应用时间序列分析
补充

李东风

2012年秋季学期

目录

Chapter 1

时间序列

1.1 微积分复习

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。

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

$$\int_a^b f'(x)dx\le f(b)-f(a)$$

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。

1.2 傅立叶级数复习

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$\sigma$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

$$T_n(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \{a_n \cos(nx) + b_n \sin(nx)\}$$

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 §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$(0<\alpha\le 1)$李普希兹条件：

$$|f(x_0\pm t)-f(x_0)|\le Lt^{\alpha},\ 0<t\le\delta$$

(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

$$\lim_{n\to\infty}\int_{-\pi}^{\pi}|f(x)-S_n(f,x)|^2dx=0$$

1.3 随机过程

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。

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》。

正态过程存在定理 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《时间序列分析》P5。

复值正态分布: 实部和虚部为联合正态分布。

1.4 时间序列与Hilbert空间

平方可积函数的Hilbert空间$L^2(d\lambda)$: 定义在有限区间(a,b)上的平方可积函数全体，内积为

$$< f, g >= \int_a^b fgd\lambda$$

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

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

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。

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$\Longleftrightarrow$ $\{\xi_t\}$在$t = 0$均方连续$\Longleftrightarrow$ $\gamma(\tau)$在$\mathbb{R}$连续$\Longleftrightarrow$ $\gamma(\tau)$在$\tau = 0$连续。

谱函数存在性定理 证明见谢衷洁《时间序列分析》P21定理1.6。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Hilberb空间$H_\xi$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。

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《时间序列分析》PP30–40。

定理 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《时间序列分析》P49。

1.5 三角级数求和

P.27**例**3.1**的推导**

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

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 2

自回归模型

2.1 复变复习

若复变函数$f(z)$在点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**。

区域指连通开集。

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。

2.2 滤波的收敛意义

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

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

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 §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。

2.3 Y-W方程

Y-W方程中如果$\Gamma_{p+1} > 0$则$a_1, \ldots, a_p, \sigma^2$唯一确定, 其中$a_1, \ldots, a_p$满足最小相位条件（定理2.4.1），$\sigma^2 > 0$。

反过来，如果$a_1, \ldots, a_p$满足最小相位条件，$\sigma^2 > 0$, Y-W方程中的$\gamma_0, \gamma_1, \ldots, \gamma_p$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$\gamma_0, \gamma_1, \ldots, \gamma_p$, 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：

$$\begin{aligned} \Gamma_n \boldsymbol{a}_n =& \boldsymbol{\gamma}_n \\ \gamma_0 - \boldsymbol{a}_n^T \boldsymbol{\gamma}_n =& \sigma_n^2 \end{aligned}$$

其中$\boldsymbol{\gamma}_n = (\gamma_1, \ldots, \gamma_n)^T$。假设$\boldsymbol{a}_n = (a_{n1}, a_{n2}, \ldots, a_{nn})^T$ 和$\sigma_n^2 > 0$给定，$\boldsymbol{a}_n$满足最小相位条件，则满足上述Y-W方程的$\gamma_0, \gamma_1, \ldots, \gamma_n$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 3

滑动平均模型与自回归滑动平均模型

3.1 MA模型

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》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《时间序列分析》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 §3.2 Proposition 3.2.1. 用了新息分解，正交分解来证明。$b_1, b_2, \ldots, b_q$是关于新息张成的空间上的投影系数。

3.2 广义ARMA

设

$$A(\mathscr{B})X_t=B(\mathscr{B})\varepsilon_t,\ t\in\mathbb{Z},$$

其中$A(z)\neq 0, |z|=1$, A与B没有公共根。则Lorent级数

$$\frac{B(z)}{Z(z)}=\sum_{j=-\infty}^{\infty}\psi_j z^j,\ \rho^{-1}<|z|<\rho(\rho>1)$$

