^2 非负，但有下面的 ARCH 模型定义和下一节的平稳解结果。

定义 29.1 (ARCH 模型)**.** 设 $\{v_t\}$ 是独立同分布零均值标准白噪声 WN(0,1)，非负常数 $\alpha_0, \alpha_1, \dots, \alpha_p$ 满足 $\alpha_1 + \dots + \alpha_p < 1$ 且 $\alpha_0 > 0$, $\alpha_p > 0$，则如下模型

$$\begin{cases} \varepsilon_t = [\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2]^{1/2} v_t, \\ v_t\text{与}\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}\text{相互独立}, \end{cases} \tag{29.1}$$

称为 ARCH(p) 模型。如果 $\{\varepsilon_t\}$ 是严平稳白噪声且满足上述模型，则称其为 ARCH(p) 序列。

记

$$\sigma_t = [\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2]^{1/2}.$$

则模型(29.1)可以表述为

$$\begin{cases} \varepsilon_t = \sigma_t v_t, \\ \sigma_t = [\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2]^{1/2}, \\ v_t\text{与}\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}\text{相互独立}. \end{cases} \tag{29.2}$$

对 ARCH(p) 序列，来证明

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 = \text{Var}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots).$$

记 $\mathscr{F}_{t-1} = \sigma(\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\})$，则 v_t 与 $\mathscr{F}_{t-1}$ 独立而 σ_t^2 关于 $\mathscr{F}_{t-1}$ 可测，故 v_t 与 σ_t 独立。于是，

$$\begin{aligned} &\text{Var}(\varepsilon_t | \mathscr{F}_{t-1}) = E(\varepsilon_t^2 | \mathscr{F}_{t-1}) \\ =&E\left\{\sigma_t^2 v_t^2 | \mathscr{F}_{t-1}\right\} = \sigma_t^2 E(v_t^2 | \mathscr{F}_{t-1}) \\ =&\sigma_t^2 E(v_t^2) = \sigma_t^2. \end{aligned}$$

对 ARCH(p) 序列，设 $E v_t^4 < \infty$，$E \varepsilon_t^4 < \infty$，来证明 $\{\varepsilon_t^2\}$ 满足如下 AR(p) 模型：

$$(\varepsilon_t^2 - \sigma^2) = \alpha_1 (\varepsilon_{t-1}^2 - \sigma^2) + \dots + \alpha_p (\varepsilon_{t-p}^2 - \sigma^2) + \eta_t,$$

其中 η_t 为独立同分布零均值白噪声列。事实上，令

$$\eta_t = \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2(v_t^2 - 1),$$

注意 v_t 与 $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ 独立，v_t 与 σ_t 独立，所以有

$$\begin{aligned}
E(\eta_t) =& E(\sigma_t^2(v_t^2 - 1)) = E(\sigma_t^2)E(v_t^2 - 1) = 0, \\
E(\eta_t \eta_{t+k}) =& E[\sigma_t^2(v_t^2 - 1)\sigma_{t+k}^2(v_{t+k}^2 - 1)] = E(\sigma_t(v_t^2 - 1)\sigma_{t+k})E(v_{t+k}^2 - 1) = 0, \\
E(\eta_t^2) =& E[\sigma_t^4(v_t^2 - 1)^2] = E[\sigma_t^4]E[(v_t^2 - 1)^2] = E(\sigma_1^4)E[(v_1^2 - 1)^2].
\end{aligned}$$

这里利用了 $\{\varepsilon_t\}$ 严平稳，所以 $E\sigma_t^4$ 不依赖于 t。上式说明了 $\{\eta_t\}$ 是零均值白噪声列，从而

$$\varepsilon_t^2 = \sigma_t^2 + \eta_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \eta_t,$$

这是一个 AR(p) 模型。

29.3 ARCH 模型平稳解

引理 29.1. 设 $\alpha_0, \alpha_1, \dots$ 是非负常数，$\{u_t\}$ 是独立同分布非负随机变量序列，$E(u_t) = 1$。如果 $\alpha_0 > 0$，$c = \sum_{j=1}^\infty \alpha_j < 1$，则有以下结论：

(*1*) 有唯一的严平稳遍历序列 $\{Y_t\}$ 满足模型

$$\begin{cases} Y_t = \sigma_t^2 u_t, \text{ 其中} \sigma_t^2 = \alpha_0 + \sum_{j=1}^\infty \alpha_j Y_{t-j}. \\ E(Y_t) < \infty, \ \{Y_{t-1}, Y_{t-2}, \dots\} \text{与} u_t \text{独立}. \end{cases} \tag{29.3}$$

(*2*) 满足(29.3)的严平稳遍历序列可表示成

$$Y_t = \alpha_0 \sum_{n=1}^\infty A_t(n), \ E(Y_t) = \frac{\alpha_0}{1-c}, \ t \in \mathbb{N}. \tag{29.4}$$

其中 $A_t(0) = u_t$，对 $n \geq 1$，有

$$A_t(n) = \sum_{i_1, i_2, \dots, i_n \geq 1} \alpha_{i_1} \alpha_{i_2} \dots \alpha_{i_n} u_t u_{t-i_1} u_{t-(i_1+i_2)} \dots u_{t-(i_1+\dots+i_n)}; \tag{29.5}$$

若进一步假设 $c\sqrt{E(u_t^2)} < 1$，则 $E(Y_t^2) < \alpha_0 E(u_t^2)/(1 - c\sqrt{E(u_t^2)})^2$.

参见 Fan J Q, Yao Q W. Nonlinear time series: non-parametric and parametric methods. Springer-Verlag. 2005. 证明见 (何书元, 2023) 附录 1.5.

定理 29.1. 对于 *ARCH* 模型(29.2)，设 $c = \sum_{j=1}^{p} \alpha_j$，$u_t = v_t^2$，$A_t(n)$ 按(29.5)定义，其中当 $j > p$ 时 $\alpha_j = 0$。则

（*1*）存在严平稳遍历白噪声 $\{\varepsilon_t\}$ 满足 *ARCH* 模型(29.2)，其中

$$\varepsilon_t = \left[\alpha_0 \sum_{n=0}^{\infty} A_t(n)\right]^{1/2} u_t,\ E\varepsilon_t = 0,\ E\varepsilon_t^2 = \frac{\alpha_0}{1-c}; \tag{29.6}$$

（*2*）如果 $\{e_t\}$ 也是模型(29.2)的严平稳解，则 $e_t^2 = \varepsilon_t^2$, *a.s.;*

（*3*）如果 $c\sqrt{E(v_1^4)} < 1$，则

$$E(\varepsilon_t^4) < \alpha_0^2 E(v_1^4) / \left[1 - c\sqrt{E(v_1^4)}\right]^2.$$

证明：（1）取 $u_t = v_t^2$，则 $\{u_t\}$ 是独立同分布非负随机变量序列且 $E(u_t) = 1$。令 Y_t 为(29.4)的定义，则 $\{Y_t\}$ 符合引理29.1中的结论 (1)，即

$$\begin{cases} Y_t = \left(\alpha_0 + \sum_{j=1}^{p} \alpha_j Y_{t-j}\right) u_t, \\ E(Y_t) < \infty, \quad u_t\text{与}\{Y_{t-1}, Y_{t-2}, \dots\}\text{独立}. \end{cases} \tag{29.7}$$

取

$$\varepsilon_t = \left(\alpha_0 + \sum_{j=1}^{p} \alpha_j Y_{t-j}\right)^{1/2} v_t, \tag{29.8}$$

则 $\varepsilon_t^2 = Y_t$，从而可知 $E(\varepsilon_t^2) < \infty$。由定理4.1可知 $\{\varepsilon_t\}$ 是严平稳遍历序列。将(29.8)中的 Y_t 替换成 ε_t^2，可知 $\{\varepsilon_t\}$ 满足如下模型

$$\varepsilon_t = \left(\alpha_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2\right)^{1/2} v_t.$$

根据 Y_t 的定义可以看出 v_t 与 $\{Y_{t-1}, Y_{t-2}, \dots\}$ 独立，而 $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ 仅依赖于 $\{v_{t-1}, Y_{t-1}, Y_{t-2}, \dots\}$，所以 v_t 与 $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ 独立。

只要证明 $\{\varepsilon_t\}$ 是白噪声列。前面已证明 $E(\varepsilon_t^2) < \infty$，于是

$$\begin{aligned}
E(\varepsilon_t) =& E\left[\left(\alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2\right)^{1/2} v_t\right] \\
=& E\left[\left(\alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2\right)^{1/2}\right] E(v_t) \\
=& 0. \\
E(\varepsilon_t^2) =& E(Y_t) = \frac{\alpha_0}{1-c} = \frac{\alpha_0}{1-\sum_{j=1}^p \alpha_j}. \\
E(\varepsilon_s \varepsilon_t) =& E\left[\left(\alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{s-j}^2\right)^{1/2} v_s \left(\alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2\right)^{1/2} v_t\right] \\
=& E\left[\left(\alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{s-j}^2\right)^{1/2} v_s \left(\alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2\right)^{1/2}\right] E(v_t) \\
=& 0 \quad (s < t).
\end{aligned}$$

即 $\{\varepsilon_t\}$ 是零均值白噪声列，且是严平稳遍历序列，满足 ARCH 模型。

（2）如果 $\{e_t\}$ 也是模型(29.4)的严平稳解，则 $\{Y_t = e_t^2\}$ 是模型(29.7)的唯一严平稳解，根据 ε_t 定义可知 $e_t^2 = \varepsilon_t^2$，a.s.

（3）由引理29.1结论 (3)，若 $c\sqrt{E(v_1^4)} < 1$, 即 $c\sqrt{E(u_t^2)} < 1$, 则

$$E(\varepsilon_t^4) = E(Y_t^2) < \alpha_0^2 E(u_t^2) / \left[1 - c\sqrt{E(u_t^2)}\right]^2 = \alpha_0^2 E(v_1^4) / \left[1 - c\sqrt{E(v_1^4)}\right]^2.$$

由定理29.1结论 (3) 可知，当 $c\sqrt{E(v_1^4)} < 1$, 则 $E(\varepsilon_t^4) < \infty$，根据上一节的讨论可知这时 $\{\varepsilon_t^2\}$ 满足如下 AR(p) 模型：

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \eta_t, \tag{29.9}$$

其中 $\eta_t = \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2(v_t^2 - 1)$ 是零均值白噪声列,

$$\text{Var}(\eta_t) = E(\eta_t^2) = E(\sigma_t^4) E[(v_t^2 - 1)^2].$$

对随机变量 ξ，若 $E(\xi^4) < \infty$，则定义

$$\kappa_\xi = \frac{E(X - E(X))^4}{[\text{Var(X)}]^2},$$

称 κ_ξ 为 ξ 的**峰度**。对正态分布的 ξ，有 $\kappa_\xi = 3$，称 $\kappa_\xi - 3$ 为 ξ 的**超额峰度**。这个指标度量了随机变量分布的厚尾性，其样本的表现是异常值比正态分布更多。ARCH 模型主要用于金融资产收益率模型，金融资产收益率常常体现出厚尾分布特性，而 ARCH 模型一般是厚尾的。

例 29.1. 设 $\{\varepsilon_t\}$ 是 ARCH 模型(29.2)的严平稳解，κ_ε 是 ε_t 的峰度，κ_v 是 v_t 的峰度，则：

（1）$\kappa_\varepsilon \geq \kappa_v$，且等号成立当且仅当 σ_t^2 为常数值，即没有条件异方差性的情形；

（2）当 $E(v_1^4) < 1, E(\varepsilon_t^4) < 1$ 时，作为 AR(p) 序列的 $\{\varepsilon_t^2\}$ 序列的自相关系数都是非负的；特别地当 $\alpha_j > 0$ 时 $\rho_{kj} > 0, k = 1, 2, \dots$。

证明：（1）$\varepsilon_t = \sigma_t v_t$，$E(\varepsilon_t) = 0$，$E(v_t^2) = 1$，

$$E(\varepsilon_t^2) = E(\sigma_t^2 v_t^2) = E(\sigma_t^2)E(v_t^2) = E(\sigma_t^2).$$

从而

$$\begin{aligned}
E(\varepsilon_t^4) =& E(\sigma_t^4 v_t^4) = E(\sigma_t^4)E(v_t^4) \\
\geq& [E(\sigma_t^2)]^2 E(v_t^4), \\
\kappa_\varepsilon =& \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2} \\
\geq& \frac{[E(\sigma_t^2)]^2 E(v_t^4)}{[E(\sigma_t^2)]^2} \\
=& E(v_t^4) = \kappa_v.
\end{aligned}$$

等式成立的条件是等式 $E(\sigma_t^4) = [E(\sigma_t^2)]^2$，这当且仅当 $\sigma_t^2 = c$, a.s.

（2）这时 $E(v_t^4) < \infty$，$E(\varepsilon_t^4) < \infty$，所以 $\{\varepsilon_t^2\}$ 满足 AR(p) 模型(29.9)，令 $A(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$，有

$$\varepsilon_t^2 - E(\varepsilon_t^2) = A^{-1}(\mathscr{B})(\eta_t) = \sum_{j=0}^{\infty} c_j \eta_{t-j}.$$

记 $s = \sum_{j=1}^p \alpha_j t^j$，则

$$A^{-1}(t) = \frac{1}{1-s} = \sum_{k=0}^{\infty} s^k = \sum_{k=0}^{\infty} (\sum_{j=1}^{p} \alpha_j t^j)^k, \ t \in (0, 1].$$

则 $\{c_j\}$ 满足

$$\sum_{j=0}^{\infty} c_j t^j = \sum_{k=0}^{\infty} (\sum_{j=1}^{p} \alpha_j t^j)^k, \ t \in (0,1].$$

可见 $c_j \geq 0, j = 0, 1, 2, ...$，故 $\{\varepsilon_t^2\}$ 的协方差函数

$$\gamma_{\varepsilon_t^2}(k) = \sigma_\eta^2 \sum_{n=0}^{\infty} c_n c_{n+k} \geq 0.$$

即 $\{\varepsilon_t^2\}$ 的相关系数都是非负的。如果对某个 j_0 满足 $\alpha_{j_0} > 0$，则 $\sum_{k=0}^{\infty}(\sum_{j=1}^{p} \alpha_j t^j)^k$ 中 t^{kj_0} 系数为正值，从而 $c_{kj_0} > 0$，这时

$$\gamma_{\varepsilon_t^2}(kj_0) \geq \sigma_\eta^2 \sum_{m=0}^{\infty} c_{mj_0} c_{mj_0+kj_0} = \sigma_\eta^2 \sum_{m=0}^{\infty} c_{mj_0} c_{(m+k)j_0} > 0.$$

上面的例子说明 ARCH 模型对应的 $\{\varepsilon_t\}$ 序列通常是厚尾分布的，而且 $\{\varepsilon_t^2\}$ 都是正相关的。

如果金融收益率序列 $\{r_t\}$ 是平稳列，有 Wold 分解，其新息序列为 $\{\varepsilon_t\}$，设其满足鞅差条件：

$$E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = 0,$$

则可设 $\{\varepsilon_t\}$ 满足 ARCH(p) 模型，这个模型保证了波动率聚集效应（即 $\{\varepsilon_t^2\}$ 序列相关性为正相关），厚尾分布，条件方差随时间变化，称为 ARCH 效应。对实际数据建模时，可以对平稳数据先建立 ARMA 模型，将这样的模型看做是关于条件期望 $E(r_t | r_{t-1}, r_{t-2}, \dots)$ 的模型，然后将模型的残差看做是新息序列 $\{\varepsilon_t\}$，可以通过检验 $\{\varepsilon_t^2\}$ 是否白噪声列来检验是否具有 ARCH 效应，如果有 ARCH 效应则对 $\{\varepsilon_t\}$ 建立 ARCH(p) 模型，作为条件方差 $\text{Var}(r_t | r_{t-1}, r_{t-2}, \dots)$ 的模型。

29.4 ARCH 模型参数估计

办法是设定 v_t 的分布，称为“条件分布”，然后定义似然函数，进行最大似然估计。对 v_t 可以考虑使用正态分布、t 分布等类型，为了进行模型选择，可以计算 AIC 准则值，以及对拟合残差进行残差诊断。

29.5 GARCH 模型

ARCH 模型容易理解，有严平稳遍历解，但是在实际数据建模时，往往需要比较高阶才能良好拟合。为此，类似于从 AR 模型推广到 ARMA 模型，将 ARCH 模型推广为如下的 GARCH 模型（广义自回归条件异方差模型）。仍设 $\{r_t\}$ 为平稳的资产收益率，$\{\varepsilon_t\}$ 为其新息，满足 $E(\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\dots)=0$。

定义 29.2 (GARCH 模型)**.** 设 $\{v_t\}$ 为独立同分布 WN(0,1) 列，非负常数 α_i, β_j 满足条件

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1, \quad \alpha_0\alpha_p\beta_q > 0. \tag{29.10}$$

称

$$\begin{cases} \varepsilon_t = \sigma_t v_t, \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \\ v_t\text{与}\{\varepsilon_{t-1},\varepsilon_{t-2},\dots\}\text{独立}. \end{cases} \tag{29.11}$$

为 GARCH(p,q) 模型，其中 $\sigma_t^2=\mathrm{Var}(\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\dots)$。如果 $\{\varepsilon_t\}$ 是严平稳白噪声，满足模型(29.11)，则称 $\{\varepsilon_t\}$ 是 GARCH(p,q) 序列。

下面讨论模型(29.11)的严平稳解存在性。记 $h=\max(p,q)$，当 $i>p$ 时定义 $\alpha_i=0$, 当 $j>q$ 时定义 $\beta_j=0$。定义两个多项式

$$\alpha(t)=\sum_{i=1}^p \alpha_i t^i, \quad \beta(t)=\sum_{j=1}^q \beta_j t_j,\ t\in[-1,1].$$

由(29.10)，可以展开 $\frac{1}{1-\beta(t)}$ 为

$$\frac{1}{1-\beta(t)}=\sum_{k=0}^\infty [\beta(t)]^k,$$

于是

$$[1-\beta(t)]^{-1}\alpha(t)=\alpha(t)\sum_{k=0}^\infty [\beta(t)]^k=\sum_{n=1}^\infty c_n t^n. \tag{29.12}$$

因为 $\alpha(t)$ 和 $\beta(t)$ 都是非负系数的多项式，所以 $\sum_{k=0}^\infty [\beta(t)]^k$ 也是非负系数的，从而 c_n 都非负，且存在正值。

由(29.10)可知 $\alpha(1)+\beta(1)<1$，所以

$$0<c=\sum_{n=1}^\infty c_n = \frac{\alpha(1)}{1-\beta(1)}<1. \tag{29.13}$$

定理 29.2. 在 *GARCH(p,q)* 模型中，记 $u_t = v_t^2$，$\{c_n\}$，c 分别由(29.12)和(29.13)定义。引入

$$\begin{cases} A_t(0) = u_t, \\ A_t(n) = \sum_{i_1,i_2,\dots,i_n \ge 1} c_{i_1} c_{i_2} \dots c_{i_n} u_t u_{t-i_1} u_{t-i_2} \dots u_{t-i_n}. \end{cases} \tag{29.14}$$

则有以下结果：

（*1*）*GARCH* 模型(29.11)有如下的严平稳遍历解

$$\begin{cases} \varepsilon_t = \left[\alpha_0 \sum_{n=0}^{\infty} A_t(n)\right]^{1/2} v_t, \\ E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = \frac{\alpha_0}{1-\sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j}. \end{cases} \tag{29.15}$$

（*2*）如果 $\{e_t\}$ 也是(29.11)的严平稳解，则 $e_t^2 = \varepsilon_t^2$，*a.s.*

（*3*）如果 $c\sqrt{E(v_1^4)} < 1$，则 $E(\varepsilon_t^4) < \alpha_0^2 E(v_1^4)/(1 - c\sqrt{E(v_1^4)})^2$.

证明：（1）令 $c_0 = \alpha_0/(1-\beta(1))$。$\{c_n\}$, $\{u_t\}$ 满足引理29.1的条件，所以模型

$$\begin{cases} Y_t = \left(c_0 + \sum_{j=1}^{\infty} c_j Y_{t-j}\right) u_t, \\ u_t \text{与} \{Y_{t-1}, Y_{t-2}, \dots\} \text{独立} \end{cases}$$

的唯一的严平稳遍历解是

$$Y_t = c_0 \sum_{n=0}^{\infty} A_t(n).$$

并且 $E(Y_t) = c_0/(1-c) < \infty$. 定义

$$\varepsilon_t = \left(c_0 + \sum_{j=1}^{\infty} c_j Y_{t-j}\right)^{1/2} v_t, \tag{29.16}$$

则 $\varepsilon_t^2 = Y_t$，由定理4.1可知 $\{\varepsilon_t\}$ 是严平稳遍历序列，由 $A_t(n)$ 定义可以看出 v_t 与 $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ 独立，于是

$$E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = E\left(c_0 + \sum_{j=1}^{\infty} c_j \varepsilon_{t-j}^2\right)^{1/2} E(v_t) = 0,$$

又

$$\begin{aligned}\sigma_t^2 =& \text{Var}(\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\dots)\\ =& E(\varepsilon_t^2|\varepsilon_{t-1},\varepsilon_{t-2},\dots)\\ =& E(Y_t^2|\varepsilon_{t-1},\varepsilon_{t-2},\dots)\\ =& \left(c_0+\sum_{j=1}^{\infty}c_j\varepsilon_{t-j}^2\right)E(v_t^2|\varepsilon_{t-1},\varepsilon_{t-2},\dots)\\ =& \left(c_0+\sum_{j=1}^{\infty}c_j\varepsilon_{t-j}^2\right)E(v_t^2)\\ =& c_0+\sum_{j=1}^{\infty}c_j\varepsilon_{t-j}^2.\end{aligned}$$

由定理4.1可知 $\{\sigma_t\}$ 也是严平稳遍历序列。上式表明 ARCH(p,q) 模型是一种无穷阶的 ARCH 模型。用推移算子可以将上式写成

$$\sigma_t^2 = c_0 + [1-\beta(\mathscr{B})]^{-1}\alpha(\mathscr{B})\varepsilon_t^2.$$

由 $c_0=\frac{\alpha_0}{1-\beta(1)}=[1-\beta(\mathscr{B})]^{-1}\alpha_0$ 可见

$$\sigma_t^2=[1-\beta(\mathscr{B})]^{-1}\alpha_0+[1-\beta(\mathscr{B})]^{-1}\alpha(\mathscr{B})\varepsilon_t^2=[1-\beta(\mathscr{B})]^{-1}(\alpha_0+\alpha(\mathscr{B}))\varepsilon_t^2.$$

在上式两边作用 $1-\beta(\mathscr{B})$ 得

$$[1-\beta(\mathscr{B})]\sigma_t^2=\alpha_0+\alpha(\mathscr{B})\varepsilon_t^2,$$

可写成

$$\sigma_t^2=\alpha_0+\alpha(\mathscr{B})\varepsilon_t^2+\beta(\mathscr{B})\sigma_t^2,$$

即

$$\sigma_t^2=\alpha_0+\alpha_1\varepsilon_{t-1}^2+\dots+\alpha_p\varepsilon_{t-p}^2+\beta_1\sigma_{t-1}^2+\dots+\beta_q\sigma_{t-q}^2.$$

即严平稳遍历序列 $\{\varepsilon_t\}$ 满足 GARCH 模型(29.11)，

$$E(\varepsilon_t^2)=E(Y_t)=\frac{c_0}{1-c}=\frac{\alpha_0/(1-\beta(1))}{1-\frac{\alpha(1)}{1-\beta(1)}}=\frac{\alpha_0}{1-\beta(1)-\alpha(1)}.$$

与定理29.1证明类似可见 $\{\varepsilon_t\}$ 是白噪声列。

（2）和（3）的证明与定理29.1证明类似。

令 $h=\max(p,q)$，

$$A(z)=1-\alpha(z)-\beta(z), \quad B(z)=1-\beta(z),$$

则

$$A(z)=1-\sum_{j=1}^{h}(\alpha_j+\beta_j)z^j, \quad B(z)=1-\sum_{j=1}^{q}\beta_j z^j.$$

条件(29.10)保证 $A(t)$ 满足最小相位性，$B(t)$ 满足可逆性。事实上，记 $a_j=\alpha_j+\beta_j$，则 $a_j\geq 0$，$\sum_{j=1}^h a_j=\sum_{j=1}^p \alpha_j+\sum_{j=1}^q \beta_j<1$。于是对任意复数 z 满足 $|z|\leq 1$，有

$$\left|\sum_{j=1}^{h} a_j z^j\right| \leq \sum_{j=1}^{h} a_j |z|^j \leq \sum_{j=1}^{h} a_j < 1,$$

所以 $1-\sum_{j=1}^h a_j z^j$ 在 $|z|\leq 1$ 时没有零点，即 $A(z)$ 满足最小相位条件，类似可知 $B(z)$ 满足可逆条件。

命题 29.1. *如果 $E(v_1^4)<\infty$, $E(\varepsilon_1^4)<\infty$，则 GARCH(p,q) 序列 $\{\varepsilon_t^2\}$ 满足如下 ARMA(p,q) 模型：*

$$A(\mathscr{B})\varepsilon_t^2=\alpha_0+B(\mathscr{B})\eta_t, \tag{29.17}$$

其中 $\{\eta_t\}$ 是严平稳的零均值白噪声列。

证明： 令 $\eta_t=\varepsilon_t^2-\sigma_t^2$，则 $\{\eta_t\}$ 为严平稳列，$\varepsilon_t^2=\sigma_t^2+\eta_t$，代入到(29.11)中可得

$$\begin{aligned}
\varepsilon_t^2 =&\sigma_t^2+\eta_t \\
=&\alpha_0+\alpha(\mathscr{B})\varepsilon_t^2+\beta(\mathscr{B})\sigma_t^2+\eta_t \\
=&\alpha_0+\alpha(\mathscr{B})\varepsilon_t^2+\beta(\mathscr{B})(\varepsilon_t^2-\eta_t)+\eta_t \\
=&\alpha_0+[\alpha(\mathscr{B})+\beta(\mathscr{B})]\varepsilon_t^2+\eta_t-\beta(\mathscr{B})\eta_t \\
=&\alpha_0+[1-A(\mathscr{B})]\varepsilon_t^2+B(\mathscr{B})\eta_t,
\end{aligned}$$

这就是模型(29.17)，且满足最小相位条件和可逆性条件。只要再证明 $\{\eta_t\}$ 是零均值白噪声列。

$$\eta_t=\varepsilon_t^2-\sigma_t^2=\sigma_t^2(v_t^2-1),$$

由 v_1 四阶矩有限和 ε_1 四阶矩有限可知 η_t 二阶矩有限，由 $\{\eta_t\}$ 严平稳可知 $\{\eta_t\}$ 宽平稳。由 v_t 与 $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ 独立，σ_t 由 $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ 决定，可知 v_t 与 σ_t 独立，从而

$$\begin{aligned} E(\eta_t) =& E(\sigma_t^2) E(v_t^2 - 1) = 0, \\ E(\eta_t \eta_{t+k}) =& E(\sigma_t^2 (v_t^2 - 1) \sigma_{t+k}^2) E(v_{t+k}^2 - 1) = 0. \end{aligned}$$

由于 $E(\varepsilon_1^4) < \infty$, $E(v_1^4) < \infty$，必有 $E(\sigma_1^4) < \infty$，这是因为

$$\sigma_t^2 = E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots),$$

由条件 Jensen 不等式有

$$\sigma_t^4 = [E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)]^2 \leq E(\varepsilon_t^4 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \text{ a.s.},$$

从而 $E(\sigma_t^4) \leq E(\varepsilon_t^4) < \infty$。由 $\{\sigma_t\}$ 的严平稳性可知

$$E(\eta_t^2) = E(\sigma_t^4) E[(v_t^2 - 1)^2] < \infty,$$

且不依赖于 t，从而 $\{\eta_t\}$ 是严平稳的零均值白噪声列。

GARCH 序列也具有厚尾性。

命题 29.2. 设 $\{\varepsilon_t\}$ 是 *GARCH(p, q)* 序列，κ_ε 是 ε_t 的峰度，κ_v 是 v_t 的峰度，则

(*1*) $\kappa_\varepsilon \geq \kappa_v$，等号成立当且仅当 σ_t 为常数 *(a.s.)*;

(*2*) 如果 $E(v_1^4) < \infty$, $E(\varepsilon_1^4) < \infty$，则 $\{\varepsilon_t^2\}$ 的自相关系数 $\rho_k = Corr(\varepsilon_t^2, \varepsilon_{t+k}^2)$ 都非负，且当 $\alpha_j > 0$ 时所有的 $\rho_{kj} > 0$，$k = 1, 2, \dots$。

证明略。

Part VII

多元时间序列分析

Chapter 30

多元平稳序列介绍

30.1 多维平稳序列的概念

沿时间变化的量经常有多个，互相之间有相关性。如第 1 章的北京地区洪涝灾害受灾面积 X_t 和成灾面积 Y_t 有正相关，可以看成向量值时间序列

$$(X_t, Y_t),\ t \in \mathbb{N}$$

向量值的时间序列称为多维（多元）时间序列。仅介绍多维平稳序列。

定义 30.1 (多维平稳序列)**.** 称 m 维随机序列 $X_t = (X_{1t}, X_{2t}, \dots, X_{mt})^T$, $t \in \mathbb{Z}$ 是**平稳序列**，如果对任何 $t, n \in \mathbb{Z}$,

- (1) $EX_t = \mu = (\mu_1, \mu_2, \dots, \mu_m)^T$ 与 t 无关；
- (2) $\Gamma(n) = E[(X_{t+n} - \mu)(X_t - \mu)^T]$ 与 t 无关。

这时称 $\{\Gamma(n), n \in \mathbb{Z}\}$ 为平稳序列 $\{X_t\}$ 的自协方差函数（矩阵）。

定义

$$\gamma_{jk}(n) = E[(X_{j,t+n} - \mu_j)(X_{kt} - \mu_k)] \tag{30.1}$$

则

$$\Gamma(n) = (\gamma_{jk}(n))_{j,k=1,2,\dots,m}. \tag{30.2}$$

定义自相关系数

$$\rho_{jk}(n) = \frac{\gamma_{jk}(n)}{\sqrt{\gamma_{jj}(0)\gamma_{kk}(0)}} \tag{30.3}$$

称

$$R(n) = (\rho_{jk}(n))_{j,k=1,2,\ldots,m}$$

为 $\{X_t\}$ 的自相关系数（矩阵）。

把 $\{X_t\}$ 标准化得

$$Y_t = \left(\frac{X_{1t}-\mu_1}{\sqrt{\gamma_{11}(0)}}, \frac{X_{2t}-\mu_2}{\sqrt{\gamma_{22}(0)}}, \ldots, \frac{X_{mt}-\mu_m}{\sqrt{\gamma_{mm}(0)}}\right)^T$$

则 $\{R(n)\}$ 是平稳序列 $\{Y_t\}$ 的自协方差函数。

定理 30.1. 对任何 $n \in \mathbb{Z}$,

- *(1)* $\Gamma(-n) = [\Gamma(n)]^T$;
- *(2)* $|\gamma_{jk}(n)| \le [\gamma_{jj}(0)\gamma_{kk}(0)]^{1/2}$;
- *(3)* 非负定性：对任何实向量 $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}^m$,

$$\sum_{k=1}^{n}\sum_{j=1}^{n} \alpha_k^T \Gamma(k-j)\alpha_j \ge 0.$$

这些性质 $\{R(n)\}$ 也满足。

证明: (1)

$$\begin{aligned}\Gamma(-n) =& E\left[(X_{t-n}-\mu)(X_t-\mu)^T\right] \\ =& \left\{E\left[(X_t-\mu)(X_{t-n}-\mu)^T\right]\right\}^T \\ =& \left[\Gamma(n)\right]^T\end{aligned}$$

(2) 这就是 Schwarz 不等式。

(3) 记 $Y_t = X_t - \mu$, 则

$$
\begin{aligned}
&E\left[\sum_{j=1}^{n}\alpha_j^T Y_j\right]^2\\
=&E\sum_{k=1}^{n}\sum_{j=1}^{n}(\alpha_k^T Y_k)(\alpha_j^T Y_j)\\
=&\sum_{k=1}^{n}\sum_{j=1}^{n}\alpha_k^T E[Y_k Y_j^T]\alpha_j\\
=&\sum_{k=1}^{n}\sum_{j=1}^{n}\alpha_k^T \Gamma(k-j)\alpha_j \geq 0
\end{aligned}
$$

定义 30.2 (平稳相关). 设 $\{X_t\}$ 和 $\{Y_t\}$ 是两个平稳列，如果 $\mathrm{Cov}(X_s, Y_t) = r_{s-t}$ 对任意 s,t 成立，则称 $\{X_t\}$ 和 $\{Y_t\}$ **平稳相关**。

