第六章　条件期望与均方估值

陈斌

Outline

条件期望

$$
\begin{aligned}
E(X \mid Y=y) &= \int_{-\infty}^{+\infty} x dP(X \leq x \mid Y=y) \\
&= \int_{-\infty}^{+\infty} x dF(x \mid Y=y) = \int_{-\infty}^{+\infty} x dF(x \mid y)
\end{aligned}
$$

称为X 关于$\{Y=y\}$ 的**条件数学期望.**
一般情况,设X,Y的联合分布为$F(x,y)$

$$
dF(x|y) = \begin{cases} \text{离散} & \Pr(X=x|Y=y) \\ \text{连续} & f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_{\mathbb{R}} f(x,y)dx} \end{cases}
$$

设 X, Y 为离散型随机变量，分布为 $\Pr\{X=x_i, Y=y_j\}$.
在$Y=y_j$ 条件下，X 的均值为

$$E(X|Y=y_j) = \sum_i x_i \Pr\{X=x_i|Y=y_j\}.$$

当 Y 的取值变动时， $EX|Y$ 也相应变动，
故 $EX|Y$ 是 Y 的函数，是一个随机变量。

由全概率公式 $\Pr(X=x_i)=\sum_j \Pr(X=x_i|Y=y_j)\Pr(Y=y_j)$

$$
\begin{aligned}
\text{故 } EX &= \sum_i x_i \Pr(X=x_i) = \sum_i x_i \sum_j \Pr(X=x_i|Y=y_j)\Pr(Y=y_j) \\
&= \sum_j \Pr(Y=y_j)\cdot\sum_i x_i \Pr(X=x_i|Y=y_j) \\
&= \sum_j \Pr(Y=y_j) E(X|Y=y_j) = EE(X|Y) \quad \text{全期望公式}
\end{aligned}
$$

定义:

$EX|Y \triangleq \int_{\mathbb{R}} x dF(x|y)$ 称为 X 关于 Y 的条件数学期望。

$EX|Y \triangleq g(Y)$ 是 Y 的函数，当 $Y=y$ 时其值为 $E(X|Y=y)$.

$$
\begin{aligned}
\textbf{注: } EE(X\mid Y) &= Eg(Y) = \int g(y)dF_Y(y) \\
&= \int E(X\mid Y=y)dF_Y(y) = \int E(X\mid Y=y)f_Y(y)dy \\
EE(X\mid Y) &= \sum_j E\left(X\mid Y=y_j\right)\Pr\left(Y=y_j\right)
\end{aligned}
$$

求条件期望的方法：$EY \mid X = ?$

$$E(Y \mid X = x) \triangleq \int_{\mathbb{R}} y dF(y \mid x) = \int_{\mathbb{R}} y f_{Y|X}(y \mid x) dy$$

1. 求 $f(x, y)$
2. 求 $f_X(x) = \int_{\mathbb{R}} f(x, y) dy$
3. 用 $f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)}$，求出条件密度
4. $E(Y \mid X = x) = \int_{\mathbb{R}} y f_{Y|X}(y \mid x) dy \Rightarrow E(Y \mid X)$

设 $(X_1, X_2) \sim N(\mu_1, \mu_2, r, \sigma_1^2, \sigma_2^2)$ 二元联合高斯向量

条件概率密度: $f(x_1|x_2) = \frac{f(x_1,x_2)}{f(x_2)}$

$$
\begin{aligned}
f(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}[(\frac{x_1-\mu_1}{\sigma_2})^2 \right. \\
&\quad \left. -2r(\frac{x_1-\mu_1}{\sigma_1})(\frac{x_2-\mu_2}{\sigma_2}) + (\frac{x_2-\mu_2}{\sigma_2})^2]\right\} \\
f_2(x_2) &= \int_{\mathbb{R}} f(x_1, x_2) dx_1 = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right\} \\
\text{故 } f(x_1|x_2) &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-r^2}} \\
&\quad \cdot \exp\left\{-\frac{1}{2(1-r^2)\sigma_1^2}[x_1-\mu_1-\frac{\sigma_1}{\sigma_2}r(x_2-\mu_2)]^2\right\}
\end{aligned}
$$