收敛，于是有唯一平稳解

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

3.3 ARMA模型的谱

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《时间序列分析》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

$$\frac{A(z)}{B(z)}=\sum_{j=0}^{\infty}d_j z^j,\ |z|\le\rho_2\ (\rho_2>1)$$

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。

3.4 ARMA模型与Hilbert空间

参考谢衷洁《时间序列分析》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

$$\sigma\psi_j =< X_t, \varepsilon_{t-j}/\sigma >,\ j = 0, 1, 2, \dots$$

可逆ARMA模型中的白噪声ε_t是新息(Wold序列)：

$$\varepsilon_t = X_t - L(X_t|H_X(t-q))$$

(参考谢衷洁《时间序列分析》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

$$\begin{aligned} d_0 =& 1 \\ d_j =& -\sum_{k=1}^{j} \psi_k d_{j-k},\ j = 1, 2, \dots \end{aligned}$$

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

$$\sum_{j=0}^{\infty}\left(\sum_{k=0}^{j} \psi_k d_{j-k}\right) z^j = 1$$

3.5 ARIMA不平稳

ARIMA(p,d,q)模型没有平稳解。
讨论 当$d=1$时，问题为，$\{\xi_t\}$是ARMA(p,q)序列, $\{X_t\}$满足

$$X_t - X_{t-1} = \xi_t \tag{3.1}$$

来证明(3.1)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 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

$$\begin{aligned}
&\mathrm{Var}(X_t - X_0) = \mathrm{Var}(\sum_{k=1}^{t}\xi_t) = \sum_{k=1}^{t}\sum_{j=1}^{t}\gamma_{k-j} \\
=&\sum_{k=1}^{t}\sum_{j=1}^{t}\int_{-\pi}^{\pi} e^{i(k-j)\lambda} f(\lambda)d\lambda = \int_{-\pi}^{\pi}\left|\sum_{k=1}^{t} e^{ik\lambda}\right|^2 f(\lambda)d\lambda \\
\geq& c\int_{-\pi}^{\pi}\left|\sum_{k=1}^{t} e^{ik\lambda}\right|^2 d\lambda = c\sum_{k=1}^{t}\sum_{j=1}^{t}\int_{-\pi}^{\pi} e^{i(k-j)\lambda}d\lambda \\
=&c\sum_{k=1}^{t} 2\pi \quad (\text{仅}k-j=0\text{时积分不为零}) = 2\pi ct
\end{aligned}$$

所以$\{X_t\}$不平稳。

情形3 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：

$$A(\mathscr{B})(1-\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t$$

两边没有公因子所以$B(1)\neq 0$。于是

$$\begin{aligned}
&\mathrm{Var}(X_t - X_0) = \mathrm{Var}(\sum_{k=1}^{t}\xi_t) = \sum_{k=1}^{t}\sum_{j=1}^{t}\gamma_{k-j} \\
=&\sum_{k=1-t}^{t-1}(t-|k|)\gamma_k \\
=&t\sum_{k=1-t}^{t-1}\gamma_k - \sum_{k=1-t}^{t-1}|k|\gamma_k
\end{aligned}$$

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$\mathrm{Var}(X_t - X_0) \to +\infty,\ t \to +\infty$，$\{X_t\}$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。具体证明待查。

Chapter 4

均值和自协方差函数的估计

4.1 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 5

时间序列的预报

5.1 预报

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。

5.1.1 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。

5.1.2 预报误差方差单调性

σ_k^2单调上升，$\sigma_{k,m}^2$关于m单调下降但是关于k不是单调上升的。反例如下。令

$$X_t = \frac{1}{2} X_{t-2} + \varepsilon_t,\ \{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2).$$

此AR(2)模型平稳解为

$$X_t = \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j \varepsilon_{t-2j},\ t \in \mathbb{Z}.$$