例 30.1 (多维白噪声). 若 m 维平稳序列 $\{X_t\}$ 满足

$$EX_1 = \mu, \quad \Gamma(n) = Q\delta_n$$

就称 $\{X_t\}$ 是m **维白噪声**，简记为 $\mathrm{WN}(\mu, Q)$。

这时 $\forall k, j$, 只要 $n \neq 0$，就有

$$E[(X_{kt} - \mu_k)(X_{j,t+n} - \mu_j)] = 0$$

另外，每个分量是一维白噪声。特别地，当 Q 是对角阵时，各分量是 m 个互不相关的白噪声。

例 30.2 (多维线性平稳列). 设 $\{\varepsilon_t, t \in \mathbb{Z}\}$ 是 m 维 $\mathrm{WN}(0,Q)$, $Q = E[\varepsilon_t \varepsilon_t^T]$。如果实矩阵列 $\{C_j\}$ 满足

$$\sum_{j=-\infty}^{\infty} C_j Q C_j^T < \infty$$

就称

$$X_t = \sum_{j=-\infty}^{\infty} C_j \varepsilon_{t-j},\ t \in \mathbb{Z} \tag{30.4}$$

为m **维平稳线性序列**。

这时可以证明(30.4)中每个分量都是均方收敛的，并且

$$EX_t = 0, \tag{30.5}$$

$$\Gamma(n) = E[X_{t+n}X_t^T] = \sum_{j=-\infty}^{\infty} C_{j+n}QC_j^T \tag{30.6}$$

例 30.3 (多维 MA 模型). 设 $\{\varepsilon_t\}$ 是 m 维 WN(0,Q)。如果 $B_1, B_2, \dots, B_q$ 是 $m \times m$ 实矩阵，满足

$$\det(I_m + B_1 z + B_2 z^2 + \cdots + B_q z^q) \neq 0,\ |z| < 1$$

就称

$$X_t = \varepsilon_t + B_1\varepsilon_{t-1} + \cdots + B_q\varepsilon_{t-q},\ t \in \mathbb{Z}$$

为一个m **维 MA(q) 序列**。

若进一步要求

$$\det(I_m + B_1 z + B_2 z^2 + \cdots + B_q z^q) \neq 0,\ |z| \leq 1$$

则称 $\{X_t\}$ 为一个可逆的 MA(q) 序列。

矩阵系数多项式：

设 $A_1, A_2, \dots, A_p, B_1, B_2, \dots, B_q$ 是 $p+q$ 个 $m \times m$ 实矩阵，记

$$A(z) = I_m - A_1 z - A_2 z^2 - \cdots - A_p z^p$$
$$B(z) = I_m + B_1 z + B_2 z^2 + \cdots + B_q z^q$$

形如 $A(z), B(z)$ 的多项式被称为**矩阵系数多项式**，实际是 m^2 维向量值函数，每个分量是多项式。

如果对任意满足

$$A(z) = C(z)A_1(z),\ B(z) = C(z)B_1(z)$$

的 $m \times m$ 矩阵系数多项式 $C(z)$ 必有 $\det(C(Z)) =$ 常数，就称 $A(z)$ 和 $B(z)$ 是**左互素**的。

例 30.4 (多维 ARMA 的不可识别性). 设 $\{\varepsilon_t\}$ 是 WN(0,Q)。称 m 维平稳序列 $\{X_t\}$ 满足 m 维平稳可逆的 ARMA(p,q) 模型，如果对任何 $t \in \mathbb{Z}$，

$$X_t = \sum_{j=1}^{p} A_j X_{t-j} + \varepsilon_t + \sum_{j=1}^{q} B_j \varepsilon_{t-j} \tag{30.7}$$

其中多项式 $A(z), B(z)$ 满足

$$\begin{aligned}&(1)A(z), B(z)\text{左互素};\\&(2)\det(A(z)B(z)) \neq 0,\ |z| \leq 1.\end{aligned}$$

用推移算子把模型方程写成

$$A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t,\ t \in \mathbb{Z} \tag{30.8}$$

对于一维 ARMA，自协方差函数可以唯一决定模型参数；对于多维 ARMA 不能。

在模型(30.8)中，如果 $\det(A(z)) = c$ 是常数，则 $A^{-1}(z)$ 仍然是一个矩阵多项式，并且 $\det(A^{-1}(z)) = c^{-1}$。于是

$$X_t = A^{-1}(\mathscr{B})B(\mathscr{B})\varepsilon_t = B_1(\mathscr{B})\varepsilon_t,\ t \in \mathbb{Z}$$

其中 $B_1(z) = A^{-1}(z)B(z)$ 是矩阵系数多项式，满足

$$\det(B_1(z)) = c^{-1}\det(B(z)) \neq 0,\ |z| \leq 1$$

就有了两个模型。所以 ARMA 模型参数是不唯一的。

30.2 多维平稳序列的均值和自协方差函数的估计

30.2.1 均值的估计

设 $\{X_t\}$ 是 m 维平稳序列, $X_1, X_2, \dots, X_N$ 是观测值。均值 μ 的估计为

$$\hat{\mu}_N = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_m)^T = \frac{1}{N}\sum_{t=1}^{N} X_t.$$

由分量的相应结论可得：

定理 30.2. 如果 $\{X_t\}$ 的每个分量序列 $\{X_{jt} : t \in \mathbb{Z}\}$ 都是严平稳遍历序列，则当 $N \to \infty$ 时，

$$\widehat{\mu}_N \to \mu, \ a.s.$$

定理 30.3. 如果自协方差函数 $\Gamma(n) \to 0$, 当 $n \to \infty$，则

$$E|\hat{\mu}_N - \hat{\mu}|^2 \to 0, \text{ 当 } N \to \infty \text{ 时}$$

其中 $|\hat{\mu}_N - \mu|^2 = \sum_{j=1}^m (\hat{\mu}_j - \mu_j)^2$。

这是均方收敛，推出依概率收敛，即 $\widehat{\mu}_N$ 是 μ 的相合估计。

证明:

$$\begin{aligned}
E|\hat{\mu}_N - \hat{\mu}|^2 &= \sum_{j=1}^m E(\hat{\mu}_j - \mu_j)^2 \\
&= \sum_{j=1}^m E\left(\frac{1}{N}\sum_{t=1}^N X_{jt} - \mu_j\right)^2 \\
&= \sum_{j=1}^m \frac{1}{N^2} E\left(\sum_{t=1}^N (X_{jt} - \mu_j)\right)^2 \\
&= \sum_{j=1}^m \frac{1}{N^2} \sum_{l=1}^N \sum_{k=1}^N \gamma_{jj}(l-k) \\
&= \sum_{j=1}^m \frac{1}{N^2} \sum_{k=1-N}^{N-1} (N-|k|)\gamma_{jj}(k) \\
&\le \sum_{j=1}^m \frac{1}{N} \sum_{k=1-N}^{N-1} |\gamma_{jj}(k)| \to 0, \ (N \to \infty)
\end{aligned}$$

定理 30.4. 如果

$$\sum_{k=0}^{\infty} |\gamma_{jj}(k)| < \infty, \ j = 1, 2, \dots, m$$

则当 $N \to \infty$ 时

$$NE|\hat{\mu}_N - \mu|^2 \to \sum_{j=1}^m \sum_{k=-\infty}^{\infty} \gamma_{jj}(k).$$

证明按定理30.3的证明，有

$$
\begin{aligned}
& NE|\hat{\mu}_N - \mu|^2 \\
=& \sum_{j=1}^{m} \frac{1}{N} \sum_{k=1-N}^{N-1} (N - |k|)\gamma_{jj}(k) \\
=& \sum_{j=1}^{m} \sum_{k=1-N}^{N-1} \gamma_{jj}(k) - \sum_{j=1}^{m} \frac{1}{N} \sum_{k=1-N}^{N-1} |k|\gamma_{jj}(k) \\
\to & \sum_{j=1}^{m} \sum_{k=-\infty}^{\infty} \gamma_{jj}(k).
\end{aligned}
$$

定理 30.5. 设 $\{\varepsilon_t\}$ 是 m 维独立同分布的 $WN(0, Q)$，$m \times m$ 矩阵 $C_n = (c_{j,k}(n))$ 的每个元素 $\{c_{j,k}(n)\}$ 对 $n \in \mathbb{Z}$ 绝对可和，m 维平稳序列 $\{X_t\}$ 为线性平稳列

$$
X_t = \mu + \sum_{j=-\infty}^{\infty} C_j \varepsilon_{t-j}, \ t \in \mathbb{Z}
$$

如果

$$
\Sigma = \left(\sum_{j=-\infty}^{\infty} C_j \right) Q \left(\sum_{j=-\infty}^{\infty} C_j^T \right) \neq 0
$$

则

$$
\sqrt{N}(\hat{\mu}_N - \mu) \xrightarrow{d} N(0, \Sigma)
$$

见 (Brockwell & Davis, 1987)。

30.2.2 自协方差函数的估计

设 $\{X_t\}$ 是 m 维平稳序列, $X_1, X_2, \dots, X_N$ 是观测值。自协方差函数 $\Gamma(n)$ 的估计为

$$
\begin{cases}
\hat{\Gamma}(n) = \frac{1}{N} \sum_{t=1}^{N-n} (X_{t+n} - \hat{\mu}_N)(X_t - \hat{\mu}_N)^T, & 0 \leq n \leq N-1 \\
\hat{\Gamma}(-n) = \hat{\Gamma}^T(n), & 1 \leq n \leq N-1
\end{cases}
$$

是一维情况的推广，具有良好统计性质。

若 $\hat{\gamma}_{jk}(n)$ 为 $\hat{\Gamma}(n)$ 的第 (j,k) 元素，则相关系数估计为

$$
\hat{\rho}_{jk}(n) = \frac{\hat{\gamma}_{jk}(n)}{\sqrt{\hat{\gamma}_{jj}(0)\hat{\gamma}_{kk}(0)}} \tag{30.9}
$$

自相关系数矩阵的估计为

$$\hat{R}(n) = (\hat{\rho}_{jk}(n)). \tag{30.10}$$

利用严平稳序列的遍历定理可得：

定理 30.6. *在定理30.5的条件下，对固定的 n，当 $N \to \infty$ 时*

$$\hat{\Gamma}(n) \to \Gamma(n),\ a.s., \quad \hat{R}(n) \to R(n),\ a.s.$$

关于 $\hat{R}(n)$ 的渐近分布有 (见 (Brockwell & Davis, 1987) 定理 11.2.2)：

定理 30.7. *设 $\{Z_{1t}\}$ 和 $\{Z_{2t}\}$ 都是一维独立同分布的零均值白噪声，彼此相互独立，$\{a_j\}$ 和 $\{b_j\}$ 是绝对可和的实数列，*

$$X_{1t} = \sum_{j=-\infty}^{\infty} a_j Z_{1,t-j},\ t \in \mathbb{N}_+,$$
$$X_{2t} = \sum_{j=-\infty}^{\infty} b_j Z_{2,t-j},\ t \in \mathbb{N}_+,$$

则

- *(1) 对 $k \geq 0$, 当 $N \to \infty$ 时,*

$$\sqrt{N}\hat{\rho}_{12}(k) \xrightarrow{d} N(0, \sigma_{11})$$

- *(2) 对 $h, k \geq 0$ 且 $h \neq k$,*

$$\sqrt{N}(\hat{\rho}_{12}(h), \hat{\rho}_{12}(k)) \xrightarrow{d} N(0, \Sigma)$$

  其中

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{pmatrix}$$
$$\sigma_{11} = \sum_{j=-\infty}^{\infty} \rho_{11}(j)\rho_{22}(j)$$
$$\sigma_{12} = \sum_{j=-\infty}^{\infty} \rho_{11}(j)\rho_{22}(j+k-h).$$

30.3 VAR 模型

定义 30.3. 设 $\{\varepsilon_t\}$ 是 m 维 WN(0,Q), $A_1, A_2, \ldots, A_p$ 是 $m \times m$ 实矩阵，使得

$$\det\left(I_m - \sum_{j=1}^{p} A_j z^j\right) \neq 0,\ |z| \leq 1 \tag{30.11}$$

称如下的模型

$$X_t = \sum_{j=1}^{p} A_j X_{t-j} + \varepsilon_t,\ t \in \mathbb{Z} \tag{30.12}$$

是一个 m 维 AR(p) 模型；如果平稳序列 $\{X_t\}$ 满足 m 维 AR(p) 模型，就称 $\{X_t\}$ 是一个 m 维的 AR(p) 序列。

记

$$A(z) = I_m - \sum_{j=1}^{p} A_j z^j$$

模型为

$$A(\mathscr{B}) X_t = \varepsilon_t,\ t \in \mathbb{Z} \tag{30.13}$$

用 $A(z)$ 的伴随矩阵表示 $A^{-1}(z)$ 后，可知 $A^{-1}(z)$ 的每个元素在 $|z| \leq 1$ 有泰勒展式，于是

$$A^{-1}(z) = \sum_{j=0}^{\infty} C_j z^j,\ |z| \leq 1. \tag{30.14}$$

其中 $C_n = (c_{jk}(n))$, $\{c_{jk}(n), n \in \mathbb{N}_+\}$ 以负指数速度趋于零。类似一维情况可得

$$X_t = A^{-1}(\mathscr{B})\varepsilon_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j},\ t \in \mathbb{Z} \tag{30.15}$$

是模型(30.12)的唯一平稳解。

例 30.5 (一维 AR 的马氏化)**.** 将一维 AR(p) 序列表示成 VAR。

一维 AR(p)

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t,\ t \in \mathbb{Z}, \tag{30.16}$$

$$\{\varepsilon_t\}\text{为 WN}(0,\sigma^2) \tag{30.17}$$

令

$$\begin{aligned}
\boldsymbol{X}_t =&(X_t, X_{t-1}, \dots, X_{t-p+1})^T \\
\boldsymbol{\varepsilon}_t =&(\varepsilon_t, 0, \dots, 0)^T \\
A =& \begin{pmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}
\end{aligned}$$

则得到 p 维 AR(1) 模型：

$$\boldsymbol{X}_t = A\boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t,\ t \in \mathbb{Z}. \tag{30.18}$$

$$\tag{30.19}$$

其中

$$\begin{aligned}
&\det(I_p - Az) \\
=&\det \begin{pmatrix} 1-a_1 z & -a_2 z & \cdots & -a_{p-1} z & -a_p z \\ -z & 1 & \cdots & 0 & 0 \\ 0 & -z & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -z & 1 \end{pmatrix} \\
=&1 - a_1 z - a_2 z^2 - \cdots - a_p z^p \neq 0,\ |z| \leq 1.
\end{aligned}$$

计算上面行列式时，按 $j = 2, 3, \dots, p$ 的顺序把第 j 列乘以 z^{j-1} 后加到第一列，则第一列只有 $(1,1)$ 元素非零。

30.3.1 多维 AR 序列的自协方差函数

由平稳解(30.15)可知

$$E[\boldsymbol{X}_t \boldsymbol{\varepsilon}_{t+n}^T] = 0, n \geq 1.$$

这体现出模型有因果性。于是对 $n \geq 0$

$$\begin{aligned}
\Gamma(n) =& E(X_{t+n}X_t^T) && (30.20)\\
=& E\left[\left(\sum_{j=1}^{p} A_j X_{t+n-j} + \varepsilon_{t+n}\right) X_t^T\right] && (30.21)\\
=& \sum_{j=1}^{p} A_j \Gamma(n-j) + E(\varepsilon_{t+n} X_t^T) && (30.22)
\end{aligned}$$

从而

$$A(\mathscr{B})\Gamma(n) = 0, \ n \geq 1. \tag{30.23}$$

另外

$$E(\varepsilon_t X_t^T) = E\left(\sum_{j=0}^{\infty} \varepsilon_t \varepsilon_{t-j}^T C_j^T\right) = QC_0^T = Q \tag{30.24}$$

总之有 m 维情况下自协方差函数矩阵的 Yule-Walker 方程

$$\begin{cases} \Gamma(0) = \sum_{j=1}^{p} A_j \Gamma(-j) + Q, \\ \Gamma(n) = \sum_{j=1}^{p} A_j \Gamma(n-j), \qquad n \geq 1 \end{cases} \tag{30.25}$$

把(30.25)的第二式写成分块矩阵形式：

$$\begin{pmatrix} \Gamma^T(1) \\ \Gamma^T(2) \\ \vdots \\ \Gamma^T(n) \end{pmatrix} = \begin{pmatrix} \Gamma(0) & \Gamma(1) & \cdots & \Gamma(p-1) \\ \Gamma(-1) & \Gamma(0) & \cdots & \Gamma(p-2) \\ \vdots & \vdots & & \vdots \\ \Gamma(-p+1) & \Gamma(-p+2) & \cdots & \Gamma(0) \end{pmatrix} \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_p^T \end{pmatrix} \tag{30.26}$$

系数矩阵是向量

$$(X_p^T, X_{p-1}^T, \dots, X_1^T)^T$$

的协方差矩阵，如果它正定则 $A_1, A_2, \dots, A_p$ 和 Q 可以由

$$\Gamma(0), \Gamma(1), \dots, \Gamma(p)$$

唯一决定，且满足最小相位条件(30.11)。

30.3.2 多维 AR 序列的参数估计

30.3.2.1 Y-W 方法

在 Y-W 方程(30.25)中，把 $\Gamma(n)$ 用样本自协方差函数 $\hat{\Gamma}(n)$ 代替，得到样本 Y-W 方程，可解得模型的 Y-W 估计

$$(\hat{A}_1, \hat{A}_2, \dots, \hat{A}_p, \hat{Q}).$$

解法有类似一维时 Levinson 递推那样的递推公式，见 (谢衷洁, 1990)。只要 Y-W 方程中的系数矩阵正定，则 Y-W 估计满足最小相位条件(30.11)。

30.3.2.2 最小二乘

把观测值 $X_1, X_2, \dots, X_N$ 满足的模型写成

$$X_t^T = \sum_{j=1}^{p} X_{t-j}^T A_j^T + \varepsilon_t^T,\ t = p+1, p+2, \dots, N \tag{30.27}$$

引入

$$X_n = \begin{pmatrix} X_p^T & X_{p-1}^T & \cdots & X_1^T \\ X_{p+1}^T & X_p^T & \cdots & X_2^T \\ \vdots & \vdots & & \vdots \\ X_{n-1}^T & X_{n-2}^T & \cdots & X_{n-p} \end{pmatrix}_{(n-p)\times mp}$$

$$Y_n = \begin{pmatrix} X_{p+1}^T \\ X_{p+2}^T \\ \vdots \\ X_n^T \end{pmatrix}_{(n-p)\times m} \qquad E_n = \begin{pmatrix} \varepsilon_{p+1}^T \\ \varepsilon_{p+1}^T \\ \vdots \\ \varepsilon_n^T \end{pmatrix}_{(n-p)\times m}$$

$$A = \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ X_p^T \end{pmatrix}_{mp\times m}$$

可以把(30.27)写成

$$Y_n = X_n A + E_n$$

求 $A_1, A_2, \dots, A_p$ 的最小二乘估计 $\hat{A}$，就是求 $\hat{A}$ 使得

$$S(A) = (Y_n - X_n A)^T (Y_n - X_n A)$$

最小。这里 $S(A)$ 是 $m\times m$ 矩阵。对两个对称矩阵 A 和 B，当且仅当 $A-B$ 非负定且不等于零矩阵时称 $A>B$。$\hat{A}$ 满足方程

$$(X_n^T X_n)\hat{A} = X_n^T Y_n$$

当 $(X_n^T X_n)$ 可逆时

$$\hat{A} = (X_n^T X_n)^{-1} X_n^T Y_n$$

事实上，若 $\hat{A}$ 满足下式

$$(X_n^T X_n)\hat{A} = X_n^T Y_n$$

则对任何与 $\hat{A}$ 同阶的 B，有

$$\begin{aligned}
&(Y_n - X_n B)^T (Y_n - X_n B)\\
=&(Y_n - X_n\hat{A} + X_n\hat{A} - X_n B)^T\\
&\cdot (Y_n - X_n\hat{A} + X_n\hat{A} - X_n B)\\
=&(Y_n - X_n\hat{A})^T (Y_n - X_n\hat{A}) + (X_n(\hat{A}-B))^T (X_n(\hat{A}-B))\\
\geq&(Y_n - X_n\hat{A})^T (Y_n - X_n\hat{A})
\end{aligned}$$

30.3.3 VAR 的预测

例30.5证明了一维 AR(p) 模型可以化为 p 维 AR(1) 模型。进一步地，任何一个 m 维 AR(p) 模型可以写成一个 mp 维的 AR(1) 模型。考虑 m 维 AR(p) 模型

$$X_t = \sum_{j=1}^{p} A_j X_{t-j} + \varepsilon_t,\ t\in\mathbb{Z} \tag{30.28}$$

记

$$Y_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{pmatrix}_{mp\times 1}, \quad \eta_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{mp\times 1}$$

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_m & 0 & \cdots & 0 & 0 \\ 0 & I_m & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I_m & 0 \end{pmatrix}_{mp\times mp}$$

则 m 维 AR(p) 模型(30.28)可以写成 mp 维 AR(1):

$$Y_t = AY_{t-1} + \eta_t,\ t \in \mathbb{Z}$$

其中，系数矩阵 A 满足

$$\begin{aligned}
&\det(I_p - Az) && (30.29)\\
=&\det\begin{pmatrix} I_p - A_1 z & -A_2 z & \cdots & -A_{p-1} z & -A_p z \\ -I_p z & I_p & \cdots & 0 & 0 \\ 0 & -I_p z & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -I_p z & I_p \end{pmatrix} && (30.30)\\
=&\det(I_p - A_1 z - A_2 z^2 - \cdots - A_p z^p) \neq 0,\ |z| \leq 1. && (30.31)
\end{aligned}$$

且 $\{\eta_t\}$ 为 mp 维零均值白噪声。

于是，只需要研究多维 AR(1) 的预报问题。对 m 维随机向量 $Y = (Y_1, Y_2, \dots, Y_m)^T$ 定义

$$L(Y|Z) = [L(Y_1|Z), L(Y_2|Z), \dots, L(Y_m|Z)]$$

其中 $L(Y_k|Z)$ 是 Z 对 Y_k 的最佳线性预测。设 $\{X_t\}$ 是 m 维 AR(1) 序列，满足

$$X_t = AX_{t-1} + \varepsilon_t,\ t \in \mathbb{Z}.$$

考虑用 $X_1, X_2, \dots, X_n$ 对 X_{n+k} 进行最佳线性预测。

利用平稳解的因果性，即 $E(X_t \varepsilon_{n+k}) = 0 (k \geq 1)$ 可得

$$\begin{aligned}
&L(X_{n+1}|X_1, X_2, \dots, X_n)\\
=&L(AX_n + \varepsilon_{n+1}|X_1, X_2, \dots, X_n)\\
=&L(AX_n|X_1, X_2, \dots, X_n)\\
=&A\,L(X_n|X_1, X_2, \dots, X_n)\\
=&AX_n
\end{aligned}$$

对 $k \ge 1$ 有

$$
\begin{aligned}
&L(X_{n+k}|X_1, X_2, \dots, X_n) \\
=&L(AX_{n+k-1} + \varepsilon_{n+k}|X_1, X_2, \dots, X_n) \\
=&L(AX_{n+k-1}|X_1, X_2, \dots, X_n) \\
=&A\,L(X_{n+k-1}|X_1, X_2, \dots, X_n) \\
=&\cdots = A^k X_n
\end{aligned}
$$

总之有

$$
\begin{aligned}
&L(X_{n+k}|X_1, X_2, \dots, X_n) \\
=&L(X_{n+k}|X_n) = A^k X_n,\ n, k \in \mathbb{N}_+
\end{aligned}
$$

30.4 多维平稳序列的谱分析

30.4.1 多维平稳序列的谱函数

设 $\{X_t = (X_{1t}, X_{2t})^T:\ t \in \mathbb{Z}\}$ 是一个 2 维零均值平稳序列。对给定复数 z 定义

$$Y_t = X_{1t} + zX_{2t},\ t \in \mathbb{Z}$$

易见 $EY_t = 0$,

$$
\begin{aligned}
\gamma_z(k) \overset{\triangle}{=}& E(Y_{t+k}\bar{Y}_t) = E[(X_{1,t+k} + zX_{2,t+k})(X_{1t} + \bar{z}X_{2t})] \\
=&\gamma_{11}(k) + z\gamma_{21}(k) + \bar{z}\gamma_{12}(k) + |z|^2\gamma_{22}(k)
\end{aligned}
$$

都不依赖于 t，所以 $\{Y_t\}$ 是一个复值平稳序列。设 $\{Y_t\}$ 有谱函数 F_z, 则对 $n \in \mathbb{Z}$,

$$\gamma_z(n) = \int_{-\pi}^{\pi} e^{in\lambda} dF_z(\lambda),\ F_z(-\pi) = 0.$$

当 $z = \pm 1, \pm i$ 时得

$$
\begin{aligned}
\gamma_1(n) =& \gamma_{11}(n) + \gamma_{21}(n) + \gamma_{12}(n) + \gamma_{22}(n) \\
\gamma_{-1}(n) =& \gamma_{11}(n) - \gamma_{21}(n) - \gamma_{12}(n) + \gamma_{22}(n) \\
\gamma_i(n) =& \gamma_{11}(n) + i\gamma_{21}(n) - i\gamma_{12}(n) + \gamma_{22}(n) \\
\gamma_{-i}(n) =& \gamma_{11}(n) - i\gamma_{21}(n) + i\gamma_{12}(n) + \gamma_{22}(n)
\end{aligned}
$$

可以解出

$$\gamma_{12}(n) = \frac{1}{4}[\gamma_1(n) - \gamma_{-1}(n) + i\gamma_i(n) - i\gamma_{-i}(n)]$$
$$\gamma_{21}(n) = \frac{1}{4}[\gamma_1(n) - \gamma_{-1}(n) - i\gamma_i(n) + i\gamma_{-i}(n)]$$

用 $F_{11}(\lambda)$ 和 $F_{22}(\lambda)$ 分别表示 $\{X_{1t}\}$ 和 $\{X_{2t}\}$ 的谱函数，并引入

$$F_{12}(\lambda) = \frac{1}{4}[F_1(\lambda) - F_{-1}(\lambda) + iF_i(\lambda) - iF_{-i}(\lambda)]$$
$$F_{21}(\lambda) = \frac{1}{4}[F_1(\lambda) - F_{-1}(\lambda) - iF_i(\lambda) + iF_{-i}(\lambda)]$$

引入矩阵函数

$$F(\lambda) = \begin{pmatrix} F_{11}(\lambda) & F_{12}(\lambda) \\ F_{21}(\lambda) & F_{22}(\lambda) \end{pmatrix}, \ \lambda \in [-\pi, \pi] \tag{30.32}$$

则有

$$\Gamma(n) = E(X_{t+n}X_t^T) = \int_{-\pi}^{\pi} e^{in\lambda} dF(\lambda) \tag{30.33}$$
$$= \begin{pmatrix} \int_{-\pi}^{\pi} e^{in\lambda} dF_{11}(\lambda) & \int_{-\pi}^{\pi} e^{in\lambda} dF_{12}(\lambda) \\ \int_{-\pi}^{\pi} e^{in\lambda} dF_{21}(\lambda) & \int_{-\pi}^{\pi} e^{in\lambda} dF_{22}(\lambda) \end{pmatrix}, \ n \in \mathbb{Z} \tag{30.34}$$

称 $F(\lambda)$ 为 2 维平稳序列 $\{X_t\}$ 的**谱函数矩阵**。由于矩阵 $F(\lambda)$ 的每个元素都是 $[-\pi, \pi]$ 上分布函数的线性组合，所以都是有界变差右连续函数。另外 $F(\lambda)$ 还是 Hermite 矩阵：

$$F^*(\lambda) = F(\lambda)$$

其中星号表示共轭转置。

还可证明 $F(\lambda)$ 关于 λ 单调不减，即 $\forall \lambda_1 < \lambda_2, \lambda_1, \lambda_2 \in [-\pi, \pi]$, $F(\lambda_2) - F(\lambda_1)$ 非负定。

当 $F(\lambda)$ 的每个元素的实部和虚部都是连续函数，并且除去有限个点外导函数连续时，称

$$f(\lambda) = \begin{pmatrix} f_{11}(\lambda) & f_{12}(\lambda) \\ f_{21}(\lambda) & f_{22}(\lambda) \end{pmatrix} \tag{30.35}$$
$$= \begin{pmatrix} F'_{11}(\lambda) & F'_{12}(\lambda) \\ F'_{21}(\lambda) & F'_{22}(\lambda) \end{pmatrix}, \ \lambda \in [-\pi, \pi] \tag{30.36}$$

为 $\{X_t\}$ 的**谱密度矩阵**, 称 $f_{12}(\lambda)$ 是 $\{X_{1t}\}$ 和 $\{X_{2t}\}$ 的**互谱密度**。

因为 $F(\lambda)$ 是 Hermite 矩阵，并且单调不减，所以 $f(\lambda)$ 是 Hermite 非负定的。这时

$$\Gamma(n) = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda \tag{30.37}$$

$$\triangleq \begin{pmatrix} \int_{-\pi}^{\pi} e^{in\lambda} f_{11}(\lambda) d\lambda & \int_{-\pi}^{\pi} e^{in\lambda} f_{12}(\lambda) d\lambda \\ \int_{-\pi}^{\pi} e^{in\lambda} f_{21}(\lambda) d\lambda & \int_{-\pi}^{\pi} e^{in\lambda} f_{22}(\lambda) d\lambda \end{pmatrix}, \ n \in \mathbb{Z} \tag{30.38}$$

定理 30.8. 设 m 维平稳序列 $\{X_t\}$ 有自协方差函数 $\{\Gamma(n)\}$，则存在唯一的 $m \times m$ 函数矩阵

$$F(\lambda) = (F_{jk}(\lambda)), \ \lambda \in [-\pi, \pi],$$

使得

- *(1)* $\Gamma(n) = \int_{-\pi}^{\pi} e^{in\lambda} dF(\lambda)$, $n \in \mathbb{Z}$;
- *(2)* $F^*(\lambda) = F(\lambda)$, $F(-\pi) = 0$, 当 $\lambda_1 < \lambda_2$ 时 $F(\lambda_2) - F(\lambda_1)$ 非负定;
- *(3)* $F(\lambda)$ 中的每个元素 $F_{jk}(\lambda)$ 是有界变差和右连续的。

$F(\lambda)$ 叫做 $\{X_t\}$ 的**谱函数矩阵**。

如果 $F(\lambda)$ 的每个元素 $F_{jk}(\lambda)$ 的实部和虚部都是连续函数，且除去有限个点外导函数连续，就称

$$f(\lambda) = (f_{jk}(\lambda)) \triangleq (F'_{jk}(\lambda))$$

为 $\{X_t\}$ 的**谱密度矩阵**。这时

$$\Gamma(n) = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda \tag{30.39}$$

$$= \left(\int_{-\pi}^{\pi} e^{in\lambda} f_{jk}(\lambda) d\lambda \right)_{m \times m}, \ n \in \mathbb{Z} \tag{30.40}$$

称实变复值函数是绝对连续函数，如果其实部和虚部都是绝对连续函数。当 $F(\lambda)$ 的每个元素都是绝对连续函数时，$f(\lambda) = (F'_{jk}(\lambda))$ 就是 $\{X_t\}$ 的**谱密度矩阵**。

定理 30.9. 设 m 维平稳序列 $\{X_t\}$ 有自协方差函数 $\Gamma(n) = (\gamma_{jk}(n))$, 如果

$$\sum_{n=-\infty}^{\infty} |\gamma_{jk}(n)| < \infty, \ j, k = 1, 2, \ldots, m$$

则

$$f(\lambda) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\Gamma(n)e^{-in\lambda},\ \lambda\in[-\pi,\pi]$$

是 $\{X_t\}$ 的**谱密度矩阵**。

30.4.2 多维平稳序列的谱表示

设 $\{X_t = (X_{1t}, X_{2t}, \dots, X_{mt})^T$ 是 m 维零均值平稳序列，则其每个分量 $\{X_j(t)\}$ 是一位零均值平稳序列，有谱表示

$$X_{jt} = \int_{-\pi}^{\pi} e^{it\lambda} dZ_j(\lambda),\ t\in\mathbb{Z} \tag{30.41}$$

中 $\{Z_j(\lambda)\}$ 是 $[-\pi,\pi]$ 上右连续的正交增量过程。

记

$$Z(\lambda) = (Z_1(\lambda), Z_2(\lambda), \dots, Z_m(\lambda))^T,\ \lambda\in[-\pi,\pi], \tag{30.42}$$

有

$$X_t = \int_{-\pi}^{\pi} e^{in\lambda} dZ(\lambda)\ \lambda\in[-\pi,\pi]. \tag{30.43}$$

这称为 m 维平稳序列 $\{X_t\}$ 的**谱表示**。

(30.43)中的 $Z(\lambda)$ 有下列性质：

- (1) 正交增量性：对 $-\pi\le\lambda_1<\lambda_2\le\lambda_3<\lambda_4\le\pi$,

$$E\left[(Z(\lambda_2)-Z(\lambda_1))(Z(\lambda_4)-Z(\lambda_3))^*\right]=0.$$

- (2) 右连续性：当 $\delta\downarrow 0$ 时

$$E\left[(Z(\lambda+\delta)-Z(\lambda))(Z(\lambda+\delta)-Z(\lambda))^*\right]\to 0.$$

- (3) $F(\lambda)=E[Z(\lambda)Z^*(\lambda)]$ 是 $\{X_t\}$ 的谱函数矩阵，满足

$$F(\lambda_2)-F(\lambda_1) = E\left[(Z(\lambda_2)-Z(\lambda_1))(Z(\lambda_2)-Z(\lambda_1))^*\right],\ \lambda_1<\lambda_2$$