当 $X_2 = x_2$ 时，X_1服从$N(\mu_1 + \frac{\sigma_1}{\sigma_2}r(x_2-\mu_2), \sigma_1^2(1-r^2))$

故由条件期望定义,

$EX_1|X_2 = \mu_1 + \frac{\sigma_1}{\sigma_2}r(X_2-\mu_2) = \mu_1 + \frac{C(X_1,X_2)}{C(X_2,X_2)}(X_2-\mu_2)$

$EEX_1|X_2 = EX_1 = \mu_1$

性质:

1. $EC|X = C$, C为常数
2. $EX|X = X$; $Eg(X)|X = g(X)$, g为确定性函数
3. 线性 $E(C_1X_1 + C_2X_2)|Y = C_1EX_1|Y + C_2EX_2|Y$

性质: $Eg(Y)X|Y = g(Y) \cdot EX|Y$

$$\text{证:}\quad \forall y,\ E\{g(Y) \cdot X|Y = y\} = {}^{1}\int_{\mathbb{R}} g(y) \cdot x dF(X|Y = y)$$
$$= g(y) \int_{\mathbb{R}} x dF(X|Y = y) = g(y)E(X|Y = y)$$

[1] 一般$E[q(X, Y) \mid Y = y] = \int q(x, y) dF(x \mid Y = y)$

性质:

1. $EC|X = C, \quad C$为常数
2. $EX|X = X$; $Eg(X)|X = g(X), \quad g$为确定性函数
3. 线性 $E(C_1X_1 + C_2X_2)|Y = C_1EX_1|Y + C_2EX_2|Y$

性质: $Eg(Y)X|Y = g(Y) \cdot EX|Y$

$$\begin{aligned} \text{证:} \quad \forall y,\ E\{g(Y)\cdot X|Y=y\} &= {}^{1}\int_{\mathbb{R}} g(y)\cdot x dF(X|Y=y) \\ = g(y)\int_{\mathbb{R}} x dF(X|Y=y) &= g(y)E(X|Y=y) \end{aligned}$$

性质: $EEX|Y = EX, (EEX|Y \triangleq \int_{\mathbb{R}} E(X|Y=y)dF_Y(y))$ 全期望公式
证:考虑连续情形，离散情形已证。

$$\begin{aligned} E(X|Y=y) &= \int_{\mathbb{R}} x dF(X|Y=y) = \int_{\mathbb{R}} x f(x|y)dx \\ EEX|Y &= \int_{\mathbb{R}}\int_{\mathbb{R}} x f(x|y)dx dF_Y(y) = \int_{\mathbb{R}}\int_{\mathbb{R}} x f(x|y) f_Y(y) dx dy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} x f(x,y)dxdy = \int_{\mathbb{R}} x f_X(x)dx = EX \end{aligned}$$

[1] 一般$E[q(X,Y) \mid Y=y] = \int q(x,y)dF(x \mid Y=y)$

多元随机变量(向量)的条件期望:

$E(X|Y_1,\cdots Y_n)$ 是随机向量 Y 的函数,
当 $Y=y$ 时, 取值 $E(X|Y=y)\triangleq E(X|Y_1=y_1,\cdots Y_n=y_n)$
类似地有: 随机向量 $Y=(Y_1,\cdots Y_n)^T$

1. 线性 $EC_1X_1+C_2X_2|Y=C_1EX_1|Y+C_2EX_2|Y$
2. $E[g(Y_1,\cdots Y_n)X|Y]=g(Y_1,\cdots Y_n)EX|Y$
3. 若 X 与 $Y_1,\cdots Y_n$ 独立, 则 $EX|Y=EX$
4. $EEX|Y_1,\cdots Y_n=EX$ 作业