自协方差$\gamma_0 = \frac{4}{3}\sigma^2$, $\gamma_1 = 0$, $\gamma_2 = \frac{2}{3}\sigma^2$。

$$\begin{aligned} L(X_t|X_{t-1}) =& 0, & \sigma_{1,1}^2 = E(X_t - 0)^2 = \frac{4}{3}\sigma^2, \\ L(X_t|X_{t-2}) =& \frac{1}{2} X_{t-2}, & \sigma_{2,1}^2 = E(X_t - \frac{1}{2} X_{t-2})^2 = \sigma^2 \end{aligned}$$

这里$\sigma_{1,1}^2 > \sigma_{2,1}^2$。

5.1.3 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

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1 + 2e^{i\lambda}|^2 = \frac{\sigma^2}{2\pi}(5 + 4\cos\lambda)$$

但$\varepsilon_t \neq X_t - L(X_t|H_{t-1})$。（待证明）

5.1.4 例2.2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

$$\begin{aligned}
&2\cos\lambda_j Z_j(t-1) - Z(t-2)\\
=&\xi_j\{2\cos\lambda_j \cos[(t-1)\lambda_j] - \cos[(t-2)\lambda_j]\}\\
&+ \eta_j\{2\cos\lambda_j \sin[(t-1)\lambda_j] - \sin[(t-2)\lambda_j]\}\\
=&\xi_j\{\cos(t\lambda_j) + \cos[(t-2)\lambda_j] - \cos[(t-2)\lambda_j]\}\\
&+ \eta_j\{\sin(t\lambda_j) + \sin[(t-2)\lambda_j] - \sin[(t-2)\lambda_j]\}\\
=&Z_j(t)
\end{aligned}$$

所以

$$L(Z_j(t)|Z_j(t-1), Z_j(t-2)) = 2\cos\lambda_j Z_j(t-1) - Z_j(t-2).$$

5.2 ARMA预报

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:

$$\begin{aligned}
\hat{X}_t =& L(X_t|X_{t-1}, \dots, X_1), t = 1, 2, \dots, n\\
\hat{X}_{n+k} =& L(X_{t+k}|X_1, X_2, \dots, X_n)
\end{aligned}$$

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，$n+1$及以后用递推预报。

5.3 思考问题

5.3.1 AR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？

$$\begin{aligned}
&L(X_{t+k}|X_1, \dots, X_n)\\
&\qquad = L(X_{t+k}|X_n, X_{n-1}, \dots, X_{n-p+1}
\end{aligned}$$

解方程

$$\Gamma_p \boldsymbol{b} = (\gamma_k, \gamma_{k+1}, \dots, \gamma_{k+p-1})^T$$

得k步预报均方误差

$$\sigma_k^2 = \gamma_0 - b_1\gamma_k - b_2\gamma_{k+1} - \dots - b_p\gamma_{k+p-1}$$

有没有简化公式？

Chapter 6

ARMA模型的参数估计

6.1 概率极限

6.1.1 依概率收敛

定理 设$\{a_n\}$为实数列，则$a_n \xrightarrow{\text{P}} a$ 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$a_n \xrightarrow{\text{P}} a$。$\forall \delta > 0$, 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 设$\xi_n \xrightarrow{\text{P}} \xi$, $\eta_n \xrightarrow{\text{P}} \eta$, 则

$$\xi_n + \eta_n \xrightarrow{\text{P}} \xi + \eta.$$

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 设$\xi_n \xrightarrow{\text{P}} \xi$, a为常数, 则

$$a\xi_n \xrightarrow{\text{P}} a\xi.$$

证明 $a = 0$时显然。当$a \neq 0$时，$\forall \varepsilon > 0$,

$$\begin{aligned}
&P(|a\xi_n - a\xi| \geq \varepsilon) \\
= &P(|\xi_n - \xi| \geq \frac{\varepsilon}{|a|}) \to 0 \quad (n \to \infty)
\end{aligned}$$

□

推论 设$\{\xi_n\}$, $\{\eta_n\}$为两个随机序列，a, b, c为常数，若$\xi_n \xrightarrow{\text{P}} \xi$, $\eta_n \xrightarrow{\text{P}} \eta$, 则$a\xi_n + b\eta_n + c \xrightarrow{\text{P}} a\xi + b\eta + c$。