满足上述 (1), (2), (3) 的 m 维随机过程 $\{Z(\lambda)\}$ 叫做**右连续的正交增量过程**。

定理 30.10. 对 m 维零均值平稳序列 $\{X_t\}$，有右连续的正交增量过程 $\{Z(\lambda)\}$ 使得 $Z(-\pi)=0$ 和(30.43)成立。如果正交增量过程 $\{\xi(\lambda)\}$ 也满足上述的条件，则

$$P(\xi(\lambda)=Z(\lambda))=1,\ \lambda\in[-\pi,\pi].$$

例 30.6 (多维线性平稳序列的谱密度)**.** 设 $\{\varepsilon_t\}$ 是 m 维 WN(0, Q), $Q=E[\varepsilon_t\varepsilon_t^T]$, $\{A_j\}$ 是一列 $m\times m$ 实值矩阵，满足

$$\sum_{j=-\infty}^{\infty}A_jQA_j^T<\infty,$$

则

$$X_t=\sum_{j=-\infty}^{\infty}A_j\varepsilon_{t-j},\ t\in\mathbb{Z}$$

是 m 维零均值平稳序列。

自协方差函数为

$$\Gamma(n)=E[X_{t+n}X_n]=\sum_{j=-\infty}^{\infty}A_{j+n}QA_j^T,\ n\in\mathbb{Z}$$

本例中，对两个实矩阵 A,B，用 $A\le B$ 表示 A 和 B 的对应元素比较全部成立小于等于关系，用 $[A]$ 表示把 A 的所有元素都取绝对值组成的矩阵。这时必有

$$[AB]\le[A][B]$$

如果要求 $\sum_{j=-\infty}^{\infty}[A_j]<\infty$, 则

$$\begin{aligned}\sum_{n=-\infty}^{\infty}[\Gamma(n)]&=\sum_{n=-\infty}^{\infty}\left[\sum_{j=-\infty}^{\infty}A_{j+n}QA_j^T\right]\\&\le\sum_{n=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}[A_{j+n}][Q][A_j^T]\\&\le\left(\sum_{n=-\infty}^{\infty}[A_n]\right)[Q]\left(\sum_{j=-\infty}^{\infty}[A_j^T]\right)<\infty\end{aligned}$$

从而 $\{X_t\}$ 有谱密度

$$\begin{aligned}
f(\lambda) =& \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\Gamma(n)e^{-in\lambda}\\
=& \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}A_{j+n}QA_j^T e^{-in\lambda}\\
=& \frac{1}{2\pi}\left(\sum_{k=-\infty}^{\infty}A_k e^{-ik\lambda}\right)Q\left(\sum_{j=-\infty}^{\infty}A_j e^{-ij\lambda}\right)^*
\end{aligned}$$

例 30.7 (VARMA 序列谱密度)**.** 考虑平稳可逆的 m 维 ARMA(p,q) 模型的谱密度。模型为

$$A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t,\ t\in\mathbb{Z},$$

$\det(A(z))A^{-1}(z)$ 是 $A(z)$ 的伴随矩阵，仍是矩阵系数多项式。于是 $A^{-1}(z)B(z)$ 的每个元素是有理多项式。将 $A^{-1}(z)B(z)$ 的每个元素进行泰勒展开后得 $A^{-1}(z)B(z)$ 泰勒展开式

$$A^{-1}(z)B(z) = \sum_{j=0}^{\infty}C_j z^j,\ |z|\le 1.$$

其中的系数矩阵满足

$$\sum_{j=0}^{\infty}[C_j] < \infty.$$

由例30.6，平稳解

$$X_t = A^{-1}(\mathscr{B})B(\mathscr{B})\varepsilon_t = \sum_{j=0}^{\infty}C_j\varepsilon_{t-j},\ t\in\mathbb{Z}$$

有谱密度矩阵

$$f(\lambda) =\frac{1}{2\pi}\left(\sum_{k=-\infty}^{\infty}C_k e^{-ik\lambda}\right)Q\left(\sum_{j=-\infty}^{\infty}C_j e^{-ij\lambda}\right)^*$$

例 30.8 (二次相干函数)**.** 设二维零均值平稳序列 $\{X_t\}$ 有谱表示

$$X_t = \int_{-\pi}^{\pi}e^{it\lambda}d\begin{pmatrix} Z_1(\lambda)\\ Z_2(\lambda)\end{pmatrix}$$

取 $\lambda \in [-\pi, \pi]$, 对充分小的正数 $\Delta\lambda$，定义

$$\Delta Z_k(\lambda) = Z_k(\lambda + \Delta\lambda) - Z_k(\lambda),\ k = 1, 2$$

则有

$$\begin{aligned}
&F(\lambda + \Delta\lambda) - F(\lambda) \\
=&E\left[\begin{pmatrix} \Delta Z_1(\lambda) \\ \Delta Z_2(\lambda) \end{pmatrix} (\Delta\bar{Z}_1(\lambda), \Delta\bar{Z}_2(\lambda))\right] \\
=&\begin{pmatrix} E|\Delta Z_1(\lambda)|^2 & E[\Delta Z_1(\lambda)\Delta\bar{Z}_2(\lambda)] \\ E[\Delta\bar{Z}_1(\lambda)\Delta Z_2(\lambda)] & E|\Delta Z_2(\lambda)|^2 \end{pmatrix}
\end{aligned}$$

如果 $F'(\lambda)$ 连续，则有

$$dF(\lambda) = f(\lambda)d\lambda = \begin{pmatrix} E|dZ_1(\lambda)|^2 & E[dZ_1(\lambda)d\bar{Z}_2(\lambda)] \\ E[d\bar{Z}_1(\lambda)dZ_2(\lambda)] & E|dZ_2(\lambda)|^2 \end{pmatrix}$$

$dZ_1(\lambda)$ 和 $dZ_2(\lambda)$ 的相关系数为

$$\begin{aligned}
\rho_{12}(\lambda) =&\frac{E[dZ_1(\lambda)d\bar{Z}_2(\lambda)]}{\sqrt{E|dZ_1(\lambda)|^2 E|dZ_2(\lambda)|^2}} \\
=&\frac{f_{12}(\lambda)d\lambda}{\sqrt{f_{11}(\lambda)d\lambda \cdot f_{22}(\lambda)d\lambda}} \\
=&\frac{f_{12}(\lambda)}{\sqrt{f_{11}(\lambda)f_{22}(\lambda)}}
\end{aligned}$$

其中规定 $0/0 = 0$。

通常称

$$K_{12}^2(\lambda) = |\rho_{12}(\lambda)|^2 = \frac{|f_{12}(\lambda)|^2}{f_{11}(\lambda)f_{22}(\lambda)}$$

为 $\{X_t, t \in \mathbb{Z}\}$ 的**二次相干函数**。$K_{12}^2(\lambda)$ 描述了 $\{X_{1t}\}$ 和 $\{X_{2t}\}$ 在角频率 λ 处的线性相关性强弱。如果 $K_{12}^2(\lambda) = 1$, 说明 $\{X_{1t}\}$ 和 $\{X_{2t}\}$ 在角频率 λ 处线性相关。

例 30.9 (线性滤波的二次相干函数). 设平稳序列 $\{X_{1t}\}$ 有谱密度 $f(\lambda)$, $\{h_j\}$ 是绝对可和线性滤波器，输出过程为

$$X_{2t} = \sum_{j=-\infty}^{\infty} h_j X_{1,t-j},\ t \in \mathbb{Z}$$

则 $\{X_{2t}\}$ 有谱密度（见定理6.5）

$$f_{22}(\lambda) = |h(\lambda)|^2 f(\lambda)$$

其中

$$h(\lambda) = \sum_{j=-\infty}^{\infty} h_j e^{-ij\lambda}.$$

计算得

$$\begin{aligned}
\gamma_{12}(n) =& E(X_{1,t+n}X_{2t}) \\
=& \sum_{j=-\infty}^{\infty} h_j E(X_{1,t+n}X_{1,t-j}) \\
=& \sum_{j=-\infty}^{\infty} h_j \gamma_{11}(n+j) \\
=& \sum_{j=-\infty}^{\infty} h_j \int_{-\pi}^{\pi} e^{i(n+j)\lambda} f(\lambda) d\lambda \\
=& \int_{-\pi}^{\pi} e^{in\lambda} h(-\lambda) f(\lambda) d\lambda
\end{aligned}$$

所以 $\{X_{1t}\}$ 和 $\{X_{2t}\}$ 的互谱密度为

$$f_{12}(\lambda) = h(-\lambda) f(\lambda),\ \lambda \in [-\pi, \pi].$$

于是 2 维平稳序列 $X_t = (X_{1t}, X_{2t})^T$ 有退化的谱密度矩阵

$$f(\lambda) = \begin{pmatrix} f(\lambda) & h(-\lambda)f(\lambda) \\ h(\lambda)f(\lambda) & |h(\lambda)|^2 f(\lambda) \end{pmatrix}$$

这时二次相干函数

$$K^2(\lambda) \equiv 1,\ \forall \lambda \in [-\pi, \pi].$$

30.4.3 谱密度矩阵的估计

二维零均值平稳序列 $\{X_t\}$ 的观测值 $X_1, X_2, \dots, X_N$ 的周期图定义为

$$I_N(\lambda) = \frac{1}{2\pi N} \left(\sum_{j=1}^{N} X_j e^{-ij\lambda} \right) \left(\sum_{j=1}^{N} X_j e^{-ij\lambda} \right)^* .$$

周期图 $I_n(\lambda)$ 的值是 2×2 方阵，可以对周期图平滑得到谱密度矩阵估计。

设 $\lambda_j = 2j\pi/N$, 对 $\lambda \in [0,\pi]$, 用 $g(N,\lambda)$ 表示 $\{\lambda_j : 0 \le j \le N/2\}$ 中距离 λ 最近的 λ_j（若左右两个距离相等取左边一个）。取 $M_N = O(\sqrt{N})$ 且 $\lim_{N\to\infty} M_N = \infty$。设 $W_N(k)$ 为满足下列条件的实值权函数：

$$
\begin{aligned}
&(1) W_N(k) = W_N(-k),\ W_N(k) \ge 0,\ |k| \le M_N;\\
&(2) \sum_{|k|\le M_N} W_N(k) = 1;\\
&(3) \lim_{N\to\infty} \sum_{|k|\le M_N} W_N^2(k) = 0.
\end{aligned}
$$

谱密度矩阵的平滑周期图估计为

$$
\begin{aligned}
\hat f(\lambda) =& \sum_{|k|\le M_N} W_N(k) I_N(g(N,\lambda)+\lambda_k),\ \lambda\in[0,\pi],\\
\hat f(\lambda) =& [\hat f(-\lambda)]^T,\ \lambda \in [-\pi, 0].
\end{aligned}
$$

设

$$
\hat f(\lambda) = \begin{pmatrix} \hat f_{11} & \hat f_{12} \\ \hat f_{21} & \hat f_{22} \end{pmatrix}
$$

二次相干函数 $K_{12}^2(\lambda)$ 的估计定义为

$$
\hat K_{12}^2(\lambda) = \frac{|\hat f_{12}(\lambda)|^2}{\hat f_{11}(\lambda)\hat f_{22}(\lambda)}.
$$

30.5 附录：补充

30.5.1 多维 ARMA 不可辨识的例子

只要 $\det(A(z))$ 为不依赖于 z 的非零常数，则多维 ARMA 模型参数不能从 $\{\Gamma(n)\}$ 唯一确定。

例如，考虑 2 维 $VAR(1)$。模型为

$$
X_t = AX_{t-1} + \varepsilon_t
$$

其中

$$
A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}
$$

特征多项式的行列式为

$$\begin{aligned}\det(I-Az)=&\begin{vmatrix}1-a_{11}z & -a_{12}z\\ -a_{21}z & 1-a_{22}z\end{vmatrix}\\ =&1-(a_{11}+a_{22})z+(a_{11}a_{22}-a_{12}a_{21})z^2\end{aligned}$$

令 z 和 z^2 系数等于零，则

$$a_{22}=-a_{11},\ a_{12}a_{21}=-a_{11}^2$$

取 $a_{11}\neq 0$, $a_{22}=-a_{11}$, 取 $a_{12}\neq 0$，取 $a_{21}=-\frac{a_{11}^2}{a_{12}}$ 则有 $\det(I-Az)=1$ 不依赖于 z。这时

$$\begin{aligned}(I-Az)^{-1}=&\begin{pmatrix}1-a_{22}z & a_{21}z\\ a_{12}z & 1-a_{11}z\end{pmatrix}\\ =&I+\begin{pmatrix}-a_{22} & a_{21}\\ a_{12} & -a_{11}\end{pmatrix}z\end{aligned}$$

令

$$B=\begin{pmatrix}-a_{22} & a_{21}\\ a_{12} & -a_{11}\end{pmatrix}$$

则模型可以写成

$$X_t=(I-A\mathscr{B})^{-1}\varepsilon_t=(I+B\mathscr{B})\varepsilon_t$$

又可以写成一个 2 维 MA(1) 模型。

Part VIII

附录

附录 A

时间序列与 R

A.1 本课程的软件需求

金融时间序列分析课需要学生掌握一定的运用统计软件进行建模分析的能力。作为基础，学生需要重点掌握 R 软件如下功能：

- 基本数据类型，日期和日期时间类型，数据输入输出
- Rmd 格式文件运用
- 时间序列类型 zoo, xts, timeSeries 类型
- quantmod 包

关于 R 软件基本使用，详见李东风的 R 软件教程，第 1、2 章，第 9 章。

关于 Rmd 格式文件, 详见李东风的 R 软件教程，第 19–21 章。

关于 R 软件与金融时间序列分析的较详细的介绍，参见李东风的《金融时间序列分析》研究生课程讲义。

参考:

- CRAN 时间序列任务视点
- CRAN 金融任务视点
- CRAN 计量经济任务视点

A.2 基本 R 使用

R 是一个数据分析、统计建模和作图的软件，其中包含一门计算机语言称为 R 语言，此语言与通常的 C、C++、Java 等编程语言相比，支持更多的数据类型，如向量、矩阵，并提供了多种统计和数学计算方法。

为了在金融时间序列计算中使用 R 语言，这里简单介绍一些最基本的使用方法。

R 软件是一个开源软件，可以免费地从其网站http://www.r-project.org提供的镜像网站下载安装。另外，RStudio 是一个 R 软件的集成开发环境，在该软件中可以更方便地使用 R 软件，虽然 RStudio 是商业软件，但非商业用户可以免费地使用。

R 可以在命令行运行，每次运行一条命令，结果显示在命令下方。也可以将多条 R 命令写成一个源程序文件，然后运行整个文件。在 RStudio 软件中也可以选中若干行程序，运行选中的部分。

A.2.1 四则运算

R 中用 `+ - * / ^` 分别表示加、减、乘、除、乘方。数值可以写成如 `123`, `-123`, `123.45`, `1.23E-5` 这样的形式，其中 `1.23E-5` 表示 1.23×10^{-5}。

A.2.2 字符串

R 中可以用双撇号`"` 或者单撇号`'`将文字内容包裹起来，称为字符串，如`"two.sided"`，`" 收益率"`。

A.2.3 向量

R 中最基本的单位是向量，单个数值是长度为 1 的向量。简单的向量可以用 `c(...)` 定义，如

```
x <- c(1.1, 1.02, 1.4, 0.9)
x
## [1] 1.10 1.02 1.40 0.90
```

长度相同的两个向量之间可以做四则运算，结果是对应的元素进行四则运算。单

个元素可以和一个向量进行四则运算，结果是该元素与向量的每一个元素进行四则运算。如

```
x + 100
## [1] 101.10 101.02 101.40 100.90
```

A.2.4 矩阵

R 支持矩阵类型，矩阵是有 n 行、p 列的数值排列成的存储结果，在第 i 行、第 j 列有一个数值元素。矩阵 `A` 的第 (i, j) 元素表示为 `A[i,j]`，矩阵的第 j 列作为一个向量，可以表示为 `A[,j,drop=TRUE]`。`nrow(A)` 求 `A` 的行数，`ncol(A)` 求 `A` 的列数。

两个相同形状的矩阵可以进行加、减、乘、除四则运算，结果是对应的元素进行四则运算。单个元素可以与矩阵进行四则运算，结果是该元素与矩阵的每个元素进行四则运算。

两个向量 `x1` 和 `x2`，可以用 `cbind(x1, x2)` 合并为两列的矩阵。

A.2.5 数据框

R 数据框是和矩阵类似的存储结构，但是允许不同列的数据类型不同，比如，一列是日期，一列是评级（字符串），一列是收益率（数值）。

数据框的每一列都有列名，用 `names(d)` 或者 `colnames(d)` 访问。将数据框 `d` 的某一列取出作为一个向量，用 `d[["列名"]]` 的格式。

A.2.6 扩展包

R 的许多功能由不同的扩展包（package）提供。为了调用某个扩展包的功能，需要先用 `library()` 命令载入该扩展包，如

```
library(xts)
```

某些扩展包还允许用包名::函数名() 的格式直接调用其中的函数。

A.2.7 日期和日期时间

R 日期可以保存为 Date 类型，一般用整数保存，数值为从 1970-1-1 经过的天数。

R 中用一种叫做 POSIXct 和 POSIXlt 的特殊数据类型保存日期和时间，可以仅包含日期部分，也可以同时有日期和时间。技术上，POSIXct 把日期时间保存为从 1970 年 1 月 1 日零时到该日期时间的时间间隔秒数。日期时间会涉及到所在时区、夏时制等问题，比较复杂。

lubridate 是一个提供了许多日期和日期时间功能的扩展包。对于`"2018-07-02"`这样的字符型的日期，可以用 `lubridate::ymd("2018-07-02")` 这样的方法转换成 Date 类型，该函数也可以输入字符型向量，向量的每个元素是一个日期字符串，结果为 Date 类型元素组成的向量。

如果 `year`, `mon`, `day` 分别是数值类型的年、月、日，可以用 `lubridate::make_date(year, mon, day)` 转换成 Date 类型。

`lubridate::ymd_hms()` 可以将`"2018-07-02 13:15:59"` 这样的日期和时间字符串转换成 POSIXct 类型，`lubridate::make_datetime()` 可以将单独的整数表示的年、月、日、时、分、秒转换成 POSIXct 类型。

`lubridate` 包的 `year()`, `month()`, `mday()` 可以从日期或日期时间中取出年、月、日，`hour()`、`minute()`、`second()` 可以从日期时间中取出时、分、秒。

更详细的用法见R 软件教程。

A.2.8 读入时间序列数据

常见的数据形式是用文本格式保存，数据之间用空行、空格或者逗号分隔。

A.2.8.1 单个向量读入

最简单的情况是仅有一个序列，数据仅包含序列的数值而不包含时间标签，这时数据可以保存为用空行和空格分隔的文本格式数据，例如文件 `sp500-2691.txt` 是标准普尔 500 指数月超额收益率从 1926 年 1 月到 1991 年 12 月的数据，内容为（节选）：

```
0.0225 -0.044 -0.0591 0.0227 0.0077 0.0432 0.0455 0.0171 0.0229 -0.0313 0.0223 0.0
-0.0208 0.0477 0.0065 0.0172 0.0522 -0.0094 0.065 0.0445 0.0432 -0.0531 0.0678 0.0
```

```
-0.0051 -0.0176 0.1083 0.0324 0.0127 -0.0405 0.0125 0.0741 0.024 0.0145 0.1199 0.0029
0.0571 -0.0058 -0.0023 0.0161 -0.0428 0.1124 0.0456 0.098 -0.0489 -0.1993 -0.1337 0.0253
```

可以用如下程序读入成 R 向量：

```
x <- scan("data/sp500-2691.txt")
```

A.2.8.2 带有时间标签的序列

有些序列带有时间标签，同一行的数据之间用空格或者制表符分隔。文件 `d-ibm-0110.txt` 为 IBM 股票从 2001-1-2 到 2010-12-31 的日简单收益率，部分数据如下：

```
date           return
20010102      -0.002206
20010103       0.115696
20010104      -0.015192
20010105       0.008719
20010108      -0.004654
20010109      -0.010688
20010110       0.009453
```

可以用如下程序读入成一个 R 数据框：

```
d <- read.table(
  "data/d-ibm-0110.txt", header=TRUE,
  colClasses=c("character", "numeric"))
```

其中的日期可以用如下程序转换成 R 的 Date 类型：

```
d[["date"]] <- lubridate::ymd(d[["date"]])
```

有些序列的年月日是分开输入的，比如，文件 `m-unrate.txt` 为美国从 1948 年 1 月到 2010 年 9 月的经季节调整的月失业率百分数，部分数据如：

```
Year mon dd  rate
1948 01 01   3.4
1948 02 01   3.8
1948 03 01   4.0
```

```
1948 04 01   3.9
1948 05 01   3.5
```

可以用如下程序读入为 R 数据框，并生成日期列：

```r
d <- read.table(
  "data/m-unrate.txt", header=TRUE,
  colClasses=rep("numeric", 4))
d[["date"]] <- lubridate::make_date(d[["Year"]], d[["mon"]], d[["dd"]])
```

同一个有多个序列时，也可以用如上方法读入为 R 数据框。

A.2.8.3 读入 csv 格式

CSV 文件也是一种文本格式的数据文件，相对而言更规范一些，所以经常用于在不同的数据管理和数据分析软件之间传递数据。数据如：

```
"date","ibm","sp"
"19260130",-0.010381,0.022472
"19260227",-0.024476,-0.043956
"19260331",-0.115591,-0.059113
"19260430",0.089783,0.022688
"19260528",0.036932,0.007679
"19260630",0.068493,0.043184
"19260731",0,0.045455
```

可以用如下程序读入为 R 数据框：

```r
library(readr)
d <- read_csv(
  "m-ibmsp-2611.csv",
  col_types=cols(
    date=col_date(format="%Y%m%d"),
    .default=col_double()
  )
)
```

A.3 生成时间序列数据

R 软件中有些时间序列分析函数需要特定的时间序列类型数据作为输入。常用的有 ts 类型、xts 类型等。

ts 类型用于保存一元或者多元的等间隔时间序列，如月度、季度、年度数据。生成方法如

```
ts(x, start=c(2001, 1), frequency=12)
```

其中 x 是向量或者矩阵，取矩阵值时矩阵的每一列是一个序列。frequency 对月度数据是 12，对季度数据是 4，对年度数据可以缺省（值为 1）。start=c(2001,1) 表示序列开始时间是 2001 年 1 月。如果是年度数据，用如 ts(x, start=2001) 即可。

例如，文件 m-unrate.txt 为美国从 1948 年 1 月到 2010 年 9 月的经季节调整的月失业率百分数，转换成 ts 类型的程序如下：

```
d <- read.table(
  "data/m-unrate.txt", header=TRUE,
  colClasses=rep("numeric", 4))
unrate <- ts(d[["rate"]], start=c(1948,1), frequency=12)
```

xts 也是一种时间序列数据类型，既可以保存等间隔时间序列数据，也可以保存不等间隔的时间序列数据，并且 xts 类型的数据访问功能更为方便。读入方法例如

```
xts(x, date)
```

其中 x 是向量、矩阵或数据框，date 是日期或者日期时间。x 取矩阵或者数据框时每列是一个时间序列。

例如，文件 d-ibm-0110.txt 为 IBM 股票从 2001-1-2 到 2010-12-31 的日简单收益率，读入数据并转换为 xts 类型的程序如下：

```
library(xts)
d <- read.table(
  "data/d-ibm-0110.txt", header=TRUE,
  colClasses=c("character", "numeric"))
```

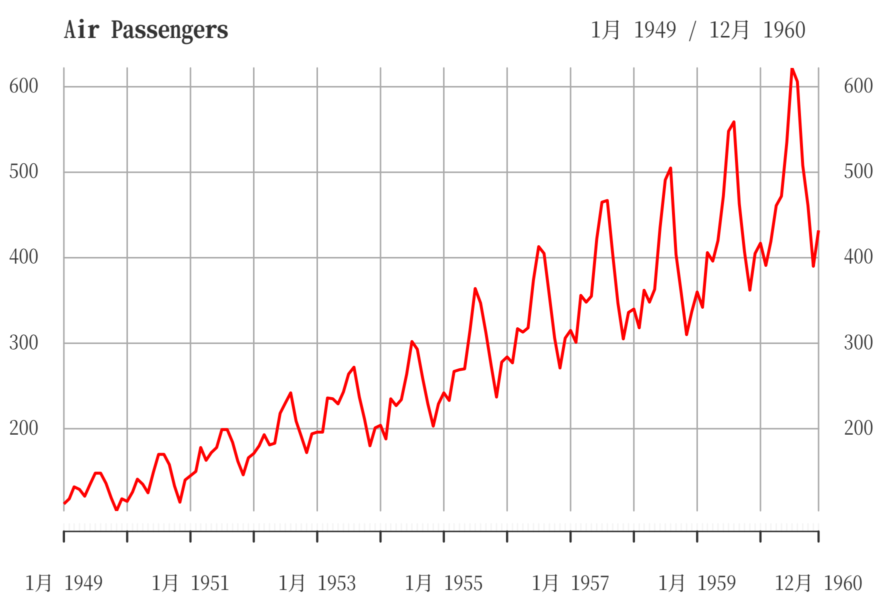
```
ibmrtn <- xts(d[["return"]], lubridate::ymd(d[["date"]]))
```

有了 xts 类型的变量 x 后，可以用 coredate(x) 返回 x 的不包含时间的纯数据；用 index(x) 返回 x 的时间标签。

A.4 时间序列图形

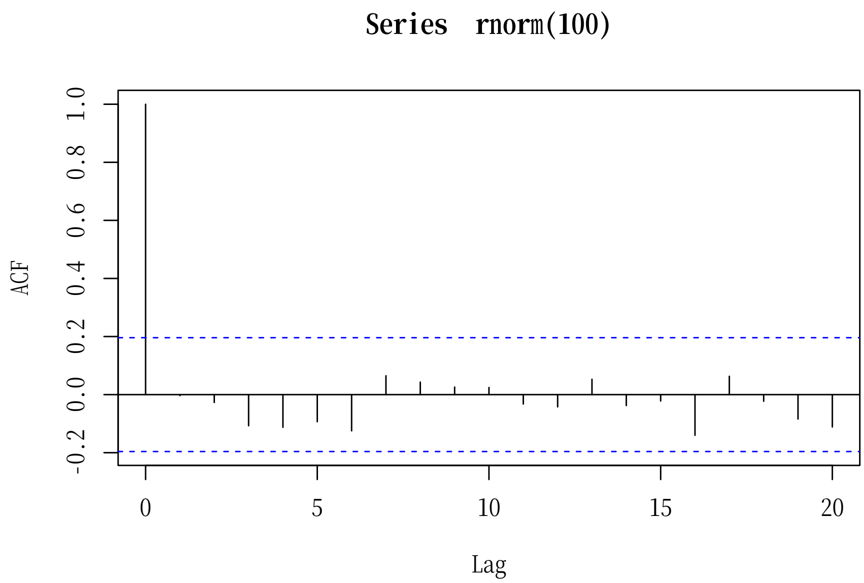
对 ts 类型的时间序列 x，用 plot(x) 作时间序列图。对 xts 类型的时间序列，作图如：

```
library(xts)
xts.ap <- as.xts(AirPassengers)
plot(xts.ap, main="Air Passengers",
     major.ticks="years", minor.ticks=NULL,
     grid.ticks.on="years",
     col="red")
```



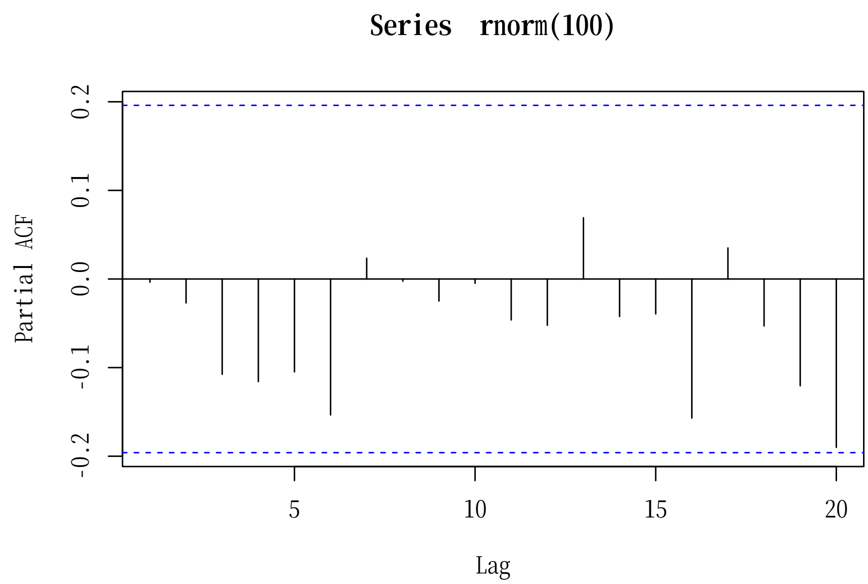
ACF 图示例：

```
set.seed(1)
acf(rnorm(100))
```



PACF 示例：

```
set.seed(1)
pacf(rnorm(100))
```



A.5 时间序列基本统计

对于一元时间序列，仍可以用处理向量的函数如 `mean`, `sd`, `var`, `min`, `max` 等函数进行统计。长度可以用 `length` 求。

用 `diff(x)` 求一阶差分。

用 `acf(x)` 作样本自相关函数图，自动给出上下两条界限作为白噪声情况下近似 95% 界限。用 `lag.max=` 指定需要的最大滞后阶数。用 `type="covariance"` 可以返回样本自协方差估计。`acf()` 的返回值是一个列表，对于一元时间序列，`c(acf(x)$acf)` 返回从滞后 0 到滞后 `lag.max` 的自相关函数值为一个向量，`c(acf(x, type="covariance")$acf)` 返回自协方差函数值为一个向量。

用 `pacf(x)` 作偏相关函数图。自动给出上下两条界限用来则建立 AR 模型时帮助判断何处偏相关截尾。

A.6 ARIMA 模型有关

可以模拟生成 ARMA 或者 ARIMA 模型的模拟数据。用 set.seed(n) 在模拟前设置种子编号以保证模拟结果可重复。用 arima.sim() 函数模拟生成 AR、MA、ARMA 或者 ARIMA 模型数据。格式为

```
arima.sim(model, n, rand.gen = rnorm, innov = rand.gen(n, ...),
          n.start = NA, start.innov = rand.gen(n.start, ...),
          ...)
```

其中 n 是要输出的观测个数，model 是一个列表，列表元素 ar 为 AR 部分的系数（写在方程右侧的格式)，列表元素 ma 为 MA 部分的系数，列表元素 order 如果存在，是 (p, d, q) 三个元素的向量，p 是 AR 阶数，q 是 MA 阶数，d 是差分阶数。三个列表元素都可以省略。白噪声默认为零均值标准正态白噪声，如果需要白噪声方程为 σ^2，只要把生成的序列乘以 σ。在模拟 ARMA 时输出的是零均值序列。

可选参数可以提供一个用来产生白噪声列的随机数函数，默认为 rnorm 函数。可选参数 innov 可以规定为用户自己提供的白噪声列（R 向量格式)。可选参数 n.start 是为了达到平稳需要预热模拟的步数，默认是程序自动选择。可选参数 start.innov 可以提供预热阶段的白噪声输入。

例如，模拟如下的 AR(2) 模型，产生长度为 300 的样本：

$$X_t = 0.2X_{t-1} + 0.5X_{t-2} + \varepsilon_t,\ \varepsilon_t \sim \text{WN}(0, 2^2)$$

程序为:

```
x <- 2.0*arima.sim(model=list(ar=c(0.2, 0.5)), n=300)
```

又如，模拟如下的 MA(2) 模型, 产生长度为 300 的样本：

$$X_t = \varepsilon_t - 0.36\varepsilon_{t-1} + 0.85\varepsilon_{t-2},\ \varepsilon_t \sim \text{WN}(0, 2^2)$$

程序为：

```
x <- 2.0*arima.sim(model=list(ma=c(-0.36, 0.85)), n=300)
```

再比如，模拟如下的 ARMA(4,2) 模型，产生长度为 300 的样本：

$$X_t = -0.9X_{t-1} - 1.4X_{t-2} - 0.7X_{t-3} - 0.6X_{t-4} + \varepsilon_t + 0.5\varepsilon_{t-1} - 0.4\varepsilon_{t-2},\ \varepsilon_t \sim \text{WN}(0, 2^2)$$

程序为：

```
x <- 2.0*arima.sim(model=list(
  ar=c(-0.9, -1.4, -0.7, -0.6),
  ma=c(0.5, -0.4)), n=300)
```

模拟如下的 ARIMA(1,1,1) 模型，产生长度为 300 的样本：

$$Y_t = X_t - X_{t-1}, \quad X_t = -0.9X_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1},\ \varepsilon_t \sim \text{WN}(0, 2^2)$$

程序为：

```
x <- 2.0*arima.sim(model=list(
  ar=c(-0.9), ma=c(0.5), order=c(1,1,1)), n=300)
```

