应用时间序列分析
课程介绍

主讲老师：席瑞斌

本课件基于李东风老师课件修改

2017年秋季学期

预报问题

- 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$\{X_t\}$都可以按第一章的(1.4)式分解成趋势项$\{T_t\}$、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。

预报问题

- 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$\{X_t\}$都可以按第一章的(1.4)式分解成趋势项$\{T_t\}$、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。

预报问题

- 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$\{X_t\}$都可以按第一章的(1.4)式分解成趋势项$\{T_t\}$、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。

预报问题

- 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$\{X_t\}$都可以按第一章的(1.4)式分解成趋势项$\{T_t\}$、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。

预报问题

- 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$\{X_t\}$都可以按第一章的(1.4)式分解成趋势项$\{T_t\}$、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。

预报问题

- 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$\{X_t\}$都可以按第一章的(1.4)式分解成趋势项$\{T_t\}$、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。

预报问题

- 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$\{X_t\}$都可以按第一章的(1.4)式分解成趋势项$\{T_t\}$、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。

最佳线性预测定义

- **定义1.1** 设Y 和$\mathbf{X}=(X_1,\ldots,X_n)^T$是均值为零, 方差有限的随机变量(向量). 如果$\mathbf{a}\in\mathbb{R}^n$使得对任何的$\mathbf{b}\in\mathbb{R}^n$, 有

$$\mathrm{E}(Y-\mathbf{a}^T\mathbf{X})^2\le\mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2.$$

则称$\mathbf{a}^T\mathbf{X}$是用$X_1,X_2,\ldots,X_n$对Y进行预测的**最佳线性预测**, 记做$L(Y|\mathbf{X})$或$\hat{Y}$. 于是

$$\hat{Y}=L(Y|\mathbf{X})=\mathbf{a}^T\mathbf{X}. \tag{1.2}$$

- **定义1.2** 如果$\mathrm{E}Y=b,\mathrm{E}\mathbf{X}=\mu$, 定义

$$L(Y|\mathbf{X})=L(Y-b|\mathbf{X}-\mu)+b, \tag{1.3}$$

并称$L(Y|\mathbf{X})$ 是用$X_1,X_2,\cdots,X_n$对Y进行预测时的最佳线性预测.

最佳线性预测定义

- **定义1.1** 设Y 和$\mathbf{X}=(X_1,\ldots,X_n)^T$是均值为零，方差有限的随机变量(向量). 如果$\mathbf{a}\in\mathbb{R}^n$使得对任何的$\mathbf{b}\in\mathbb{R}^n$，有

$$\mathrm{E}(Y-\mathbf{a}^T\mathbf{X})^2\le \mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2.$$

则称$\mathbf{a}^T\mathbf{X}$是用$X_1,X_2,\ldots,X_n$对Y进行预测的**最佳线性预测**，记做$L(Y|\mathbf{X})$或$\hat{Y}$. 于是

$$\hat{Y}=L(Y|\mathbf{X})=\mathbf{a}^T\mathbf{X}. \tag{1.2}$$

- **定义1.2** 如果$\mathrm{E}Y=b,\mathrm{E}\mathbf{X}=\boldsymbol{\mu}$，定义

$$L(Y|\mathbf{X})=L(Y-b|\mathbf{X}-\boldsymbol{\mu})+b, \tag{1.3}$$

并称$L(Y|\mathbf{X})$ 是用$X_1,X_2,\cdots,X_n$对Y进行预测时的最佳线性预测.

最佳线性预测定义II

- 以下总设随机变量均值为零。
- 用$\Gamma = \mathrm{E}(\mathbf{X}\mathbf{X}^T)$表示$\mathbf{X}$的协方差阵。
- 用$\Sigma_{\mathbf{X}Y} = \mathrm{E}(\mathbf{X}Y)$表示$\mathbf{X}$和$Y$的协方差向量。

最佳线性预测定义II

- 以下总设随机变量均值为零。
- 用$\mathbf{\Gamma} = \mathrm{E}(\mathbf{X}\mathbf{X}^T)$表示$\mathbf{X}$的协方差阵。
- 用$\Sigma_{\mathbf{X}Y} = \mathrm{E}(\mathbf{X}Y)$表示$\mathbf{X}$和$Y$的协方差向量。

最佳线性预测定义II

- 以下总设随机变量均值为零。
- 用$\Gamma = \mathrm{E}(\mathbf{X}\mathbf{X}^T)$表示$\mathbf{X}$的协方差阵。
- 用$\Sigma_{\mathbf{X}Y} = \mathrm{E}(\mathbf{X}Y)$表示$\mathbf{X}$和$Y$的协方差向量。

性质1

- **性质1.** 如果$\mathbf{a} \in \mathbb{R}^n$, 使得

$$\Gamma \mathbf{a} = \Sigma_{\mathbf{X}Y}, \tag{1.5}$$

则

$$L(Y|\mathbf{X}) = \mathbf{a}^T \mathbf{X},$$

并且有

$$E(Y - L(Y|\mathbf{X}))^2 = EY^2 - E[L(Y|\mathbf{X})]^2 = EY^2 - \mathbf{a}^T \Gamma \mathbf{a}. \tag{1.6}$$

如果Γ和$\Sigma_{\mathbf{X}Y}$已知, 以$\mathbf{a}$ 为未知数的线性方程组(1.5)被称为**预测方程**.

性质1—II

证明： 对任何$\mathbf{b} \in \mathbb{R}^n$,

$$
\begin{aligned}
& E(Y-\mathbf{b}^T\mathbf{X})^2 \\
=& E[Y-\mathbf{a}^T\mathbf{X}+(\mathbf{a}^T-\mathbf{b}^T)\mathbf{X}]^2 \\
=& E(Y-\mathbf{a}^T\mathbf{X})^2+E[(\mathbf{a}^T-\mathbf{b}^T)\mathbf{X}]^2+2E[(\mathbf{a}^T-\mathbf{b}^T)\mathbf{X}(Y-\mathbf{a}^T\mathbf{X})] \\
=& E(Y-\mathbf{a}^T\mathbf{X})^2+E[(\mathbf{a}^T-\mathbf{b}^T)\mathbf{X}]^2 \\
& +2(\mathbf{a}^T-\mathbf{b}^T)[E(\mathbf{X}Y)-E(\mathbf{X}\mathbf{X}^T)\mathbf{a}] \\
=& E(Y-\mathbf{a}^T\mathbf{X})^2+E[(\mathbf{a}^T-\mathbf{b}^T)\mathbf{X}]^2 \\
\geq& E(Y-\mathbf{a}^T\mathbf{X})^2.
\end{aligned}
$$

所以，$\mathbf{a}^T\mathbf{X}$是Y的最佳线性预测. 利用(1.5)得到

$$
\begin{aligned}
& E[Y-L(Y|\mathbf{X})]^2 = E(Y-\mathbf{a}^T\mathbf{X})^2 \\
=& EY^2+\mathbf{a}^T E(\mathbf{X}\mathbf{X}^T)\mathbf{a}-2\mathbf{a}^T E(\mathbf{X}Y) \\
=& EY^2+\mathbf{a}^T\Gamma\mathbf{a}-2\mathbf{a}^T\Gamma\mathbf{a} \\
=& EY^2-\mathbf{a}^T\Gamma\mathbf{a}.
\end{aligned}
$$

性质1—III

- 注意：$\mathbf{a}$是预测方程的解等价于

$$\mathrm{E}((Y - \mathbf{a}^T\mathbf{X})\mathbf{X}) = \Sigma_{\mathbf{X}Y} - \Gamma\mathbf{a} = 0$$

即$Y - \mathbf{a}^T\mathbf{X}$与$\mathbf{X}$正交。(注意$\mathbf{a}^T\mathbf{X}$是标量)

- 性质说明$Y - \mathbf{a}^T\mathbf{X}$与$\mathbf{X}$正交则$\mathbf{a}^T\mathbf{X} = L(Y|\mathbf{X})$。

性质1—III

- 注意：$\mathbf{a}$是预测方程的解等价于

$$\mathrm{E}((Y-\mathbf{a}^T\mathbf{X})\mathbf{X}) = \Sigma_{\mathbf{X}Y} - \Gamma\mathbf{a} = 0$$

  即$Y-\mathbf{a}^T\mathbf{X}$与$\mathbf{X}$正交。(注意$\mathbf{a}^T\mathbf{X}$是标量)
- 性质说明$Y-\mathbf{a}^T\mathbf{X}$与$\mathbf{X}$正交则$\mathbf{a}^T\mathbf{X} = L(Y|\mathbf{X})$。

性质2

- **性质2.**

(1) 如果$\Gamma = E(\mathbf{X}\mathbf{X}^T)$可逆，则$\mathbf{a} = \Gamma^{-1}E(\mathbf{X}Y)$ 使得$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X}$。

(2) 预测方程$\Gamma\mathbf{a} = E(\mathbf{X}Y)$总有解.

(3) 如果$\det(\Gamma) = 0$，取正交矩阵A使得

$$A\Gamma A^T = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, 0, \ldots, 0), \quad \lambda_j > 0, j = 1, \ldots, r.$$

定义$\mathbf{Z} = A\mathbf{X} = (Z_1, Z_2, \ldots, Z_n)^T = (Z_1, Z_2, \ldots, Z_r, 0, \ldots, 0)^T$ 和$\xi = (Z_1, Z_2, \cdots, Z_r)^T$,则$E(\xi\xi^T)$正定，并且当取

$$\alpha = [E(\xi\xi^T)]^{-1}E(\xi Y) \qquad (1.7)$$

时,$L(Y|\mathbf{X}) = L(Y|\xi) = \alpha^T\xi$.

性质2

- **性质2.**

(1) 如果$\Gamma = E(\mathbf{X}\mathbf{X}^T)$可逆，则$\mathbf{a} = \Gamma^{-1}E(\mathbf{X}Y)$ 使得$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X}$。

(2) 预测方程$\Gamma\mathbf{a} = E(\mathbf{X}Y)$总有解.