定理 设$\xi_n \xrightarrow{\text{P}} a$, a为常数，函数$g(\cdot)$在a连续，则

$$g(\xi_n) \xrightarrow{\text{P}} g(a).$$

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$\xi_n\xrightarrow{\mathrm{P}}a$, 则

$$\begin{aligned}\xi_n^2 &\xrightarrow{\mathrm{P}} a^2\\ 1/\xi_n &\xrightarrow{\mathrm{P}} 1/a\ (\text{只要}a\neq 0)\\ \sqrt{\xi_n} &\xrightarrow{\mathrm{P}} \sqrt{a}\ (\text{只要}a\geq 0)\end{aligned}$$

定理 设$\xi_n\xrightarrow{\mathrm{P}}\xi$, $g(\cdot)$为连续函数，则

$$g(\xi_n)\xrightarrow{\mathrm{P}}g(\xi).$$

证明参考Tucker, H.G.(1967), A Graduate Course in Probability, New York: Academic Press.

定理 设$\xi_n\xrightarrow{\mathrm{P}}\xi$, $\eta_n\xrightarrow{\mathrm{P}}\eta$, 则

$$\xi_n\eta_n\xrightarrow{\mathrm{P}}\xi\eta.$$

证明

$$\xi_n\eta_n=\frac{1}{2}\left[\xi_n^2+\eta_n^2-(\xi_n-\eta_n)^2\right]$$

其中$\xi_n^2\xrightarrow{\mathrm{P}}\xi^2$, $\eta_n^2\xrightarrow{\mathrm{P}}\eta^2$, $\xi_n-\eta_n\xrightarrow{\mathrm{P}}\xi-\eta$, $(\xi_n-\eta_n)^2\xrightarrow{\mathrm{P}}(\xi-\eta)^2$, 所以

$$\xi_n\eta_n\xrightarrow{\mathrm{P}}\frac{1}{2}\left[\xi^2+\eta^2-(\xi-\eta)^2\right]=\xi\eta$$

6.1.2 依分布收敛

定理 设$\xi_n\xrightarrow{\mathrm{P}}\xi$, 则$\xi_n\xrightarrow{\mathrm{d}}\xi$。
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

$$\overline{\lim_{n\to\infty}}F_n(x)\leq F(x+\epsilon).$$

另一方面，

$$\begin{aligned}P(\xi_n > x) =& P(\xi_n > x,\ |\xi_n - \xi| < \epsilon) + P(\xi_n > x,\ |\xi_n - \xi| \geq \epsilon)\\ \leq& P(\xi > x - \epsilon) + P(|\xi_n - \xi| \geq \epsilon),\\ \overline{\lim_{n\to\infty}}\, [1 - F_n(x)] \leq& 1 - F(x - \epsilon),\\ \underline{\lim_{n\to\infty}}\, F_n(x) \geq& F(x - \epsilon),\end{aligned}$$

总之有

$$F(x-\epsilon) \leq \underline{\lim_{n\to\infty}}\, F_n(x) \leq \overline{\lim_{n\to\infty}}\, F_n(x) \leq F(x+\epsilon),$$

令$\epsilon \to 0+$则可知

$$\lim_{n\to\infty} F_n(x) = F(x).$$

□

定理 若a是常数，$\xi_n \xrightarrow{\text{d}} a$, 则$\xi_n \xrightarrow{\text{P}} a$。
证明 记

$$F(x) = \begin{cases} 1, & x \geq a \\ 0, & x < a \end{cases}$$

则

$$P(\xi_n \leq x) \to F(x),\ \forall x \neq a.$$

$\forall \delta > 0$,

$$\begin{aligned}& P(|\xi_n - a| > \delta)\\ =& P(\xi_n > a + \delta) + P(\xi_n < a - \delta)\\ =& 1 - P(\xi_n \leq a + \delta) + P(\xi_n < a - \delta)\\ \leq& 1 - P(\xi_n \leq a + \delta) + P(\xi_n \leq a - \delta)\\ \to& 0 \quad (n \to \infty)\end{aligned}$$