为了建立 AR 模型，并用 AIC 定阶，用 `ar(x, method="mle")`。例如，对航空乘客数据对数差分序列：

```
ar(diff(log(AirPassengers)), method="mle")
```

```
##
## Call:
## ar(x = diff(log(AirPassengers)), method = "mle")
##
## Coefficients:
##       1        2        3        4        5        6        7        8
## -0.2212  -0.2835  -0.2550  -0.2973  -0.2300  -0.2621  -0.2571  -0.3274
##       9       10       11       12
## -0.2195  -0.2934  -0.1980   0.6020
##
## Order selected 12  sigma^2 estimated as  0.001643
```

为了指定 (p, q) 估计 ARMA(p, q) 模型，用 `arima(x, order=(p,0,q))`。比如，对航空乘客数据对数值差分序列建立 ARMA(1,1) 模型：

```r
arima(diff(log(AirPassengers)), order=c(1,0,1))
```

```
##
## Call:
## arima(x = diff(log(AirPassengers)), order = c(1, 0, 1))
##
## Coefficients:
##           ar1     ma1  intercept
##       -0.5826  0.8502     0.0098
## s.e.   0.1285  0.0856     0.0099
##
## sigma^2 estimated as 0.0102:  log likelihood = 124.8,  aic = -241.61
```

上述输出对应的模型为：

$$
\begin{aligned}
X_t =& 0.0098 + Y_t \\
Y_t =& -0.5826 X_{t-1} + \varepsilon_t + 0.8502\varepsilon_{t-1}, \ \varepsilon_t \sim \text{WN}(0, 0.0102)
\end{aligned}
$$

注意 `ar1` 系数是在模型方程等号右侧的，`sigma^2` 是白噪声方差 σ^2 的估计。

为了指定 (p,d,q) 估计 ARIMA 模型，用 `arima(x, order=(p,d,q))`。比如，对航空乘客数据对数值建立 ARIMA(1,1,1) 模型：

```r
arima(log(AirPassengers), order=c(1,1,1))
```

```
##
## Call:
## arima(x = log(AirPassengers), order = c(1, 1, 1))
##
## Coefficients:
##           ar1     ma1
##       -0.5780  0.8482
## s.e.   0.1296  0.0865
##
## sigma^2 estimated as 0.01027:  log likelihood = 124.31,  aic = -242.63
```

上述模型应该写成：

$$
\begin{aligned}
Y_t =& X_t - X_{t-1} \\
Y_t =& -0.5780X_{t-1} + \varepsilon_t + 0.8482\varepsilon_{t-1}, \ \varepsilon_t \sim \text{WN}(0, 0.01027)
\end{aligned}
$$

A.7 白噪声检验

用 `Box.test(x, lag=12, type="Ljung")` 作 Ljung-Box 白噪声检验，用 `lag=` 指定采用多少个样本自相关系数的平方和计算卡方统计量。对 ARMA 模型的输出，可以加 `fitdf=` 指定 ARMA 模型的系数个数用来进行自由度调整，必须满足 `lag>fitdf`。

例如：

```
set.seed(101)
Box.test(rt(100, 5), lag=1, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  rt(100, 5)
## X-squared = 1.1389, df = 1, p-value = 0.2859
```

A.8 单位根检验

fUnitRoots 包的 adfTest() 函数可以执行单位根 ADF 检验。tseries 包的 adf.test() 函数也可以执行单位根 ADF 检验。

`adfTest(x, lags=1, type="nc")` 作 ADF 检验，对应的 AR 阶数为 1（用 `lags=1` 指定）。用 `type="nc"` 表示不带漂移的随机游动，用 `type="c"` 表示带有漂移的随机游动，用 `type="ct"` 表示带有常数项和线性时间因素的随机游动。

如：

```
library(fUnitRoots)
set.seed(101)
adfTest(log(AirPassengers), type="c")
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 1
##   STATISTIC:
##     Dickey-Fuller: -2.0185
##   P VALUE:
##     0.3072
##
## Description:
##  Mon Nov 15 10:49:13 2021 by user: Lenovo
```

ADF 检验的零假设是有单位根，对立假设是没有单位根。上述检验的 p 值 0.3072 不显著，表示有单位根。因为序列有明显增长趋势所以用了 `type="c"` 选项，表示有漂移的随机游动。

附录 B

数学分析

时间序列分析用到了数学分析、复分析、实变函数、泛函分析、测度论、概率论、随机过程、数理统计、多元统计分析中的一些结果。这里对一些数学知识进行整理。

B.1 极限

定理 B.1 (分部求和公式)**.** 若 $\{x_k, k = m, m+1, \dots, n\}$, $\{y_k, k = m, m+1, \dots, n+1\}$ 是数列，则

$$
\begin{aligned}
\sum_{k=m}^{n} x_k (y_{k+1} - y_k) =& [x_n y_{n+1} - x_m y_m] - \sum_{k=m+1}^{n} y_k (x_k - x_{k-1}) \\
=& [x_n y_{n+1} - x_m y_m] - \sum_{k=m}^{n-1} y_{k+1} (x_{k+1} - x_k)
\end{aligned}
$$

证明：

$$
\begin{aligned}
\text{左边} =& \sum_{k=m}^{n} x_k y_{k+1} - \sum_{k=m}^{n} x_k y_k \\
\text{右边} =& x_n y_{n+1} - x_m y_m - \sum_{k=m+1}^{n} x_k y_k + \sum_{k=m+1}^{n} x_{k-1} y_k \\
=& x_n y_{n+1} + \sum_{s=m}^{n-1} x_s y_{s+1} - \sum_{k=m}^{n} x_k y_k \\
=& \sum_{s=m}^{n} x_s y_{s+1} - \sum_{k=m}^{n} x_k y_k \\
=& \text{左边}
\end{aligned}
$$

定理 B.2 (Kronecker 引理)**.** 设 $\{x_n, n \in \mathbb{N}_+\}$ 是复数列，$\sum_{n=1}^{\infty} x_n$ 收敛到复数 s。实数列 $\{b_n\}$ 满足 $0 < b_1 \leq b_2 \leq \dots$ 且 $\lim_{n\to\infty} b_n = \infty$，则

$$
\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} b_k x_k = 0.
$$

证明： 记 $S_n = \sum_{j=1}^{n} x_j$, $S_0 = 0$, $y_{n+1} = S_n$。由分部求和公式有

$$
\begin{aligned}
& \sum_{k=1}^{n} b_k x_k = \sum_{k=1}^{n} b_k (S_k - S_{k-1}) = \sum_{k=1}^{n} b_k (y_{k+1} - y_k) \\
=& b_n y_{n+1} - b_1 y_1 - \sum_{k=1}^{n-1} y_{k+1} (b_{k+1} - b_k) \\
=& b_n S_n - \sum_{k=1}^{n-1} S_k (b_{k+1} - b_k)
\end{aligned}
$$

于是

$$
\frac{1}{b_n} \sum_{k=1}^{n} b_k x_k = S_n - \frac{1}{b_n} \sum_{k=1}^{n-1} (b_{k+1} - b_k) S_k
$$

由于 $S_n \to s$, $\forall \varepsilon > 0$，存在 N 使得 $n \geq N$ 时 $|S_n - s| < \varepsilon/2$。将上式右边变

成

$$
\begin{aligned}
&S_n-\frac{1}{b_n}\sum_{k=1}^{N-1}(b_{k+1}-b_k)S_k-\frac{1}{b_n}\sum_{k=N}^{n-1}(b_{k+1}-b_k)S_k\\
=&S_n-\frac{1}{b_n}\sum_{k=1}^{N-1}(b_{k+1}-b_k)S_k-\frac{1}{b_n}\sum_{k=N}^{n-1}(b_{k+1}-b_k)s-\frac{1}{b_n}\sum_{k=N}^{n-1}(b_{k+1}-b_k)(S_k-s)\\
=&S_n-\frac{1}{b_n}\sum_{k=1}^{N-1}(b_{k+1}-b_k)S_k-\frac{b_n-b_N}{b_n}s-\frac{1}{b_n}\sum_{k=N}^{n-1}(b_{k+1}-b_k)(S_k-s)
\end{aligned}
$$

当 $n\to\infty$ 时，第一项和第三项分别趋于 s 和 $-s$，可以消去；第二项趋于 0，第四项的绝对值小于等于 $\frac{1}{2}\varepsilon\frac{b_n-b_N}{b_n}\le\frac{1}{2}\varepsilon$，所以存在 $N_2>N$ 使得 $n>N_2$ 时四项之和绝对值小于 ε。Knonecker 引理证毕。

定理 B.3 (Stolz 定理). 设实数列 $\{a_n\}$、$\{b_n\}$ 满足

(1) $\{b_n\}$ 严格单调递增;

(2) $\lim_{n\to\infty}b_n=+\infty$;

(3) $\lim_{n\to\infty}\frac{a_{n+1}-a_n}{b_{n+1}-b_n}=L$ 有意义，L 为有限实数、$+\infty$ 或 $-\infty$;

则有

$$
\lim_{n\to\infty}\frac{a_n}{b_n}=\lim_{n\to\infty}\frac{a_{n+1}-a_n}{b_{n+1}-b_n}=L
$$

这是与微积分中洛必达法则类似的数列极限定理。

证明：当 L 为有限实数时，由条件 (3) 和条件 (1) 可知，$\forall\epsilon>0,\exists N_1>0$，当 $n>N_1$ 时

$$
\left|\frac{a_{n+1}-a_n}{b_{n+1}-b_n}-L\right|<\epsilon
$$

从而

$$
L-\epsilon<\frac{a_{n+1}-a_n}{b_{n+1}-b_n}<L+\epsilon
$$

$$
(L-\epsilon)(b_{n+1}-b_n)<a_{n+1}-a_n<(L+\epsilon)(b_{n+1}-b_n)\quad(*)
$$

由条件 (2)，$\exists N_2>N_1$，当 $n>N_2$ 时 $b_n>\epsilon>0$。

当 $n > N_2$ 时，从 $N_2 + 1$ 到 n 对 (*) 式累加，有

$$(L-\epsilon)(b_{n+1}-b_{N_2+1}) < a_{n+1}-a_{N_2+1} < (L+\epsilon)(b_{n+1}-b_{N_2+1})$$

于是

$$L-\epsilon < \frac{a_{n+1}-a_{N_2+1}}{b_{n+1}-b_{N_2+1}} < L+\epsilon$$

由 $b_{n+1} > \epsilon > 0$，得

$$L-\epsilon < \frac{\frac{a_{n+1}}{b_{n+1}}-\frac{a_{N_2+1}}{b_{n+1}}}{1-\frac{b_{N_2+1}}{b_{n+1}}} < L+\epsilon$$

令 $n \to \infty$，因为

$$\lim_{n\to\infty}\frac{a_{N_2+1}}{b_{n+1}} = 0, \quad \lim_{n\to\infty}\frac{b_{N_2+1}}{b_{n+1}} = 0$$

所以存在 $N_3 > N_2$，$n > N_3$ 时

$$|L+\epsilon|\cdot\left|\frac{b_{N_2+1}}{b_{n+1}}\right| < \epsilon, \quad |L-\epsilon|\cdot\left|\frac{b_{N_2+1}}{b_{n+1}}\right| < \epsilon, \quad \left|\frac{a_{N_2+1}}{b_{n+1}}\right| < \epsilon$$

于是

$$L-2\epsilon < \frac{a_{n+1}}{b_{n+1}}-\frac{a_{N_2+1}}{b_{n+1}} < L+2\epsilon$$

$$L-3\epsilon < \frac{a_{n+1}}{b_{n+1}} < L+3\epsilon$$

即

$$\left|\frac{a_{n+1}}{b_{n+1}}-L\right| < 3\epsilon$$

即

$$\lim_{n\to\infty}\frac{a_n}{b_n} = L.$$

L 为无穷时的证明略。

推论 B.1. 如果数列 $a_n \to 0(n \to \infty)$，则

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}a_i = 0$$

证明：由 Stolz 定理，记 $S_n = \sum_{i=1}^{n} a_i$，则

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}a_i &= \lim_{n\to\infty}\frac{S_n}{n}\\ &= \lim_{n\to\infty}\frac{S_n - S_{n-1}}{n-(n-1)}\\ &= \lim_{n\to\infty}a_n = 0\end{aligned}$$

B.2 微积分

定理 B.4 (微积分基本定理)**.** *(1) 若 $f(x)$ 是定义在 $[a,b]$ 上的 Riemann 可积函数且在 $x = x_0$ 处连续，则函数*

$$F(x) = \int_a^x f(t)dt, \quad x \in [a,b]$$

在 $x = x_0$ 处可微且 $F'(x_0) = f(x_0)$。

(2) 若 $f(x)$ 是定义在 $[a,b]$ 上的可微函数，$f'(x)$ 在 $[a,b]$ 上是 Riemann 可积函数，则 $f(x)$ 是其导函数的不定积分：

$$\int_a^x f'(t)dt = f(x) - f(a), \quad x \in [a,b]$$

对 Lebesgue 积分也有类似结论。

定理 B.5 (Lebesgue 定理)**.** *若 $f(x)$ 是定义在 $[a,b]$ 上的单调上升实值函数，则 $f(x)$ 的不可微点集为零测集且有*

$$\int_a^b f'(x)dx \le f(b) - f(a)$$

勒贝格积分与黎曼积分关系

黎曼积分是按照 x 的区间进行分割，当细分小区间长度趋于零时的极限（如果存在）。勒贝格积分是按照函数值 y 的区间进行分割，用简单函数的积分逼近一般函数的积分。

定理 B.6. 对闭区间 $[a,b]$ 上的有界函数 f，如果黎曼可积，则 f 必为 $Borel$ 可测函数且勒贝格可积，积分值相等。

定理 B.7. 对闭区间 $[a,b]$ 上的有界函数 f，f 黎曼可积的充分必要条件是 f 在 $[a,b]$ 中的不连续点组成的集合为勒贝格零测集。

推论： $[a,b]$ 上仅有有限个不连续点的函数是黎曼可积的，也是勒贝格可积的，两种积分相等。

定义 B.1 (有界变差函数)**.** 设 $f(x)$ 是定义在 $[a,b]$ 上的实值函数，作分划 Δ_t: $a=x_0<x_1<\cdots<x_n=b$ 以及相应的和

$$\nu_\Delta=\sum_{i=1}^{n}|f(x_i)-f(x_{i-1})|$$

令

$$\bigvee_a^b(f)=\sup\{\nu_\Delta:\Delta\text{为 }[a,b]\text{ 的任一分划}\}$$

并称它为 f 在 $[a,b]$ 上的全变差。若

$$\bigvee_a^b(f)<+\infty$$

则称 $f(x)$ 是 $[a,b]$ 上的有界变差函数，其全体记为 $BV([a,b])$。

有界变差函数有界，$BV([a,b])$ 构成一个线性空间。

B.3 数值级数

设 $\{a_n,n=1,2,\ldots\}$ 为实数列，考虑 $\sum_{n=1}^{\infty}a_n$。称

$$S_n=\sum_{i=1}^{n}a_i$$

为部分和序列。如果 S_n 有实数值极限 S，则称级数 $\sum_{n=1}^{\infty}a_n$ 收敛到 S；如果 S_n 极限为 $+\infty$ 或 $-\infty$，则称级数 $\sum_{n=1}^{\infty}a_n$ 发散到 $+\infty$ 或 $-\infty$；如果 S_n 极限不存在，则称级数 $\sum_{n=1}^{\infty}a_n$ 发散。

如果 $\sum_{n=1}^{\infty}|a_n|$ 收敛到有限值，则称级数 $\sum_{n=1}^{\infty}a_n$ 绝对收敛，绝对收敛推出收敛。

如果级数 $\sum_{n=1}^{\infty}a_n$ 收敛，则 $\lim_{n\to\infty}a_n=0$。

对正项级数 $\sum_{n=1}^{\infty}a_n$ 和 $\sum_{n=1}^{\infty}b_n$，如果 $\lim_{n\to\infty}\frac{a_n}{b_n}$ 为有限的非零实数值，即 a_n 和 b_n 同阶，则两个级数同时收敛或者同时发散。

达朗倍尔判别法：设 $\sum_{n=1}^{\infty}a_n$ 是正项级数，若

$$\varlimsup_{n\to\infty}\frac{a_{n+1}}{a_n}=r<1,$$

则 $\sum_{n=1}^{\infty}a_n$ 收敛；若

$$\varliminf_{n\to\infty}\frac{a_{n+1}}{a_n}=r>1,$$

则 $\sum_{n=1}^{\infty}a_n$ 发散；$r=1$ 时不能判断。

哥西判别法：设 $\sum_{n=1}^{\infty}a_n$ 是正项级数，若

$$\varlimsup_{n\to\infty}a_n^{1/n}=\rho,$$

则当 $\rho<1$ 时级数收敛，当 $\rho>1$ 时级数发散。

如果级数 $\sum_{n=1}^{\infty}a_n$ 绝对收敛，则任意改变求和次序，级数仍绝对收敛，且收敛到相同值；否则，改变求和次序可能发散或收敛到不同的结果。

二重级数：对数列 $\{a_{ij},i=1,2,\ldots,j=1,2,\ldots\}$，令 $S_i=\sum_{j=1}^{\infty}a_{ij}$，若每个 S_i 收敛，且 $\sum_{i=1}^{\infty}S_i$ 收敛，则级数 $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}a_{ij}$ 收敛到 $\sum_{i=1}^{\infty}S_i$。

如果其中 $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|a_{ij}|<\infty$，则可交换次序

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}a_{ij}=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}a_{ij}.$$

级数乘法：设级数 $\sum_{n=1}^{\infty}a_n$ 和 $\sum_{n=1}^{\infty}b_n$ 至少有一个绝对收敛，则

$$\left(\sum_{n=1}^{\infty}a_n\right)\left(\sum_{n=1}^{\infty}b_n\right)=\sum_{n=1}^{\infty}c_n,$$

其中

$$c_n=\sum_{i=1}^{n}a_ib_{n+1-i}.$$

常用求和公式：

$$
\begin{aligned}
1+2+3+\cdots+n &= \sum_{k=1}^{n} k = \frac{1}{2}n(n+1). \\
1^2+2^2+3^2+\cdots+n^2 &= \sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1). \\
1^3+2^3+3^3+\cdots+n^3 &= \sum_{k=1}^{n} k^3 = \left(\frac{1}{2}n(n+1)\right)^2. \\
1^4+2^4+3^4+\cdots+n^4 &= \sum_{k=1}^{n} k^4 = \frac{1}{30}n(n+1)(2n+1)(3n^2+3n-1).
\end{aligned}
$$

B.4 函数项级数

设 $f_n(x)$ 是定义在区间 I 上的函数，$n=1,2,\ldots$，称 $\{f_n(x), n=1,2,\ldots\}$ 为函数序列。如果在 I 的一个非空子集 I_1 上

$$\lim_{n\to\infty} f_n(x) = f(x),\ \forall x \in I_1,$$

则称 $f(x)$ 在 I_1 上是函数序列的极限函数。

考虑函数级数 $\sum_{n=1}^{\infty} u_n(x)$，$u_n(x)$ 是区间 I 上的函数，如果其部分和序列

$$S_n(x) = \sum_{i=1}^{n} u_i(x)$$

在 $I_1 \subset I$ 中收敛到极限函数 $S(x)$，则称级数 $\sum_{n=1}^{\infty} u_n(x)$ 在 I_1 中收敛到 $S(x)$。

使得级数 $\sum_{n=1}^{\infty} u_n(x)$ 收敛的点 x 称为收敛点，否则称为发散点。所有收敛点组成的集合称为收敛区域，所有发散点组成的集合称为发散区域。

如果 $\sum_{n=1}^{\infty} |u_n(x)|$ 在 I_1 中收敛，则称 $\sum_{n=1}^{\infty} u_n(x)$ 在 I_1 中绝对收敛，绝对收敛推出收敛。

级数 $\sum_{n=1}^{\infty} u_n(x)$ 与极限 $\lim_{n\to\infty} \sum_{i=1}^{n} u_i(x)$ 是同一问题。对其中函数的微分、积分、极限等操作能否与求和或者极限运算交换次序？在一致收敛条件下可以。

一致收敛：设 $f_n(x)$ 在区间 I_1 有极限函数 $f(x)$，如果任给 $\epsilon>0$，都存在一个不依赖于 x 的正整数 N，当 $n\geq N$ 时，对任意 $x\in I_1$ 都有

$$|f_n(x)-f(x)|<\epsilon,$$

则称 $f_n(x)$ 在区间 I_1 一致收敛到 $f(x)$。类似地，如果级数的部分和序列一致收敛，则称级数一致收敛。

$f_n(x)$ 在区间 I_1 一致收敛到 $f(x)$，当且仅当

$$\lim_{n\to\infty}\sup_{x\in I_1}|f_n(x)-f(x)|=0.$$

对函数级数 $\sum_{n=1}^{\infty}u_n(x)$，如果 $\sum_{i=n+1}^{\infty}u_i(x)$ 一致收敛到 0，则函数级数一致收敛。

极限次序交换：设函数 $f_n(x), n=1,2,...$ 定义在 $[a,b]$ 上，$x_0\in[a,b]$ 且 $f_n(x)$ 在 $[a,b]\backslash\{x_0\}$ 上一致收敛到 $f(x)$，设 $\lim_{x\to x_0}f_n(x)$ 存在，则

$$\lim_{x\to x_0}\lim_{n\to\infty}f_n(x)=\lim_{n\to\infty}\lim_{x\to x_0}f_n(x).$$

极限与求和号交换次序：设函数 $u_n(x), n=1,2,...$ 定义在 $[a,b]$ 上，$x_0\in[a,b]$ 且 $\sum_{n=1}^{\infty}u_n(x)$ 在 $[a,b]\backslash\{x_0\}$ 上一致收敛到 $S(x)$，设 $\lim_{x\to x_0}u_n(x)$ 存在，则

$$\lim_{x\to x_0}\sum_{n=1}^{\infty}u_n(x)=\sum_{n=1}^{\infty}\lim_{x\to x_0}u_n(x).$$

如果 $f_n(x)$ 是闭区间 $[a,b]$ 上的连续函数，$f_n(x)$ 在 $[a,b]$ 上一致收敛到 $f(x)$，则 $f(x)$ 也是闭区间 $[a,b]$ 上的连续函数。

如果 $u_n(x)$ 是闭区间 $[a,b]$ 上的连续函数，$\sum_{n=1}^{\infty}u_n(x)$ 在在 $[a,b]$ 上一致收敛到 $S(x)$，则 $S(x)$ 也是闭区间 $[a,b]$ 上的连续函数。

如果 $u_n(x)$ 是开区间 (a,b) 上的连续函数，$\sum_{n=1}^{\infty}u_n(x)$ 在 (a,b) 内每一个闭区间上都一致收敛到 $S(x)$，则 $S(x)$ 也是开区间 (a,b) 上的连续函数。

积分号下取极限：设 $f_n(x)$ 是闭区间 $[a,b]$ 上的连续函数，$f_n(x)$ 在 $[a,b]$ 上一致收敛到 $f(x)$，则

$$\lim_{n\to\infty}\int_a^b f_n(x)\,dx=\int_a^b\lim_{n\to\infty}f_n(x)\,dx.$$

积分与求和号交换次序：设 $u_n(x)$ 是闭区间 $[a,b]$ 上的连续函数，级数 $\sum_{n=1}^{\infty} u_n(x)$ 在在 $[a,b]$ 上一致收敛到 $S(x)$，则

$$\int_a^b \sum_{n=1}^{\infty} u_n(x)\,dx = \sum_{n=1}^{\infty} \int_a^b u_n(x)\,dx.$$

微分与求和号交换次序：设 $u_n(x)$ 在闭区间 $[a,b]$ 上可微，$\sum_{n=1}^{\infty} u_n'(x)$ 一致收敛，且 $\sum_{n=1}^{\infty} u_n(x)$ 至少在某一个点 x_0 上收敛，则 $\sum_{n=1}^{\infty} u_n(x)$ 在 $[a,b]$ 上一致收敛，且

$$\left(\sum_{n=1}^{\infty} u_n(x)\right)' = \sum_{n=1}^{\infty} u_n'(x).$$

B.5 幂级数

形如

$$\sum_{n=0}^{\infty} a_n (x-x_0)^n$$

的函数项级数称为幂级数，其中 x_0 是任意给定实数，$\{a_n\}$ 是实数列。实际上只要考虑

$$\sum_{n=0}^{\infty} a_n x^n. \tag{B.1}$$

幂级数(B.1)的收敛区域只有如下三种情况：

1. 整个实数轴；
2. 关于原点对称的有限区间 $(-R,R)$，可含端点；
3. 只在 $x=0$ 处收敛。

令

$$\rho = \varlimsup_{n\to\infty} |a_n|^{1/n},$$

当 $0 \le \rho < \infty$ 时，幂级数(B.1)在 $|x| < \frac{1}{\rho}$ 绝对收敛；当 $0 < \rho < \infty$，$|x| > \frac{1}{\rho}$ 时，幂级数(B.1)发散。称 $R = \frac{1}{\rho}$ 为幂级数的收敛半径，$(-R,R)$ 为幂级数的收敛区间。当 $\rho = 0$ 时，收敛区间是 $(-\infty,\infty)$；当 $\rho = \infty$ 时，仅在 $x = 0$ 处收敛。收敛区间端点处是否收敛不确定。

若幂级数(B.1)在 $x = x_1 \ne 0$ 处收敛，则它在 $|x| < |x_1|$ 处绝对收敛；如果级数在 $x = x_0$ 处发散，则它在 $|x| > |x_0|$ 处也发散。

如果

$$\lim_{n\to\infty}\frac{|a_{n+1}|}{|a_n|}=l,$$

则幂级数(B.1)的收敛半径 $R=1/l$（包括 $l=0$ 和 $l=\infty$ 的情况）。

设幂级数(B.1)的收敛半径 $R>0$，则对任意 $0<r<R$，级数在 $[-r,r]$ 上一致收敛，称为在 $(-R,R)$ 内闭一致收敛。

幂级数(B.1)在收敛区间 $(-R,R)$ 内是连续函数。

幂级数微分：幂级数(B.1)在收敛区间 $(-R,R)$ 内可微，且微分与求和号可交换：

$$\left(\sum_{n=0}^{\infty}a_nx^n\right)'=\sum_{n=1}^{\infty}na_nx^{n-1},$$

右边的级数与(B.1)有相同的收敛半径。

幂级数积分：设幂级数(B.1)收敛半径 $R>0$，则积分号与求和号可交换：

$$\int_0^x\sum_{n=0}^{\infty}a_nt^n\,dt=\sum_{n=0}^{\infty}\frac{a_n}{n+1}x^{n+1},$$

右边的级数与(B.1)有相同的收敛半径。

泰勒展开：设函数 $f(x)$ 在 $I=(x_0-\delta,x_0+\delta)$ 上有任意阶导数，且存在正常数 M 使得

$$|f^{(n)}(x)|\le M^n,\ \forall x\in I,\ n=1,2,\dots,$$

则对 $x\in I$ 有

$$f(x)=\sum_{n=0}^{\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n.$$

B.6 傅立叶级数

考虑复数域上的希尔伯特空间 $L^2[-\pi,\pi]=(L^2[-\pi,\pi],\mathscr{B},U)$, 其中 $\mathscr{B}$ 是 $[-\pi,\pi]$ 上的 Borel 集组成的 σ 域，U 是 $[-\pi,\pi]$ 上的 Lebegue 测度。定义内积为

$$<f,g>=E(f\bar{g})=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(x)\bar{g}(x)dx.$$

这时 $\{e_n=e^{inx},n\in\mathbb{Z}\}$ 构成标准正交基。如果 $f\in L^2[-\pi,\pi]$ 且

$$<f,e_j>=0,\ \forall j\in\mathbb{Z}$$

则

$$f(x)=0,\ \text{a.e.}$$

对 $f \in L^2[-\pi,\pi]$, 令

$$S_n f=\sum_{j=-n}^{n} < f, e_j > e_j,$$

其中

$$< f, e_j >= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ijx} f(x)dx$$

叫做 f 的 Fourier 系数，Fourier 系数列必平方可和。$S_n f$ 叫做 f 的 n 阶 Fourier 逼近，$S_n f$ 是 f 在 $\overline{\mathrm{sp}}\{e_j, |j|\le n\}$ 上的投影。

$S_n f$ 均方极限存在且等于 f。$S_n f$ 的极限写成函数级数

$$Sf=\sum_{j=-\infty}^{\infty} < f, e_j > e_j.$$

$$L^2[-\pi,\pi]=\overline{\mathrm{sp}}\{e_j, j\in\mathbb{Z}\}.$$

$$\|f\|^2=\sum_{j=-\infty}^{\infty} |< f, e_j >|^2.$$

$$< f, g >= \sum_{j=-\infty}^{\infty} < f, e_j > \cdot \overline{< g, e_j >}.$$

若 $f(x)$ 是以 2π 为周期的连续函数，则任给 $\varepsilon>0$，存在三角多项式

$$T_n(x)=\frac{a_0}{2}+\sum_{j=1}^{n}\left\{a_j\cos(jx)+b_j\sin(jx)\right\}$$

使得

$$|f(x)-T_n(x)|<\varepsilon,\ \forall x\in(-\infty,\infty)$$

事实上，

$$n^{-1}(S_0 f+S_1 f+\ldots S_{n-1}f)\to f$$

在 $[-\pi,\pi]$ 一致收敛 $(n\to\infty)$。

若 $f(x)$ 是以 2π 为周期的连续函数，且 $f' \in L^2[-\pi,\pi]$，则 $S_n f$ 不仅均方收敛到 f，而且绝对一致收敛到 f。(见 (Brockwell & Davis, 1987)§2.8, §2.11)。

对于以 2π 为周期的函数 $f(x)$，如果在 $[-\pi,\pi]$ 上可积（有瑕点时绝对可积），则可以计算

$$\begin{aligned} a_n =& \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx,\ n=0,1,2,\ldots \\ b_n =& \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx,\ n=1,2,\ldots \end{aligned}$$

并形式地写出函数级数

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}\{a_n\cos(nx) + b_n\sin(nx)\}$$

但不能保证级数收敛且收敛到 $f(x)$。

如果 $f(x)$ 在 $x=x_0$ 处满足 α 级 $(0<\alpha\leq 1)$ 李普希兹条件：

$$|f(x_0 \pm t) - f(x_0)| \leq Lt^{\alpha},\ 0<t\leq\delta$$

(其中 $L>0, \delta>0$)，则 $f(x)$ 的傅立叶级数在 x_0 处收敛到 $f(x)$。

若 $f(x)$ 在 $[a,b]$ 逐段可微（除了有限个点外可微，在这些点上有左右导数），则其傅立叶级数在每个 $x=x_0$ 处均收敛到

$$S_0 = \frac{f(x_0+0)+f(x_0-0)}{2}$$

当然，除去不可微的有限个点之外都收敛到 $f(x_0)$。

若对点 x_0 存在 $h>0$ 使得 $f(x)$ 在 $[x_0-h,x_0]$ 和 $[x_0,x_0+h]$ 分别单调, 则 $f(x)$ 的傅立叶级数在 x_0 收敛到

$$\frac{f(x_0+0)+f(x_0-0)}{2}$$

若 $f(x)$ 逐段单调，则其傅立叶级数对任意 x_0 均收敛到

$$\frac{f(x_0+0)+f(x_0-0)}{2}$$

若 $f(x)$ 在区间 $[-\pi,\pi]$ 上平方可积，则 $\forall\varepsilon>0$，存在三角多项式 $T(x)$ 使得

$$\int_{-\pi}^{\pi}|f(x)-T(x)|^2dx < \varepsilon$$

若 $f(x)$ 在区间 $[-\pi,\pi]$ 上黎曼可积或在广义积分意义下平方可积，设 $S_n(f,x)$ 为其傅立叶级数的部分和，则

$$\lim_{n\to\infty}\int_{-\pi}^{\pi}|f(x)-S_n(f,x)|^2dx=0$$

B.7 参变积分

定理 B.8 (参变积分连续性（一）). 设二元函数 $f(x,y)$ 是 $[a,b]\times[\alpha,\beta]$ 上的连续函数，则

$$g(x)=\int_{\alpha}^{\beta}f(x,y)\,dy$$

是 $[a,b]$ 上的连续函数。

推论 (积分号下取极限) 在定理条件下对 $x_0\in[a,b]$ 有

$$\lim_{x\to x_0}\int_{\alpha}^{\beta}f(x,y)\,dy=\int_{\alpha}^{\beta}\lim_{x\to x_0}f(x,y)\,dy.$$

如果是广义积分或者瑕积分则需要更强的条件。

定理 B.9 (参变积分连续性（二）). 设二元函数 $f(x,y)$ 是 $[a,b]\times[\alpha,\beta]$ 上的连续函数，$\phi(x)$, $\psi(x)$ 是 $[a,b]$ 上的连续函数且取值于 $[\alpha,\beta]$，则

$$g(x)=\int_{\phi(x)}^{\psi(x)}f(x,y)\,dy$$

是 $[a,b]$ 上的连续函数。

定理 B.10 (积分号下求导). 设 $f(x,y)$ 和 $\frac{\partial f(x,y)}{\partial x}$ 都是 $[a,b]\times[\alpha,\beta]$ 上的连续函数，则

$$g(x)=\int_{\alpha}^{\beta}f(x,y)\,dy$$

在 $[a,b]$ 上可微，且

$$g'(x)=\frac{\partial}{\partial x}\int_{\alpha}^{\beta}f(x,y)\,dy=\int_{\alpha}^{\beta}\frac{\partial f(x,y)}{\partial x}\,dy.$$