(3) 如果$\det(\Gamma) = 0$，取正交矩阵A使得

$$A\Gamma A^T = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, 0, \ldots, 0), \quad \lambda_j > 0, j = 1, \ldots, r.$$

定义$\mathbf{Z} = A\mathbf{X} = (Z_1, Z_2, \ldots, Z_n)^T = (Z_1, Z_2, \ldots, Z_r, 0, \ldots, 0)^T$ 和$\xi = (Z_1, Z_2, \cdots, Z_r)^T$,则$E(\xi\xi^T)$正定，并且当取

$$\alpha = [E(\xi\xi^T)]^{-1}E(\xi Y) \qquad (1.7)$$

时,$L(Y|\mathbf{X}) = L(Y|\xi) = \alpha^T\xi$.

性质2

- **性质2.**
  (1) 如果$\Gamma = E(\mathbf{X}\mathbf{X}^T)$可逆，则$\mathbf{a} = \Gamma^{-1}E(\mathbf{X}Y)$ 使得$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X}$。
  (2) 预测方程$\Gamma\mathbf{a} = E(\mathbf{X}Y)$总有解.
  (3) 如果$\det(\Gamma) = 0$，取正交矩阵A使得

$$A\Gamma A^T = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, 0, \ldots, 0), \quad \lambda_j > 0, j = 1, \ldots, r.$$

定义$\mathbf{Z} = A\mathbf{X} = (Z_1, Z_2, \ldots, Z_n)^T = (Z_1, Z_2, \ldots, Z_r, 0, \ldots, 0)^T$ 和$\xi = (Z_1, Z_2, \cdots, Z_r)^T$,则$E(\xi\xi^T)$正定，并且当取

$$\alpha = [E(\xi\xi^T)]^{-1}E(\xi Y) \tag{1.7}$$

时,$L(Y|\mathbf{X}) = L(Y|\xi) = \alpha^T\xi$.

性质2

- **性质2.**
  (1) 如果$\Gamma = E(\mathbf{X}\mathbf{X}^T)$可逆，则$\mathbf{a} = \Gamma^{-1}E(\mathbf{X}Y)$ 使得$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X}$。
  (2) 预测方程$\Gamma\mathbf{a} = E(\mathbf{X}Y)$总有解.
  (3) 如果$\det(\Gamma) = 0$，取正交矩阵A使得

  $$A\Gamma A^T = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, 0, \ldots, 0), \quad \lambda_j > 0, j = 1, \ldots, r.$$

  定义$\mathbf{Z} = A\mathbf{X} = (Z_1, Z_2, \ldots, Z_n)^T = (Z_1, Z_2, \ldots, Z_r, 0, \ldots, 0)^T$ 和$\boldsymbol{\xi} = (Z_1, Z_2, \cdots, Z_r)^T$,则$E(\boldsymbol{\xi}\boldsymbol{\xi}^T)$正定，并且当取

  $$\boldsymbol{\alpha} = [E(\boldsymbol{\xi}\boldsymbol{\xi}^T)]^{-1}E(\boldsymbol{\xi}Y) \tag{1.7}$$

  时,$L(Y|\mathbf{X}) = L(Y|\boldsymbol{\xi}) = \boldsymbol{\alpha}^T\boldsymbol{\xi}$.

性质2—II

- 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，$L(Y|\mathbf{X})$可以通过$\mathbf{X}$ 的基表示。

性质2—II

- 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，$L(Y|\mathbf{X})$可以通过$\mathbf{X}$ 的基表示。

性质2—III

证明: 仅需证明$\det(\Gamma)=0$时第三和第二条成立。

$$\begin{aligned}\mathrm{E}(\mathbf{Z}\mathbf{Z}^T) =& \mathrm{E}(A\mathbf{X}\mathbf{X}^TA^T)=A\Gamma A^T\\ =& \mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_r,0,\ldots,0)\end{aligned}$$

故$Z_{r+1}=\cdots=Z_n=0$。且$\mathrm{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T)=\mathrm{diag}(\lambda_1,\ldots,\lambda_r)$ 是正定阵。当$\boldsymbol{\alpha}$按(1.7)定义时,有

$$\begin{pmatrix}\lambda_1 & \cdots & 0\\ & \ddots & \\ 0 & \cdots & \lambda_r\end{pmatrix}\boldsymbol{\alpha}=E(\boldsymbol{\xi}Y).$$

性质2—IV

注意$A\Gamma A^T$与$\lambda_1,\ldots,\lambda_r$的关系，可以导出

$$\begin{aligned} A\Gamma A^T \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix} &= \mathrm{diag}(\lambda_1,\ldots,\lambda_r,0,\ldots,0) \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{E}(\boldsymbol{\xi} Y) \\ \mathbf{0} \end{pmatrix} = \mathrm{E}\left(\begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{0} \end{pmatrix} Y \right) \\ &= \mathrm{E}(\mathbf{Z}Y) = \mathrm{E}(A\mathbf{X}Y) = A\mathrm{E}(\mathbf{X}Y) = A\Sigma_{\mathbf{X}Y} \end{aligned}$$

两边同乘以A^T，记$\mathbf{a} = A^T \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix}$有

$$\Gamma \mathbf{a} = \Sigma_{\mathbf{X}Y}$$

由性质1知

$$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X} = (\boldsymbol{\alpha}^T,0,\ldots,0)A\mathbf{X} = \boldsymbol{\alpha}^T\boldsymbol{\xi}$$

这在证明了第三条的同时也证明了第二条。

性质3

- **性质3.** 尽管$\mathbf{a}$由$\Gamma\mathbf{a} = E(\mathbf{X}Y)$决定时可以不惟一，但$L(Y|\mathbf{X})$总是(a.s.)惟一的.
- **证明：** 预测方程总有解，设$\mathbf{a}$为预测方程的一个解，则由性质1对$\forall \mathbf{b} \in \mathbb{R}^n$有

$$\mathrm{E}(Y - \mathbf{b}^T\mathbf{X})^2 = \mathrm{E}(Y - \mathbf{a}^T\mathbf{X})^2 + \mathrm{E}((\mathbf{a} - \mathbf{b})^T\mathbf{X})^2$$

若还有$\mathbf{b}$使得$L(Y|\mathbf{X}) = \mathbf{b}^T\mathbf{X}$则$\mathrm{E}(Y - \mathbf{b}^T\mathbf{X})^2 = \mathrm{E}(Y - \mathbf{a}^T\mathbf{X})^2$，于是$\mathrm{E}((\mathbf{a} - \mathbf{b})^T\mathbf{X})^2 = 0$，即$\mathbf{b}^T\mathbf{X} = \mathbf{a}^T\mathbf{X}$，a.s.

性质3

- **性质3.** 尽管$\mathbf{a}$由$\Gamma\mathbf{a}=E(\mathbf{X}Y)$决定时可以不惟一，但$L(Y|\mathbf{X})$总是(a.s.)惟一的.
- **证明**: 预测方程总有解，设$\mathbf{a}$为预测方程的一个解，则由性质1对$\forall\mathbf{b}\in\mathbb{R}^n$有

$$\mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2=\mathrm{E}(Y-\mathbf{a}^T\mathbf{X})^2+\mathrm{E}((\mathbf{a}-\mathbf{b})^T\mathbf{X})^2$$

  若还有$\mathbf{b}$使得$L(Y|\mathbf{X})=\mathbf{b}^T\mathbf{X}$则$\mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2=\mathrm{E}(Y-\mathbf{a}^T\mathbf{X})^2$，于是$\mathrm{E}((\mathbf{a}-\mathbf{b})^T\mathbf{X})^2=0$, 即$\mathbf{b}^T\mathbf{X}=\mathbf{a}^T\mathbf{X}$，a.s.

性质4

- **性质4.**
  (1) 如果$E(\mathbf{X}Y)=0$,则$L(Y|\mathbf{X})=0$.
  (2) 如果$Y=\sum_{j=1}^{n} b_j X_j$,则$L(Y|\mathbf{X})=Y$.
- 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) $\forall \mathbf{b}$

$$\begin{aligned}\mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2 =&\mathrm{E}Y^2+\mathbf{b}^T\Gamma\mathbf{b}-2\mathbf{b}^T\mathrm{E}(\mathbf{X}Y)\\ =&\mathrm{E}Y^2+\mathbf{b}^T\Gamma\mathbf{b}\geq \mathrm{E}Y^2=\mathrm{E}(Y-0)^2\end{aligned}$$

  所以$L(Y|X)=0$。
  (2) 这时$Y=\mathbf{b}^T\mathbf{X}$, $\mathrm{E}(Y-\mathbf{b}\mathbf{X})^2=0$，所以$L(Y|\mathbf{X})=\mathbf{b}^T\mathbf{X}=Y$。

性质4

- **性质4.**
  (1) 如果$E(\mathbf{X}Y)=0$,则$L(Y|\mathbf{X})=0$.
  (2) 如果$Y=\sum_{j=1}^{n} b_jX_j$,则$L(Y|\mathbf{X})=Y$.
- 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) $\forall \mathbf{b}$

$$
\begin{aligned}
\mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2 =& \mathrm{E}Y^2 + \mathbf{b}^T\Gamma\mathbf{b} - 2\mathbf{b}^T\mathrm{E}(\mathbf{X}Y) \\
=& \mathrm{E}Y^2 + \mathbf{b}^T\Gamma\mathbf{b} \geq \mathrm{E}Y^2 = \mathrm{E}(Y-0)^2
\end{aligned}
$$

  所以$L(Y|X)=0$。
  (2) 这时$Y=\mathbf{b}^T\mathbf{X}$, $\mathrm{E}(Y-\mathbf{b}\mathbf{X})^2=0$，所以$L(Y|\mathbf{X})=\mathbf{b}^T\mathbf{X}=Y$。