□

依分布收敛不一定依概率收敛。比如，设$X \sim \text{N}(0,1)$，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$p_n(X)$有可积的上界，则根据控制收敛定理可知，ξ_n的分布函数收敛。

定理 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$M(t) = Ee^{t\xi}$, $t \in [-h_1, h_1]$, $0 < h_1 \le h$, 若$\lim_{n\to\infty} M_n(t) = M(t), \forall t \in [-h_1, h_1]$, 则$\xi_n \overset{\mathrm{d}}{\to} \xi$。

证明略。

定理 设$\xi_n \overset{\mathrm{d}}{\to} \xi$, $\eta_n \overset{\mathrm{P}}{\to} 0$, 则

$$\xi_n + \eta_n \overset{\mathrm{d}}{\to} \xi.$$

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

$$\begin{aligned}
&P(\xi_n + \eta_n \le x_0) - F(x_0)\\
\le&[P(\xi_n \le x_0 + \delta) - F(x_0 + \delta)] + [F(x_0 + \delta) - F(x_0)] + P(|\eta_n| \ge \delta) \qquad (*)
\end{aligned}$$

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

$$\begin{aligned}
&P(\xi_n + \eta_n \ge x_0) - F(x_0)\\
\ge&P(\xi_n \le x_0 - \delta) - F(x_0 - \delta) + F(x_0 - \delta) - F(x_0) - P(|\eta_n| > \delta) \qquad (**)
\end{aligned}$$

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$\xi_n \xrightarrow{\mathrm{d}} \xi$, $\eta_n \xrightarrow{\mathrm{P}} 0$, 存在$n_1$使得$\forall n \geq n_1$，

$$\begin{aligned} P(\xi_n \leq x_0+\delta) - F(x_0+\delta) <& \frac{\varepsilon}{3} \\ P(|\eta_n| \geq \delta) <& \frac{\varepsilon}{3} \end{aligned}$$

由(*)式, 当$n \geq n_1$时

$$P(\xi_n+\eta_n \leq x_0) - F(x_0) < \varepsilon$$

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$n_2 \geq n_1$使得$\forall n \geq n_2$有

$$\begin{aligned} P(\xi_n \leq x_0-\delta) - F(x_0-\delta) >& -\frac{\varepsilon}{3} \\ P(|\eta_n| \geq \delta) <& \frac{\varepsilon}{3} \end{aligned}$$

由(**)可得当$n \geq n_2$时

$$P(\xi_n+\eta_n \leq x_0) - F(x_0) > -\varepsilon$$

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 设$\xi_n \xrightarrow{\mathrm{d}} \xi$, $\eta_n \xrightarrow{\mathrm{P}} 1$, 则$\eta_n\xi_n \xrightarrow{\mathrm{d}} \xi$。
证明与上一定理类似。

定理 设$\xi_n \xrightarrow{\mathrm{d}} \xi$, $g(\cdot)$是定义在ξ的支撑集上的连续函数，则

$$g(\xi_n) \xrightarrow{\mathrm{d}} g(\xi).$$

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(Slutsky) 设$\xi_n \xrightarrow{\mathrm{d}} \xi$, $A_n \xrightarrow{\mathrm{P}} a$, $B_n \xrightarrow{\mathrm{P}} b$, 则

$$A_n + B_n\xi_n \xrightarrow{\mathrm{d}} a + b\xi.$$

证明略。

定理 设$\{F_1, F_2, \ldots, \}$是分布函数列（没有$F(-\infty) = 0$和$F(+\infty) = 1$的限制），若对某一个$\delta > 0$和$r_0 > 0$，数列$\{\int_R |x|^{r_0+\delta} dF_n(x) : n \geq 1\}$有界且$F_n$依分布收敛到$F$，则对任意$r \in [0, r_0]$和整数$k \in (0, r_0]$，当$n \to \infty$时有

$$\int_R x^k dF_n(x) \to \int_R x^k dF(x) \qquad \int_R |x|^r dF_n(x) \to \int_R |x|^r dF(x)$$