定理 B.11 (参变积分求导). 设 $f(x,y)$ 和 $\frac{\partial f(x,y)}{\partial x}$ 都是 $[a,b]\times[\alpha,\beta]$ 上的连续函数，$\phi(x)$, $\psi(x)$ 是 $[a,b]$ 上的可微函数且取值于 $[\alpha,\beta]$，则

$$g(x) = \int_{\phi(x)}^{\psi(x)} f(x,y)\,dy$$

在 $[a,b]$ 上可微，且

$$\begin{aligned} g'(x) =& \frac{\partial}{\partial x}\int_{\phi(x)}^{\psi(x)} f(x,y)\,dy \\ =& \int_{\phi(x)}^{\psi(x)} \frac{\partial f(x,y)}{\partial x}\,dy + f(x,\psi(x))\psi'(x) - f(x,\phi(x))\phi'(x). \end{aligned}$$

B.8 向量和矩阵的微分

B.8.1 关于向量的微分

对 $f:\mathbb{R}^p\to\mathbb{R}$, 记 $\frac{\partial f(x)}{\partial x}$ 为 f 的 p 个一阶偏导数组成的列向量，称为 f 的梯度，记一阶偏导数组成的行向量为 $\frac{\partial f(x)}{\partial x^T}$。

记 $\frac{\partial^2 f(x)}{\partial x\partial x^T}$ 为 f 的二阶偏导数组成的 $p\times p$ 矩阵，称为 f 的海色阵 (Hessian)。

设 a 为 p 维列向量，A 为 $p\times p$ 对称阵，则

$$\begin{aligned} &\frac{\partial(a^T x)}{\partial x} = a, \quad \frac{\partial(x^T a)}{\partial x} = a, \\ &\frac{\partial(x^T A x)}{\partial x} = 2Ax, \\ &\frac{\partial^2(x^T A x)}{\partial x\partial x^T} = 2A. \end{aligned}$$

B.8.2 关于矩阵的微分

设 $f(X)$ 是以矩阵 $X=(x_{ij})_{m\times n}$ 为自变量的实值函数，关于各矩阵元素可导，记 $\frac{\partial f(X)}{\partial X}$ 表示 f 关于每个元素 x_{ij} 的偏导数组成的矩阵，即

$$\left(\frac{\partial f(X)}{\partial x_{ij}}\right)_{m\times n}.$$

性质：

对 $X_{m\times n}$，

$$\frac{\partial f(X)}{\partial X^T} = \left(\frac{\partial f(X)}{\partial X}\right)^T.$$

对 $X_{m\times n}$ 和 $A_{n\times m}$，

$$\frac{\partial \mathrm{tr}(XA)}{\partial X} = \frac{\partial \mathrm{tr}(AX)}{\partial X} = A^T,$$

对 $X_{m\times n}$ 和 $A_{m\times n}$，

$$\frac{\partial \mathrm{tr}(X^TA)}{\partial X} = \frac{\partial \mathrm{tr}(AX^T)}{\partial X} = A.$$

对 $X_{m\times n}$, $A_{p\times m}$, $B_{n\times p}$,

$$\frac{\partial \mathrm{tr}(AXB)}{\partial X} = \frac{\partial \mathrm{tr}(BAX)}{\partial X} = A^TB^T.$$

对 $X_{m\times n}$ 和对称阵 $A_{n\times n}$，

$$\frac{\partial \mathrm{tr}(XAX^T)}{\partial X} = 2XA.$$

对 $X_{m\times n}$, $A_{n\times m}$, $B_{n\times m}$,

$$\frac{\partial \mathrm{tr}(XAXB)}{\partial X} = B^TX^TA^T + A^TX^TB^T.$$

对 $X_{m\times n}$, $A_{n\times n}$, $B_{m\times m}$,

$$\frac{\partial \mathrm{tr}(XAX^TB)}{\partial X} = B^TXA^T + BXA.$$

对 $X_{m\times n}$, $B_{m\times m}$,

$$\frac{\partial \mathrm{tr}(X^TXB)}{\partial X} = X(B+B^T).$$

对可逆的 $m\times m$ 矩阵 X，有

$$\begin{aligned}\frac{\partial \log\det(X)}{\partial X} &= (X^T)^{-1},\\ \frac{\partial \det(X^{-1})}{\partial X} &= -\frac{1}{\det(X)}(X^T)^{-1},\\ \frac{\partial \det(X^{-1})}{\partial X^{-1}} &= -\det(X)X^T.\end{aligned}$$

附录 C

概率论

C.1 测度与概率

代数: 设 Ω 为集合，$\mathscr{F}$ 是 Ω 的子集组成的非空集合，称为集类。如果 $\mathscr{F}$ 满足以下条件：

(1) 对“余”运算封闭：$\forall A \in \mathscr{F}$ 都有 $A^c \in \mathscr{F}$；

(2) 对“可数和”运算封闭：若 $\{A_n : n \geq 1\} \subset \mathscr{F}$，则 $\bigcup_{n=1}^{\infty} A_n \subset \mathscr{F}$；

则称 $\mathscr{F}$ 为 域或者 代数。由定义易见 $\Omega \in \mathscr{F}$，$\emptyset \in \mathscr{F}$。$\{\Omega, \emptyset\}$ 构成最小的 域。Ω 的所有子集组成的集合构成最大的 域。

称 $(\Omega, \mathscr{F})$ 为可测空间。

若 Ω 为实数域中的区间 I，包含其中的所有闭区间的最小的 域称为 I 上的 Borel 域, 记作 $\mathscr{B}(I)$。实数域上的 Borel 域记作 $\mathscr{B}$。

设 $(\Omega, \mathscr{F})$ 为可测空间，定义在 $\mathscr{F}$ 上的实值函数 $\mu(\cdot)$ 称为一个测度，如果：

(1) $P(A) \geq 0, \forall A \in \mathscr{F}$;

(2) 对 $\{A_n : n \geq 1\} \subset \mathscr{F}$，如果 $\{A_n\}$ 互不相交，则 $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$,

则称 $\mu(\cdot)$ 是可测空间 $(\Omega, \mathscr{F})$ 上的一个**测度**。

对 $\mathscr{B}$ 或 $\mathscr{B}(I)$，定义测度 $\mu(\cdot)$ 使得

$$\mu([a,b]) = b - a,$$

称 $\mu(\cdot)$ 为 Lebesgue（勒贝格）测度，这是区间长度的推广。

在 $\mathscr{F}$ 上定义函数 $P(\cdot)$，满足：

(1) $P(A) \geq 0, \forall A \in \mathscr{F}$;

(2) $P(\Omega) = 1$;

(3) 对 $\{A_n : n \geq 1\} \subset \mathscr{F}$，如果 $\{A_n\}$ 互不相交，则 $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$；

则称 $P(\cdot)$ 是可测空间 $(\Omega, \mathscr{F})$ 上的**概率测度**，称 $(\Omega, \mathscr{F}, P)$ 为**概率空间**。概率测度是满足 $P(\Omega) = 1$ 的测度。

设 $X = X(\omega)$ 是定义在 Ω 上的函数，如果 $\forall x \in (-\infty, \infty)$，事件 $\{\omega : X(\omega) \leq x\} \in \mathscr{F}$，则称 X 为测度空间 $(\Omega, \mathscr{F})$ 上的**可测函数**；对概率空间 $(\Omega, \mathscr{F}, P)$，称 X 为**随机变量**。随机变量是可测空间 $(\Omega, \mathscr{F})$ 到可测空间 $(\mathbb{R}, \mathscr{B})$ 的可测映射。

对 $B \in \mathscr{B}$，令 $P_X(B) = P(X \in B)$，则 $(\mathbb{R}, \mathscr{B}, P_X)$ 构成概率空间，称为随机变量 X 的样本概率空间。

随机变量 X 的**分布函数**为

$$F(x) = P(X \leq x),\ x \in (-\infty, \infty)$$

分布函数是单调不减右连续函数，$\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$。满足这样条件的函数必为某个随机变量的分布函数。

设 $g(\cdot)$ 为 Borel 可测函数，则

$$Eg(X) = \int_{-\infty}^{\infty} g(x) dF(x)$$

只要右侧的积分存在。

C.2 矩不等式

C.2.1 高阶矩存在则低阶矩存在

对 $0 < r_1 < r_2$，若 $E|X|^{r_2} < \infty$, 必有 $E|X|^{r_1} < \infty$。事实上，

$$
\begin{aligned}
|x|^{r_1} & \begin{cases} \leq 1, & |x| \leq 1 \\ \leq |x|^{r_2}, & |x| > 1 \end{cases} \\
& \leq 1 + |x|^{r_2}
\end{aligned}
$$

所以 $E|X|^{r_1} \leq 1 + E|X|^{r_2}$。

C.2.2 C_r 不等式

记

$$
c_r = \begin{cases} 1, & 0 < r \leq 1 \\ 2^{r-1}, & r > 1 \end{cases}
$$

则对随机变量 X, Y 有

$$
E|X+Y|^r \leq c_r(E|X|^r + E|Y|^r)
$$

称上述不等式为 C_r 不等式。

下面证明。对 $0 < r \leq 1$，

$$
|a+b|^r \leq (|a| + |b|)^r \leq |a|^r + |b|^r
$$

事实上，只要证明 $(|a| + |b|)^r - |a|^r - |b|^r \leq 0$，不妨设 $a > 0$, $b > 0$, 只要证明 $f(x) = 1 - [x^r + (1-x)^r] \geq 0$, $x \in [0,1]$。易见 $f(0) = f(1) = 0$, $f''(x) = r(1-r)(x^{r-2} + (1-x)^{r-2}) > 0, \forall 0 < x < 1$,所以 $f(x) \leq 0, 0 \leq x \leq 1$。

对 $r > 1$，有

$$
(a+b)^r \leq 2^{r-1}(|a|^r + |b|^r)
$$

事实上，不妨设 $a > 0, b > 0$，只要证明 $(a+b)^r - 2^{r-1}(a^r + b^r) < 0$, 只要证对 $0 \leq x \leq 1$，$f(x) = 1 - 2^{r-1}(x^r + (1-x)^r) \leq 0$。易见 $f(0) < 0, f(1) < 0$，$f''(x) = -2^{r-1}r(r-1)[x^{r-2} + (1-x)^{r-2}] < 0, x \in (0,1)$，$f(x)$ 是凹函数，有唯一的最大值点，令 $f'(x) = 0$ 得 $x = \frac{1}{2}$ 为最大值点，而 $f(\frac{1}{2}) = 0$，所以 $f(x) \leq 0$, $x \in [0,1]$。

利用 C_r 记号则有如下 c_r 不等式：

$$(a+b)^r \le c_r(|a|^r + |b|^r)$$

从而对随机变量 X, Y 有

$$E|X+Y|^r \le c_r(E|X|^r + E|Y|^r)$$

C.2.3 Hölder 不等式

对 $p>1$, $q>1$, $\frac{1}{p}+\frac{1}{q}=1$，随机变量 X, Y，有如下的 Hölder 不等式:

$$E|XY| \le (E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}}$$

记 $\|X\|_p = (E|X|^p)^{\frac{1}{p}}$，不等式可以写成

$$E|XY| \le \|X\|_p \, \|Y\|_q$$

特别地，当 $p=q=2$ 时为如下的 Schwarz 不等式：

$$E|XY| \le \sqrt{EX^2 EY^2}$$

证明可以利用如下的 Young 不等式：

设 $p>1$, $q>1$, $\frac{1}{p}+\frac{1}{q}=1$, $a>0$, $b>0$，有

$$ab \le \frac{1}{p}a^p + \frac{1}{q}b^q$$

等号成立当且仅当 $a^p = b^q$。

Young 不等式证明: $\frac{1}{q} = 1 - \frac{1}{p} = \frac{p-1}{p}$, $q = \frac{p}{p-1}$。固定 $b>0$，令

$$f(a) = \frac{1}{p}a^p + \frac{1}{q}b^q - ab$$

只要证明 $f(a) \ge 0$ 且等号成立当且仅当 $a = b^{\frac{1}{p-1}}$。求导得

$$f'(a) = a^{p-1} - b$$

令 $f'(a) = 0$ 解得 $a^* = b^{\frac{1}{p-1}}$。对 $a = a^*$ 有

$$\begin{aligned} f(a^*) =& \frac{1}{p}b^{\frac{p}{p-1}} + \frac{p-1}{p}b^{\frac{p}{p-1}} - b^{\frac{1}{p-1}}b \\ =& 0 \end{aligned}$$

注意到 $a < a^*$ 时 $f'(a) < 0$, $a > a^*$ 时 $f'(a) > 0$，所以 $a = a^*$ 是 $f(a)$ 的唯一的严格最小值点，故 $f(a) > f(a^*) = 0$，对 $a \neq a^*$，而 $f(a^*) = 0$ 是唯一的一个取等号的点。Young 不等式证毕。

Hölder 不等式证明：

对 $p > 1$, $q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$，取

$$a = \frac{|X|}{(E|X|^p)^{\frac{1}{p}}}, b = \frac{|Y|}{(E|Y|^q)^{\frac{1}{q}}}$$

利用 Young 不等式有

$$\frac{|XY|}{(E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}}} \leq \frac{1}{p}\frac{|X|^p}{E|X|^p)} + \frac{1}{q}\frac{|Y|^q}{E|Y|^q}$$

所以

$$|XY| \leq \frac{1}{p}|X|^p(E|X|^p)^{\frac{1}{p}-1}(E|Y|^q)^{\frac{1}{q}} + \frac{1}{q}|Y|^q(E|Y|^q)^{\frac{1}{q}-1}(E|X|^p)^{\frac{1}{p}}$$

两边取期望得

$$\begin{aligned} E|XY| \leq& \frac{1}{p}(E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}} + \frac{1}{q}(E|Y|^q)^{\frac{1}{q}}(E|X|^p)^{\frac{1}{p}} \\ =&(E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}} \end{aligned}$$

得证。

C.2.4 Minkowski 不等式

对 $p \geq 1$，记 $\|X\|_p = (E|X|^p)^{\frac{1}{p}}$，有

$$\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$$

证明

$p = 1$ 时 $E|X+Y| \leq E|X| + E|Y|$ 显然。对 $p > 1$，利用 Hölder 不等式，有

$$\begin{aligned} \|X+Y\|_p^p =&(E|X+Y|^p)^{\frac{1}{p}} \leq E\left(|X| \cdot |X+Y|^{p-1}\right) + E\left(|Y| \cdot |X+Y|^{p-1}\right) \\ \leq&\|X\|_p \left(E|X+Y|^{(p-1)q}\right)^{\frac{1}{q}} + \|Y\|_p \left(E|X+Y|^{(p-1)q}\right)^{\frac{1}{q}} \end{aligned}$$

注意到 $\frac{1}{q} = \frac{p-1}{p}$, $q = \frac{p}{p-1}$, $(p-1)q = p$，所以

$$\|X+Y\|_p^p \leq \left(\|X\|_p + \|Y\|_p\right)\|X+Y\|_p^{\frac{p}{q}}$$

而

$$p - \frac{p}{q} = p(1 - \frac{1}{q}) = p\frac{1}{p} = 1$$

所以

$$\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$$

成立。

C.2.5 Jensen 不等式

定理 C.1. 设 $g(\cdot)$ 为凸函数，X 为随机变量，$Eg(X)$ 和 EX 存在，则

$$g(EX) \leq Eg(X)$$

等号成立当且仅当存在常数 c 使得 $X = c$, *a.s.*

证明略。

C.2.6 $\log E|X|^p$ 是 p 的凸函数

对 $p \geq 0$，$g(p) = \log E|X|^p$ 是 p 的凸函数。

利用 Schwarz 不等式，设 $0 \leq p_1 < p_2$，有

$$\begin{aligned}\left(E|X|^{\frac{p_1+p_2}{2}}\right)^2 =& \left\{E\left(|X|^{\frac{p_1}{2}}|X|^{\frac{2_1}{2}}\right)\right\}^2 \\ \leq& E|X|^{p_1} \cdot E|X|^{p_2}\end{aligned}$$

取对数得

$$2\log E|X|^{\frac{p_1+p_2}{2}} \leq \log E|X|^{p_1} + \log EE|X|^{2_1}$$

即

$$g(\frac{1}{2}p_1 + \frac{1}{2}p_2) \leq \frac{1}{2}g(p_1) + \frac{1}{2}g(p_2)$$

易见 $g(p)$ 关于 p 连续，所以 $g(p) = \log E|X|^p$ 是 $p \geq 0$ 的凸函数。

C.3 可测函数极限

随机变量是概率空间 $(\Omega, \mathscr{F}, P)$ 上的可测函数。设 $X = X(\omega)$ 是 $(\Omega, \mathscr{F}, P)$ 上的可测函数，允许取 $+\infty$ 和 $-\infty$ 值。称 X **a.s. 有限**，如果 $P(X = \pm\infty) = 0$。

设 $X(\omega)$ 和 $Y(\omega)$ 是 Ω 上的函数，如果存在零概率事件 $N \in \mathscr{F}$，使得 $\{\omega: X(\omega) \neq Y(\omega)\} \subset N$，称 X 和 Y **a.s. 相等**或者 a.e. 相等（几乎处处相等）。

如果 $Y(\omega)$ 是 Ω 上的函数，$X(\omega)$ 是随机变量，使得 Y 与 X a.s. 相等，则称 Y **a.s. 可测**或几乎处处可测。注意这时 $\{\omega: Y(\omega) \neq X(\omega)\}$ 包含于零概率集合但是本身不一定是可测集，所以几乎处处可测不一定可测，Y 不一定是随机变量 ($(\Omega, \mathscr{F})$ 可测函数)，但是在等价性的角度来看可以看成是随机变量。

设 X_n 是随机变量序列且 $P(X_n = \pm\infty) = 0$，$X(\omega)$ 是 Ω 上的函数，若存在零概率集合 $N \in \mathscr{F}$ 使得 $\{\omega: X_n(\omega) \not\to X(\omega)\} \subset N$，称 X_n 几乎处处收敛到 X，或者 a.s. 收敛到 X, X 一定是几乎处处可测的。事实上，令

$$\tilde{X}(\omega) = \begin{cases} \lim_n X_n(\omega), & \omega \notin N \\ 0, & \omega \in N \end{cases}$$

则 $\tilde{X}$ 可测且与 X 几乎处处相等。

如果 $\forall \omega \in \Omega$，$X_n(\omega) \to X(\omega)$，$X_n$ 可测，则 X 可测。

如果将几乎处处相等的 Ω 上的函数看成同一个函数，则随机变量序列的几乎处处极限如果存在就是唯一的。当 $\mathscr{F}$ 对 P 是完全的，几乎处处极限 X 是可测的（即随机变量）。

所谓完全 (完备) 测度，是指如果 $N \in \mathscr{F}$ 使得 $P(N) = 0$，则对任何 $A \subset N$ 都有 $A \in \mathscr{F}$，称 $\mathscr{F}$ 对测度 P 是完全的。

参见朱成熹《测度论基础》，科学出版社 1986。

C.4 多元期望和方差阵的性质

设 A, B, C 为矩阵，M 为随机矩阵，有

$$E(AMB + C) = AE(M)B + C.$$

设 a, b 为实值向量，A, B 为实值矩阵，X, Y, Z 为实值随机向量，则

$$
\begin{aligned}
\mathrm{Var}(a^T X) =& a^T \mathrm{Var}(X) a \\
\mathrm{Var}(AX + b) =& A \mathrm{Var}(X) A^T \\
\mathrm{Cov}(X + Y, Z) =& \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z) \\
\mathrm{Var}(X + Y) =& \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, X) \\
\mathrm{Cov}(AX, BY) =& A \mathrm{Cov}(X, Y) B^T.
\end{aligned}
$$

附录 D

线性代数

D.1 行列式

对 n 阶方阵 A，行列式为

$$\det(A) = \sum_{j_1 j_2 \dots j_n} (-1)^{\tau(j_1 j_2 \dots j_n)} a_{1j_1} a_{2j_2} \dots a_{nj_n}.$$

其中的求和对所有的 $n!$ 个 $(1, 2, \dots, n)$ 的全排列 $(j_1, j_2, \dots, j_n)$ 进行，$\tau(j_1 j_2 \dots j_n)$ 是排列 $j_1 j_2 \dots j_n$ 的逆序数，即

$$\tau(j_1 j_2 \dots j_n) = \#\{(j_i, j_k) : 1 \leq i < k \leq n, j_i > j_k\}.$$

行列式可以看成是关于 n^2 个自变量的 n 次多项式函数。

对 n 阶方阵 A 的元素 a_{ij}，将第 i 行和第 j 列删去后得到的 $n-1$ 阶行列式 M_{ij} 称为 a_{ij} 的余子式，而 $A_{ij} = (-1)^{i+j} M_{ij}$ 称为 a_{ij} 的代数余子式。行列式按一行或一列展开的公式为

$$\begin{aligned}
\det(A) =& \sum_{k=1}^{n} a_{ik} A_{ik},\ \forall i \in \{1, 2, \dots, n\} \\
=& \sum_{l=1}^{n} a_{lj} A_{lj},\ \forall j \in \{1, 2, \dots, n\}.
\end{aligned}$$

范德蒙 (Vandrmonde) 行列式：方阵 A 元素为 $a_{ij}=c_j^{i-1}$, 则

$$\det(A)=\prod_{i<j}(c_j-c_i).$$

克莱姆 (Cramer) 法则：设 A 为 n 阶方阵，b 为 n 维向量，方程组 $Ax=b$ 当且仅当 $\det(A)\neq 0$ 时存在唯一解，且第 j 个未知数的解等于将 A 的第 j 列替换成 b 后的矩阵行列式，除以 $\det(A)$ 的结果。这样，n 个方程、n 个未知数的方程组在 $\det(A)\neq 0$ 时的解都是有理分式形式，分子和分母的多项式次数为 n。

伴随矩阵：对方阵 A，设元素 a_{ij} 的代数余子式为 A_{ij}，令矩阵 $A^*=(A_{ij})^T_{n\times n}$，称 A^* 为 A 的伴随矩阵，有

$$AA^*=A^*A=\det(A)I,$$

当 A 可逆时有

$$A^{-1}=\frac{1}{\det(A)}A^*.$$

D.2 线性空间和内积空间

实数域上的线性空间某个集合 H 如果定义了如下的加法运算“+”和数乘运算“·”，使得

$$\begin{aligned}
&(1)\text{若}x,y\in H,\text{则}x+y\in H,\text{且}\\
&\quad x+y=y+x\\
&\quad (x+y)+z=x+(y+z),\text{其中}z\in H\\
&\quad \text{存在零元素}0,\text{使得}x+0=x,\forall x\in H\\
&\quad \forall x\in H,\text{存在负元素}z\text{使得}x+z=0\\
&(2)\text{对标量}\alpha,\beta\in\mathbb{R},\text{向量}x,y\in H,\text{数乘结果}\alpha\cdot x\in H,\text{且}\\
&\quad (\alpha+\beta)\cdot x=\alpha\cdot x+\beta\cdot x\\
&\quad \alpha\cdot(x+y)=\alpha\cdot x+\alpha\cdot y\\
&\quad (\alpha\beta)\cdot x=\alpha\cdot(\beta\cdot x)\\
&\quad 1\cdot x=x
\end{aligned}$$

则称 H 为实数域 $\mathbb{R}$ 上的线性空间。

内积: 实数域 $\mathbb{R}$ 上的线性空间 H 中向量 x 和 y 的实值二元函数 $< x, y >$ 称为一个内积，如果满足如下条件：

$$
\begin{aligned}
&(1) < x, y > = < y, x >, \forall x, y \in H \\
&(2) < x + y, z > = < x, z > + < y, z >, \forall x, y, z \in H \\
&(3) < \alpha x, y > = \alpha < x, y >, \forall \alpha \in \mathbb{R}, x, y \in H \\
&(4) < x, x > \geq 0, \forall x \in H \\
&(5) < x, x > = 0 \iff x = 0.
\end{aligned}
$$

定义了内积的线性空间称为**内积空间**。n 维欧式空间 $\mathbb{R}^n$ 是内积空间，内积空间是 $\mathbb{R}^n$ 的推广。

从内积可以导出向量的模（长度、范数）：

$$
\|x\| = \sqrt{< x, x >}
$$

$\|x\| = 0$ 当且仅当 $x = 0$。

实数域上的内积空间的内积和内积对应的模总满足 Cauchy-Schwarz 不等式：

$$
| < x, y > | \leq \|x\| \, \|y\|
$$

等号成立当且仅当 $y = \alpha x$ 或 $x = \beta y$。证明参见 (Brockwell & Davis, 1987) P.44 §2.1 式 (2.1.4) 的证明。

内积导出的模满足三角不等式：

$$
\|x + y\| \leq \|x\| + \|y\|
$$

这可以用 Cauchy-Schwarz 不等式证明。

范数（模）的定义：实数域 $\mathbb{R}$ 上的线性空间 H 上的实值函数 $\| \bullet \|$ 称为一个范数（模），如果满足如下条件：

$$
\begin{aligned}
&(1) \|x\| \geq 0, \forall x \in H \\
&(2) \|x\| = 0 \iff x = 0, \forall x \in H \\
&(3) \|x + y\| \leq \|x\| + \|y\|, \forall x, y \in H \\
&(4) \|\alpha x\| = |\alpha| \, \|x\|, \forall \alpha \in \mathbb{R}, x \in H
\end{aligned}
$$

定义了范数（模）的线性空间称为度量空间或者赋范空间。

由内积导出的模满足以上的一般范数定义。

在定义了范数以后，可以定义空间中的元素极限。对 x_n, x，称 $\lim_{n\to\infty} x_n = x$，如果 $\lim_{n\to\infty} \|x_n - x\| = 0$。

如果 H 是内积空间，内积有如下的连续性：若 $x_n, y_n, x, y \in H$，且 $\lim_{n\to\infty} x_n = x$, $\lim_{n\to\infty} y_n = y$, 则

$$(1)\|x_n\| \to \|x\|,\ n当 \to \infty$$
$$(2) < x_n, y_n > \to < x, y >,\ 当n \to \infty$$

证明与前面关于 L^2 的证明相同。

附录 E

实变函数和泛函分析

参考：(郭懋正, 2005)

E.1 集合运算

用 $\mathbb{R}$ 表示实数域，$\overline{\mathbb{R}}$ 表示 $\mathbb{R} \cup \{+\infty, -\infty\}$。

对集合 X，用 2^X 表示 X 的所有子集组成的集合（集合族），称为 X 的**幂集**。

集合上极限：

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega：\omega\text{在无穷多个}A_n\text{中出现}\}$$

集合下极限：

$$\underline{\lim_{n\to\infty}} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \{\omega：\text{从某个}n\text{开始}\omega\text{属于后续所有的}A_n\}$$

直积（笛卡尔积）：对集合 A, B，$A \times B = \{(x, y) : x \in A, y \in B\}$。类似可定义 $X_1 \times X_2 \cdots \times X_n$，$X^n$，$X^T$（其中 T 是 $\mathbb{R}$ 的子集）。

上确界：对 $A \subset \mathbb{R}$，m 是 A 的一个上界，且对 A 的任意上界 m' 都有 $m \leq m'$，则称 m 为 A 的上确界，记为 $\sup A$。如果 A 没有有限的上界则令 $\sup A = +\infty$。类似定义下确界。

数列上极限：

$$\overline{\lim_{n\to\infty}}\, a_n = \lim_{n\to\infty} \sup_{m\geq n} a_m,$$

可以等于 $\pm\infty$。类似定义下极限。类似定义函数上极限和下极限。

集合中元素的个数，分为三种情况：

- 有限个；
- 无限可数个，称为可列个；
- 无限不可数个。

$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) : x_i \in \mathbb{R}, i = 1, 2, \dots, n\}$，模为

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

两个点 x, y 的距离为 $d(x, y) = \|x - y\|$。

序列 $\{x_n\} \subset \mathbb{R}^n$ 收敛到极限 x，定义为 $\lim_{n\to\infty} \|x_n - x\| = 0$。

$\mathbb{R}^n$ 中的点集 A 中点的分类：

- **内点**x：x 的一个邻域均属于 A。
- **边界点**x：$x \in A$ 且 x 的任意一个邻域均与 A^c 都有非空交集。其中，若 x 的一个邻域中仅有 x 属于 A 称 x 为孤立点。
- **聚点**x：存在 $x_n \in A$，$x_n \neq x$ 使得 $\lim_{n\to\infty} x_n = x$。

A 与 A 的所有内点组成的集合的并集称为 A 的**闭包**，记为 $\overline{A}$。

$\mathbb{R}^n$ 中的**闭集**：A 包含其所有的句点。即 A 对极限运算封闭。有限多个闭集的并集是闭集，任意多个闭集的交集是闭集。闭包是闭集。

$\mathbb{R}^n$ 中的**开集**：所有点都是内点的集合。

生成的 代数：设全集为 X, $\mathcal{A}$ 是 X 的一些子集组成的集合族，称包含 $\mathcal{A}$ 的所有的 代数的交集为 $\mathcal{A}$ 生成的 代数。

Borel 代数：$\mathbb{R}^n$ 中所有开集所生成的 代数称为 Borel 代数，记为 $\mathscr{B}^n$，其中的集合称为 Borel 集。Borel 集合的余集、可列并、可列交、上极限、下极限运算结果均是 Borel 集。

$\mathbb{R}$ 中任一非空开集是至多可数个互不相交的开区间的并集。$\mathbb{R}^n$ 中任意非空开集是至多可数个互不相交的 n 为半开矩体的并集。半开矩体是 $[a_1, b_1) \times [a_2, b_2) \times \dots [a_n, b_n)$ 这样的集合。

康拓集（Cantor set）：从 $[0,1]$ 中每次中间的一段，保留端点，重复操作，极限情况下得到的集合。是非空有界闭集，每个点都是聚点，没有内点，无穷不可数。

函数 $f:\mathbb{R}^n \to \mathbb{R}$ 连续，当且仅当对任意 $\lambda \in \mathbb{R}$，$\{x : f(x) > \lambda\}$ 都是开集；当且仅当对任意 $\lambda \in \mathbb{R}$，$\{x : f(x) \le \lambda\}$ 都是闭集。

E.2 序关系

二元关系：设 X 和 Y 是两个集合，E 是 $X \times Y$ 的子集，称 E 是 $X \times Y$ 的一个二元关系，$X \times X$ 上的二元关系成为 X 的二元关系。对 $x \in X, y \in Y$，如果 $(x,y) \in E$，就称 x 和 y 满足二元关系 E。

例如，$X = \mathbb{R}$，$E = \{(x,y) : x \le y\}$，则二元关系 E 就是小于等于关系。

序关系：设 X 是集合, 如果 E 为满足如下序公理的二元关系：

- （1）自反性：$\forall x \in X$, x 与 x 自身满足 E；
- （2）反对称：如果 a 与 b 满足 E, b 与 a 也满足 E，则必有 $a = b$;
- （3）传递性：若 a 与 b 满足 E, b 与 c 满足 E，则 a 与 c 满足 E。

就称 E 为 X 上的序关系，x 和 y 满足 E，记为 $x \preceq y$，即 $x \preceq y$ 当且仅当 $(x,y) \in E$。称 $(X, \preceq)$ 为偏序空间，简称序空间，并将赋予了序关系的集合 X 称为有序集或偏序集。

$x \preceq y$ 可以等价地写成 $y \succeq x$。如果 $x \preceq y$ 而且 $x \ne y$，可以写成 $x \prec y$ 或 $y \succ x$。

$\le$ 和 $\ge$ 是序关系，$=$ 也是序关系。$<$ 和 $>$ 不是序关系。

全序空间：给定偏序空间 $(X, \preceq)$，如果 $\forall x, y \in X$，在 $x \preceq y$ 和 $y \preceq x$ 两者中至少有一个成立，则称 $(X, \preceq)$ 是全序空间，X 是全序集合。

全序集合中任意两个元素 x, y 之间的关系必为 $x \prec y$, $x = y$，$x \succ y$ 三者之一，这是实数比较 $\le$ 的自然推广。

一般的偏序则不需要任意两个元素可比，典型代表是 X 为集合，2^X 为 X 的所有子集组成的集合族，2^X 中 $\subset$ 是一个序关系，但不构成全序，两个子集可以没有包含关系。

对于偏序空间 $(X,\preceq)$，设 $B\subset X$，如果 $(B,\preceq)$ 构成全序空间，则称 B 为 X 的全序子集。设 A 是 X 的全序子集，且 X 中没有以 A 为真子集的全序子集，就称 A 为 X 的最大全序子集。这个定义不保证最大全序子集唯一。

Zorn 公理：每个偏序集有最大全序子集。

给定偏序集 X 和全序子集 $B\subset X$，若 $a\in X$ 使得对 $\forall x\in B$，都有 $x\preceq a$，则称 a 为 B 的一个上界。若 $b\in B$ 是 B 的一个上界，称 b 为 B 的最大元。对 $a\in X$，如果对 X 中任意能与 a 比较的 x 都有 $x\preceq a$，称 a 为 X 的一个极大元。