性质4

- **性质4.**
  (1) 如果$E(\mathbf{X}Y)=0$,则$L(Y|\mathbf{X})=0$.
  (2) 如果$Y=\sum_{j=1}^{n} b_j X_j$,则$L(Y|\mathbf{X})=Y$.
- 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) $\forall \mathbf{b}$

$$
\begin{aligned}
E(Y-\mathbf{b}^T\mathbf{X})^2 &= EY^2+\mathbf{b}^T\Gamma\mathbf{b}-2\mathbf{b}^T E(\mathbf{X}Y) \\
&= EY^2+\mathbf{b}^T\Gamma\mathbf{b}\ge EY^2=E(Y-0)^2
\end{aligned}
$$

  所以$L(Y|X)=0$。
  (2) 这时$Y=\mathbf{b}^T\mathbf{X}$, $E(Y-\mathbf{b}\mathbf{X})^2=0$，所以$L(Y|\mathbf{X})=\mathbf{b}^T\mathbf{X}=Y$。

性质4

- **性质4.**
  (1) 如果$E(\mathbf{X}Y)=0$,则$L(Y|\mathbf{X})=0$.
  (2) 如果$Y=\sum_{j=1}^{n} b_j X_j$,则$L(Y|\mathbf{X})=Y$.
- 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) $\forall \mathbf{b}$

$$
\begin{aligned}
E(Y-\mathbf{b}^T\mathbf{X})^2 =& EY^2+\mathbf{b}^T\Gamma\mathbf{b}-2\mathbf{b}^T E(\mathbf{X}Y) \\
=& EY^2+\mathbf{b}^T\Gamma\mathbf{b} \geq EY^2 = E(Y-0)^2
\end{aligned}
$$

  所以$L(Y|X)=0$。
  (2) 这时$Y=\mathbf{b}^T\mathbf{X}$, $E(Y-\mathbf{b}\mathbf{X})^2=0$，所以$L(Y|\mathbf{X})=\mathbf{b}^T\mathbf{X}=Y$。

性质4

- **性质4.**
  (1) 如果$E(\mathbf{X}Y)=0$,则$L(Y|\mathbf{X})=0$.
  (2) 如果$Y=\sum_{j=1}^{n}b_jX_j$,则$L(Y|\mathbf{X})=Y$.
- 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) $\forall \mathbf{b}$

$$\begin{aligned}\mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2 =&\mathrm{E}Y^2+\mathbf{b}^T\Gamma\mathbf{b}-2\mathbf{b}^T\mathrm{E}(\mathbf{X}Y)\\ =&\mathrm{E}Y^2+\mathbf{b}^T\Gamma\mathbf{b}\geq \mathrm{E}Y^2=\mathrm{E}(Y-0)^2\end{aligned}$$

  所以$L(Y|X)=0$。
  (2) 这时$Y=\mathbf{b}^T\mathbf{X}$, $\mathrm{E}(Y-\mathbf{b}\mathbf{X})^2=0$，所以$L(Y|\mathbf{X})=\mathbf{b}^T\mathbf{X}=Y$。

性质4

- **性质4.**
  (1) 如果$E(\mathbf{X}Y)=0$,则$L(Y|\mathbf{X})=0$.
  (2) 如果$Y=\sum_{j=1}^{n} b_j X_j$,则$L(Y|\mathbf{X})=Y$.
- 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) $\forall \mathbf{b}$

$$
\begin{aligned}
\mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2 =& \mathrm{E}Y^2+\mathbf{b}^T\Gamma\mathbf{b}-2\mathbf{b}^T\mathrm{E}(\mathbf{X}Y) \\
=& \mathrm{E}Y^2+\mathbf{b}^T\Gamma\mathbf{b} \geq \mathrm{E}Y^2=\mathrm{E}(Y-0)^2
\end{aligned}
$$

  所以$L(Y|X)=0$。
  (2) 这时$Y=\mathbf{b}^T\mathbf{X}$, $\mathrm{E}(Y-\mathbf{b}\mathbf{X})^2=0$，所以$L(Y|\mathbf{X})=\mathbf{b}^T\mathbf{X}=Y$。

性质4

- **性质4.**
  (1) 如果$E(\mathbf{X}Y)=0$,则$L(Y|\mathbf{X})=0$.
  (2) 如果$Y=\sum_{j=1}^{n} b_j X_j$,则$L(Y|\mathbf{X})=Y$.
- 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) $\forall \mathbf{b}$

$$\begin{aligned}\mathrm{E}(Y-\mathbf{b}^T\mathbf{X})^2 =&\mathrm{E}Y^2+\mathbf{b}^T\Gamma\mathbf{b}-2\mathbf{b}^T\mathrm{E}(\mathbf{X}Y)\\ =&\mathrm{E}Y^2+\mathbf{b}^T\Gamma\mathbf{b}\geq \mathrm{E}Y^2=\mathrm{E}(Y-0)^2\end{aligned}$$

  所以$L(Y|X)=0$。
  (2) 这时$Y=\mathbf{b}^T\mathbf{X}$, $\mathrm{E}(Y-\mathbf{b}\mathbf{X})^2=0$，所以$L(Y|\mathbf{X})=\mathbf{b}^T\mathbf{X}=Y$。

性质5

- 设$Y_1, Y_2, \cdots, Y_m$是随机变量, b_j是常数. 如果$Y = \sum_{i=1}^{m} b_i Y_i$, 则

$$L(Y|\mathbf{X}) = \sum_{j=1}^{m} b_j L(Y_j|\mathbf{X}).$$

- 性质5说明求最佳线性预测的运算$L(\cdot|\mathbf{X})$是一种线性运算.

性质5

- 设$Y_1, Y_2, \cdots, Y_m$是随机变量，b_j是常数．如果$Y = \sum_{i=1}^{m} b_i Y_i$，则

$$L(Y|\mathbf{X}) = \sum_{j=1}^{m} b_j L(Y_j|\mathbf{X}).$$

- 性质5说明求最佳线性预测的运算$L(\cdot|\mathbf{X})$是一种线性运算．

性质5—II

- **证明**: 设$\mathbf{a}_i$为$\Gamma\mathbf{a}_i = \mathrm{E}(\mathbf{X}Y_i)$的解$(i = 1, 2, \ldots, m)$，则$L(Y_i|\mathbf{X}) = \mathbf{a}_i^T\mathbf{X}$。取$\mathbf{a} = \sum_{i=1}^m b_i\mathbf{a}_i$，则

$$\begin{aligned}\Gamma\mathbf{a} =& \Gamma\left(\sum_{i=1}^m b_i\mathbf{a}_i\right) = \sum_{i=1}^m b_i(\Gamma\mathbf{a}_i) = \sum_{i=1}^m b_i\mathrm{E}(\mathbf{X}Y_i)\\ =& \mathrm{E}\left(\mathbf{X}\sum_{i=1}^m b_iY_i\right) = \mathrm{E}(\mathbf{X}Y)\end{aligned}$$

由性质1即知

$$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X} = \sum_{i=1}^m b_i\mathbf{a}_i^T\mathbf{X} = \sum_{i=1}^m b_iL(Y_i|\mathbf{X})$$

性质6

- **性质6.** 设$\mathbf{X} = (X_1, X_2, \cdots, X_n)^T$, $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_m)^T$. 如果$E(\mathbf{X}\mathbf{Z}^T) = 0$($\mathbf{X}$与$\mathbf{Z}$不相关), 则有

$$L(Y|\mathbf{X}, \mathbf{Z}) = L(Y|\mathbf{X}) + L(Y|\mathbf{Z}).$$

性质6—II

- **证明:** 记$\xi = \begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix}$，记$\Sigma_{XX} = \mathrm{E}(\mathbf{X}\mathbf{X}^T)$, $\Sigma_{ZZ} = \mathrm{E}(\mathbf{Z}\mathbf{Z}^T)$,

  则

$$\Sigma \triangleq \mathrm{E}(\xi\xi^T) = \begin{pmatrix} \Sigma_{XX} & \mathbf{0} \\ \mathbf{0} & \Sigma_{ZZ} \end{pmatrix}$$

  设$\mathbf{a}$, $\mathbf{b}$使得$\Sigma_{XX}\mathbf{a} = \mathrm{E}(\mathbf{X}Y)$, $\Sigma_{ZZ}\mathbf{b} = \mathrm{E}(\mathbf{Z}Y)$,

  则$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X}$，$L(Y|\mathbf{Z}) = \mathbf{b}^T\mathbf{Z}$, 取$\mathbf{c} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$，则

$$\begin{aligned} \Sigma\mathbf{c} = & \begin{pmatrix} \Sigma_{XX}\mathbf{a} \\ \Sigma_{ZZ}\mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathrm{E}(\mathbf{X}Y) \\ \mathrm{E}(\mathbf{Z}Y) \end{pmatrix} \\ = & \mathrm{E}\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix} Y\right) = \mathrm{E}(\xi Y) \end{aligned}$$

  由性质1

$$\begin{aligned} L(Y|\mathbf{X}, \mathbf{Z}) = & L(Y|\xi) = \mathbf{c}^T\xi \\ = & \mathbf{a}^T\mathbf{X} + \mathbf{b}^T\mathbf{Z} = L(Y|\mathbf{X}) + L(Y|\mathbf{Z}) \end{aligned}$$

性质7

- **性质7** 设$\tilde{Y} = \mathbf{b}^T\mathbf{X}$是$\mathbf{X}$的线性组合，则$\tilde{Y} = L(Y|\mathbf{X})$的充分必要条件是

$$E(X_j(Y - \tilde{Y})) = 0,\ 1 \le j \le n. \tag{1.8}$$

即

$$\mathrm{E}(\mathbf{X}(Y - \mathbf{b}^T\mathbf{X})) = 0$$

- $L(Y|\mathbf{X})$可以看成Y在$\mathbf{X}$张成的空间上的投影，此性质即投影应满足的性质。
- 注意

$$\mathrm{E}(\mathbf{X}(Y - \mathbf{b}^T\mathbf{X})) = \mathrm{E}(\mathbf{X}Y) - \Gamma\mathbf{b}.$$