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$E0 = 0$。

6.1.3 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

$$\sup_n P(|\xi_n| > M) \leq \varepsilon$$

就称时间序列$\{\xi_n\}$是**依概率有界的**，记做$\xi_n = O_p(1)$。依概率有界就是除去一个概率任意小的集合后序列有界。

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$\xi_n = O_p(\eta_n)$。

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$n \geq N$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$j = 1, 2, \ldots, N$，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

$$\begin{aligned} P(|\xi_n| > M') &< \varepsilon, \ \forall n \in \mathbb{N}_+, \\ \sup_{n \in \mathbb{N}_+} P(|\xi_n| > M') &\leq \varepsilon, \end{aligned}$$

满足原定义。 □

若$\xi_n \stackrel{\mathrm{Pr}}{\to} 0 \ (n \to \infty)$ 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 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

$$\mathrm{Pr}(|\xi_n/c_n| > \delta) < \varepsilon$$

取$M \geq \delta$使得

$$\Pr(\max_{1\leq n\leq N} |\xi_n/c_n| > M) < \varepsilon,$$

则

$$\Pr(|\xi_n/c_n| > M) < \varepsilon, \quad n = 1, 2, \ldots$$

即$\xi_n/c_n = O_p(1)$. □

定理 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

$$P\left(\left|\frac{\xi_n}{c_n}\right| > \delta\right) = P(|\xi_n| > |c_n|\delta) \leq P(|\xi_n| > M) < \varepsilon.$$

□

定理 对随机变量ξ，$\xi = O_p(1)$.

定理 若存在非负随机变量ξ使得$|\xi_n| \leq \xi$ a.s.则$\xi_n = O_p(1)$。
证明 由$P(|\xi_n| \leq \xi) = 1$，$\forall \varepsilon > 0$, $\exists M > 0$使$P(\xi > M) < \varepsilon$，于是$P(|\xi_n| > M) \leq P(\xi > M) < \varepsilon$, $\forall n \in \mathbb{N}_+$。

定理 若$\{\xi_n\}$同分布，则$\xi_n = O_p(1)$。
证明 $\forall \varepsilon > 0$, $\exists M > 0$使$\Pr(|\xi_1| > M) < \varepsilon$。由同分布性知$\Pr(|\xi_t| > M) = \Pr(|\xi_1| > M) < \varepsilon$。

定理

$$\begin{aligned}
O_p(1) \pm O_p(1) =& O_p(1) \\
O_p(1) \cdot O_p(1) =& O_p(1) \\
O_p(1) \pm o_p(1) =& O_p(1) \\
O_p(1) \cdot o_p(1) =& o_p(1).
\end{aligned}$$

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

$$\begin{aligned}
&\overline{\lim_{n\to\infty}}\, P(|\xi_n\eta_n| > \delta) \\
\leq& \overline{\lim_{n\to\infty}}\, P(|\xi_n\eta_n| > \delta, |\xi_n| \leq M) + \overline{\lim_{n\to\infty}}\, P(|\xi_n\eta_n| > \delta, |\xi_n| > M) \\
\leq& \overline{\lim_{n\to\infty}}\, P(|\eta_n| > \frac{\delta}{M}) + \overline{\lim_{n\to\infty}}\, P(|\xi_n| > M) \\
\leq& \epsilon,
\end{aligned}$$