Zorn 引理：设 X 为非空偏序集，若 X 的每个全序子集在 X 中有上界，则 X 必有极大元。

E.3 勒贝格测度

对 $\mathbb{R}^n$ 中的开矩体

$$I=[a_1,b_1]\times[a_2,b_2]\times...[a_n,b_n],$$

定义其体积为

$$|I|=\prod_{i=1}^{n}(b_i-a_i).$$

设 $E\subset\mathbb{R}^n$，$\{I_k\}$ 是开矩体列，$E\subset\bigcup_{k=1}^{\infty}I_k$（称 $\{I_k\}$ 是 E 的一个覆盖），则

$$m^*(E)=\inf\left\{\sum_{k=1}^{\infty}|I_k|:\ I_k\text{是开矩体},E\subset\bigcup_{k=1}^{\infty}I_k\right\}$$

称为 E 的外测度。外测度非负，单调，满足次可加性 $m^*(E)(\bigcup_{k=1}^{\infty}E_k)\leq\sum_{k=1}^{\infty}m^*(E_k)$，平移不变性 $m^*(E+\{x\})=m^*(E),\forall x\in\mathbb{R}^n$。

$E+\{x\}=\{y+x:y\in E\}$。

勒贝格可测：设 $E\subset\mathbb{R}^n$，若对 $\mathbb{R}^n$ 到任意子集 T 都有

$$m^*(T)=m^*(T\cap E)+m^*(T\cap E^c),$$

则称 E 是勒贝格可测集，简称为可测集，记为 $\mathscr{M}$ 或 $\mathscr{M}_n$。当 E 可测时记 $m(E)=m^*(E)$，称为 E 的勒贝格测度。

所有开矩体可测，且 $m(I) = |I|$。

所有可测集构成 $\mathbb{R}^n$ 的一个 代数。Borel 集都可测，即 $\mathscr{B}^n \subset \mathscr{M}^n$，若 E 可测，必存在 Borel 集 F, G 使得 $F \subset E \subset G$ 且 $m(F) = m(E) = m(G)$，所以可测集与 Borel 集可以近似等同看待。

设 X 是集合，$\mathscr{F}$ 是 X 中子集组成的 代数，称 $(X, \mathscr{F})$ 为**可测空间**；如果集合函数 $\mu : \mathscr{F} \to [0, \infty]$ 满足：

(1) $\mu(\emptyset) = 0$;

(2) 若 $A_n \in \mathscr{F}$, $n = 1, 2, ...$ 互不相交，则有 可加性

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n),$$

则称 μ 为 $(X, \mathscr{F})$ 上的一个测度，称 $(X, \mathscr{F}, \mu)$ 为测度空间。

如果对任意的零测集 A(满足 $\mu(A) = 0$)，其子集都可测，称这样的测度空间是**完备**的。

如果 $\mu(X) < \infty$，称 μ 为有限测度。如果 $X = \bigcup_{n=1}^{\infty} A_n$，其中 $\{A_n\}$ 互不相交，且 $\mu(A_n) < \infty$，称 μ 为 有限测度。

测度对单独增集合列和单调减集合都有连续性：

$$\lim_{n\to\infty} \mu(A_n) = \mu(\lim_{n\to\infty} \mu(A_n).$$

关于 $x \in \mathbb{R}^n$ 的一个论断 $S(x)$ 在可测集 E 上几乎处处成立，记为 $S(x)$ a.e.$[E]$，定义为存在集合 $E_0 \subset E$，$m(E_0) = 0$，使得 $S(x)$ 对所有 $x \in E \backslash E_0$ 成立。注意不需要 $S(x)$ 在 E_0 上都不成立，所以不需要测度完备。当 $E = \mathbb{R}^n$ 时记 $S(x)$ a.e.。

E.4 勒贝格可测函数

考虑从 $\mathbb{R}^n$ 到 $\overline{\mathbb{R}}$ 的广义的函数，并定义 $0 \times \infty = 0$。

可测函数：设 E 是 $\mathbb{R}^n$ 中的可测集，$f : E \to \overline{\mathbb{R}}$，如果 $\forall t \in \mathbb{R}$，$\{x \in \mathbb{R}^n : x \in E, f(x) > t\}$ 都是可测集，则称 f 为 E 上的（勒贝格）可测函数。记 $\mathscr{M}(E)$ 为 E 上的可测函数的全体。

可测与如下条件等价：$\forall t \in \mathbb{R}$，$\{x \in \mathbb{R}^n : x \in E, f(x) > t\}$ 都可测；$\{x \in \mathbb{R}^n : x \in E, f(x) > t\}$ 都可测；$\{x \in \mathbb{R}^n : x \in E, f(x) \geq t\}$ 都可测；$\{x \in \mathbb{R}^n : x \in E, f(x) \leq t\}$ 都可测。

简单函数：设 $E = \bigcup_{i=1}^m E_i$, E_i 互不相交且可测，$\alpha_i \in \mathbb{R}$，

$$\psi(x) = \sum_{i=1}^m \alpha_i I_{E_i}(x)$$

称为 E 上的简单函数。当每个 E_i 是矩体时称为阶梯函数。可测集 E 上的简单函数都可测。

可测集 E 上非负函数 $f : E \to \overline{\mathbb{R}}$ 可测的充分必要条件是存在 E 上的非负简单函数列 $\{\psi_k(x)\}$ 使得 $\psi_k(x) \leq \psi_{k+1}(x)$, $\forall x \in E$，且 $\lim_{k\to\infty} \psi_k(x) = f(x)$, $\forall x \in E$。

可测函数的运算：若 f, g 是 E 上的可测函数，$c \in \mathbb{R}$，则 $cf(x)$ 可测，$|f(x)|$ 可测，$f(x)g(x)$ 可测；$f(x) + g(x)$ 和 $f(x)/g(x)$ 在其有定义的集合上可测（除去 $\infty - \infty$, 除以 0 等无定义的点）。

可测集 E 上的连续函数都是可测函数。

对可测集 E 上的可测函数列 $f_k(x)$，$\sup_k f_k(x)$, $\inf_k f_k(x)$，$\overline{\lim}_k f_k(x)$, $\underline{\lim}_k f_k(x)$ 都可测；如果 $\lim_k f_k(x) = f(x)$ 存在则 $f(x)$ 可测；如果存在零测集 E_0 使得

$$\lim_{k\to\infty} f_k(x) = f(x), \ x \in E \backslash E_0,$$

称 $f_k(x)$ 几乎处处收敛到 $f(x)$，这时 $f(x)$ 可测。

可测集 E 上的可测函数 $f(x)$ 可测，当且仅当 $f^+(x)$ 和 $f^-(x)$ 都可测。

$$f^+(x) = \max\{f(x), 0\}, \quad f^-(x) = \max\{-f(x), 0\}.$$

如果 $f(x) : \mathbb{R} \to \mathbb{R}$ 连续，$g(x) : E \to \mathbb{R}$ 可测，$E \subset \mathbb{R}^n$ 可测，则 $f(g(x))$ 是 E 上可测函数。

如果可测集 E 上的函数 f 和 g 几乎处处相等，则它们同时可测或同时不可测。

一般的测度空间 $(X, \mathcal{F}, \mu)$ 上的可测函数有类似的定义和性质。

E.5 可测函数极限

一致收敛：设 E 为点集，对 $E \to \mathbb{R}$ 的 f_k 和 f，$\forall \varepsilon > 0$, 存在 K 使得 $k \geq K$ 时 $\forall x \in E$ 有 $|f(x_k) - f(x)| < \varepsilon$。

几乎一致收敛：设 E 为 $\mathbb{R}^n$ 的可测集，$\forall \delta > 0$，存在 $E_\delta \subset E$ 使得 $m(E \backslash E_\delta) < \delta$，在 E_δ 上 f_k 一致收敛到 f。

一致收敛推出几乎一致收敛。

点点收敛：设 E 为点集，对 $E \to \mathbb{R}$ 的 f_k 和 f，若 $\lim_{k\to\infty} f_k(x) = f(x)$, $\forall x \in E$，称 $f_k(x)$ 在 E 上点点收敛到 $f(x)$。如果 f_k 都可测，则 f 也可测。

一致收敛推出点点收敛。

几乎处处收敛：设 E 为 $\mathbb{R}^n$ 中点集，$f(x), f_k(x)$ 是 $E \to \overline{\mathbb{R}}$ 函数，存在 $Z \subset E$ 使得 $m(Z) = 0$, $\forall x \in E \backslash Z$ 有 $\lim_k f_k(x) = f(x)$，则称 $f_k(x)$ 几乎处处收敛到 $f(x)$，记为 $\lim_k f_k(x) = f(x)$, a.e.[E]。如果 E 可测，f_k 都可测，则 f 也可测。

几乎一致收敛推出几乎处处收敛。反过来，如果 $m(E) < \infty$，f_k 和 f 都可测且几乎处处有限，则几乎处处收敛推出几乎一致收敛，这是叶戈罗夫定理。

依测度收敛：设 E 为可测集，f, f_k 为 $E \to \overline{\mathbb{R}}$ 函数，可测且几乎处处有限，$\forall \varepsilon > 0$ 有

$$\lim_{k\to\infty} m(\{x \in E : |f_k(x) - f(x)| > \varepsilon\}) = 0.$$

如果两个函数都是同一个函数列的依测度极限，则这两个函数几乎处处相等。即在几乎处处意义下依测度收敛的极限是唯一的。

几乎处处收敛推出依测度收敛：设 E 可测，$m(E) < \infty$，$f_k(x)$ 是 E 上几乎处处有限的可测函数列，$f_k(x)$ 几乎处处收敛到 $f(x)$，则 $f(x)$ 几乎处处有限且可测，并且 $f_k(x)$ 依测度收敛到 $f(x)$。

依测度收敛有子列几乎处处收敛（Riesz 定理）：设 E 可测，$m(E) < \infty$，$f_k(x)$ 是 E 上几乎处处有限的可测函数列，$f_k(x)$ 依测度收敛到几乎处处有限的可测函数 $f(x)$，则存在子列 $f_{k_i}(x)$ 使得 $f_{k_i}(x)$ 几乎处处收敛到 $f(x)$。

依测度收敛当且仅当是依测度基本列：

$$\lim_{k,j\to\infty} m(\{x \in E : |f_k(x) - f_j(x)| > \varepsilon\}) = 0.$$

对可测集 $E \subset \mathbb{R}^n$ 上的几乎处处有限的可测函数 f，必存在 $\mathbb{R}^n$ 上的连续函数列 $\{g_k(x)\}$ 使得 $g_k(x)$ 在 E 上几乎处处收敛到 $f(x)$。

一般的测度空间 $(X, \mathcal{F}, \mu)$ 上的可测函数收敛性类似。

E.6 勒贝格积分

对可测集 $E \subset \mathbb{R}^n$ 的非负可测简单函数 $h(x) : E \to \overline{\mathbb{R}}$:

$$h(x) = \sum_{j=1}^{m} \alpha_j I_{E_j}(x),$$

定义勒贝格积分为

$$\int_E h(x)\, dx = \sum_{j=1}^{m} \alpha_j m(E_j).$$

对可测集 $E \subset \mathbb{R}^n$ 的非负可测函数 $f(x) : E \to \overline{\mathbb{R}}$，定义

$$\int_E f(x)\, dx = \sup \left\{ \int_E h(x)\, dx : \ h(x)\text{是}E\text{上非负可测简单函数，且}h(x) \leq f(x), \forall x \in E \right\}.$$

若 $\int_E f(x)\, dx < \infty$，称 $f(x)$ 在 E 上是勒贝格可积的，记为 $f \in L(E)$。

如果 E 可测，A 是 E 的可测子集，$f(x)$ 是 E 上非负可测函数，则

$$\int_A f(x)\, dx = \int_E f(x) I_A(x)\, dx.$$

Levi 定理（非负单调收敛定理）：设 $\{f_k(x)\}$ 是可测集 E 上的非负可测函数列，$f_k(x) \leq f_{k+1}(x) (\forall x \in E)$，且 $\lim_k f_k(x) = f(x), \forall x \in E$，则

$$\lim_{k\to\infty} \int_E f_k(x)\, dx = \int_E f(x)\, dx.$$

对可测集 E 上可测函数 $f(x)$，若 $f^+(x)$ 和 $f^-(x)$ 至少一个可积，则定义

$$\int_E f(x)\, dx = \int_E f^+(x)\, dx - \int_E f^-(x)\, dx.$$

如果 $f^+(x)$ 和 $f^-(x)$ 都可积，则称 $f(x)$ 在 E 上勒贝格可积，记 $f(x) \in L(E)$。

$f(x)$ 可积当且仅当 $|f(x)|$ 可积，且这时

$$\left| \int_E f(x)\, dx \right| \leq \int_E |f(x)|\, dx.$$

设 E 是可测集：

- 如果 $f = g$ a.e.$[E]$，$f \in L(E)$ 则 $g \in L(E)$ 且 $\int_E f(x)\,dx = \int_E g(x)\,dx$；
- 如果 $f(x)$ 可测，而 $g(x)$ 非负可积，$f(x) \le g(x), \forall x \in E$，则 $f(x)$ 可积且 $\left|\int_E f(x)\,dx\right| \le \int_E g(x)\,dx$；
- 若 $m(E) < \infty$，则 E 上任意几乎处处有界的可测函数是可积的；
- 若 $f, g \in L(E)$, 且 $f(x) \le g(x)$, a.e.$[E]$，则 $\int_E f(x)\,dx \le \int_E g(x)\,dx$；
- 若 $f, g \in L(E)$, $\alpha, \beta \in \mathbb{R}$，则（线性性质）

$$\int_E (\alpha f(x) + \beta g(x))\,dx = \alpha \int_E f(x)\,dx + \beta \int_E g(x)\,dx.$$

如果 $m(E) = 0$，则 E 上的可测函数 $f(x)$ 都可积且 $\int_E f(x)\,dx = 0$。

若可测集 E 上的可测函数 $f(x)$ 满足 $\int_E |f(x)|\,dx < \infty$，则 $|f(x)| < \infty$, a.e.$[E]$。

若可测集 E 上的可测函数 $f(x)$ 满足 $\int_E |f(x)|\,dx = 0$，则 $f(x) = 0$, a.e.$[E]$。

积分的绝对连续性：对可测集 E 上的可测函数 $f(x)$，$\forall \varepsilon > 0$，必存在 $\delta > 0$，E 的子集 A 只要满足 $m(A) < \delta$ 就一定有

$$\left|\int_A f(x)\,dx\right| \le \int_A |f(x)|\,dx < \varepsilon.$$

可积函数的连续函数逼近：对可测集 E 上的可测函数 $f(x)$，存在 $\mathbb{R}^n$ 上有紧支集的连续函数列 $g_k(x)$ 使得

$$\begin{aligned} &(1)\ \lim_{k\to\infty} \int_E |f(x) - g_k(x)|\,dx = 0; \\ &(2)\ \lim_{k\to\infty} g_k(x) = f(x),\ \text{a.e}[E]. \end{aligned}$$

与黎曼积分的关系：设 $f(x)$ 为闭区间 $[a, b]$ 上的有界函数，若 $f(x)$ 在 $[a, b]$ 黎曼可积，必勒贝格可积且积分相等；$f(x)$ 在 $[a, b]$ 黎曼可积当且仅当 $f(x)$ 的不连续点为零测集。

一般的测度空间 $(X, \mathcal{F}, \mu)$ 上也可以类似定义勒贝格积分。

E.7 极限与积分的交换

非负单调收敛定理（Levi 定理）：若可测集 E 上的非负可测函数 $f_k(x)$ 满足 $f_k(x) \le f_{k+1}(x), \forall x, \forall k$，$f(x) = \lim_{k\to\infty} f_k(x)$，a.e.$[E]$，则

$$\lim_{k\to\infty}\int_E f_k(x)\,dx = \int_E f(x)\,dx.$$

勒贝格基本定理：设 $\{f_k(x)\}$ 是可测集 E 上的非负可测函数列，则

$$\sum_{k=1}^{\infty}\int_E f_k(x)\,dx = \int_E \sum_{k=1}^{\infty} f_k(x)\,dx.$$

Fatou 引理：设 $\{f_k(x)\}$ 是可测集 E 上的非负可测函数列，则

$$\int_E \varliminf_{k\to\infty} f_k(x)\,dx \le \varliminf_{k\to\infty}\int_E f_k(x)\,dx.$$

控制收敛定理：设 E 为可测集，$\{f_k(x)\}$ 是 E 上可测函数列，$f_k(x)$ 在 E 上几乎处处收敛到 $f(x)$ 或者依测度收敛到 $f(x)$，若存 E 上非负可积函数 $g(x)$ 使得 $f_k(x) \le g(x)$, a.e. $[E]$，$\forall k$，则 $f_k(x)$ 和 $f(x)$ 在 E 上可积且

$$\lim_{k\to\infty}\int_E f_k(x)\,dx = \int_E f(x)\,dx.$$

有界收敛定理：若 $m(E) < \infty$，$\{f_k(x)\}$ 是 E 的可测函数列，存在常数 M 使得

$$|f_k(x)| \le M, \forall x \in E, \forall k,$$

且 $f_k(x)$ 几乎处处或者依测度收敛到函数 $f(x)$，则 $f_k(x)$ 和 $f(x)$ 可积且

$$\lim_{k\to\infty}\int_E f_k(x)\,dx = \int_E f(x)\,dx.$$

逐项积分：设 E 为 $\mathbb{R}^n$ 的可测集，$f_k(x) \in L(E)$，若

$$\sum_{k=1}^{\infty}\int_E |f_k(x)|\,dx < \infty,$$

则级数 $\sum_{k=1}^{\infty} f_k(x)$ 几乎处处收敛，记为 $f(x)$, 则 $f(x) \in L(E)$ 且

$$\sum_{k=1}^{\infty}\int_E f_k(x)\,dx = \int_E f(x)\,dx.$$

参变积分：设 E 为 $\mathbb{R}^n$ 的可测集，$f(x,y): E\times[a,b]\to\overline{\mathbb{R}}$, 对每个 $y\in[a,b]$，$f(\cdot,y)$ 是 E 上的可积函数，则

$$\phi(y)=\int_E f(x,y)\,dx$$

是定义于 $[a,b]$ 的有限实值函数，称为区间 $[a,b]$ 上的参变积分。

对参变积分：

(1) 若存在 $g(x)\in L(E)$ 使得

$$f(x,y)\leq g(x),\ \forall x\in E,\ y\in[a,b].$$

且 $\lim_{y\to y_0} f(x,y)$ 在 E 上 a.e. 收敛，就有

$$\lim_{y\to y_0}\phi(y)=\int_E \lim_{y\to y_0} f(x,y)\,dx;$$

(2) 若存在 $g(x)\in L(E)$ 使得

$$f(x,y)\leq g(x),\ \forall x\in E,\ y\in[a,b].$$

且对 a.e.$[E]$ 的 x 函数 $f(x,\cdot)$ 在 y_0 处连续，则 $\phi(y)$ 在 y_0 处连续；

(3) 若 $\frac{\partial f(x,y)}{\partial y}$ 存在，存在 $g(x)\in L(E)$ 使得 $|\frac{\partial f(x,y)}{\partial y}|\leq g(x)$, $\forall x\in E, y\in[a,b]$，则

$$\phi'(y)=\int_E \frac{\partial f(x,y)}{\partial y}\,dx.$$

E.8 重积分与累次积分

设 A 和 B 分别为 $\mathbb{R}^p$ 和 $\mathbb{R}^q$ 的可测集，则 $A\times B$ 为 $\mathbb{R}^{p+q}$ 的可测集且

$$m(A\times B)=m(A)m(B).$$

Tonelli 定理：设 $f(x,y):\mathbb{R}^{p+q}\to\overline{\mathbb{R}}$ 非负可测，则

(1) 对几乎处处的 $x\in\mathbb{R}^p$，$f(x,\cdot)$ 是 $\mathbb{R}^q$ 的非负可测函数；

(2) $F_f(x)=\int_{\mathbb{R}^q} f(x,y)\,dy$ 在 $\mathbb{R}^p$ 上几乎处处有定义且非负可测；

(3) 重积分与累次积分相等，累次积分次序可交换：

$$\int_{\mathbb{R}^{p+q}} f(x,y)\,dxdy = \int_{\mathbb{R}^p} dx \int_{\mathbb{R}^q} f(x,y)\,dy = \int_{\mathbb{R}^q} dy \int_{\mathbb{R}^p} f(x,y)\,dx.$$

Fubini 定理：设 $f(x,y) : \mathbb{R}^{p+q} \to \overline{\mathbb{R}}$ 可积，则

(1) 对几乎处处的 $x \in \mathbb{R}^p$，$f(x,\cdot)$ 是 $\mathbb{R}^q$ 的可积函数；

(2) $F_f(x) = \int_{\mathbb{R}^q} f(x,y)\,dy$ 在 $\mathbb{R}^p$ 上几乎处处有定义且可积；

(3) 重积分与累次积分相等，累次积分次序可交换：

$$\int_{\mathbb{R}^{p+q}} f(x,y)\,dxdy = \int_{\mathbb{R}^p} dx \int_{\mathbb{R}^q} f(x,y)\,dy = \int_{\mathbb{R}^q} dy \int_{\mathbb{R}^p} f(x,y)\,dx.$$

注意，即使累次积分都存在且相等，$f(x,y)$ 也不一定可积；Fubini 定理在使用时，为了验证 $f(x,y)$ 是否可积，可以先用 Tonelli 定理计算 $|f(x,y)|$ 的累次积分从而判断可积性。

E.9 L^p 空间

E.9.1 定义

本节中设 E 为 $\mathbb{R}^n$ 中的可测集。

设 $f(x)$ 为 E 上的可测函数，$1 \le p < \infty$，记

$$\|f\|_p = \left(\int_E |f(x)|\,dx\right)^{\frac{1}{p}},$$

称 $\|f\|_p$ 为 f 的 L^p 范数或 L^p 模。令

$$L^p(E) = \left\{f : f\text{为}E\text{上的可测函数且}\|f\|_p < \infty\right\},$$

称 $L^p(E)$ 为 E 上的 L^p 空间。

设 $f(x)$ 为 E 上的可测函数，如果存在正数 M 使得 $|f(x)| \le M$, a.e.$[E]$，称 $f(x)$ 为本性有界的，取所有这样的 M 的下确界，记为 $\|f\|_\infty$，用 $L^\infty(E)$ 表示 E 上所有的本性有界函数组成的集合。对 $f \in L^\infty(E)$，有

$$\|f\|_\infty = \inf_{\triangle\subset E,\, m(\triangle)=0}\ \sup_{x\in E\setminus\triangle} |f(x)|.$$

对一般的测度空间 $(X,\mathscr{F},\mu)$，也可以类似定义其 L^p 空间 $L^p(X,\mathscr{F},\mu)$。

下面设 $1\le p,q\le\infty$，$\frac{1}{p}+\frac{1}{q}=1$，称 p,q 为共轭指数，$p=1$ 时 $q=\infty$，$p=2$ 时 $q=2$，$1<p<\infty$ 时 $q=\frac{p}{p-1}$。

Hölder 不等式：对 $f\in L^p(E)$, $g\in L^q(E)$，有

$$\left|\int_E f(x)g(x)\,dx\right|\le\int_E|f(x)g(x)|\,dx=\|fg\|_1\le\|f\|_p\,\|g\|_q.$$

当 $p=q=2$ 时即 Schwarz 不等式。

若 $m(E)<\infty$，则对 $p_1<p_2$ 有 $\|f\|_{p_2}<\infty$ 推出 $\|f\|_{p_1}<\infty$，$\lim_{p\to\infty}\|f\|_p=\|f\|_\infty$。

E.9.2 距离空间

$L^p(E)$ 对加法运算和数乘封闭，构成实数域上的向量空间（线性空间）。

对一般的实数域上的向量空间（线性空间）X，函数 $\|\cdot\|:X\to\mathbb{R}$ 称为 X 的一个范数，如果它满足如下的三条：

- （1）齐次性：$\|\alpha x\|=|\alpha|\cdot\|x\|,\ \forall\alpha\in\mathbb{R},x\in X$；
- （2）三角不等式：$\|x+y\|\le\|x\|+\|y\|,\ \forall x,y\in X;$
- （3）正定性：$\|x\|\ge 0$，$\forall x\in X$，$\|x\|=0$ 当且仅当 x 是向量空间 X 的零元素。

将 $L^p(E)$ 中几乎处处相等的函数看成空间中的同一个元素，则 $\|\cdot\|$ 满足上述的范数的条件，$\|f\|_p=0$ 当且仅当 $f(x)=0$, a.e.$[E]$。

距离空间设 X 为实数域上的向量空间，如果函数 $d:X\times X\to\mathbb{R}$ 满足如下性质：

- （1）正定性：$d(x,y)\ge 0,\ \forall x,y\in X,\ d(x,y)=0$ 当且仅当 $x=y;$
- （2）对称性：$d(x,y)=d(y,x),\ \forall x,y\in X$；
- （3）三角不等式：$d(x,y)\le d(x,z)+d(z,y),\ \forall x,y,z\in X$。

对 $1\le p\le\infty$，在 $L^p(E)$ 中定义

$$d(f,g)=\|f-g\|_p,$$

则 $L^p(E)$ 构成距离空间。有了距离以后就可以定义开集、闭集等拓扑，定义序列极限。

对 $L^p(E)$ 中的序列 $\{f_m\}$ 和函数 f，如果

$$\|f_m - f\| \to 0,\ m \to \infty,$$

则称 f_m 依 L^p 意义收敛到 f，记为 $f_m \overset{L^p}{\to} f$，对 $p = 2$ 称为均方收敛。

$$\|f_m - f_k\|_p \to 0,\ m, k \to \infty,$$

称 $\{f_m\}$ 为 $L^p(E)$ 的基本列。如果 $\{f_m\}$ 收敛 $f \in L^p(E)$，$\{f_m\}$ 必为基本列；反之，如果 $\{f_m\}$ 为基本列，必存在 $f \in L^p(E)$ 使得 $f_m \overset{L^p}{\to} f$。称 $L^p(E)$ 是完备的距离空间。完备的距离空间中的聚点都在空间内。

如果考虑闭区间 $[a, b]$ 上黎曼可积函数组成的距离空间，就不是完备的。这说明勒贝格积分比黎曼积分更优秀。

L^p 收敛有如下性质：

- (1) 唯一性：如果序列收敛到两个函数 f 和 g 则 $f = g$，a.e.$[E]$；
- (2) 如果 $f_m \overset{L^p}{\to} f$ 则 $\|f_m\|_p \to \|f\|$；
- (3) 极限对加法和数乘封闭，若 $f_m \overset{L^p}{\to} f$，$g_m \overset{L^p}{\to} g$，$\alpha \in \mathbb{R}$，则 $f_m + g_m \overset{L^p}{\to} f + g$, $\alpha f_m \overset{L^p}{\to} \alpha f$。

E.9.3 收敛之间的关系

L^p 收敛推出依测度收敛：对 $1 \le p < \infty$，如果 $f_m \overset{L^p}{\to} f$，则 f_m 依测度收敛到 f。

对 $1 \le p \le \infty$，若 $f_m \overset{L^p}{\to} f$，则存在子列几乎处处收敛到 f。

控制收敛定理：设 $1 \le p \le \infty$, f_m, f 为 E 上的可测函数，f_m 几乎处处收敛到 f 或者依测度收敛到 f，且存在非负可测函数 $g \in L^p(E)$ 使得

$$|f_m(x)| \le g(x),\ \text{a.e.}[E],$$

则 $f_m \overset{L^p}{\to} f$。

E.9.4 可分性

一个距离空间 (X, d) 中，如果 $F \subset G \subset X$，且对任意 $x \in G$ 存在 $\{x_n\} \subset F$ 使得 $d(x_n, x) \to 0$，称 F 在 G 中稠密，或称 G 中的元素可以用 F 中的元素逼近。一个距离空间如果有可数的稠密子集，称其为可分距离空间。

对 $f \in L^p(E)$，可以用简单可测函数逼近；可以用 $\mathbb{R}^n$ 中有紧支集的连续函数逼近；可以用 $\mathbb{R}^n$ 中有紧支集的阶梯函数逼近。

对 $1 \le p < \infty$，$L^p(E)$ 有可数稠密子集，从而是可分距离空间，可以用紧支集阶梯函数组成可数稠密子集。

$L^\infty(E)$ 不一定可分。

E.10 L^2 内积空间

E.10.1 定义

本节中设 E 是 $\mathbb{R}^n$ 的可测集。

范数（模）仅能引出长度、距离、收敛的概念，但是无法引出角度、正交（垂直）的概念。$L^2(E)$ 的模与 $\mathbb{R}^n$ 的模最接近，也可以定义内积构成内积空间。对 $f, g \in L^2(E)$，定义

$$\langle f, g\rangle = \int_E f(x)g(x)\,dx,$$

称为 f 和 g 的内积。由 Schwarz 不等式可知内积是有限实数值：

$$|\langle f, g\rangle| \le \|f\|_2 \cdot \|g\|_2.$$

对一般的实数域上的向量空间 X，函数 $K(x, y)$ $X \times X \to \mathbb{R}$ 如果满足以下要求：

- (1) 双线性：$K(x, y)$ 分别关于 x 和 y 是线性函数；
- (2) 对称性：$K(x, y) = K(y, x)$;
- (3) 正定性：$K(x, x) \ge 0$, $K(x, x) = 0$ 当且仅当 $x = 0$，

则称 $K(x, y)$ 是 X 的一个**内积**，记为 $\langle x, y\rangle$。$L^2(E)$ 中的 $\langle f, g\rangle$ 满足内积的这三条要求。定义了内积的向量空间称为**内积空间**。

从内积可以定义模 $\|x\| = \sqrt{\langle x, y\rangle}$，从模可以定义距离 $d(x, y) = \|x - y\|$，所以内积空间是距离空间。

E.10.2 性质

简记 $\|f\|_2$ 为 $\|f\|$。

弱收敛：对 $f_m, f \in L^2(E)$，如果对任意 $g \in L^2(E)$ 都有

$$\langle f_m, g\rangle \to \langle f, g\rangle,$$

即 $\langle f_m - f, g\rangle \to 0$，则称 f_m 弱收敛到 f。L^2 收敛必弱收敛；如果 f_m 弱收敛到 f 且 $\|f_m\| \to \|f\|$ 则 $f_m \overset{L^p}{\to} f$。

弱收敛可以用稠密子集判断：对 $f_m, f \in L^2(E)$，如果 $\{\|f_m\|\}$ 有界，且存在 $L^2(E)$ 的稠密子集 G 使得对任意 $g \in G$ 都有 $\langle f_m, g\rangle \to \langle f, g\rangle$，则 f_m 弱收敛到 f。

正交：在内积空间中，如果 $\langle x, y\rangle = 0$ 则称 x 与 y 正交（或垂直），记为 $x \perp y$。对 $L^2(E)$ 即

$$\int_E f(x)g(x)\,dx = 0.$$

如果 f 和 g 都是非负函数则内积等于 0 当且仅当 $f = g = 0$, a.e.$[E]$。

标准正交系：设函数列 $\{\phi_a\}_{a\in A}$ 中任意两个元素正交，则称 $\{\phi_a\}_{a\in A}$ 为正交系；如果进一步还有 $\|\phi_a\| = 0,\ \forall a \in A$，称 $\{\phi_a\}_{a\in A}$ 为标准正交系。

$L^2(E)$ 中的任意标准正交系必为可数集合。

例：$L^2[-\pi, \pi]$ 的一个标准正交系为三角函数列

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, ...$$

设 $\{\phi_k(x)\}_{k=1}^{\infty}$ 为 $L^2(E)$ 的一个标准正交系，对 $f \in L^2(E)$，称

$$c_k = \langle f, \phi_k\rangle = \int_E f(x)\phi_k(x)\,dx$$

为 f（关于正交系 $\{\phi_k\}$）的傅立叶 (Fourier) 系数，称

$$\sum_{k=1}^{\infty} c_k\phi_k(x)$$

为 f（关于正交系 $\{\phi_k\}$）的傅立叶级数，记为 $f \sim \sum_{k=1}^{\infty} c_k\phi_k(x)$，注意傅立叶级数不一定收敛，即使收敛其极限也不一定等于 $f(x)$。

Bessel 不等式：在以上定义中

$$\sum_{k=1}^{\infty} c_k^2 \le \|f\|^2.$$

Riesz-Fischer 定理：设 $\{\phi_k(x)\}_{k=1}^{\infty}$ 为 $L^2(E)$ 的一个标准正交系，实数列 $\{c_k\}$ 平方可和（即 $\sum_{k=1}^{\infty} c_k^2 < \infty$），则级数 $\sum_{k=1}^{\infty} c_k\phi_k(x)$ 在 $L^2(E)$ 中收敛（均方收敛），记极限为 f，则 $c_k = \langle f, \phi_k \rangle$，且 $\|f\|^2 = \sum_{k=1}^{\infty} c_k^2$。

设 $\{\phi_k(x)\}_{k=1}^{\infty}$ 为 $L^2(E)$ 的一个标准正交系，令

$$G = \left\{ \sum_{k=1}^{\infty} c_k\phi_k(x) : \sum_{k=1}^{\infty} c_k^2 < \infty \right\},$$