即残差与自变量正交等价于系数**b**满足预测方程。

性质7

- **性质7** 设$\tilde{Y} = \mathbf{b}^T\mathbf{X}$是$\mathbf{X}$的线性组合，则$\tilde{Y} = L(Y|\mathbf{X})$的充分必要条件是

$$E(X_j(Y - \tilde{Y})) = 0,\ 1 \le j \le n. \tag{1.8}$$

即

$$\mathrm{E}(\mathbf{X}(Y - \mathbf{b}^T\mathbf{X})) = 0$$

- $L(Y|\mathbf{X})$可以看成Y在$\mathbf{X}$张成的空间上的投影，此性质即投影应满足的性质。
- 注意

$$\mathrm{E}(\mathbf{X}(Y - \mathbf{b}^T\mathbf{X})) = \mathrm{E}(\mathbf{X}Y) - \Gamma\mathbf{b}.$$

即残差与自变量正交等价于系数$\mathbf{b}$满足预测方程。

性质7

- **性质7**　设$\tilde{Y}=\mathbf{b}^T\mathbf{X}$是$\mathbf{X}$的线性组合，则$\tilde{Y}=L(Y|\mathbf{X})$的充分必要条件是

$$E(X_j(Y-\tilde{Y}))=0,\ 1\le j\le n. \tag{1.8}$$

即

$$\mathrm{E}(\mathbf{X}(Y-\mathbf{b}^T\mathbf{X}))=0$$

- $L(Y|\mathbf{X})$可以看成Y在$\mathbf{X}$张成的空间上的投影，此性质即投影应满足的性质。
- 注意

$$\mathrm{E}(\mathbf{X}(Y-\mathbf{b}^T\mathbf{X}))=\mathrm{E}(\mathbf{X}Y)-\Gamma\mathbf{b}.$$

即残差与自变量正交等价于系数**b**满足预测方程。

性质7—II

- **证明:**
  - **必要性:** 设$\tilde{Y}$为$L(Y|\mathbf{X})$，由性质2知存在$\mathbf{a}$满足预测方程，由性质1和性质3知$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X} = \mathbf{b}^T\mathbf{X}$。两边右乘以$\mathbf{X}^T$取期望得
  $$\mathbf{a}^T\Gamma = \mathbf{b}^T\Gamma$$
  注意$\Gamma\mathbf{a} = \mathrm{E}(\mathbf{X}Y)$所以由上式得$\Gamma\mathbf{b} = \mathrm{E}(\mathbf{X}Y)$, 即条件成立。
  - **充分性:** 条件成立时$\mathbf{b}$是预测方程的解，由性质1即知$\tilde{Y} = \mathbf{b}^T\mathbf{X}$是最佳线性预测。

性质7—II

- 证明:
  - 必要性: 设$\tilde{Y}$为$L(Y|\mathbf{X})$，由性质2知存在$\mathbf{a}$满足预测方程，由性质1和性质3知$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X} = \mathbf{b}^T\mathbf{X}$。两边右乘以$\mathbf{X}^T$取期望得

$$\mathbf{a}^T\Gamma = \mathbf{b}^T\Gamma$$

    注意$\Gamma\mathbf{a} = \mathrm{E}(\mathbf{X}Y)$所以由上式得$\Gamma\mathbf{b} = \mathrm{E}(\mathbf{X}Y)$, 即条件成立。
  - 充分性: 条件成立时$\mathbf{b}$是预测方程的解，由性质1即知$\tilde{Y} = \mathbf{b}^T\mathbf{X}$是最佳线性预测。

性质7—II

- **证明**:
  - **必要性**: 设$\tilde{Y}$为$L(Y|\mathbf{X})$，由性质2知存在$\mathbf{a}$满足预测方程，由性质1和性质3知$L(Y|\mathbf{X}) = \mathbf{a}^T\mathbf{X} = \mathbf{b}^T\mathbf{X}$。两边右乘以$\mathbf{X}^T$取期望得

$$\mathbf{a}^T\Gamma = \mathbf{b}^T\Gamma$$

    注意$\Gamma\mathbf{a} = \mathrm{E}(\mathbf{X}Y)$所以由上式得$\Gamma\mathbf{b} = \mathrm{E}(\mathbf{X}Y)$, 即条件成立。
  - **充分性**: 条件成立时$\mathbf{b}$是预测方程的解，由性质1即知$\tilde{Y} = \mathbf{b}^T\mathbf{X}$是最佳线性预测。

性质8

- **性质8.** 如果

$$\hat{Y} = L(Y|X_1, X_2, \cdots, X_n),$$
$$\tilde{Y} = L(Y|X_1, X_2, \cdots, X_{n-1}),$$

则有

$$L(\hat{Y}|X_1, X_2, \cdots, X_{n-1}) = \tilde{Y},$$

并且有

$$E(Y-\hat{Y})^2 \le E(Y-\tilde{Y})^2. \tag{1.9}$$

- (1.9)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.

性质8

- **性质8.** 如果

$$\hat{Y} = L(Y|X_1, X_2, \cdots, X_n),$$
$$\tilde{Y} = L(Y|X_1, X_2, \cdots, X_{n-1}),$$

则有

$$L(\hat{Y}|X_1, X_2, \cdots, X_{n-1}) = \tilde{Y},$$

并且有

$$E(Y - \hat{Y})^2 \leq E(Y - \tilde{Y})^2. \quad (1.9)$$

- (1.9)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.

性质8—II

- 证明 $Y_0 \triangleq L(\hat{Y}|X_1, X_2, \ldots, X_{n-1})$
  是$X_1, X_2, \ldots, X_{n-1}$的线性组合，利用$Y-\hat{Y}$，$\hat{Y}-Y_0$都和$X_1, \ldots, X_{n-1}$正交，得到

  $$Y - Y_0 = (Y - \hat{Y}) + (\hat{Y} - Y_0)$$

  和$X_1, \ldots, X_{n-1}$正交. 利用性质7即
  知$Y_0 = L(Y|X_1, \ldots, X_{n-1}) = \tilde{Y}$.
- $\hat{Y}$是$X_1, \ldots, X_n$对Y的最佳线性预测
  而$\tilde{Y}$是$X_1, \ldots, X_{n-1}$的一个线性组合所以有(1.9)成立。
- 这个性质实际是投影的性质。

性质8—II

- 证明　$Y_0 \triangleq L(\hat{Y}|X_1, X_2, \ldots, X_{n-1})$
是$X_1, X_2, \ldots, X_{n-1}$的线性组合，利用$Y-\hat{Y}$，$\hat{Y}-Y_0$都和$X_1, \ldots, X_{n-1}$正交，得到

$$Y - Y_0 = (Y - \hat{Y}) + (\hat{Y} - Y_0)$$

和$X_1, \ldots, X_{n-1}$正交．利用性质7即知$Y_0 = L(Y|X_1, \ldots, X_{n-1}) = \tilde{Y}$.
- $\hat{Y}$是$X_1, \ldots, X_n$对Y的最佳线性预测
而$\tilde{Y}$是$X_1, \ldots, X_{n-1}$的一个线性组合所以有(1.9)成立。
- 这个性质实际是投影的性质。

性质8—II

- 证明 $Y_0 \triangleq L(\hat{Y}|X_1, X_2, \ldots, X_{n-1})$
  是$X_1, X_2, \ldots, X_{n-1}$的线性组合，利用$Y - \hat{Y}$，$\hat{Y} - Y_0$都和$X_1, \ldots, X_{n-1}$正交，得到

$$Y - Y_0 = (Y - \hat{Y}) + (\hat{Y} - Y_0)$$

  和$X_1, \ldots, X_{n-1}$正交．利用性质7即知$Y_0 = L(Y|X_1, \ldots, X_{n-1}) = \tilde{Y}$.
- $\hat{Y}$是$X_1, \ldots, X_n$对Y的最佳线性预测而$\tilde{Y}$是$X_1, \ldots, X_{n-1}$的一个线性组合所以有(1.9)成立。
- 这个性质实际是投影的性质。

性质9

- **性质9.**(非零均值的最佳线性预测的意义)　如果$EY=b$, $E\mathbf{X}=\mu$, 按定义$L(Y|\mathbf{X})=b+L(Y-b|\mathbf{X}-\mu)$。事实上对任何$c_0\in\mathbb{R}$, $\mathbf{c}\in\mathbb{R}^n$,

$$E[Y-L(Y|\mathbf{X})]^2\le E[Y-(c_0+\mathbf{c}^T\mathbf{X})]^2. \tag{1.10}$$

- 证明　设$L(Y|\mathbf{X})=b+\mathbf{a}^T(\mathbf{X}-\mu)$, 则

$$\begin{aligned}
&\mathrm{E}\left\{Y-c_0-\mathbf{c}^T\mathbf{X}\right\}^2\\
=&\mathrm{E}\left\{Y-b-\mathbf{a}^T(\mathbf{X}-\mu)+b+\mathbf{a}^T(\mathbf{X}-\mu)\right.\\
&\left.-\left[c_0+\mathbf{c}^T\mu+\mathbf{c}^T(\mathbf{X}-\mu)\right]\right\}^2\\
=&\mathrm{E}\left\{[Y-b-\mathbf{a}^T(\mathbf{X}-\mu)]+(b-c_0-\mathbf{c}^T\mu)+(\mathbf{a}-\mathbf{c})^T(\mathbf{X}-\mu)\right\}^2\\
=&\mathrm{E}[Y-b-\mathbf{a}^T(\mathbf{X}-\mu)]^2+(b-c_0-\mathbf{c}^T\mu)^2+(\mathbf{a}-\mathbf{c})^T\Gamma(\mathbf{a}-\mathbf{c})\\
\ge&\mathrm{E}[Y-b-\mathbf{a}^T(\mathbf{X}-\mu)]^2=\mathrm{E}[Y-L(Y|\mathbf{X})]^2
\end{aligned}$$