即$\lim_n P(|\xi_n\eta_n|>\delta)=0,\ \xi_n\eta_n=o_p(1)$。 □

定理 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 设$\xi_n\xrightarrow{\mathrm{d}}\xi$, 则$\xi_n=O_p(1)$。
证明 设ξ分布函数为$F(x)$，对$\forall\varepsilon>0$，存在$M>0$且M和$-M$为$F(x)$的连续点，使得

$$\begin{aligned}\liminf_{n\to\infty} P(|\xi_n|\le M)=&\lim_{n\to\infty} P(\xi_n\le M)-\liminf_{n\to\infty} P(\xi_n<-M)\\ \ge&\lim_{n\to\infty} P(\xi_n\le M)-\lim_{n\to\infty} P(\xi_n\le -M)\\ =&F(M)-F(-M)\ge 1-\frac{\varepsilon}{2}>1-\varepsilon\end{aligned}$$

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 设$\xi_n=O_p(1)$，$\eta_n/\xi_n=o_p(1)$, 则$\eta_n=o_p(1)$。
证明 $\forall\delta>0,\ \forall\varepsilon>0$。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$\eta_n=o_p(1)$。

6.1.4 Delta方法

定理(Delta**方法**) 设随机变量序列$\{\xi_n\}$有极限分布

$$\sqrt{n}(\xi_n-\theta)\xrightarrow{\mathrm{d}}\mathrm{N}(0,\sigma^2)$$

函数$g(x)$在θ处可微，$g'(\theta)\ne 0$。则

$$\sqrt{n}(g(\xi_n)-g(\theta))\xrightarrow{\mathrm{d}}\mathrm{N}(0,\sigma^2(g'(\theta))^2).$$

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$\mathrm{N}(0, \sigma^2(g'(\theta))^2)$，后一项是中$\sqrt{n}|\xi_n - \theta| = O_p(1)$所以是$o_p(1)$的，于是结果可得。

引理 设函数$h(x)$在$x = 0$处可微，$h'(0) = 0$。若$\xi_n = o_p(1)$则$h(\xi_n) = o_p(|\xi_n|)$。
证明 $\forall \varepsilon > 0, \forall \delta > 0$，由$h'(0) = 0$可知存在$\delta_1 > 0$使得对任意$0 < |x| \leq \delta_1$有

$$|\frac{h(x)}{x}| < \delta$$

于是

$$\begin{aligned}& P\left(\left|\frac{h(\xi_n)}{\xi_n}\right| > \delta\right) \\ =& P\left(\left|\frac{h(\xi_n)}{\xi_n}\right| > \delta,\ |\xi_n| \leq \delta_1\right) + P\left(\left|\frac{h(\xi_n)}{\xi_n}\right| > \delta,\ |\xi_n| > \delta_1\right) \\ =& P\left(\left|\frac{h(\xi_n)}{\xi_n}\right| > \delta,\ |\xi_n| > \delta_1\right) \\ \leq& P(|\xi_n| > \delta_1) \to 0 \quad (n \to \infty)\end{aligned}$$

□

引理

$$\lim_{n\to\infty} \left(1 + \frac{b}{n} + o(\frac{1}{n})\right)^{cn} = e^{bc}.$$

6.1.5 随机向量的极限

定义 设$\{\boldsymbol{\xi}_n\}$为随机向量序列，$\boldsymbol{\xi}$为随机向量，如果对任意$\delta > 0$都有

$$\lim_{n\to\infty} P(\|\boldsymbol{\xi}_n - \boldsymbol{\xi}\| > \delta) = 0,$$

则称$\boldsymbol{\xi}_n$依概率收敛到$\boldsymbol{\xi}$，记作$\boldsymbol{\xi}_n \xrightarrow{\mathrm{P}} \boldsymbol{\xi}$。

定理 设$\boldsymbol{\xi}_n = (\xi_{n1}, \ldots, \xi_{nm})^T$, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)^T$, 则$\boldsymbol{\xi}_n \xrightarrow{\mathrm{P}} \boldsymbol{\xi}$ 当且仅当$\xi_{nj} \xrightarrow{\mathrm{P}} \xi_j$, $j = 1, \ldots, m$。