均方估值

已知随机变量 X 可观测，X 与未知随机变量 Y 相关

若 X,Y 的二阶矩存在，如何使用 X 去估计 Y？

$i.e.$，寻找 X 的某个函数$g(X)$，$s.t.$ $g(X)$ 与 Y 之间的误差最小

均方准则：$\min\limits_{g} E|Y-g(X)|^2, \quad i.e. \quad \min\limits_{g} \|Y-g(X)\|$

设 g^* 满足 $\min E|Y-g(X)|^2 = E|Y-g^*(X)|^2$, 如何求 g^*?

定理： $E|Y-g(X)|^2 \geq E|Y-E(Y|X)|^2 \;\; i.e. \; g^*(X) = EY|X$

证：由于 $EY|X$ 是 X 的函数，故可令 $E(Y|X) \triangleq \hat{g}(x)$

$$
\begin{aligned}
& E|Y-g(X)|^2 = E|Y-\hat{g}(X)+\overline{\hat{g}(X)-g(X)}|^2 \\
&= E|\underline{Y-\hat{g}(X)}|^2 + E(Y-\hat{g}(X))\overline{(\hat{g}(X)-g(X))} \\
&\quad + E\overline{(Y-\hat{g}(X))}(\hat{g}(X)-g(X)) + E|\hat{g}(X)-g(X)|^2 \\
&\geq E|Y-\hat{g}(X)|^2
\end{aligned}
$$

等号在$E|\hat{g}(X)-g(X)|^2=0$时成立 $\Leftrightarrow g(X) = EY|X, \; a.e.$

(接上一页证明：)
使用全期望公式[2]

$$
\begin{aligned}
& E(Y-\hat{g}(X))\cdot(\overline{\hat{g}(X)-g(X)}) \\
= \; & {}^{3}E\{E[Y-\hat{g}(X)][\overline{\hat{g}(X)-g(X)}]|X\} \\
= \; & E\{[\overline{\hat{g}(X)-g(X)}]E[Y-\hat{g}(X)]|X\} \\
= \; & E\{[\overline{\hat{g}(X)-g(X)}][EY|X-E\hat{g}(X)|X]\} \\
= \; & 0 \qquad \text{其中} \quad EY|X=\hat{g}(X) \ \text{且} \ E\hat{g}(X)|X=\hat{g}(X)
\end{aligned}
$$

结论：$E\ Y|X$ 是 Y 的最佳均方估计，称为 Y 关于 X 的回归。

用 $EY|X$ 预测 Y 可使均方误差最小！

[2] $EEX \mid Y = EX,\ \ Eg(X)\mid X = g(X)$

[3] $Eg(Y)h(X,Y)|Y = g(Y)\cdot Eh(X,Y)|Y$

例： $X=(X_1,X_2)$ n元高斯 $N(\mu,\mathbf{B}), X_1,X_2$ 为子（行)向量

$$\mu_{X_1}=\mu_1,\ \mu_{X_2}=\mu_2,\ \mathbf{B}=\begin{pmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{pmatrix},\text{ 求 } EX_2|X_1?^4$$

解： 在二元情形，求出 $f(y|x)$ ，然后对 y 配方，得到均值。

但在 n 元情形，这种方法十分繁琐，很难得到结果。

解决方案：化 $\mathbf{B}$ 为对角(分块)矩阵，简化处理。

令$(Y_1,Y_2)^T=((X_1,X_2)\begin{pmatrix}\mathbf{I} & -\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\\ \mathbf{0} & \mathbf{I}\end{pmatrix})^T$,

显然 $Y_1=X_1,\quad Y_2=X_2-X_1\mathbf{B}_{11}^{-1}\mathbf{B}_{12}$.