性质9

- **性质9.**(非零均值的最佳线性预测的意义)　如果$EY=b$, $E\mathbf{X}=\boldsymbol{\mu}$, 按定义$L(Y|\mathbf{X})=b+L(Y-b|\mathbf{X}-\boldsymbol{\mu})$。事实上对任何$c_0\in\mathbb{R}$, $\mathbf{c}\in\mathbb{R}^n$,

$$E[Y-L(Y|\mathbf{X})]^2\le E[Y-(c_0+\mathbf{c}^T\mathbf{X})]^2. \tag{1.10}$$

- **证明**　设$L(Y|\mathbf{X})=b+\mathbf{a}^T(\mathbf{X}-\boldsymbol{\mu})$, 则

$$\begin{aligned}
&\mathrm{E}\left\{Y-c_0-\mathbf{c}^T\mathbf{X}\right\}^2\\
=&\mathrm{E}\left\{Y-b-\mathbf{a}^T(\mathbf{X}-\boldsymbol{\mu})+b+\mathbf{a}^T(\mathbf{X}-\boldsymbol{\mu})\right.\\
&\left.-\left[c_0+\mathbf{c}^T\boldsymbol{\mu}+\mathbf{c}^T(\mathbf{X}-\boldsymbol{\mu})\right]\right\}^2\\
=&\mathrm{E}\left\{[Y-b-\mathbf{a}^T(\mathbf{X}-\boldsymbol{\mu})]+(b-c_0-\mathbf{c}^T\boldsymbol{\mu})+(\mathbf{a}-\mathbf{c})^T(\mathbf{X}-\boldsymbol{\mu})\right\}^2\\
=&\mathrm{E}[Y-b-\mathbf{a}^T(\mathbf{X}-\boldsymbol{\mu})]^2+(b-c_0-\mathbf{c}^T\boldsymbol{\mu})^2+(\mathbf{a}-\mathbf{c})^T\Gamma(\mathbf{a}-\mathbf{c})\\
\ge&\mathrm{E}[Y-b-\mathbf{a}^T(\mathbf{X}-\boldsymbol{\mu})]^2=\mathrm{E}[Y-L(Y|\mathbf{X})]^2
\end{aligned}$$

性质10

- **性质10.** 设$\mathbf{X}$ 和$\mathbf{Z}$ 分别是m和n维向量，如果有实矩阵A, B使得$\mathbf{X} = A\mathbf{Z}$, $\mathbf{Z} = B\mathbf{X}$, 则$L(Y|\mathbf{X}) = L(Y|\mathbf{Z})$.
- 如果$\mathbf{X}$和$\mathbf{Z}$能互相线性表示则其预报Y能达到的下界是相同的，预报是一致的。
- 证明为习题。

性质10

- **性质10.** 设$\mathbf{X}$ 和$\mathbf{Z}$ 分别是m和n维向量，如果有实矩阵A, B使得$\mathbf{X} = A\mathbf{Z}$, $\mathbf{Z} = B\mathbf{X}$, 则$L(Y|\mathbf{X}) = L(Y|\mathbf{Z})$.
- 如果$\mathbf{X}$和$\mathbf{Z}$能互相线性表示则其预报Y能达到的下界是相同的，预报是一致的。
- 证明为习题。

性质10

- **性质10.** 设$\mathbf{X}$ 和$\mathbf{Z}$ 分别是m和n维向量，如果有实矩阵A, B使得$\mathbf{X} = A\mathbf{Z}$, $\mathbf{Z} = B\mathbf{X}$, 则$L(Y|\mathbf{X}) = L(Y|\mathbf{Z})$.
- 如果$\mathbf{X}$和$\mathbf{Z}$能互相线性表示则其预报Y能达到的下界是相同的，预报是一致的。
- 证明为习题。

预测计算例子

- 设已知ARMA(4,2)的参数。观测到$\mathbf{x}_n = (x_1, \ldots, x_n)^T$， $n = 14$。
- 用$\mathbf{x}_n$预测x_{n+k}, $k = 1, 2, \ldots, m$, $m = 7$。
- 预测方程中Γ为平稳序列的Γ_n。
- 预测方程中Σ_{XY}当$Y = X_{n+k}$时为

$$\mathbf{g}_k = \mathrm{E}(\mathbf{X}_n X_{n+k}) = (\gamma_{n+k-1}, \gamma_{n+k-2}, \ldots, \gamma_k)^T$$

- 最佳线性预测为

$$\hat{X}_{n+k} \triangleq L(X_{n+k}|\mathbf{X}_n) = (\Gamma_n^{-1}\mathbf{g}_k)^T\mathbf{X}_n = \mathbf{g}_k^T\Gamma_n^{-1}\mathbf{X}_n$$

- 预测方差为

$$\sigma^2(k) = \gamma_0 - (\Gamma_n^{-1}\mathbf{g}_k)^T\Gamma_n(\Gamma_n^{-1}\mathbf{g}_k) = \gamma_0 - \mathbf{g}_k^T\Gamma_n^{-1}\mathbf{g}_k$$

预测计算例子

- 设已知ARMA(4,2)的参数。观测到$\mathbf{x}_n = (x_1, \ldots, x_n)^T$， $n = 14$。
- 用$\mathbf{x}_n$预测x_{n+k}，$k = 1, 2, \ldots, m$，$m = 7$。
- 预测方程中Γ为平稳序列的Γ_n。
- 预测方程中Σ_{XY}当$Y = X_{n+k}$时为

$$\mathbf{g}_k = \mathrm{E}(\mathbf{X}_n X_{n+k}) = (\gamma_{n+k-1}, \gamma_{n+k-2}, \ldots, \gamma_k)^T$$

- 最佳线性预测为

$$\hat{X}_{n+k} \triangleq L(X_{n+k}|\mathbf{X}_n) = (\Gamma_n^{-1}\mathbf{g}_k)^T \mathbf{X}_n = \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{X}_n$$

- 预测方差为

$$\sigma^2(k) = \gamma_0 - (\Gamma_n^{-1}\mathbf{g}_k)^T \Gamma_n (\Gamma_n^{-1}\mathbf{g}_k) = \gamma_0 - \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{g}_k$$

预测计算例子

- 设已知ARMA(4,2)的参数。观测到$\mathbf{x}_n = (x_1, \ldots, x_n)^T$，$n = 14$。
- 用$\mathbf{x}_n$预测x_{n+k}，$k = 1, 2, \ldots, m$，$m = 7$。
- 预测方程中Γ为平稳序列的Γ_n。
- 预测方程中Σ_{XY}当$Y = X_{n+k}$时为

$$\mathbf{g}_k = \mathrm{E}(\mathbf{X}_n X_{n+k}) = (\gamma_{n+k-1}, \gamma_{n+k-2}, \ldots, \gamma_k)^T$$

- 最佳线性预测为

$$\hat{X}_{n+k} \triangleq L(X_{n+k}|\mathbf{X}_n) = (\Gamma_n^{-1}\mathbf{g}_k)^T \mathbf{X}_n = \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{X}_n$$

- 预测方差为

$$\sigma^2(k) = \gamma_0 - (\Gamma_n^{-1}\mathbf{g}_k)^T \Gamma_n (\Gamma_n^{-1}\mathbf{g}_k) = \gamma_0 - \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{g}_k$$

预测计算例子

- 设已知ARMA(4,2)的参数。观测到$\mathbf{x}_n = (x_1, \ldots, x_n)^T$， $n = 14$。
- 用$\mathbf{x}_n$预测x_{n+k}，$k = 1, 2, \ldots, m$，$m = 7$。
- 预测方程中Γ为平稳序列的Γ_n。
- 预测方程中$\Sigma_{\mathbf{X}Y}$当$Y = X_{n+k}$时为

$$\mathbf{g}_k = \mathrm{E}(\mathbf{X}_n X_{n+k}) = (\gamma_{n+k-1}, \gamma_{n+k-2}, \ldots, \gamma_k)^T$$

- 最佳线性预测为

$$\hat{X}_{n+k} \triangleq L(X_{n+k}|\mathbf{X}_n) = (\Gamma_n^{-1}\mathbf{g}_k)^T \mathbf{X}_n = \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{X}_n$$

- 预测方差为

$$\sigma^2(k) = \gamma_0 - (\Gamma_n^{-1}\mathbf{g}_k)^T \Gamma_n (\Gamma_n^{-1}\mathbf{g}_k) = \gamma_0 - \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{g}_k$$

预测计算例子

- 设已知ARMA(4,2)的参数。观测到$\mathbf{x}_n = (x_1, \ldots, x_n)^T$， $n = 14$。
- 用$\mathbf{x}_n$预测x_{n+k}，$k = 1, 2, \ldots, m$，$m = 7$。
- 预测方程中Γ为平稳序列的Γ_n。
- 预测方程中$\Sigma_{\mathbf{X}Y}$当$Y = X_{n+k}$时为

$$\mathbf{g}_k = \mathrm{E}(\mathbf{X}_n X_{n+k}) = (\gamma_{n+k-1}, \gamma_{n+k-2}, \ldots, \gamma_k)^T$$

- 最佳线性预测为

$$\hat{X}_{n+k} \triangleq L(X_{n+k}|\mathbf{X}_n) = (\Gamma_n^{-1}\mathbf{g}_k)^T \mathbf{X}_n = \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{X}_n$$

- 预测方差为

$$\sigma^2(k) = \gamma_0 - (\Gamma_n^{-1}\mathbf{g}_k)^T \Gamma_n (\Gamma_n^{-1}\mathbf{g}_k) = \gamma_0 - \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{g}_k$$