则 $G \subset L^2(E)$；对任意 $f \in L^2(E)$，令

$$\tilde{f} = \sum_{k=1}^{\infty} c_k\phi_k(x),$$

其中 $\{c_k\}$ 为 f 的傅立叶系数，则 $\tilde{f} \in G$ 且

$$\|f - \tilde{f}\| = \min_{g \in G} \|f - g\|.$$

标准正交基：设 $\{\phi_k(x)\}_{k=1}^{\infty}$ 为 $L^2(E)$ 的一个标准正交系，如果对任意 $f \in L^2(E)$ 都存在 k 使得 $\langle f, \phi_k \rangle \neq 0$，则称 $\{\phi_k(x)\}_{k=1}^{\infty}$ 为 $L^2(E)$ 的一个**完全（完备）标准正交系**。如果对任意 $f \in L^2(E)$，其傅立叶级数均收敛到 f，即

$$f = \sum_{k=1}^{\infty} \langle f, \phi_k \rangle \, \phi_k(x),$$

则称 $\{\phi_k(x)\}_{k=1}^{\infty}$ 为 $L^2(E)$ 的一个**标准正交基**。对 $L^2(E)$，完全标准正交系是标准正交基。

$L^2[-\pi, \pi]$ 中三角函数列是标准正交基。

如果 $\{\phi_k(x)\}_{k=1}^{\infty}$ 为 $L^2(E)$ 的一个标准正交系，则以下条件等价：

- (1) $\{\phi_k(x)\}_{k=1}^{\infty}$ 是标准正交基；
- (2) $\{\phi_k(x)\}_{k=1}^{\infty}$ 的有限线性组合构成的集合在 $L^2(E)$ 中稠密；
- (3) $\{\phi_k(x)\}_{k=1}^{\infty}$ 是完全的；

- (4) Parseval 等式成立：

$$\|f\|^2 = \sum_{k=1}^{\infty} \langle f, \phi_k \rangle^2, \ \forall f \in L^2(E);$$

- (5) 内积等式：

$$\langle f, g \rangle = \sum_{k=1}^{\infty} \langle f, \phi_k \rangle \cdot \langle g, \phi_k \rangle, \ \forall f, g \in L^2(E).$$

如果 $L^2(E)$ 中有可数的线性无关的函数列，使得其有限线性组合在 $L^2(E)$ 中稠密，可以用 Gram-Schmidt 正交化方法将其转化为标准正交基。

例：$L^2[-\pi, \pi]$ 上三角函数列是正交基；$L^2[-, 1]$ 的 Legendre 正交多项式是正交基；$L^2(-\infty, \infty)$ 的 Hermite 多项式是标准正交基。

E.11 卷积和傅立叶变换

E.11.1 复值函数

傅立叶变换用复值函数表示比较方便，将实值函数推广到复值函数。设 $u(x), v(x)$ 是 $\mathbb{R}^n$ 的可测集 E 上的几乎处处有限的可测函数，令 $f(x) = u(x) + iv(x)$，称 $f(x)$ 为几乎处处有限的复值可测函数，若 $\int_E |f(x)|\, dx < \infty$，称 $f(x)$ 在 E 上可积，E 上可积函数的全体仍记为 $L(E)$，

$$\left| \int_E f(x)\, dx \right| \le \int_E |f(x)|\, dx.$$

$L^2(E)$ 由几乎处处有限的复值函数且

$$\|f\|_p = \left(\int_E |f(x)|^p\, dx \right)^{1/p} < \infty$$

的函数组成。其数乘用复数域。

对范数 $\|\cdot\|_p$，齐次性中的 α 可以是复数；对复数域上 $L^2(E)$ 的内积，定义为：

$$\langle f, g \rangle = \int_E f(x) \overline{g(x)}\, dx,$$

满足如下性质：

- （1）共轭双线性：关于第一自变量是线性函数，关于第二自变量是共轭线性的；
- （2）共轭对称性：$\langle f,g\rangle=\overline{\langle g,f\rangle}$；
- （3）正定性与实数域情形相同。

实数域上 $L^2(E)$ 中的弱收敛性、正交性、傅立叶级数展开、标准正交基等性质都适用于复数域上的 $L^2(E)$ 空间。性质中的傅立叶系数 $\{c_k\}$ 变成复数列，$\sum_{k=1}^\infty c_k^2$ 改为 $\sum_{k=1}^\infty |c_k|^2$，稠密集 G 的定义中系数 $\{c_k\}$ 也变成复数列，有标准正交基时内积等式改为 $\langle f,g\rangle=\sum_{k=1}^\infty\langle f,\phi_k\rangle\,\overline{\langle g,\phi_k\rangle}$。

E.11.2 卷积

卷积：设 f,g 是 $\mathbb{R}^n$ 上的实值或复值的可测函数，如果积分

$$\int_{\mathbb{R}^n} f(x-y)g(y)\,dy$$

存在，就称此积分为函数 f 和 g 的卷积，记为 $f*g(x)$。卷积有对称性：

$$f*g=g*f.$$

Yang 不等式：若 $1\le p\le\infty$, $f\in L^p(\mathbb{R}^n)$，$g\in L^1(\mathbb{R}^n)$，则 $f*g\in L^p(\mathbb{R}^n)$，且

$$\|f*g\|_p\le\|f\|_p\|g\|_1.$$

特别地，当 f,g 都可积时，卷积存在且 $\|f*g\|_1\le\|f\|_1\|g\|_1$。所以卷积运算在 $L(\mathbb{R}^n)$ 空间中是封闭的运算。

平均连续性：对 $1\le p<\infty$ 和 $f\in L^p(\mathbb{R}^n)$，有

$$\lim_{\Delta x\to 0}\|f(x+\Delta x)-f(x)\|=0.$$

设 $f,g\in L^2(\mathbb{R}^n)$，则 $f*g(x)$ 是 $\mathbb{R}^n$ 上的有界连续函数。

设 $1\le p<\infty$, $f\in L^p(\mathbb{R})$, 非负函数 $\phi\in L^1(\mathbb{R}^n)$, $\|\phi\|_1=1$, $\phi_\varepsilon(x)=\varepsilon^{-n}\phi(\frac{x}{\varepsilon})$，则

$$\lim_{\varepsilon\to 0}\|\phi_\varepsilon*f-f\|_p=0.$$

这里 $\phi_\varepsilon*f\in L^p(\mathbb{R}^n)$, 是 f 的逼近。如果 $f\in L^2(\mathbb{R}^n)$, $\phi\in L^2(\mathbb{R}^n)$ 且 $\|\phi\|_1=1$，则 $\phi_\varepsilon*f\in L^2(\mathbb{R}^n)$ 是连续有界函数，$\phi_\varepsilon*f$ 可以看成是对不连续的 f 进行了局部平均，变成了连续函数。

如果 $\phi_\varepsilon(x)$ 有任意阶偏导数，则 $\phi_\varepsilon * f$ 也有任意阶偏导数，从而 $L^p(E)$ 中的函数可以被有任意阶偏导数的函数逼近，记 $C^\infty(\mathbb{R}^n)$ 为 $\mathbb{R}^n$ 中有任意阶偏导数的函数全体，则 $C^\infty(\mathbb{R}^n)$ 限制在 E 上的函数的集合在 $L^p(E)$ 中稠密。

E.11.3 傅立叶变换

傅立叶变换：对任意 $f \in L(\mathbb{R}^n)$，令

$$\hat{f}(t) = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(x) e^{-it\cdot x}\, dx,\ \forall t \in \mathbb{R}^n,$$

其中 $t \cdot x = \sum_{i=1}^n t_i x_i$，称 $\hat{f}$ 是 f 的傅立叶 (Fourier) 变换，可记为 $\mathscr{F}f$，显然 $\mathscr{F}$ 是线性映射：

$$\mathscr{F}(\alpha f + \beta g) = \alpha \mathscr{F} f + \beta \mathscr{F} g.$$

因为 $|e^{-it\cdot x}| = 1$，所以 $\hat{f}$ 定义右侧的积分存在有限，且

$$|\hat{f}(t)| \leq (2\pi)^{-\frac{n}{2}} \|f\|_1,\ \forall t \in \mathbb{R}^n,$$

由控制收敛定理可以证明 $\hat{f}(t)$ 连续，所以 $f \in L(\mathbb{R}^n)$ 时 $\hat{f}$ 为有界连续函数，还可以证明

$$\lim_{t\to\infty} \hat{f}(t) = 0.$$

引入记号

$$e_t(x) = e^{it\cdot x},$$

将其看成是 $x \in \mathbb{R}^n$ 的函数。对于多重指标 $\alpha = (\alpha_1, \dots, \alpha_n)$，其中 $\alpha_1, \dots, \alpha_n$ 都是非负整数，记 $|\alpha| = \sum_{i=1}^n \alpha_i$，记

$$\begin{aligned} D^\alpha &= \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}, \\ D_\alpha &= i^{-|\alpha|} D^\alpha = \left(\frac{1}{i}\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{1}{i}\frac{\partial}{\partial x_n}\right)^{\alpha_n}, \end{aligned}$$

对 $\mathbb{R}^n$ 上的多项式

$$P(z) = \sum_\alpha c_\alpha z^\alpha = \sum_\alpha c_\alpha z_1^{\alpha_1} \cdots z_n^{\alpha_n},$$

记

$$P(D) = \sum_\alpha c_\alpha D_\alpha, \quad P(-D) = \sum_\alpha (-1)^{|\alpha|} c_\alpha D_\alpha,$$

则

$$P(D)e_t = P(t)e_t.$$

对 f 和 $x, y \in \mathbb{R}^n$，记平移算子

$$\tau_x f(y) = f(y-x).$$

$L(\mathbb{R}^n)$ 上的傅立叶变换有如下性质：

- (1) $\mathcal{F}(\tau_s f) = e_{-s}\hat{f}$;
- (2) $\mathcal{F}(e_s f) = \tau_s \hat{f}$;
- (3) $\mathcal{F}(f * g) = \hat{f}\hat{g}$;
- (4) 对 $\lambda > 0$, $\mathcal{F}(f(\frac{x}{\lambda}))(t) = \lambda^n \hat{f}(\lambda t)$.

引入记号：

- $C(\mathbb{R}^n)$ 为 $\mathbb{R}^n$ 上连续函数的全体（实值或者复值）。
- $C_c(\mathbb{R}^n)$ 为 $\mathbb{R}^n$ 上有紧支集的连续函数的全体。
- $C^\infty(\mathbb{R}^n)$ 为 $\mathbb{R}^n$ 上光滑函数的全体，记有任意阶偏导数的函数的全体。
- $C_c^\infty(\mathbb{R}^n)$ 为 $\mathbb{R}^n$ 上有紧支集的光滑函数的全体。

速降函数：若 $f \in C^\infty(\mathbb{R}^n)$，对任意 $N = 0, 1, 2, ...$，有

$$\sup_{|\alpha| \le N} \sum_{x \in \mathbb{R}^n} (1 + |x|^2)^N |D^\alpha f(x)| < \infty,$$

则称 f 为速降函数，这些函数的全体记为 $\mathcal{S}(\mathbb{R}^n)$。速降函数的任意阶偏导数乘以任意多项式都可积。

$\mathcal{S}(\mathbb{R}^n)$ 在 $L^p(\mathbb{R}^n)$（$1 \le p < \infty$）中稠密。$C_c^\infty(\mathbb{R}^n)$ 在 $L^p(\mathbb{R}^n)$（$1 \le p < \infty$）中稠密。

算子 $\mathscr{F}$ 是速降函数空间 $\mathcal{F}(\mathbb{R}^n)$ 上一个一一的线性的满映射，且保持 $L^2(\mathbb{R}^n)$ 模不变。对 $f, g \in \mathcal{S}(\mathbb{R}^n)$, 多项式 P, 多重指数 α，$\mathcal{F}$ 有如下性质：

- (1) $P \cdot f$, $f \cdot g$，$f * g$, $D_\alpha f$ 都属于 $\mathcal{S}(\mathbb{R}^n)$。
- (2) $\mathcal{F}(P(D)f) = P \cdot \hat{f}$, $\mathcal{F}(P \cdot f) = P(-D)\hat{f}$;
- (3) $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$。
- (4) $\mathcal{F}(f \cdot g) = \hat{f} * \hat{g}$。

速降函数的逆傅立叶变换定理：

(1) 若 $f\in\mathcal{S}(\mathbb{R}^n)$，则

$$f(x)=(2\pi)^{-\frac{n}{2}}\int_{\mathbb{R}^n}\hat{f}(t)e^{ix\cdot t}\,dt.$$

(2) $\mathcal{F}$ 是 $\mathcal{S}(\mathbb{R}^n)$ 一一的线性的满映射，且 $\mathcal{F}^4$ 为恒等变换。

(3) 若 $f,\hat{f}\in L(\mathbb{R}^n)$，则

$$f(x)=(2\pi)^{-\frac{n}{2}}\int_{\mathbb{R}^n}\hat{f}(t)e^{ix\cdot t}\,dt,\ \text{a.e.}$$

可以将 $\mathcal{S}(\mathbb{R}^n)$ 的一一变换 $\mathcal{F}$ 拓展到 $L^2(\mathbb{R}^n)$：存在 $L^2(\mathbb{R}^n)$ 的一个一一的线性满映射 Ψ，使得

$$\Psi f=\mathcal{F},\ \forall f\in\mathcal{S}(\mathbb{R}^n),$$

且

$$\|\Psi f\|_2=\|f\|_2,\ \forall f\in L^2(\mathbb{R}^n),$$

Ψ^4 是恒等映射，称 Ψ 为 $L^2(\mathbb{R}^n)$ 上的傅立叶变换。

对 $f,g\in L^2(\mathbb{R}^n)$ 有

$$\begin{aligned}\langle\Psi f,\Psi g\rangle=&\langle f,g\rangle=\langle\Psi^{-1}f,\Psi^{-1}g\rangle,\\ \langle f,\Psi g\rangle=&\langle\Psi^{-1}f,g\rangle,\\ \int_{\mathbb{R}^n}(\Psi f)(x)g(x)\,dx=&\int_{\mathbb{R}^n}f(x)(\Psi g)(x)\,dx.\end{aligned}$$

注意 $\langle f,g\rangle=\int_{\mathbb{R}^n}f(x)\overline{g(x)}\,dx$。

E.12 距离空间

E.12.1 定义

设 X 为实数域上的向量空间，如果函数 $d:X\times X\to\mathbb{R}$ 满足如下性质：

- (1) 正定性：$d(x,y)\geq 0,\ \forall x,y\in X,\ d(x,y)=0$ 当且仅当 $x=y$;
- (2) 对称性：$d(x,y)=d(y,x),\ \forall x,y\in X$；
- (3) 三角不等式：$d(x,y)\leq d(x,z)+d(z,y),\ \forall x,y,z\in X$。

则称 $d(x,y)$ 为 X 上一个距离，(X,d) 称为一个距离空间。

从距离，可以引出球邻域、余集、开集、闭集、矩体、Borel 代数、稠密性等概念。X 的子集中 $\emptyset$ 和全集既是开集又是闭集。

例：$\mathbb{R}^n$，欧式距离。

例：$C[a,b]$ 表示定义在 $[a,b]$ 上的连续函数全体，定义距离为

$$d(f,g)=\max_{x\in[a,b]}|f(x)-g(x)|.$$

例：$C[0,\infty)$ 表示定义在 $[0,\infty)$ 上所有连续函数构成的线性空间。在其中定义距离

$$\rho(f,g)=\sum_{n=1}^{\infty}\frac{1}{2^n}\max_{t\in[0,n]}(|f(t)-g(t)|\wedge 1).$$

例：$L^p(E)$, $1\leq\infty$，距离

$$d(f,g)=\|f-g\|_p.$$

例：l^p 空间（$1\leq p<\infty$），其中元素为复数列 (或实数列)$\{x_k\}$，满足

$$\sum_{k=1}^{\infty}x_k^p<\infty.$$

距离空间 (X,d) 中的序列 $\{x_k\}$ 和元素 x 如果满足

$$d(x_k,x)\to 0,\ k\to\infty,$$

称 $\{x_k\}$ **收敛**到 x，记为 $\lim_{k\to\infty}x_k=x$，称 $\{x_k\}$ 为收敛列。极限唯一。

收敛是按距离定义的，比如在 $C[a,b]$ 中收敛是函数的一致收敛。

如果 $\lim_{k,m\to\infty}d(x_m,x_k)=0$，称 $\{x_k\}$ 为**基本列**，如果 X 的所有基本列都收敛到 X 中，则称 X 为完备的距离空间。

空间 $C[a,b]$，$L^p(E)$，l^2 是完备的；$[a,b]$ 上的多项式全体按 $\max_{x\in[a,b]}|f(x)-g(x)|$ 距离是不完备的，$[a,b]$ 上黎曼可积函数全体按 L^1 距离是不完备的。

对 $A\subset X$，如果存在元素互不相同的序列 $\{x_k\}\subset k$ 使得 $\lim_k x_k=x$，称 x 为 A 的一个**聚点**或者极限点。

同构映射：给定距离空间 (X, d_1) 和 (Y, d_2)，设 T 是 X 到 Y 的映射，如果 $\forall x_1, x_2 \in X$ 都有 $d_1(x_1, x_2) = d_2(Tx_1, Tx_2)$，就称 T 是**等距映射**；如果 T 还是满映射，称 X 与 Y **等距同构**。

两个等距同构的距离空间，其一切与距离相联系的性质（开集、闭集、收敛、连续等）都是一样的，可以视为是等价的。当 T 是 X 到 Y 的等距映射时，X 与 $TX \subset Y$ 等距同构，可以将 (X, d_1) 看成是 (Y, d_2) 的子距离空间。

稠密子集：设 (X, d) 是一个距离空间，集合 $E \subset X$ 满足：$\forall x \in X$，存在 $\{x_m\} \subset E$，使得 $d(x_m, x) \to 0$，就称 E 为 X 的稠密子集。

例：$C[a, b]$ 中多项式空间 $P[a, b]$ 稠密。$L^p(E)$ 中 $C_c^\infty(E)$(具有紧支集和任意阶偏导数的函数) 稠密。

距离空间的完备化：任给距离空间 (X_0, d_0)，必存在一个完备距离空间 (X, d)，使得 X_0 与 X 的一个稠密子集 X' 等距同构，且在等距映射之间不区分的意义下，(X, d) 是唯一确定的，这时 (X, d) 称为 (X_0, d_0) 的完备化空间。

例：在 L^1 距离下，$[a, b]$ 上黎曼可积函数空间不完备，完备化空间是 $L^1[a, b]$；$[a, b]$ 上黎曼可积函数空间在 $L^1[a, b]$ 中稠密。

E.12.2 列紧性和可分性

对于 $\mathbb{R}^n$ 空间，元素个数无穷的有界子集一定有收敛子列，一般的距离空间则不一定，比如 $L^2[a, b]$ 空间有可数的标准正交基，有界但没有收敛子序列。

有界：设 (X, d) 为距离空间，$A \subset X$，如果存在 $x_0 \in X$，$r > 0$，使得 $A \subset B(x_0, r) = \{x : d(x, x_0) < r\}$，则称 A 为有界子集。

列紧：设 (X, d) 为距离空间，$A \subset X$，如果 A 中任意点列中有一个收敛子列，极限取值于 X 中，称 A 为列紧的。如果这些收敛子列都收敛到 A 中的点，称 A 为**自列紧**的。如果 X 是列紧的，则是自列紧的，称 X 为列紧空间。

列紧空间必为完备距离空间，其中任意子集都是列紧集，任意闭子集都是自列紧集。

对 $\mathbb{R}^n$，有界集都是列紧集，自列紧集等价于有界闭集。

距离空间中子集有界不一定列紧，对完备距离空间，列紧等价于如下的完全有界。

完全有界：给定距离空间 (X,d)，$M \subset X$。设 $N \subset M$, $\varepsilon > 0$，若 $\forall x \in M$, 总存在 $y \in N$ 使得 $x \in B(y,\varepsilon)$，则称 N 是 M 的一个 ε 网；如果 N 还是一个有穷集，称 N 是 M 的一个有穷 ε 网。如果 $\forall \varepsilon$, M 都存在有穷 ε 网，称 M 是**完全有界**的。

Hausdorff 定理：对距离空间 (X,d) 及 $M \subset X$，

- （1）如果 M 在 X 中列紧，则 M 完全有界；
- （2）若 X 是完备空间，M 完全有界，则 M 列紧。

可分距离空间：距离空间如果存可数稠密子集，就称为可分的。

完全有界的距离空间是可分的。列紧必完全有界，所以列紧的距离空间是可分的。

例：$C[a,b]$ 和 $C[0,\infty)$ 是完备可分距离空间。

紧集：设 M 是距离空间 (X,d) 的一个子集，$\Sigma = \{G_l, I \in I\}$ 是 X 的开集族，如果 $M \subset \bigcup_{l\in I} G_l$，则称 Σ 是 M 的一个开覆盖；如果 M 的任意开覆盖 $\Sigma = \{G_l, I \in I\}$ 中都有有限子覆盖，即存在 $G_{l_1}, G_{l_2}, G_{l_m}$ 使得 $M \subset \bigcup_{i=1}^{m} G_{l_i}$，就称 M 为紧致集或紧集。

距离空间的紧集（紧致集）和自列紧集等价。

设 M 是一个紧距离空间（自列紧，也是紧致），距离为 ρ，$C(M)$ 是 M 上定义的所有的实值或者复制的连续函数组成的集合，定义

$$d(f,g) = \max_{x\in M} |f(x) - g(x)|,\ \forall f,g \in C(M),$$

则 $d(\cdot,\cdot)$ 是距离，且 $C(M)$ 是完备距离空间。

Arzela-Ascoli 定理：集合 $F \subset C(M)$ 是一个列紧集（紧致集）的充分必要条件为 F 一致有界且等度连续。

$F \subset C(M)$ 一致有界，是存在 $K \in \mathbb{R}$ 使得

$$|f(x)| \le K,\ \forall x \in M,\ \forall f \in F.$$

$F \subset C(M)$ 等度连续，是 $\forall \varepsilon > 0$，$\exists \delta = \delta(\varepsilon) > 0$，使得 $\forall f \in F$，以及 $x_1, x_2 \in M$，只要 $\rho(x_1,x_2) < \delta$，就有

$$|f(x_1) - f(x_2)| < \varepsilon.$$

$F\subset M$ 等度连续是其中每个函数都一致连续，且其中的 δ 不依赖于 $f\in F$。

关于 l^2 中子集列紧的条件，$L^p(E)$ 值子集列紧的条件，见 (郭懋正, 2005) pp 217-218。

E.12.3 连续映射

连续映射：设 (X,d_1) 和 (Y,d_2) 是两个距离空间，映射 $T:X\to Y$，$\forall\varepsilon>0$, $\exists\delta=\delta(x_0,\varepsilon)>0$，使得 $\forall x\in X$,

$$d_1(x,x_0)<\delta\Longrightarrow d_2(Tx,Tx_0)<\varepsilon,$$

就称 T 在点 x_0 处连续；如果 T 在 X 的所有点都连续，称 T 是 X 上的连续映射。

T 是 X 上的连续映射，当且仅当 $\forall x_0\in X$ 和 $\{x_m\}\subset X$，有

$$\lim_{m\to\infty}d_1(x_0,x_m)=0\Longrightarrow\lim_{m\to\infty}d_2(Tx_0,Tx_m)=0.$$

压缩映射：设 (X,d) 为距离空间，T 是 X 到 X 的映射，如果存在 $0<a<1$，使得

$$d(Tx_1,Tx_2)\le a\,d(x_1,x_2),\ \forall x_1,x_2\in X,$$

就称 T 是 X 上的一个压缩映射。

压缩映射原理（Banach 不动点定理）：设 (X,d) 是一个完备距离空间，T 是 X 到自身的一个压缩映射，则存在唯一的 x^* 使得 $Tx^*=x^*$，称 x^* 为 T 的不动点。

例：$X=[0,1]$，$f(x)$ 是 $[0,1]$ 上可微函数，满足

$$f(x)\in[0,1],\ \forall x\in[0,1],$$
$$|f'(x)|\le a<1,\ \forall x\in[0,1],$$

则 f 是 X 上一个压缩映射。存在 $x^*\in[0,1]$ 使得 $f(x^*)=x^*$。

E.13 Hilbert 空间

E.13.1 定义

记 $\mathbb{K}$ 为实数域 $\mathbb{R}$ 或者复数域 $\mathbb{C}$。

内积空间：设 X 是数域 $\mathbb{K}$ 上的线性空间，如果 X 上的一个二元函数 $g(x,y):X\times X\to\mathbb{K}$ 满足：

- (1) $g(\alpha x+\beta y,z)=\alpha g(x,z)+\beta g(y,z)$;
- (2) $g(x,\alpha y+\beta z)=\overline{\alpha}g(x,y)+\overline{\beta}g(x,z)$;
- (3) $g(x,x)\geq 0$, $g(x,x)=0\Leftrightarrow x=0$;
- (4) $g(x,y)=\overline{g(y,x)}$

对任意 $x,y,z\in X$, $\alpha,\beta\in\mathbb{K}$ 成立，则称 $g(x,y)$ 为 X 上的一个内积，称 $(X,g(\cdot,\cdot))$ 为一个内积空间。如果仅有一个内积，经常简记为 $\langle x,y\rangle$。

定义中 (1) 和 (2) 称为共轭双线性，当 $\mathbb{K}=\mathbb{R}$ 时称为双线性。(3) 称为正定性。(4) 称为对称性。如果 (3) 中仅要求 $g(x,x)\geq 0$，就称 $g(x,y)$ 为一个半内积，称 $X,g(\cdot,\cdot)$ 为半内积空间。

例 1. 欧式空间 $\mathbb{R}^n$ 定义内积

$$\langle x,y\rangle=\sum_{i=1}^{n}x_iy_i.$$

欧式空间 $\mathbb{C}^n$ 定义内积

$$\langle x,y\rangle=\sum_{i=1}^{n}x_i\bar{y}_i.$$

例 2. l^2 空间定义内积

$$\langle x,y\rangle=\sum_{i=1}^{\infty}x_i\bar{y}_i.$$

例 3. $C^k(\bar{\Omega})$ 是 $\mathbb{R}^n$ 中有界闭集 $\bar{\Omega}$ 上有 k 阶连续偏导数的函数的全体，定义内积

$$\langle f,g\rangle=\sum_{|\alpha|\leq k}\partial^\alpha f(x)\overline{\partial^\alpha g(x)},\ \forall f,g\in C^k(\bar{\Omega}).$$

例 4. 设 $(X,\mathscr{F},\mu)$ 为测度空间，$L^2(X,\mathscr{F},\mu)$ 是 $(X,\mathscr{F},\mu)$ 所有平方可积的复值函数全体，即其中的函数满足

$$\|f\|_2=\left(\int_X|f(x)|^2\,dx\right)^{1/2}<\infty,$$

两个函数在空间中相等定义为依 μ 几乎处处相等。定义内积

$$\langle f,g\rangle=\int_X f(x)\overline{g(x)}\,dx,$$

则 $(X, \langle\cdot,\cdot\rangle)$ 构成内积空间。

特别地，对概率测度空间 $(\Omega, \mathscr{F}, P)$，$L^2(\Omega, \mathscr{F}, P)$ 为其中二阶矩有限的复值随机变量全体，即

$$L^2(\Omega, \mathscr{F}, P) = \{\xi : \xi\text{是}(\Omega, \mathscr{F}, P)\text{上的复值随机变量，且}E(|\xi|^2) < \infty\},$$

这是一个线性空间，定义内积

$$\langle \xi, \eta \rangle = \int_\Omega \xi(\omega)\overline{\eta(\omega)}\, dP(\omega) = E(\xi\bar{\eta}),\ \forall \xi, \eta \in L^2(\Omega, \mathscr{F}, P).$$

内积导出的范数：对内积空间 $(X, \langle\cdot,\cdot\rangle)$，定义

$$\|x\| = \langle x, x \rangle^{1/2},\ \forall x \in X,$$

则 $\|\cdot\|$ 是 X 上的范数，称为内积 $\langle\cdot,\cdot\rangle$ 导出的范数（模）。

Cauchy-Schwarz 不等式：设 $\|\cdot\|$ 是内积 $\langle\cdot,\cdot\rangle$ 导出的范数，则

$$|\langle x, y \rangle| \le \|x\|\,\|y\|,\ \forall x, y \in X,$$

且等号成立当且仅当 x, y 线性相关。

实际上，Cauchy-Schwarz 不等式可以推广。

共轭范数：对内积空间 $(X, \langle\cdot,\cdot\rangle)$，设 $\|\cdot\|$ 是一个范数，定义

$$\|x\|^* = \sup_{\|y\|=1} \langle x, y \rangle,$$

则 $\|\cdot\|^*$ 也是一个范数，称为 $\|\cdot\|$ 的共轭范数。

对 $p \ge 1, q \ge 1$，若 $\frac{1}{p} + \frac{1}{q} = 1$，则 $\|\cdot\|_p$ 和 $\|\cdot\|_q$ 在 $\mathbb{R}^n$ 中互为共轭范数，特别地，$\|\cdot\|_2$ 和 $\|\cdot\|_2$ 互为共轭范数。另外，$\|\cdot\|_1$ 和 $\|\cdot\|_\infty$ 互为共轭范数。

Hölder 不等式：对内积空间 $(X, \langle\cdot,\cdot\rangle)$，设 $p \ge 1, q \ge 1$，$\frac{1}{p} + \frac{1}{q} = 1$，则对任意 $x, y \in X$，有

$$|\langle x, y \rangle| \le \|x\|_p \|y\|_q,$$

等号成立当且仅当 $|y_i| = |x_i|^{p-1}, x_i y_i \ge 0,\ i = 1, 2, \dots, n$。证明见C.2.3。

内积导出的距离：有了导出的范数，定义

$$d(x, y) = \|x - y\|,\ \forall x, y \in X,$$

则 $d(\cdot,\cdot)$ 是 X 上的距离，称为内积 $\langle\cdot,\cdot\rangle$ 导出的距离，内积空间 $(X,\langle\cdot,\cdot\rangle)$ 也是距离空间。

Hilbert 空间：对内积空间 $(X,\langle\cdot,\cdot\rangle)$，设由内积导出的距离为 $d(\cdot,\cdot)$，如果 (X,d) 构成完备距离空间，则称 $(X,\langle\cdot,\cdot\rangle$ 为 Hilbert 空间。

例 1 的欧式空间 $\mathbb{R}^n$ 和 $\mathbb{C}^n$ 是 Hilbert 空间。例 2 的 l^2 是 Hilbert 空间。例 4 的 $L^2(X,\mathcal{F},\mu)$ 是 Hilbert 空间，特别地，$L^2(E)(E$ 为 $\mathbb{R}^n$ 中可测集) 是 Hilbert 空间；$L^2(\Omega,\mathscr{F},P)$ 是 Hilbert 空间。

$\mathbb{R}^n$ 中有界闭集 $\bar{\Omega}$ 上有 k 阶连续偏导数的函数全体 $C^k(\bar{\Omega})$ 不是完备距离空间。

例：概率空间 $(\Omega,\mathscr{F},P)$ 上所有二阶矩有限的实值随机变量的集合记作 L^2，L^2 是 Hilbert 空间，内积为 $< X,Y >= E(XY)$。0 元素是 a.s. 等于零意义上，两个元素相等也是 a.s. 相等意义上。来证明完备性。

完备性证明: 设 $\{X_n\}$ 为 L^2 的 Cauchy 列，则存在 $\{n\}$ 的子序列 n_k，使得当 $n,m\geq n_k$ 时，

$$\|X_n-X_m\|\leq 2^{-3k}$$

由切比雪夫不等式

$$P(|X_n-X_m|\geq 2^{-k})\leq 2^{2k}E(X_n-X_m)^2\leq 2^{-k}$$

由单调收敛定理得

$$\begin{aligned}
&E\sum_{k=1}^{\infty}I[|X_{n_{k+1}}-X_{n_k}\geq 2^{-k}]\\
=&\sum_{k=1}^{\infty}EI[|X_{n_{k+1}}-X_{n_k}\geq 2^{-k}]\\
=&\sum_{k=1}^{\infty}P(|X_{n_{k+1}}-X_{n_k}\geq 2^{-k})\\
\leq&\sum_{k=1}^{\infty}2^{-k}<\infty
\end{aligned}$$

从而

$$\sum_{k=1}^{\infty}I[|X_{n_{k+1}}-X_{n_k}|\geq 2^{-k}]<+\infty,\ \text{a.s.}$$

存在 $\Omega^*\subset\Omega$, $P(\Omega^*)=1$，使得 $\forall\omega\in\Omega^*$，存在 K_0 使得 $k\geq k_0$ 时 $|X_{n_{k+1}}-$

$X_{n_k}| < 2^{-k}$。于是 k 充分大时

$$|X_{n_{k+m}} - X_{n_k}| \le \sum_{j=1}^{m} |X_{n_{j+1}} - X_{n_j}| \le \sum_{j=1}^{m} 2^{-(k+j)} \le 2^{-k+1}$$

于是对每个 $\omega \in \Omega^*$，存在子序列 X_{n_k} 使得 $\{X_{n_k}\}$ 是实数基本列，存在极限 X。利用 Fatou 引理，

$$\begin{aligned} E(X_n - X)^2 =& E \lim_{k\to\infty} [X_n - X_{n_k}]^2 \\ \le& \underline{\lim}_{k\to\infty} E(X_n - X_{n_k})^2 \\ \to& 0,\ \text{当} n \to \infty \end{aligned}$$