$\mu_Y=(\mu_1,\mu_2-\mu_1\mathbf{B}_{11}^{-1}\mathbf{B}_{12})^T=(\mu_{y_1},\mu_{y_2})^{T5}$

$$\mathbf{B}_Y=\begin{pmatrix}\mathbf{I} & 0\\ -\mathbf{B}_{21}\,\mathbf{B}_{11}^{-1} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{I} & 0\\ -\mathbf{B}_{21}\,\mathbf{B}_{11}^{-1} & \mathbf{I}\end{pmatrix}^T=\begin{pmatrix}\mathbf{B}_{11} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}_{22}-\mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\end{pmatrix}$$

$$Y_1\text{与}Y_2\text{独立},\quad Jacobi=\left|\begin{pmatrix}\mathbf{I} & 0\\ -\mathbf{B}_{21}\,\mathbf{B}_{11}^{-1} & \mathbf{I}\end{pmatrix}^{-1}\right|=1,$$

---

4 用X_1去估计X_2

5 $\eta=\mathbf{C}\xi, E\eta=\mathbf{C}E\xi, \mathbf{B}_\eta=\mathbf{C}\mathbf{B}_\xi\mathbf{C}^T$

故 $f_Y(y_1,y_2)=f_Y(y_1)f_Y(y_2)=f_X(x_1,x_2)$.

$$Y_1=X_1 \quad\Rightarrow\quad f_Y(y_1)=f_X(x_1).$$

故 $f_X(x_2|x_1)=\frac{f_X(x_1,x_2)}{f_X(x_1)}=\frac{f_Y(y_1,y_2)}{f_Y(y_1)}=f_Y(y_2)$,

即当x_1给定时，其概率密度可以表示如下：

$$f_{X_2}(x_2|x_1)=f_Y(y_2)=\frac{1}{(2\pi)^{\frac{n_2}{2}}|\mathbf{B}_{y_2}|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(y_2-\mu_{y_2})\mathbf{B}_{y_2}^{-1}(y_2-\mu_{y_2})\}.$$ [6]

由$Y_2=X_2-\mathbf{B}_{21}\ \mathbf{B}_{11}^{-1}X_1$，可知

$$\begin{aligned}
y_2-\mu_{y_2} &= x_2-x_1\mathbf{B}_{11}^{-1}\mathbf{B}_{12}-(\mu_2-\mu_1\mathbf{B}_{11}^{-1}\mathbf{B}_{12}) \\
&= x_2-[\mu_2+(x_1-\mu_1)\mathbf{B}_{11}^{-1}\mathbf{B}_{12}] \\
&= x_2-E(X_2|X_1=x_1),
\end{aligned}$$

故 $E(X_2|X_1)=\mu_2+(X_1-\mu_1)\mathbf{B}_{11}^{-1}\mathbf{B}_{12}$.

[6] 为高斯分布的概率密度，即$f_{X_2}(x_2|x_1)$服从高斯分布！

$E\left(X_2 \mid X_1\right)$的另外一种表示：

$$\begin{cases} X_1 = Y_1 \\ X_2 = B_{21}B_{11}^{-1}Y_1 + Y_2 \end{cases}, \quad Y_1, Y_2 \text{ 相互独立}$$

$$\begin{aligned} E\left(X_2 \mid X_1\right) &= E\left(\left(\mathbf{B}_{21}\mathbf{B}_{11}^{-1}Y_1 + Y_2\right) \mid Y_1\right) \\ &= E\left(\mathbf{B}_{21}\mathbf{B}_{11}^{-1}Y_1 \mid Y_1\right) + E\left(Y_2 \mid Y_1\right) \\ &= \mathbf{B}_{21}\mathbf{B}_{11}^{-1}Y_1 + E\left(Y_2\right) \\ &= \mathbf{B}_{21}\mathbf{B}_{11}^{-1}X_1 + E\left(X_2 - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}X_1\right) \\ &= E\left(X_2\right) + \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\left(X_1 - E\left(X_1\right)\right) \\ &= \mu_2 + \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\left(X_1 - \mu_1\right) \end{aligned}$$