预测计算例子

- 设已知ARMA(4,2)的参数。观测到$\mathbf{x}_n = (x_1, \ldots, x_n)^T$， $n = 14$。
- 用$\mathbf{x}_n$预测x_{n+k}，$k = 1, 2, \ldots, m$，$m = 7$。
- 预测方程中Γ为平稳序列的Γ_n。
- 预测方程中$\Sigma_{\mathbf{X}Y}$当$Y = X_{n+k}$时为

$$\mathbf{g}_k = \mathrm{E}(\mathbf{X}_n X_{n+k}) = (\gamma_{n+k-1}, \gamma_{n+k-2}, \ldots, \gamma_k)^T$$

- 最佳线性预测为

$$\hat{X}_{n+k} \stackrel{\triangle}{=} L(X_{n+k}|\mathbf{X}_n) = (\Gamma_n^{-1}\mathbf{g}_k)^T \mathbf{X}_n = \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{X}_n$$

- 预测方差为

$$\sigma^2(k) = \gamma_0 - (\Gamma_n^{-1}\mathbf{g}_k)^T \Gamma_n (\Gamma_n^{-1}\mathbf{g}_k) = \gamma_0 - \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{g}_k$$

预测计算例子II

- 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。$X_{n+k} - \hat{X}_{n+k} \sim \mathrm{N}(0, \sigma^2(k))$,
- 可以构造X_{n+k}的置信区间(预测区间):

$$\Pr(|X_{n+k} - \hat{X}_{n+k}|/\sigma(k) \leq 1.96) = 0.95, \quad k = 1, 2, \ldots, m$$

- 见演示。
- 对真实数据需要用$x_1, x_2, \ldots, x_N$估计$\hat{\gamma}_k$, 然后用令$\mathbf{x}_n = (x_{N-n+1}, x_{N-n+2}, \ldots, x_N)^T$, 用$\mathbf{x}_n$预报$X_{N+k}$, $k = 1, 2, \ldots$。
- 数据先减去均值再估计并预测，预测值要把均值加回去。

预测计算例子II

- 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。$X_{n+k} - \hat{X}_{n+k} \sim \mathrm{N}(0, \sigma^2(k))$,
- 可以构造X_{n+k}的置信区间(预测区间)：

$$\Pr(|X_{n+k} - \hat{X}_{n+k}|/\sigma(k) \le 1.96) = 0.95, \quad k = 1, 2, \dots, m$$

- 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$\mathbf{x}_n = (x_{N-n+1}, x_{N-n+2}, \dots, x_N)^T$, 用$\mathbf{x}_n$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。

预测计算例子II

- 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。$X_{n+k} - \hat{X}_{n+k} \sim \mathrm{N}(0, \sigma^2(k))$，
- 可以构造X_{n+k}的置信区间(预测区间)：

$$\Pr(|X_{n+k} - \hat{X}_{n+k}|/\sigma(k) \le 1.96) = 0.95, \quad k = 1, 2, \ldots, m$$

- 见演示。
- 对真实数据需要用$x_1, x_2, \ldots, x_N$估计$\hat{\gamma}_k$，然后用令$\mathbf{x}_n = (x_{N-n+1}, x_{N-n+2}, \ldots, x_N)^T$，用$\mathbf{x}_n$预报$X_{N+k}$，$k = 1, 2, \ldots$。
- 数据先减去均值再估计并预测，预测值要把均值加回去。

预测计算例子II

- 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。$X_{n+k} - \hat{X}_{n+k} \sim \mathrm{N}(0, \sigma^2(k))$,
- 可以构造X_{n+k}的置信区间(预测区间)：

$$\Pr(|X_{n+k} - \hat{X}_{n+k}|/\sigma(k) \le 1.96) = 0.95, \quad k = 1, 2, \dots, m$$

- 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$\mathbf{x}_n = (x_{N-n+1}, x_{N-n+2}, \dots, x_N)^T$, 用$\mathbf{x}_n$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。

预测计算例子II

- 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。$X_{n+k} - \hat{X}_{n+k} \sim \mathrm{N}(0, \sigma^2(k))$，
- 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$\mathbf{x}_n = (x_{N-n+1}, x_{N-n+2}, \dots, x_N)^T$，用$\mathbf{x}_n$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。

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§1.6).
- 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

$$\mathrm{E}(Y-\hat{Y})^2=\inf_{\xi\in H}\mathrm{E}(Y-\xi)^2 \tag{1.11}$$

- **定义1.3** 如果H是L^2的闭子空间, $Y\in L^2$, $\hat{Y}\in H$使得(1.11)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1.4** 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.

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§1.6).
- 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

$$\mathrm{E}(Y-\hat{Y})^2=\inf_{\xi\in H}\mathrm{E}(Y-\xi)^2 \tag{1.11}$$

- **定义1.3** 如果H是L^2的闭子空间, $Y\in L^2$, $\hat{Y}\in H$使得(1.11)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1.4** 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.

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§1.6).
- 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

$$\mathrm{E}(Y-\hat{Y})^2=\inf_{\xi\in H}\mathrm{E}(Y-\xi)^2 \tag{1.11}$$

- **定义1.3** 如果H是L^2的闭子空间, $Y\in L^2$, $\hat{Y}\in H$使得(1.11)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1.4** 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.

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§1.6).
- 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

$$\mathrm{E}(Y-\hat{Y})^2=\inf_{\xi\in H}\mathrm{E}(Y-\xi)^2 \tag{1.11}$$

- **定义1.3** 如果H是L^2的闭子空间, $Y\in L^2$, $\hat{Y}\in H$使得(1.11)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1.4** 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.

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§1.6).
- 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

$$\mathrm{E}(Y-\hat{Y})^2=\inf_{\xi\in H}\mathrm{E}(Y-\xi)^2 \tag{1.11}$$

- **定义1.3** 如果H是L^2的闭子空间，$Y\in L^2$，$\hat{Y}\in H$使得(1.11)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1.4** 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.

投影存在唯一的证明

- 取$Y_n \in H$使
$$d = \inf_{\xi \in H} \mathrm{E}(Y - \xi)^2 = \lim_{n\to\infty} \mathrm{E}(Y - Y_n)^2$$
则$(Y_n + Y_m)/2 \in H$, 并且当$n, m \to \infty$
-
$$\begin{aligned}
&E(Y_n - Y_m)^2 \\
=&E[(Y_n - Y) - (Y_m - Y)]^2 \\
&+ E[(Y_n - Y) + (Y_m - Y)]^2 \\
&- E[(Y_n + Y_m) - 2Y]^2 \\
=&2E(Y_n - Y)^2 + 2E(Y_m - Y)^2 - 4E[(Y_n + Y_m)/2 - Y]^2 \\
\leq&2E(Y_n - Y)^2 + 2E(Y_m - Y)^2 - 4d \\
\to&2d + 2d - 4d = 0. \qquad (1.12)
\end{aligned}$$

投影存在唯一的证明

- 取$Y_n \in H$使

$$d = \inf_{\xi \in H} \mathrm{E}(Y-\xi)^2 = \lim_{n\to\infty} \mathrm{E}(Y-Y_n)^2$$

则$(Y_n+Y_m)/2 \in H$, 并且当$n, m \to \infty$

-

$$\begin{aligned} & E(Y_n - Y_m)^2 \\ =& E[(Y_n - Y) - (Y_m - Y)]^2 \\ & + E[(Y_n - Y) + (Y_m - Y)]^2 \\ & - E[(Y_n + Y_m) - 2Y]^2 \\ =& 2E(Y_n - Y)^2 + 2E(Y_m - Y)^2 - 4E[(Y_n + Y_m)/2 - Y]^2 \\ \le& 2E(Y_n - Y)^2 + 2E(Y_m - Y)^2 - 4d \\ \to& 2d + 2d - 4d = 0. \end{aligned} \tag{1.12}$$

投影存在唯一的证明II

- 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(1.11).

- 如果又有$\hat{\xi} \in H$ 也使得(1.11)成立, 仿照(1.12)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..

投影存在唯一的证明II

- 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(1.11).

- 如果又有$\hat{\xi} \in H$ 也使得(1.11)成立, 仿照(1.12)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..

投影存在唯一的证明II

- 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(1.11).

- 如果又有$\hat{\xi} \in H$ 也使得(1.11)成立, 仿照(1.12)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..

投影的垂直性(正交性)

- **定理1.1** 设$Y \in L^2$, $\hat{Y} \in H$, 则$\hat{Y} = P_H(Y)$的充分必要条件是$(Y - \hat{Y})$垂直于H.
- 最佳线性预测性质7是定理1.1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

$$\begin{aligned} d \triangleq & E(Y - \hat{Y})^2 \leq E(Y - \hat{Y} - a\xi)^2 \\ = & E(Y - \hat{Y})^2 + E(a\xi)^2 - 2aE[(Y - \hat{Y})\xi] \\ = & d + a^2 - 2a^2 = d - a^2 \end{aligned}$$

由此得到$a = 0$.

投影的垂直性(正交性)

- **定理1.1** 设$Y \in L^2$, $\hat{Y} \in H$, 则$\hat{Y} = P_H(Y)$的充分必要条件是$(Y - \hat{Y})$垂直于H.
- 最佳线性预测性质7是定理1.1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

$$\begin{aligned} d \triangleq & E(Y - \hat{Y})^2 \leq E(Y - \hat{Y} - a\xi)^2 \\ = & E(Y - \hat{Y})^2 + E(a\xi)^2 - 2aE[(Y - \hat{Y})\xi] \\ = & d + a^2 - 2a^2 = d - a^2 \end{aligned}$$

由此得到$a = 0$.

投影的垂直性(正交性)

- **定理1.1** 设$Y \in L^2$, $\hat{Y} \in H$, 则$\hat{Y} = P_H(Y)$的充分必要条件是$(Y - \hat{Y})$垂直于H.
- 最佳线性预测性质7是定理1.1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$a = 0$.