由三角不等式，

$$\sqrt{EX^2} = \|X\| \le \|X_n - X\| + \|X_n\| < \infty$$

所以极限 $X \in L^2$。这就证明了 L^2 的完备性。因为 L^2 中的元素是在 a.s. 相等意义下的，所以 X 可以是 a.s. 可测的。

完备化：设内积空间 $(X, \langle\cdot,\cdot\rangle)$ 与导出的距离 d 不构成完备距离空间，则存在 Hilbert 空间 H，使得 H 上由内积导出的距离与 H 一起成为 (X, d) 的完备化距离空间，且

$$\langle x, y\rangle_X = \langle x, y\rangle_H,\ \forall x, y \in X.$$

所以，不完备的内积空间可以完备化成为 Hilbert 空间。$\mathbb{R}$ 上的 Hilbert 空间也可以嵌入到复数域上的 Hilbert 空间中。

例 3 的 $C^k(\bar{\Omega})$ 不完备，其完备化的 Hilbert 空间记为 $H^k(\bar{\Omega})$，称为索伯列夫（Sobolev）空间。

E.13.2 正交

设 H 为 Hilbert 空间，若 $< x, y >= 0$ 则称 x 和 y 正交，记作 $x \perp y$。设 A, B 是 H 的非空子集，如果 $\forall x \in A, y \in B$ 有 $x \perp y$，则称 A 与 B 正交，记作 $A \perp B$。

记

$$\text{span}(A) = \{\sum_{i=1}^{n} \alpha_i x_i : \alpha_i \in K, x_i \in A, i = 1, 2, \dots, n, n \ge 1\},$$

称 span(A) 为 A 的线性包。如果 $x \perp A$，则 $x \perp \text{span}(A)$。

记

$$A^{\perp} = \{x \in H : x \perp A\},$$

称为 A 的正交补，这是 H 的闭子空间，也是子 Hilbert 空间。

如果 $x_1, x_2, \dots, x_n$ 在 H 中两两正交，则

$$\|x_1 + x_2 + \cdots + x_n\|^2 = \|x_1\|^2 + \|x_2\|^2 + \cdots + \|x_n\|^2.$$

设 S 是 H 的子线性空间，如果 S 中的收敛序列的极限都在 S 中，称 S 为 H 的**闭子空间**，也称为子 Hilbert 空间。

设 S 是 H 的闭子空间。若 $\forall z \in S$ 都有 $x \perp z$，称 x 与 S 正交，记作 $x \perp S$。

记所有与 S 正交的元素组成集合为 $S^{\perp}$, 这也是 H 的闭子空间，称为 S 的**正交补空间**。$x \perp S$ 当且仅当 $x \in S^{\perp}$。

正交集：设 $\mathcal{E}$ 是 Hilbert 空间 H 的一个子集，如果 $\forall x, y \in \mathcal{E}$ 且 $x \neq y$ 均有 $x \perp y$，称 $\mathcal{E}$ 为正交集。如果进一步 $\forall x \in \mathcal{E}$ 有 $\|x\| = 1$，称 $\mathcal{E}$ 为标准正交集。如果 $\mathcal{E}^{\perp} = \{0\}$(其中 0 表示线性空间 H 的零元素)，即不存在 $x \neq 0$ 使得 $x \perp \mathcal{E}$，就称 $\mathcal{E}$ 是完备的正交集。

根据 Zorn 引理可以证明：

定理：非零 Hilbert 空间 H 中必存在完备正交集。

Bessel 不等式：设 $\mathcal{E} = \{e_\alpha : \alpha \in \Lambda\}$ 为 Hilbert 空间 H 的一个标准正交集，其中 Λ 是一个集合，则 $\forall x \in H$，有

$$\sum_{\alpha \in \Lambda} |\langle x, e_\alpha \rangle|^2 \leq \|x\|^2.$$

上式左边的求和中非零项至多有可数个。

正交基：设 $\mathcal{E} = \{e_\alpha : \alpha \in \Lambda\}$ 为 Hilbert 空间 H 的一个标准正交集，如果 $\forall x \in H$，有

$$x = \sum_{\alpha \in \Lambda} \langle x, e_\alpha \rangle e_\alpha,$$

则称 $\mathcal{E}$ 为 H 的一个标准正交基，简称为基，其中的 $\{\langle x, e_\alpha \rangle : \alpha \in \Lambda\}$ 称为 x 关于基 $\mathcal{E}$ 的 Fourier 系数。非零系数一定是至多可数的。

定理：设 $\mathcal{E}=\{e_\alpha:\alpha\in\Lambda\}$ 为 Hilbert 空间 H 的一个标准正交集，则以下三条命题等价：

- (1) $\mathcal{E}$ 是标准正交基；
- (2) $\mathcal{E}$ 是完备的；
- (3) Parseval 等式成立，即

$$\|x\|^2=\sum_{\alpha\in\Lambda}|\langle x,e_\alpha\rangle|^2.$$

推论：非零 Hilbert 空间 H 必有完备正交基。

E.13.3 投影

定理 E.1 (投影存在性)**.** *设 H 为 Hilbert 空间，S 是 H 的子 Hilbert 空间。*

(1) 对 $\forall y\in H$，存在唯一的 $x\in S$，使得

$$\|y-x\|=\inf_{z\in S}\|y-z\|$$

*称 x 是 y 在闭子空间 S 上的***投影***, 记作 $\mathrm{Proj}_S\, y$。*

(2) 对 $\forall y\in H$，$x\in S$ 是 y 在 S 上的投影当且仅当 $y-x\perp S$。

证明: (1) $d=\inf_{z\in S}\|y-z\|$ 必存在且 $d\geq 0$。存 $\{z_n\}\subset S$ 使得 $\|z_n-y\|\to d(n\to\infty)$。来证明 $\{z_n\}$ 是基本列。利用恒等式

$$\|x-y\|^2+\|x+y\|^2=2\|x\|^2+2\|y\|^2$$

可得

$$\begin{aligned}\|z_n-z_m\|^2=&\|(z_n-y)-(z_m-y)\|^2\\=&-\|z_n+z_m-2y\|^2+2\|z_n-y\|^2+2\|z_m-y\|^2\\=&-4\|\frac{z_n+z_m}{2}-y\|^2+2\|z_n-y\|^2+2\|z_m-y\|^2\\\leq&-4d+2\|z_n-y\|^2+2\|z_m-y\|^2\\\to&0\ (n,m\to\infty)\end{aligned}$$

所以 $\{z_n\}$ 是基本列，存在 $x\in H$ 使得 $\|z_n-x\|\to 0$。因为 S 是闭子空间所以 $x\in S$。由内积的连续性可得

$$\|y-x\|=\lim_{n\to\infty}\|y-z_n\|=d$$

再来证明唯一性。如果有 $x' \in S$ 使得 $\|y-x\| = d$，则

$$\begin{aligned}
0 \le & \|x - x'\|^2 \\
= & \|(x-y) - (x'-y)\|^2 \\
= & -\|x + x' - 2y\|^2 + 2\|x-y\|^2 + 2\|x'-y\|^2 \\
= & -4\|\frac{x+x'}{2} - y\|^2 + 2\|x-y\|^2 + 2\|x'-y\|^2 \\
\le & -4d + 2d + 2d = 0
\end{aligned}$$

即 $x' = x$。

(2) 先证明充分性。设 $x \in S$ 使得 $y - x \perp S$，则 $\forall z \in S$，有

$$\begin{aligned}
\|y-z\|^2 = & \|(y-x) + (x-z)\|^2 \\
= & \|y-x\|^2 + \|x-z\|^2 + 2 < y-x, x-z > \\
= & \|y-x\|^2 + \|x-z\|^2 \\
\ge & \|y-x\|^2
\end{aligned}$$

所以 x 是 y 在 S 的投影。

再来证明必要性。用反证法。设 $x \in S$ 使得

$$\|y-x\| = \inf_{z \in S} \|y-z\|$$

如果 $y - x \perp S$ 不成立，则存在 $z' \in S$ 使得 $a =< y-x, z' >\neq 0$，显然 $z' \neq 0$。令

$$x' = x + \frac{a}{\|z'\|^2} z$$

则 $x' \in S$，且

$$\begin{aligned}
\|y-x'\|^2 = & \|(y-x) + (x-x')\|^2 \\
= & \|(y-x) - \frac{a}{\|z'\|^2} z'\|^2 \\
= & \|y-x\|^2 + \frac{a^2}{\|z'\|^4}\|z'\|^2 - \frac{2a}{\|z'\|^2} < y-x, z' > \\
= & \|y-x\|^2 - \frac{a^2}{\|z'\|^2} \\
< & \|y-x\|^2
\end{aligned}$$

矛盾。定理证毕。

定理说明，如果需要用闭子空间 S 上的元素 x 最优地逼近 $y \in H$，$x = \operatorname{Proj}_S y$ 是这个问题的唯一的解。这里“最优逼近”是用 $\|x-y\|$ 作为两个元素的距离时距离最小的近似。最优逼近 x 的条件也可以写成

$$< y-x, z >= 0, \ \forall z \in S.$$

对 Hilbert 空间 H 和闭子空间 S，Proj_S 是从 H 的 S 的一个线性映射。记 I 为 H 上的恒等映射，则 $\forall y \in H$,

$$\|y\|^2 = \|\operatorname*{Proj}_S y\|^2 + \|(I - \operatorname*{Proj}_S)y\|^2$$

其中 $(I - \operatorname{Proj}_S)y = y - \operatorname{Proj}_S y$。

正交分解：$\forall y \in H$，存在唯一的分解

$$y = y_1 + y_2 = \operatorname*{Proj}_S y + (I - \operatorname*{Proj}_S)y$$

其中 $y_1 \in S$, $y_2 \in S^\perp$。显然 $y_1 = \operatorname{Proj}_S y$ 和 $y_2 = (I - \operatorname{Proj}_S)y$ 满足分解；如果还有 $x_1 \in S$ 和 $x_2 \in S^\perp$ 满足 $y = x_1 + x_2$，则有

$$[x_1 - \operatorname*{Proj}_S y] + [x_2 - (I - \operatorname*{Proj}_S)y] = 0$$

两边与 $x_1 - \operatorname{Proj}_S y$ 作内积得

$$\|x_1 - \operatorname*{Proj}_S y\|^2 + 0 = 0$$

所以 $x_1 = \operatorname{Proj}_S y$，即分解式唯一。记这样的分解为

$$y = y_1 \oplus y_2, \ y_1 \in S, y_2 \in S^\perp$$

还可以考虑更一般的距离问题。考虑 $\mathbb{R}^n$ 空间中的超平面

$$H = \{x \in \mathbb{R}^n : \ w^T x + b = 0\},$$

其中 w 是法向量，b 是标量，设 $p \ge 1$，令

$$d_p(y, H) = \inf_{x \in H} \|y - x\|_p,$$

称为 y 到超平面 H 的 d_p 距离，则可以证明

$$d_p(y, H) = \frac{|w^T x + b|}{\|w\|_q}.$$

E.13.4 Riesz 表示定理

线性泛函：从 Hilbert 空间 H 到数域 $\mathbb{K}$ 的映射 f 称为一个泛函。如果

$$f(\alpha x+\beta y)=\alpha f(x)+\beta f(y),\ \forall x,y\in H,\ \forall\alpha,\beta\in\mathbb{K},$$

则称 f 是 H 上的线性泛函。

对 $x_0\in H$，如果 $\|x_n-x_0\|\to 0$ 推出 $f(x_n)\to f(x_0)$，则称泛函 f 在点 x 处连续。如果 f 在所有 $x\in H$ 连续，则称 f 为连续泛函。

Hilbert 空间 H 的所有连续线性泛函组成的集合记为 H^*，这是一个线性空间。

定理：设 f 是 Hilbert 空间 H 上的一个线性泛函，则以下几条等价：

- （1）f 连续；
- （2）f 在某个点连续；
- （3）f 在 $x=0$ 处连续；
- （4）存在常数 c 使得 $\forall x\in H, |f(x)|\le c\|x\|$。

其中第 (4) 条成立时称 f 为有界线性泛函，这时令

$$\|f\|=\sup\{|f(x)|:\ \|x\|\le 1\}$$

为 f 的模。Hilbert 空间上的有界线性泛函等价于连续线性泛函，空间都是 H^*。

若 f 是 H 上的有界线性泛函（连续线性泛函），则

$$\begin{aligned}\|f\|&=\sup\{|f(x)|:\ \|x\|\le 1\}\\&=\sup\{|f(x)|:\ \|x\|=1\}\\&=\sup\{\frac{|f(x)|}{\|x\|}:\ x\ne 0\}\\&=\inf\{c>0:\ |f(x)|\le c\|x\|,\forall x\in H\}.\end{aligned}$$

设 $x_0\in H$，令

$$f_{x_0}(x)=\langle x,x_0\rangle,\ \forall x\in H,$$

则 f_{x_0} 是一个线性泛函，由

$$|f_{x_0}(x)|\le\|x_0\|\,\|x\|,$$

可知其为有界线性泛函，且 $\|f\| \leq \|x_0\|$，又 $x_0 \neq 0$ 时

$$f_{x_0}(\frac{x_0}{\|x_0\|}) = \langle \frac{x_0}{\|x_0\|}, x_0 \rangle = \|x_0\|$$

所以 $\|f_{x_0}\| = \|x_0\|$。

Riesz 表示定理：设 f 是 Hilbert 空间 H 上的有界线性泛函，则存在唯一 $x_f \in H$，使得

$$f(x) = \langle x, x_f \rangle,\ \forall x \in H$$

且 $\|f\| = \|x_f\|$。

这样，从 $f \in H^*$ 到 $x_f \in H$ 的映射定义了一个一一满映射，且为等距映射，所以 H^* 与 H 等距同构。

E.14 Hilbert 空间上的算子

泛函：设 $\mathbb{K}$ 为数域 $\mathbb{R}$ 或者 $\mathbb{C}$，X 和 Y 是 $\mathbb{K}$ 上的线性空间，D 是 X 的子空间，$A : D \to Y$ 是映射，记 $x \in D$ 的映射结果为 Ax，D 称为 A 的定义域，有时记为 $D(A)$。$R(A) = \{Ax : x \in D\}$ 称为 A 的值域。称 A 为一个算子。如果

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay,\ \forall x, y \in X,\ \alpha, \beta \in \mathbb{K},$$

则称 A 为线性算子。当 $Y = \mathbb{K}$ 时称 A 是数域 $\mathbb{K}$ 上的线性泛函，如实泛函或复泛函。

例：A 为 $n \times m$ 矩阵，$Ax, x \in \mathbb{R}^m$ 是 $\mathbb{R}^m$ 到 $\mathbb{R}^n$ 的线性算子。

例：设 Ω 是 $\mathbb{R}^n$ 的开区域，$X = Y = C^\infty(\bar{\Omega})$，

$$P(\partial) = \sum_{|\alpha| \leq m} a_\alpha(x) \partial^\alpha,$$

其中 $a_\alpha(x) \in C^\infty(\bar{\Omega})$，定义

$$A:\ u \in X \mapsto P(\partial)u,$$

即

$$(P(\partial)u)(x) = \sum_{|\alpha| \leq m} a_\alpha(x) \partial^\alpha u(x),$$

则 A 是 X 到 Y 的一个线性算子。

如果 $X = Y = L^2(\Omega)$，$D = C^m(\bar{\Omega})$，则上面的微分多项式算子仍是 X 到 Y 的线性算子，但定义域不是全空间 X。

E.14.1 连续和有界线性算子

设 H,K 是两个 Hilbert 空间，D 是 H 的线性子空间，A 是 $D\to K$ 的线性算子，按 H 中内积导出的距离，如果 Ax 在 $x=x_0$ 处连续，称算子 A 则 $x=x_0$ 处连续，如果 Ax 在所有 $x\in D$ 连续，称 Ax 是连续线性算子。

对于线性算子，Ax 连续当且仅当 Ax 在 $x=0$ 连续。

有界线性算子：设 H,K 是两个 Hilbert 空间，称线性算子 $H\to K$ 是有界线性算子，如果存在常数 $c\geq 0$ 使得

$$\|Ax\|_K\leq c\|x\|_H,\ \forall x\in H.$$

记 $\mathcal{B}(H,K)$ 为这些有界线性算子的全体，这是一个线性空间。

定理：若 $A,B\in\mathcal{B}(H,K)$, $\alpha\in\mathbb{K}$，

- （1）$A+B\in\mathcal{B}(H,K)$, 且 $\|A+B\|\leq\|A\|+\|B\|$；
- （2）$\alpha A\in\mathcal{B}(H,K)$, 且 $\|\alpha A\|=|\alpha|\,\|A\|$；
- （3）若还有 $C\in(K,J)$，则 $CA\in\mathcal{B}(H,J)$，且 $\|CA\|\leq\|C\|\,\|A\|$。

这说明 $\mathcal{B}(H,K)$ 是定义了加法 $A+B$ 和数乘 αA 的线性空间，且定义了复合运算 CA。

如果 A 是 H 到 K 的线性算子，则 A 连续当且仅当 A 有界。

E.14.2 同构

设 H 和 K 为两个 Hilbert 空间，如果存在线性算子 $U:H\to K$ 使得

$$\langle Ux,Uy\rangle_K=\langle x,y\rangle_H,\ \forall x,y\in H,$$

则称 U 是 H 到 K 的保内积算子；如果 U 还是满映射，则称 U 是 H 到 K 上的同构算子，此时 H 和 K 称为是**同构**的。

注意，保内积算子一定是单映射：对 $x,y\in H$ 且 $x\neq y$，必有 $Ux\neq Uy$，否则

$$\langle x-y,x-y\rangle_K\langle U(x-y),U(x-y)\rangle_K=\langle Ux-Uy,Ux-Uy\rangle_K=0$$

推出 $x-y=0$，矛盾。所以同构映射是 H 到 K 的一一的满线性映射且保内积。

由内积定义模和距离，设 d_H 和 d_K 分别为 H 和 K 中内积导出的距离，对线性算子 $V: H \to K$，如果

$$d_K(Vx, Vy) = d_H(x, y), \ \forall x, y \in H,$$

则称 V 是 H 到 K 的保距算子。

注意

$$\begin{aligned} d_K(Vx, Vy) =& \|Vx - Vy\|_K, \\ d_H(x, y) =& \|x - y\|_H, \end{aligned}$$

可以证明线性算子 V 为保距算子当且仅当

$$\|Vx\|_K = \|x\|, \ \forall x \in H.$$

充分性显然，对必要性，只要取 $y = 0$ 即可。

保内积条件与保距条件也是等价的：

$$\langle Ux, Uy\rangle_K = \langle x, y\rangle_H, \ \forall x, y \in H \Longleftrightarrow \|Ux\|_K = \|x\|_H, \ \forall x \in H.$$

保内积显然推出保距；反之，如果保距条件成立，则

$$\begin{aligned} &\|Ux + Uy\|^2 = \|Ux\|^2 + \|Uy\|^2 + 2\mathrm{Re}(\langle Ux, Uy\rangle_K) \\ =& \|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\mathrm{Re}(\langle x, y\rangle_H), \end{aligned}$$

因此

$$\mathrm{Re}(\langle Ux, Uy\rangle_K) = \mathrm{Re}(\langle x, y\rangle_H), \ \forall x, y \in H.$$

类似地，

$$\begin{aligned} &\|iUx + Uy\|^2 = \|Ux\|^2 + \|Uy\|^2 + 2\mathrm{Re}(\langle iUx, Uy\rangle_K) \\ =& \|ix + y\|^2 = \|x\|^2 + \|y\|^2 + 2\mathrm{Re}(\langle ix, y\rangle_H), \end{aligned}$$

但对 $z \in \mathbb{C}$，$\mathrm{Re}(iz) = -\mathrm{Im}(z)$，所以

$$\mathrm{Im}(\langle Ux, Uy\rangle_K) = \mathrm{Im}(\langle x, y\rangle_H), \ \forall x, y \in H,$$

所以保距条件成立时保内积条件成立。

E.14.3 可分 Hilbert 空间

定理 Hilbert 空间 H 为可分的（有可数的稠密子集），当且仅当存在至多可数的标准正交基 $\mathcal{E}$。若标准正交基个数为有限值 N，则 H 与 $\mathbb{K}^N$ 同构，否则与 l^2 同构。映射为 $x \in\mapsto \{\langle x, e_i\rangle, i = 1, 2, \ldots\}$, 其中 $\{e_i\}$ 是 H 的标准正交基。

E.14.4 共轭算子

若 H 和 K 是 Hilbert 空间，$u(x,z): H\times K\to\mathbb{K}$ 称为共轭双线性形式，如果 $u(x,z)$ 关于 x 为线性函数，关于 z 为共轭线性函数：

$$u(x,\alpha z+\beta w)=\bar{\alpha}u(x,z)+\bar{\beta}u(x,w),\ \forall\alpha,\beta\in\mathbb{K},\ x\in H,z,w\in K.$$

如果存在 $c\geq 0$ 使得

$$|u(x,z)|\leq c\,\|x\|\,\|z\|,\forall x\in H,z\in K,$$

则称 u 有界，称 c 是 u 的一个上界。

共轭双线性形式用来研究有界线性算子。显然，如果 $A\in\mathcal{B}(H,K)$，则 $u(x,z)=\langle Ax,z\rangle$ 是共轭双线性形式；如果 $B\in\mathcal{B}(K,H)$，则 $v(x,z)=\langle x,Bz\rangle$ 是共轭双线性形式。

定理 E.2. 设 u 是 $H\times K$ 的有界共轭双线性形式，有上界 c，则存在唯一的 $A\in\mathcal{B}(H,K)$ 和 $B\in\mathcal{B}(K,H)$，使得

$$u(x,z)=\langle Ax,z\rangle=\langle x,Bz\rangle,\ \forall x\in H,z\in K,$$

且 $\|A\|\leq c$, $\|B\|\leq c$。

上述定理基于 Riesz 表示定理。

共轭算子：若 $A\in\mathcal{B}(H,K)$，则必存在唯一的 $B\in\mathcal{B}(K,H)$ 使得

$$\langle Ax,z\rangle=\langle x,Bz\rangle,\ \forall x\in H,z\in K,$$

称 B 为 A 的共轭算子，记为 A^*。

例（矩阵的共轭转置） 设 A 为 $m\times n$ 实矩阵，定义 $A: x\in\mathbb{R}^n\mapsto Ax\in\mathbb{R}^m$，$Ax$ 中的 x 看成是 m 维列向量，则 A 是有界线性算子，有

$$\langle Ax,z\rangle=z^T(Ax)=(A^Tz)^Tx=\langle x,A^Tz\rangle,\ \forall x\in\mathbb{R}^n,z\in\mathbb{R}^m,$$

即算子 A 的共轭算子是由矩阵 A^T 生成的线性算子。

若 A 为 $m\times n$ 复矩阵，定义 $A: x\in\mathbb{C}^n\mapsto Ax\in\mathbb{C}^m$，$\mathbb{C}^n$ 中内积定义为

$$\langle x,y\rangle=y^*x,\forall x,y\in\mathbb{C}^n,$$

其中 y 看成列向量，y^* 表示列向量 y 的共轭转置，则 A 是有界线性算子，有

$$\langle Ax, z\rangle = z^*Ax = (A^*z)^*x = \langle x, A^*z\rangle,\ \forall x \in \mathbb{C}^n, z \in \mathbb{C}^m,$$

即有界线性算子 A 的共轭算子是它的共轭转置矩阵对应的线性算子。

当 $m = n$ 时，$\mathbb{C}^n$ 上的一个线性算子可以表示成 $n \times n$ 矩阵 A，其共轭算子为 A 的共轭转置。$\mathcal{B}(H)$ 上的共轭算子就是矩阵的共轭转置的推广。

同构算子充要条件：设 $U \in \mathcal{B}(H, K)$，则 U 是 H 到 K 的同构，当且仅当 U 可逆且 $U = U^*$。

通常考虑 $\mathcal{B}(H)$ 中有界线性算子的共轭算子。从 $A \in \mathcal{B}(H)$ 到 A^* 构成 $\mathcal{B}(H)$ 到 $\mathcal{B}(H)$ 的一个映射，称为对合映射。

定理：若 $A, B \in \mathcal{B}(H)$，$\alpha, \beta \in \mathbb{K}$，则

- （1）$(\alpha A + \beta B)^* = \bar{\alpha}A^* + \bar{\beta}B^*$；
- （2）$(AB)^* = B^*A^*$；
- （3）$(A^*)^* = A$。

定理：若 $A \in \mathcal{B}(H)$，则

$$\|A\| = \|A^*\| = \|A^*A\|^{1/2}.$$

例（积分算子）设 E 为 $\mathbb{R}^n$ 中勒贝格可测集，$H = L^2(E)$ 是 E 上所有平方可积函数组成的 Hilbert 空间，$k(x, y)$ 是 $E \times E$ 上平方可积函数，称为积分核。定义线性算子 K 如下

$$Kf(x) = \int_E k(x, y)f(y)\,dy,\ \forall f \in L^2(E),$$

称 K 为积分算子。考虑 $L^2(E) \times L^2(E)$ 的共轭双线性形式

$$u(f, g) = \iint_{E\times E} k(x, y)f(y)\overline{g(x)}\,dx\,dy,$$

由 Cauchy-Schwarz 不等式知此共轭双线性形式有上界 $\|k\|_2$，由定理E.2可知 K 是有界线性算子，$\|K\| \leq \|k\|_2$。此时 K^* 也是积分算子，对应的积分核为 $k^*(x, y) = \overline{k(y, x)}$。

对 $A \in \mathcal{B}(H)$，定义

$$\ker(A) = \{x \in H : Ax = 0\},$$

这是 H 的子 Hilbert 空间，称为 A 的核空间，类似于方阵的 0 特征值对应的特征向量张成的线性子空间。

定理对 $A \in \mathcal{B}(H)$，有

$$\ker(A) = R(A^*),$$

其中 $R(A^*)$ 表示 A^* 的值域 $\{A^*y : y \in H\}$。

自共轭算子、酉算子：对 $A \in \mathcal{B}(H)$，若 $A = A^*$，称 A 是自共轭算子（或自伴算子、Hermite 算子）。若 $AA^* = A^*A$，则称 A 是正规算子。如果 $AA^* = A^*A = I$，则称 A 是酉算子。

酉算子必为正规算子。自共轭算子必为正规算子。

自共轭算子是实对称方阵和 Hermite 复矩阵的推广。酉算子是实正交阵和复酉矩阵的推广。

设 U 是酉算子，即 $UU^* = U^*U = I$，则 U 可逆，是一一的满映射，且

$$\langle x, y\rangle = \langle U * Ux, y\rangle = \langle Ux, Uy\rangle,$$

所以酉算子还是保距（保内积）算子，所以酉算子是 H 到自身的同构算子。

当 H 是复 Hilbert 空间时，$\forall A \in \mathcal{B}(H)$，$B = (A + A^*)/2$，$C = (A - A^*)/(2i)$ 都是自共轭算子，且 $A = B + iC$，B 和 C 分别称为 A 的实部与虚部（注意 B 和 C 一般仍为复矩阵）。

定理设 H 是复 Hilbert 空间，$A \in \mathcal{B}(H)$，则

- （1）A 是自共轭算子 $\Longleftrightarrow \langle Ax, x\rangle \in \mathbb{R}, \forall x \in H$；
- （2）A 是正规算子 $\Longleftrightarrow \|Ax\| = \|A^*x\|, \forall x \in H \Longleftrightarrow A$ 的实部与虚部可交换。

其中 A 的实部与虚部可交换，是指 $BC = CB$。

定理若 A 是自共轭算子，则

$$\|A\| = \sup\{|\langle Ax, x\rangle| : \|x\| = 1\} = \max(|m|, |M|),$$

其中

$$M = \sup\{\langle Ax, x\rangle : \|x\| = 1\},\ m = \inf\{\langle Ax, x\rangle : \|x\| = 1\}.$$

推论：设 $A = A^*$，$\langle Ax, x\rangle = 0$ 对 $\forall x \in H$ 成立，则 $A = 0$，即 $Ax = 0$，$\forall x \in H$。

E.14.5 投影算子性质

幂等算子：对 $P \in \mathcal{B}(H)$，如果 $P^2 = P$，则称 P 为幂等算子。

设 P 为幂等算子，则 $(I-P)^2 = I - 2P + P^2 = I - P$ 也是幂等算子。易证

$$R(P) = \ker(I-P),\ \ker(P) = R(I-P),$$

投影算子的值域与核空间都是 H 的闭子空间。

投影算子：对 $P \in \mathcal{B}(H)$，如果 $P^2 = P$ 且 $P^* = P$（幂等而且自共轭的有界线性算子），则称 P 为投影算子。

设 H 为 Hilbert 空间，S 是 H 的子 Hilbert 空间，则 $P = \mathrm{Proj}_S : H \to S$ 是线性算子，定义域为 H，值域为 S，是有界算子：

$$\|Px\| \le \|x\|,$$

所以也是连续算子。当 S 非零时（S 不是仅由零元素组成），$\|P\| = 1$。

因为对 $x \in H$，$Px = \mathrm{Proj}_S x \in S$，这时 $P(Px) = Px$，即 $P^2 = P$，P 是幂等算子。

来证明 P 自共轭，只要证明

$$\langle Px, y\rangle = \langle x, Py\rangle, \forall x, y \in H.$$

对 x, y 作正交分解

$$x = x_1 \oplus x_2,\ y = y_1 \oplus y_2,$$

其中 $x_1, y_1 \in S,\ x_2, y_2 \in S^\perp$，则

$$\langle Px, y\rangle = \langle x_1, y_1 + y_2\rangle = \langle x_1, y_1\rangle = \langle x_1 + x_2, y_1\rangle = \langle x, Py\rangle.$$

所以，Proj_S 是投影算子。

反过来，设 P 是 H 的投影算子，令 $S = R(P)$，则 S 是 H 的闭子空间，对 $x = x_1 \oplus x_2 \in H$，其中 $x_1 \in S$，$x_2 \in S^\perp$，只要证明 $Px = x_1$，只要证明 $Px_1 = x_1$，$Px_2 = 0$。事实上 $x_1 \in S = R(P)$ 故存在 $y_1 \in H$ 使得 $x_1 = Py_1$，从而

$$Px_1 = P(Py_1) = P^2 y_1 = Py_1 = x_1.$$

另外，

$$\langle Px_2, Px_2\rangle = \langle x_2, P(Px_2)\rangle = \langle x_2, P^2x_2\rangle = \langle x_2, Px_2\rangle = 0$$

（因为 $x_2 \in S^\perp$ 而 $Px_2 \in S$。）

** 定理（投影算子充要条件）设 P 是 Hilbert 空间 H 的幂等算子（同时要求有界线性算子），则以下条件等价：

- （1）$P = P^*$（即投影算子）；
- （2）$\ker(P) = R(P)^\perp$;
- （3）P 是 c；
- （4）$\|P\| = 1$;
- （5）对所有 H 中的元素 x 有 $\langle Px, x\rangle \geq 0$。

当上述任何一个条件成立时，P 都是投影算子，而且是 H 到 $R(P)$ 的正交投影。

$y \in S$ 当且仅当 $\operatorname{Proj}_S y = y$。

$y \in S^\perp$ 当且仅当 $\operatorname{Proj}_S y = 0$。

设 S 为闭子空间 M 的闭子空间，则当 $y \in M$ 时 $\operatorname{Proj}_S y \in S \subset M$ 即 $\operatorname{Proj}_S y \in M$，同时 $(I - \operatorname{Proj}_S)y = y - \operatorname{Proj}_S y \in M$。对一般的 $y \in H$，有

$$\operatorname*{Proj}_S y = \operatorname*{Proj}_S \operatorname*{Proj}_M y$$

附录 F

测试

这一页面存在说明本备课笔记还是草稿状态！

§

References

Bibliography

Brockwell, P. J., & Davis, R. A. (1987). *Time series: Theory and methods*. Springer-Verlag.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods* (2nd). Springer-Verlag.

Christoffersen, P. F. (2003). *Elements of financial risk management*. Elsevier Science (USA).

Cryer, J. D., & Chan, K.-S. (2008). *Time series analysis with applications in r* (2nd Ed.). Springer.

Lutkepohl, H., & Kratzig, M. (2004). *Applied time series econometrics*. Cambridge University Press.

Paul S.P. Cowpertwait, A. V. M. (2009). *Introductory time series with r*. Springer.

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With r examples* (4th). Springer International Publishing. https://doi.org/10.1007/978-3-319-52452-8

Shumway, R. H., & Stoffer, D. S. (2019). *Time series: A data analysis approach using r*. Taylor; Francis Group, LLC.

Tsay, R. S. (2010). *Analysis of financial time series* (3rd Ed.). John Wiley & Sons, Inc.

Tsay, R. S. (2013). *金融数据分析导论：基于 r 语言*. 机械工业出版社.

Tsay, R. S. (2014). *Multivariate time series analysis with r and financial applications*. John Wiley & Sons, Inc.

何书元. (2003). *应用时间序列分析*. 北京大学出版社.

何书元. (2006). *概率论*. 北京大学出版社.

何书元. (2023). *应用时间序列分析* (2nd). 北京大学出版社.

吴喜之 & 刘苗. (2018). *应用时间序列分析——r 软件陪同* (第二版). 机械工业出版社.

李贤平. (2010). *概率论基础* (第三版). 高等教育出版社.

谢衷洁. (1990). *时间序列分析*. 北京大学出版社.

郭懋正. (2005). *实变函数与泛函分析*. 北京大学出版社.