Y 随机变量， $X_n, X_{n-1}, \cdots$ 为可观测的随机序列
（有限或无限均可，为方便假设有限）
若用 $X_n, \cdots X_1$ 的某个函数作为 Y 的估计 Y^*,
使得误差在均方意义下最小， $i.e.$

$$E|Y - g^*(X_n, \cdots X_1)|^2 = \min_g E|Y - g(X_n, \cdots X_1)|^2$$

$$Y^* \triangleq g^*(X_n, \cdots X_1)$$

定理（推广）: $Y^* = g^*(X_n, \cdots X_1) = EY|X_n, \cdots X_1$

证明方法与一维情况类似，留为作业(与Markov链一齐交)

线性均方最佳预测

$E\, Y|X$ 作为最优估计理论上很完美，但实际应用中很难求得！
在求解全局最优困难时，人们常常退一步求解"局部/条件最优"
在所有函数中求最佳估计 $\rightarrow$ 在线性函数中求最佳估计

H：所有二阶矩变量组成的线性空间，$Y \in H$，实二阶矩
H_n：包含 $X_n, X_{n-1}, \cdots$ 的线性子空间，$H_n \subset H$
H 为 Hilbert 空间，可定义内积和距离，范数。
求：$\hat{Y}^* \in H_n$ 使得误差 $Y - \hat{Y}^*$ 最小，$i.e.$

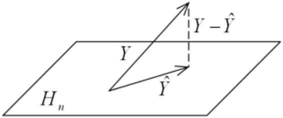
$$E|Y - \hat{Y}^*|^2 = \min_{\hat{Y} \in H_n} E|Y - \hat{Y}|^2$$

记号（定义）：

$$\|Y - \hat{Y}\| \triangleq \sqrt{E|Y - \hat{Y}|^2}, \qquad d(Y, \hat{Y}) \triangleq \|Y - \hat{Y}\| \text{ 距离}$$

$(Y, \hat{Y}) \triangleq EY\hat{Y}$ 内积，垂直 $\Leftrightarrow$ 正交 $\Leftrightarrow$ 内积为0

正交性原理



H_n 为 H 中的一个（超）平面
目标：求 $\hat{Y}^*$ 使得 $\|Y-\hat{Y}^*\|^2$ 最小
显然当 $Y-\hat{Y}^* \perp H_n$时，$\|Y-\hat{Y}^*\|$ 最小

$$\Updownarrow$$

$$\forall u \in H_n, \quad E(Y-\hat{Y}^*)u = 0$$

定理： $\hat{Y}^*$ 是 Y 对 $X_n, X_{n-1}, \cdots$ 的线性均方最佳估计， iff
$\forall u \in H_n$，满足 $E(Y-\hat{Y}^*)u = 0$
（类比初中数学中的证明线面垂直）

非退化随机变量：二维情形

例： 设有随机变量ξ,η，已知$E\eta^n = m_n, n \geq 2$，　$E\eta = 0, \xi = \eta^2$。若以η的观察值对ξ作线性估计，求其最佳估计。

解： 根据题意$\mu_\xi = E\xi = E\eta^2 = m_2 \neq 0$，

于是最佳线性估计$\hat{\xi}$应具有如下形式

$$\hat{\xi} = a\eta + b, \quad 1,\eta \text{ 张成的线性子空间}$$

所谓最佳是指$E[\xi - \hat{\xi}]^2 = E[\xi - (a\eta + b)]^2$取极小值，就是选择$a,b$值以使$E[\xi - \hat{\xi}]^2$最小。