投影的垂直性(正交性)II

- 来证明充分性。若$\hat{Y} \in H$使$Y - \hat{Y} \perp H$，则对$\forall \xi \in H$有
-
$$
\begin{aligned}
& \mathrm{E}(Y-\xi)^2 = \mathrm{E}(Y-\hat{Y}+\hat{Y}-\xi)^2 \\
= & \mathrm{E}(Y-\hat{Y})^2 + \mathrm{E}(\hat{Y}-\xi)^2 + 2\mathrm{E}[(Y-\hat{Y})(\hat{Y}-\xi)] \\
= & \mathrm{E}(Y-\hat{Y})^2 + \mathrm{E}(\hat{Y}-\xi)^2 \\
\geq & \mathrm{E}(Y-\hat{Y})^2
\end{aligned}
$$
- 即$\hat{Y} = P_H(Y)$。

投影的垂直性(正交性)II

- 来证明充分性。若$\hat{Y} \in H$使$Y - \hat{Y} \perp H$，则对$\forall \xi \in H$有
- $$\begin{aligned} & \mathrm{E}(Y-\xi)^2 = \mathrm{E}(Y-\hat{Y}+\hat{Y}-\xi)^2 \\ =& \mathrm{E}(Y-\hat{Y})^2 + \mathrm{E}(\hat{Y}-\xi)^2 + 2\mathrm{E}[(Y-\hat{Y})(\hat{Y}-\xi)] \\ =& \mathrm{E}(Y-\hat{Y})^2 + \mathrm{E}(\hat{Y}-\xi)^2 \\ \geq & \mathrm{E}(Y-\hat{Y})^2 \end{aligned}$$
- 即$\hat{Y} = P_H(Y)$。

投影的垂直性(正交性)II

- 来证明充分性。若$\hat{Y} \in H$使$Y - \hat{Y} \perp H$，则对$\forall \xi \in H$有
- $$\begin{aligned} & \mathrm{E}(Y-\xi)^2 = \mathrm{E}(Y-\hat{Y}+\hat{Y}-\xi)^2 \\ =& \mathrm{E}(Y-\hat{Y})^2 + \mathrm{E}(\hat{Y}-\xi)^2 + 2\mathrm{E}[(Y-\hat{Y})(\hat{Y}-\xi)] \\ =& \mathrm{E}(Y-\hat{Y})^2 + \mathrm{E}(\hat{Y}-\xi)^2 \\ \geq & \mathrm{E}(Y-\hat{Y})^2 \end{aligned}$$
- 即$\hat{Y} = P_H(Y)$。

最佳线性预报与投影的等价性

- 用$L^2(\mathbf{X})$表示$\mathbf{X}=(X_1,X_2,\cdots,X_n)^T$ 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(参见第1 章, 习题6.5).
- 设$\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^T=E\mathbf{X}$.
- 对任何方差有限的随机变量Y, 设$EY=b$, $\hat{Y}=L(Y|\mathbf{X})$ 由(1.3)式定义. 则有

$$Y-\hat{Y}=(Y-b)-L(Y-b|\mathbf{X}-\boldsymbol{\mu}).$$

- 利用性质7 知道

$$E[1\cdot(Y-\hat{Y})]=E(Y-b)-EL(Y-b|\mathbf{X}-\boldsymbol{\mu})=0,$$
$$E[X_i(Y-\hat{Y})]=E[(X_i-\mu_i)(Y-\hat{Y})]+\mu_iE(Y-\hat{Y})=0.$$

即得到$(Y-\hat{Y})$垂直于$H\triangleq L^2(\mathbf{X})$.

- 由定理1.1知道

$$L(Y|\mathbf{X})=P_H(Y).$$

最佳线性预报与投影的等价性

- 用$L^2(\mathbf{X})$表示$\mathbf{X}=(X_1,X_2,\cdots,X_n)^T$ 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(参见第1 章, 习题6.5).
- 设$\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^T=E\mathbf{X}$.
- 对任何方差有限的随机变量Y, 设$EY=b$, $\hat{Y}=L(Y|\mathbf{X})$ 由(1.3)式定义. 则有

$$Y-\hat{Y}=(Y-b)-L(Y-b|\mathbf{X}-\boldsymbol{\mu}).$$

- 利用性质7 知道

$$E[1\cdot(Y-\hat{Y})]=E(Y-b)-EL(Y-b|\mathbf{X}-\boldsymbol{\mu})=0,$$

$$E[X_i(Y-\hat{Y})]=E[(X_i-\mu_i)(Y-\hat{Y})]+\mu_iE(Y-\hat{Y})=0.$$

即得到$(Y-\hat{Y})$垂直于$H\triangleq L^2(\mathbf{X})$.

- 由定理1.1知道

$$L(Y|\mathbf{X})=P_H(Y).$$

最佳线性预报与投影的等价性

- 用$L^2(\mathbf{X})$表示$\mathbf{X}=(X_1,X_2,\cdots,X_n)^T$ 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(参见第1 章, 习题6.5).
- 设$\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^T=E\mathbf{X}$.
- 对任何方差有限的随机变量Y, 设$EY=b$, $\hat{Y}=L(Y|\mathbf{X})$ 由(1.3)式定义. 则有

$$Y-\hat{Y}=(Y-b)-L(Y-b|\mathbf{X}-\boldsymbol{\mu}).$$

- 利用性质7 知道

$$E[1\cdot(Y-\hat{Y})]=E(Y-b)-EL(Y-b|\mathbf{X}-\mu)=0,$$
$$E[X_i(Y-\hat{Y})]=E[(X_i-\mu_i)(Y-\hat{Y})]+\mu_iE(Y-\hat{Y})=0.$$

即得到$(Y-\hat{Y})$垂直于$H\triangleq L^2(\mathbf{X})$.
- 由定理1.1知道

$$L(Y|\mathbf{X})=P_H(Y).$$

最佳线性预报与投影的等价性

- 用$L^2(\mathbf{X})$表示$\mathbf{X} = (X_1, X_2, \cdots, X_n)^T$ 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(参见第1 章, 习题6.5).
- 设$\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)^T = E\mathbf{X}$.
- 对任何方差有限的随机变量Y, 设$EY = b$, $\hat{Y} = L(Y|\mathbf{X})$ 由(1.3)式定义. 则有

$$Y - \hat{Y} = (Y - b) - L(Y - b|\mathbf{X} - \boldsymbol{\mu}).$$

- 利用性质7 知道

$$E[1 \cdot (Y - \hat{Y})] = E(Y - b) - EL(Y - b|\mathbf{X} - \boldsymbol{\mu}) = 0,$$
$$E[X_i(Y - \hat{Y})] = E[(X_i - \mu_i)(Y - \hat{Y})] + \mu_i E(Y - \hat{Y}) = 0.$$

即得到$(Y - \hat{Y})$垂直于$H \triangleq L^2(\mathbf{X})$.
- 由定理1.1知道

$$L(Y|\mathbf{X}) = P_H(Y).$$

最佳线性预报与投影的等价性

- 用$L^2(\mathbf{X})$表示$\mathbf{X}=(X_1,X_2,\cdots,X_n)^T$ 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(参见第1 章, 习题6.5).
- 设$\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^T=E\mathbf{X}$.
- 对任何方差有限的随机变量Y, 设$EY=b$, $\hat{Y}=L(Y|\mathbf{X})$ 由(1.3)式定义. 则有

$$Y-\hat{Y}=(Y-b)-L(Y-b|\mathbf{X}-\boldsymbol{\mu}).$$

- 利用性质7 知道

$$E[1\cdot(Y-\hat{Y})]=E(Y-b)-EL(Y-b|\mathbf{X}-\boldsymbol{\mu})=0,$$
$$E[X_i(Y-\hat{Y})]=E[(X_i-\mu_i)(Y-\hat{Y})]+\mu_iE(Y-\hat{Y})=0.$$

即得到$(Y-\hat{Y})$垂直于$H\triangleq L^2(\mathbf{X})$.

- 由定理1.1知道

$$L(Y|\mathbf{X})=P_H(Y).$$

最佳线性预报与投影的等价性II

- 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$\|\xi\| = \sqrt{\mathrm{E}\xi^2}$, $\forall \xi \in L^2$。$\|\xi\|$是ξ的长度。

最佳线性预报与投影的等价性II

- 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

$$L(Y|1, X_j, j \in T) \quad \text{或} \quad L(Y|H)$$

表示$P_H(Y)$, 这里T是一个可列的指标集.

- 下面记$\|\xi\| = \sqrt{\mathrm{E}\xi^2}, \forall \xi \in L^2$。$\|\xi\|$是$\xi$的长度。

投影算子的性质

定理1.2 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. $\|L(Y|H)\| \le \|Y\|$.
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$Y \in L^2$,

$$\|Y - L(Y|M)\| \le \|Y - L(Y|H)\|.$$

(对应于最佳线性预测性质8)

投影算子的性质

定理1.2 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. $\|L(Y|H)\| \le \|Y\|$.
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$Y \in L^2$,

$$\|Y - L(Y|M)\| \le \|Y - L(Y|H)\|.$$

(对应于最佳线性预测性质8)

投影算子的性质

定理1.2 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. $\|L(Y|H)\| \le \|Y\|$.
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$Y \in L^2$,

$$\|Y - L(Y|M)\| \le \|Y - L(Y|H)\|.$$

(对应于最佳线性预测性质8)

投影算子的性质

定理1.2 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. $\|L(Y|H)\|\le\|Y\|$.
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$P_HP_M=P_MP_H=P_H$, 并且对$Y\in L^2$,

$$\|Y-L(Y|M)\|\le\|Y-L(Y|H)\|.$$

(对应于最佳线性预测性质8)

投影算子的性质

定理1.2 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. $\|L(Y|H)\| \le \|Y\|$.
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$Y \in L^2$,

$$\|Y - L(Y|M)\| \le \|Y - L(Y|H)\|.$$

(对应于最佳线性预测性质8)

投影算子的性质

定理1.2 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)