定义 设$\{\boldsymbol{\xi}_n\}$为随机向量序列，$\boldsymbol{\xi}_n$分布函数为$F_n(\boldsymbol{x})$, $\boldsymbol{\xi}$为随机向量，有分布函数$F(\boldsymbol{x})$, $F(\cdot)$的任意连续点$\boldsymbol{x}$均有

$$\lim_{n\to\infty} F_n(\boldsymbol{x}) = F(\boldsymbol{x}),$$

则称$\boldsymbol{\xi}_n$依分布收敛到$\boldsymbol{\xi}$或依分布收敛到$F(\cdot)$，记作$\boldsymbol{\xi}_n \xrightarrow{\text{d}} \boldsymbol{\xi}$ 或$\boldsymbol{\xi}_n \xrightarrow{\text{d}} F(\cdot)$。

定理 设$\{\boldsymbol{\xi}_n\}$为随机向量序列，$\boldsymbol{\xi}$为随机向量，$\boldsymbol{\xi}_n \xrightarrow{\text{d}} \boldsymbol{\xi}$, 函数$g(\boldsymbol{x})$是定义于$\boldsymbol{\xi}$的支撑集上的连续函数，则$g(\boldsymbol{\xi}_n)$依分布收敛于$g(\boldsymbol{\xi})$。

推论：随机向量依分布收敛，则相应分量依分布收敛。

定理 设$\boldsymbol{\xi}_n$有矩母函数$M_n(\boldsymbol{t}) = Ee^{\boldsymbol{t}^T\boldsymbol{\xi}_n}$, $\boldsymbol{\xi}$有矩母函数$M(t) = Ee^{\boldsymbol{t}^T\boldsymbol{\xi}}$, 若$\lim_{n\to\infty} M_n(\boldsymbol{t}) = M(\boldsymbol{t})$, $\|t\| \leq h(h>0)$, 则$\boldsymbol{\xi}_n \xrightarrow{\text{d}} \boldsymbol{\xi}$。

定理(随机向量的中心极限定理) 设独立同分布随机向量序列$\{\boldsymbol{\xi}_n\}$具有共同的期望$\boldsymbol{\mu}$和协方差阵Σ，Σ正定，设共同的矩母函数$M(\boldsymbol{t})$在$\boldsymbol{0}$的一个开邻域存在，令

$$\boldsymbol{\eta}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu}) = \sqrt{n}(\bar{\boldsymbol{\xi}} - \boldsymbol{\mu}),$$

则$\boldsymbol{\eta}_n$依分布收敛到$\mathrm{N}_m(\boldsymbol{0}, \Sigma)$分布。

定理 设m维随机向量序列$\{\boldsymbol{\xi}_n\}$渐近$\mathrm{N}_m(\boldsymbol{\mu}, \Sigma)$分布，$A, \boldsymbol{b}$为非随机的矩阵和向量，则$A\boldsymbol{\xi}_n + \boldsymbol{b}$渐近$\mathrm{N}_m(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^T)$分布。

定理 设m维随机向量序列$\{\boldsymbol{\xi}_n\}$满足

$$\sqrt{n}(\boldsymbol{\xi}_n - \boldsymbol{\mu}_0) \xrightarrow{\text{d}} \mathrm{N}_m(\boldsymbol{0}, \Sigma),$$

设$\boldsymbol{g}(\boldsymbol{x})$为一个$\mathbb{R}^m$到$\mathbb{R}^k$的变换$(k \leq m)$, 把各个一阶偏导数组成一个矩阵

$$B = \left(\frac{\partial g_i(\boldsymbol{x})}{\partial x_j}\right)_{\substack{i=1,\dots,k;\\ j=1,\dots,m}},$$

设在$\boldsymbol{\mu}_0$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$\boldsymbol{x} = \boldsymbol{\mu}_0$处的值为$B_0$，则

$$\sqrt{n}(\boldsymbol{g}(\boldsymbol{\xi}_n) - \boldsymbol{g}(\boldsymbol{\mu}_0)) \xrightarrow{\text{d}} \mathrm{N}_m(\boldsymbol{0}, B_0\Sigma B_0^T).$$

6.2 问题

问题 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。