现利用正交性原理求解；

$$E\{\xi - a\eta - b\} = 0 \Rightarrow b = E\xi = m_2$$

$$E\{[\xi - (a\eta + b)]\eta\} = 0 \Rightarrow E\xi\eta = aE\eta^2$$

$$\text{即} \quad a = \frac{E\xi\eta}{E\eta^2} = \frac{E\eta^3}{E\eta^2} = \frac{m_3}{m_2}$$

$$\text{故} \quad \hat{\xi} = \frac{m_3}{m_2}\eta + m_2$$

定理： $\hat{Y}^*$ 是 Y 对 $X_n, X_{n-1}, \cdots$ 的线性均方最佳估计， iff
$\forall u \in H_n$，满足 $E(Y-\hat{Y}^*)u = 0$

讨论一：若 $X_n = C$ **常数**

令 $\hat{Y}^* = \sum a_k X_k$，

则 $E\left(Y-\hat{Y}^*\right)C = 0 \Rightarrow EY = E\hat{Y}^* = \sum a_k EX_k$　无偏估计

$\Rightarrow Y - \sum a_k X_k = Y - EY - \left(\hat{Y}^* - E\hat{Y}^*\right)$

$= (Y - EY) - \sum a_k (X_k - EX_k)$

即 $\{X_k\}$ 对 Y 的最佳估计 $\Leftrightarrow (Y-EY)$ 对 $\{X_k - EX_k\}$ 的最佳估计

故为讨论方便, 不妨设 $Y, X_n, X_{n-1}, \ldots 0$ 均值

定理： $\hat{Y}^*$ 是 Y 对 $X_n, X_{n-1}, \cdots$ 的线性均方最佳估计， iff
$\forall u \in H_n$，满足 $E(Y - \hat{Y}^*)u = 0$

讨论一：若$X_n = C$ **常数**

令$\hat{Y}^* = \sum a_k X_k$，

则$E\left(Y - \hat{Y}^*\right)C = 0 \Rightarrow EY = E\hat{Y}^* = \sum a_k EX_k$ 无偏估计

$\Rightarrow Y - \sum a_k X_k = Y - EY - \left(\hat{Y}^* - E\hat{Y}^*\right)$

$= (Y - EY) - \sum a_k (X_k - EX_k)$

即$\{X_k\}$对Y 的最佳估计$\Leftrightarrow (Y - EY)$ 对$\{X_k - EX_k\}$ 的最佳估计

故为讨论方便, 不妨设$Y, X_n, X_{n-1}, \ldots 0$ 均值

讨论二：不妨设$Y, X_n, X_{n-1}, \ldots 0$ **均值**

若 Y 与 $X_n, X_{n-1}, \cdots$不相关，则无法用 $X_n, \cdots$ 估计 Y

$EYX_i = EYEX_i = 0 \Leftrightarrow Y \perp X_i \Leftrightarrow Y \perp H_n \Leftrightarrow \hat{Y}^* = 0$

定理： $\hat{Y}^*$ 是 Y 对 $X_n, X_{n-1}, \cdots$ 的线性均方最佳估计，iff $\forall u \in H_n$，满足 $E(Y-\hat{Y}^*)u = 0$

定理： $\hat{Y}^*$ 是 Y 对 $X_n, X_{n-1}, \cdots$ 的线性均方最佳估计， iff
$\forall u \in H_n$，满足 $E(Y - \hat{Y}^*)u = 0$

讨论三：

设 $Y - \hat{Y}^* = e$, $\hat{Y}^* = \sum_{k=0}^{\infty} \alpha_k X_{n-k}$

由于 $X_i \in H_n$, 故 $EeX_i = 0$, $i = n, n-1, \cdots$

$i.e.\ E(Y - \sum_{k=0}^{\infty} \alpha_k X_{n-k})X_i = 0$

$\Leftrightarrow \ EYX_i = \sum_k \alpha_k EX_{n-k}X_i$

$\Leftrightarrow R(Y, X_i) = \sum_k \alpha_k R(X_{n-k}, X_i)\ i = n, n-1 \cdots$ 线性方程组