定理1.2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

$$\begin{aligned}\|Y\|^2 =& \mathrm{E}[(Y - L(Y|H)) + L(Y|H)]^2 \\ =& \mathrm{E}[Y - L(Y|H)]^2 + \mathrm{E}[L(Y|H)]^2 \\ &+ 2\mathrm{E}[(Y - L(Y|H))L(Y|H)] \\ =& \mathrm{E}[Y - L(Y|H)]^2 + \mathrm{E}[L(Y|H)]^2 \\ =& \|Y - L(Y|H)\|^2 + \|L(Y|H)\|^2\end{aligned}$$

定理1.2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

$$\begin{aligned}
\|Y\|^2 =& \mathrm{E}[(Y - L(Y|H)) + L(Y|H)]^2 \\
=& \mathrm{E}[Y - L(Y|H)]^2 + \mathrm{E}[L(Y|H)]^2 \\
&+ 2\mathrm{E}[(Y - L(Y|H))L(Y|H)] \\
=& \mathrm{E}[Y - L(Y|H)]^2 + \mathrm{E}[L(Y|H)]^2 \\
=& \|Y - L(Y|H)\|^2 + \|L(Y|H)\|^2
\end{aligned}$$

定理1.2证明II

- (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$Y \perp H$。

定理1.2证明II

- (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$Y \perp H$。

定理1.2证明II

- (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$Y \perp H$。

定理1.2证明III

- (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

$$\|Y - P_H(Y)\|^2 \geq \|Y - P_M(Y)\|^2$$

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

$$M = \overline{\mathrm{sp}}\{g(\mathbf{X}):\ \mathrm{E}g^2(\mathbf{X}) < \infty, g(\cdot)\text{是可测函数}\} \tag{1.14}$$

- 考虑用M中的元素逼近Y。
- **定义1.5** 设M由(1.14)定义.
  用$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$对$Y$进行预测时, 称

$$L(Y|M) \triangleq P_M(Y) \tag{1.15}$$

为Y的最佳预测.

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

$$M = \overline{\mathrm{sp}}\{g(\mathbf{X}):\ \mathrm{E}g^2(\mathbf{X}) < \infty, g(\cdot)\text{是可测函数}\} \tag{1.14}$$

- 考虑用M中的元素逼近Y。
- **定义1.5** 设M由(1.14)定义.
  用$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$对$Y$进行预测时, 称

$$L(Y|M) \triangleq P_M(Y) \tag{1.15}$$

为Y的最佳预测.

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

$$M = \overline{\mathrm{sp}}\{g(\mathbf{X}):\ \mathrm{E}g^2(\mathbf{X}) < \infty, g(\cdot)\text{是可测函数}\} \tag{1.14}$$

- 考虑用M中的元素逼近Y。
- **定义1.5** 设M由(1.14)定义.
用$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$对$Y$进行预测时, 称

$$L(Y|M) \triangleq P_M(Y) \tag{1.15}$$

为Y的最佳预测.

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

$$M = \overline{\text{sp}}\{g(\mathbf{X}):\ \mathrm{E}g^2(\mathbf{X}) < \infty, g(\cdot)\text{是可测函数}\} \tag{1.14}$$

- 考虑用M中的元素逼近Y。
- **定义1.5** 设M由(1.14)定义.
  用$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$对$Y$进行预测时, 称

$$L(Y|M) \triangleq P_M(Y) \tag{1.15}$$

为Y的最佳预测.

最佳预测II

- 最佳预测$L(Y|M)$实际上是概率论中的条件数学期望$E(Y|\mathbf{X})$.
- $L^2(\mathbf{X})$是M的子空间，由定理1.2的(6)得

$$\|Y - L(Y|M)\| \le \|Y - L(Y|\mathbf{X})\|$$

  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$L^2(\mathbf{X})$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.

最佳预测II

- 最佳预测$L(Y|M)$实际上是概率论中的条件数学期望$E(Y|\mathbf{X})$.
- $L^2(\mathbf{X})$是M的子空间，由定理1.2的(6)得

$$\|Y - L(Y|M)\| \leq \|Y - L(Y|\mathbf{X})\|$$

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$L^2(\mathbf{X})$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.

最佳预测II

- 最佳预测$L(Y|M)$实际上是概率论中的条件数学期望$E(Y|\mathbf{X})$.
- $L^2(\mathbf{X})$是M的子空间，由定理1.2的(6)得

$$\|Y - L(Y|M)\| \leq \|Y - L(Y|\mathbf{X})\|$$

  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$L^2(\mathbf{X})$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.

最佳预测II

- 最佳预测$L(Y|M)$实际上是概率论中的条件数学期望$E(Y|\mathbf{X})$.
- $L^2(\mathbf{X})$是M的子空间，由定理1.2的(6)得

$$\|Y - L(Y|M)\| \leq \|Y - L(Y|\mathbf{X})\|$$

  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$L^2(\mathbf{X})$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.

正态分布时最佳线性预测与最佳预测的等价性

- **定理1.3** 如果$(X_1, X_2, \ldots, X_n, Y)^T$ 服从联合正态分布N(μ, Σ), M由(1.14)定义, 则

$$L(Y|M) = L(Y|X_1, X_2, \cdots, X_n). \tag{1.16}$$

- **证明** 设$\hat{Y} = L(Y|X_1, X_2, \ldots, X_n)$, 则$(Y - \hat{Y})$与$\mathbf{X}$正交. 由于$E(Y - \hat{Y}) = 0$, 所以$Y - \hat{Y}$与$\mathbf{X}$不相关. 由正态分布的性质知道, $Y - \hat{Y}$与$\mathbf{X}$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1.1 知道(1.16)成立.

正态分布时最佳线性预测与最佳预测的等价性

- **定理1.3** 如果$(X_1, X_2, \ldots, X_n, Y)^T$ 服从联合正态分布N(μ, Σ), M由(1.14)定义，则

$$L(Y|M) = L(Y|X_1, X_2, \cdots, X_n). \tag{1.16}$$

- **证明** 设$\hat{Y} = L(Y|X_1, X_2, \ldots, X_n)$, 则$(Y - \hat{Y})$与$\mathbf{X}$正交. 由于$E(Y - \hat{Y}) = 0$, 所以$Y - \hat{Y}$与$\mathbf{X}$不相关. 由正态分布的性质知道, $Y - \hat{Y}$与$\mathbf{X}$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1.1 知道(1.16)成立.

例子

- **例1.2** 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

$$
\begin{aligned}
\mathrm{E}(Y|X) =& \mathrm{E}(3\varepsilon^2\eta - \eta^3|\eta) \\
=& 3\eta - \eta^3 = 3X - X^3
\end{aligned}
$$

- 容易验证$Y - (3\eta - \eta^3) = 3(\varepsilon^2 - 1)\eta$垂直于

$$
M = \overline{sp}\left\{g(X) : Eg^2(X) < \infty, g(x) \text{ 是可测函数}\right\}.
$$

- 于是, 从$3\eta - \eta^3 \in M$ 知道: $L(Y|M) = 3X - X^3$.

例子

- **例1.2** 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

$$\begin{aligned} \mathrm{E}(Y|X) =& \mathrm{E}(3\varepsilon^2\eta - \eta^3|\eta) \\ =& 3\eta - \eta^3 = 3X - X^3 \end{aligned}$$

- 容易验证$Y - (3\eta - \eta^3) = 3(\varepsilon^2 - 1)\eta$垂直于

$$M = \overline{sp}\left\{g(X) : Eg^2(X) < \infty, g(x) \text{ 是可测函数}\right\}.$$

- 于是, 从$3\eta - \eta^3 \in M$ 知道: $L(Y|M) = 3X - X^3$.

例子

- **例1.2** 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

$$
\begin{aligned}
\mathrm{E}(Y|X) =& \mathrm{E}(3\varepsilon^2\eta - \eta^3|\eta) \\
=& 3\eta - \eta^3 = 3X - X^3
\end{aligned}
$$

- 容易验证$Y - (3\eta - \eta^3) = 3(\varepsilon^2 - 1)\eta$垂直于

$$M = \overline{sp}\left\{g(X) : Eg^2(X) < \infty, g(x) \text{ 是可测函数}\right\}.$$

- 于是，从$3\eta - \eta^3 \in M$ 知道：$L(Y|M) = 3X - X^3$.

例子

- **例1.2** 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

$$\begin{aligned}\mathrm{E}(Y|X) =& \mathrm{E}(3\varepsilon^2\eta - \eta^3|\eta) \\ =& 3\eta - \eta^3 = 3X - X^3\end{aligned}$$

- 容易验证$Y - (3\eta - \eta^3) = 3(\varepsilon^2 - 1)\eta$垂直于

$$M = \overline{sp}\left\{g(X) : Eg^2(X) < \infty, g(x) \text{ 是可测函数}\right\}.$$

- 于是, 从$3\eta - \eta^3 \in M$ 知道: $L(Y|M) = 3X - X^3$.

例子

- **例1.2** 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

$$
\begin{aligned}
\mathrm{E}(Y|X) &= \mathrm{E}(3\varepsilon^2\eta - \eta^3|\eta) \\
&= 3\eta - \eta^3 = 3X - X^3
\end{aligned}
$$

- 容易验证$Y - (3\eta - \eta^3) = 3(\varepsilon^2 - 1)\eta$垂直于

$$M = \overline{sp}\left\{g(X) : Eg^2(X) < \infty, g(x) \text{ 是可测函数}\right\}.$$

- 于是, 从$3\eta - \eta^3 \in M$ 知道: $L(Y|M) = 3X - X^3$.

例子

- **例1.2** 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

$$
\begin{aligned}
\mathrm{E}(Y|X) =& \mathrm{E}(3\varepsilon^2\eta - \eta^3|\eta) \\
=& 3\eta - \eta^3 = 3X - X^3
\end{aligned}
$$

- 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：$L(Y|M) = 3X - X^3$.

例子

- **例1.2** 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

$$\begin{aligned} \mathrm{E}(Y|X) =& \mathrm{E}(3\varepsilon^2\eta - \eta^3|\eta) \\ =& 3\eta - \eta^3 = 3X - X^3 \end{aligned}$$

- 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：$L(Y|M) = 3X - X^3$.