定理：$\hat{Y}^*$ 是 Y 对 $X_n, X_{n-1}, \cdots$ 的线性均方最佳估计， iff
$\forall u \in H_n$，满足 $E(Y-\hat{Y}^*)u = 0$

讨论三：

设 $Y-\hat{Y}^* = e$，$\hat{Y}^* = \sum_{k=0}^{\infty} \alpha_k X_{n-k}$

由于 $X_i \in H_n$，故 $EeX_i = 0$，$i = n, n-1, \cdots$

$i.e.\ E(Y - \sum_{k=0}^{\infty} \alpha_k X_{n-k})X_i = 0$

$\Leftrightarrow\ EYX_i = \sum_k \alpha_k EX_{n-k}X_i$

$\Leftrightarrow R(Y, X_i) = \sum_k \alpha_k R(X_{n-k}, X_i)\ i = n, n-1 \cdots$ 线性方程组

若$EY = E\hat{Y}^* = \sum \alpha_k EX_{n-k}$

$\Rightarrow EYX_i - EYEX_i = \sum \alpha_k \left(EX_{n-k}X_i - EX_{n-k}EX_i\right)$

$\Leftrightarrow C\left(Y, X_i\right) = \sum \alpha_k C\left(X_{n-k}, X_i\right) \quad i = n, n-1 \cdots$ 线性方程组

若已知相关矩阵(或协方差矩阵)，可解出 α_k

从而求出 $\hat{Y}^*$，即最佳均方线性估计

若 $X_n, X_{n-1}, \cdots$ 两两正交，$R(X_j, X_i) = 0,\ j \neq i$，
则$\alpha_{n-i} = \frac{R(Y,X_i)}{R(X_i,X_i)}$[7]

若 $\{X_n\}$ 是二阶矩过程， $R(\tau)$ 是相关函数
已知 $X_n, X_{n-1}, \cdots X_{n-p}$，求$X_{n+1}$的估计 $\hat{X}^*_{n+1}$

$$令\ \hat{X}^*_{n+1} = \sum_{k=0}^{p} \alpha_k X_{n-k}$$

$$R\left(X_{n+1}, X_i\right) = \sum \alpha_k R\left(X_{n-k}, X_i\right) \quad i = n, n-1, \ldots, n-p$$

$$\begin{cases} \alpha_0 R(n,n)+\alpha_1 R(n-1,n)+\cdots+\alpha_p R(n-p,n)=R(n+1,n) \\ \alpha_0 R(n,n-1)+\alpha_1 R(n-1,n-1)+\cdots+\alpha_p R(n-p,n-1)=R(n+1,n-1) \\ \vdots \\ \alpha_0 R(n,n-p)+\alpha_1 R(n-1,n-p)+\cdots+\alpha_p R(n-p,n-p)=R(n+1,n-p) \end{cases}$$

解出 $\alpha_0, \cdots \alpha_p$，即得 $\hat{X}^*_{n+1}$

结论：可见，线性最优均方估计有系统的方法可以解出

[7] $R(Y, X_i) = \sum \alpha_k R(X_{n-k}, X_i),\ i = n, n-1 \cdots$

若$\{X_n\}$ 是实宽平稳过程，$R(\tau)$ 是相关函数；
已知$X_n, X_{n-1}, \cdots X_{n-p}$，求$X_{n+1}$ 的估计$\hat{X}^*_{n+1}$

$$\text{令}\hat{X}^*_{n+1} = \sum_{k=0}^{p} \alpha_k X_{n-k}$$

$$\begin{aligned}
R\left(X_{n+1}, X_i\right) &= \sum_{k} \alpha_k R\left(X_{n-k}, X_i\right) \quad i = n, n-1, \ldots, n-p \\
R(n+1-i) &= \sum_{k=0}^{p} \alpha_k R(n-k-i) \\
&= \sum_{k=0}^{n-i} \alpha_k R(n-k-i) + \sum_{k=n-i+1}^{p} \alpha_k R(k+i-n)
\end{aligned}$$

上述方程可写为:

$$\begin{cases} \alpha_0 R(0)+\alpha_1 R(1)+\cdots+\alpha_p R(p)=R(1) \\ \alpha_0 R(1)+\alpha_1 R(0)+\alpha_2 R(1)+\cdots+\alpha_p R(p-1)=R(2) \\ \vdots \\ \alpha_0 R(p)+\alpha_1 (p-1)+\alpha_2 R(p-2)+\cdots+\alpha_p R(0)=R(p+1) \end{cases}$$

$i.e.$ 约尔-沃克方程

$$\begin{pmatrix} R(0) & R(1) & \cdots & R(p) \\ R(1) & R(0) & \cdots & R(p-1) \\ \vdots & & \ddots & \vdots \\ R(p) & R(p-1) & \cdots & R(0) \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ R(p+1) \end{pmatrix}$$

上式左边矩阵对称，称为托普利兹方阵

解方程组求得$\alpha = (\alpha_0, \alpha_1, \cdots \alpha_p)$, $\hat{X}^*_{n+1} = \sum_{k=0}^p \alpha_k X_{n-k}$

高斯马尔可夫过程

定义： $\{\xi(t), t \in T\}, \quad t_1 < t_2 < \cdots < t_m < t_{m+1} \in T$

$f(x_{m+1}|x_1, x_2, \cdots x_m) = f(x_{m+1}|x_m)$，称Markov过程

定理： 设 $\{\xi(t)\}$ 为实 0 均值高斯Markov过程

则 $R(t_1, t_3) = \frac{R(t_1,t_2)R(t_2,t_3)}{R(t_2,t_2)} \quad t_1 < t_2 < t_3$

证: $\quad R(t_1, t_3) = E\xi(t_1)\xi(t_3)$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} x_1 x_3 f(x_1, x_2, x_3) dx_1 dx_2 dx_3$$

其中 $f(x_1, x_2, x_3) = f(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_2) \quad (Markov)$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} x_1 f(x_1) f(x_2|x_1) \cdot [\int_{\mathbb{R}} x_3 f(x_3|x_2) dx_3] dx_2 dx_1$$

其中 $\int_{\mathbb{R}} x_3 f(x_3|x_2) dx_3 = E[\xi(t_3)|\xi(t_2) = x_2]$

$$\xi(t) \text{ 高斯 } \Rightarrow \xi(t_3), \xi(t_2) \text{ 二元高斯}$$

$$E\xi(t_3)|\xi(t_2) = \mu_3 + \frac{\sigma_3}{\sigma_2} r[\xi(t_2) - \mu_2]$$

$$E[\xi(t_3)|\xi(t_2)^8 = x_2] = \frac{\sigma_3 \sigma_2 r}{\sigma_2^2} x_2 = \frac{R(t_2, t_3)}{R_(t_2, t_2)} x_2$$

$$\begin{aligned} R(t_1, t_3) &= \int_{\mathbb{R}} \int_{\mathbb{R}} x_1 f(x_1) f(x_2|x_1) x_2 dx_2 dx_1 \cdot \frac{R(t_2, t_3)}{R(t_2, t_2)} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} x_1 x_2 f(x_1, x_2) dx_2 dx_1 \frac{R(t_2, t_3)}{R(t_2, t_2)} = \frac{R(t_1, t_2) R(t_2, t_3)}{R(t_2, t_2)} \end{aligned}$$

注：本定理实为充要条件，
即对于实高斯过程， $R(t_1, t_3) \cdots \Rightarrow Markov$

定理： $\{\xi(t)\}$ 0均值高斯过程，均方连续，平稳
则: $\xi(t)$ $Markov$ $\Leftrightarrow$ $R(\tau) = \exp(-\mu|\tau|)$ $\mu > 0$ 常数
林元烈《应用随机过程》或教材

[8] 0均值