*五、条件数学期望，最佳线性预测

在第三章§2我们曾经引进了条件分布函数的概念，现在要相应地引进条件数学期望的概念，并说明它的应用.

为方便起见，我们讨论两个随机变量ξ与η的场合，假定它们具有密度函数$p(x,y)$，并以$p(y|x)$记已知$\xi=x$的条件下，η的条件密度函数，以$p_1(x)$记ξ的密度函数.

定义 4.2.7 在$\xi=x$的条件下，η的**条件数学期望**定义为

$$E\{\eta|\xi=x\}=\int_{-\infty}^{\infty}yp(y|x)\mathrm{d}y \tag{4.2.46}$$

[例 11] 若（ξ，η）服从二元正态分布，则由（3.2.29）知

$$p(y|x)=\frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2\sigma_2^2(1-r^2)}\cdot\left[y-\left(a_2+r\frac{\sigma_2}{\sigma_1}(x-a_1)\right)\right]^2\right\}$$

这是正态分布$N\left(a_2+r\frac{\sigma_2}{\sigma_1}(x-a_1),\ \sigma_2^2(1-r^2)\right)$，因此

$$E\{\eta|\xi=x\}=a_2+r\frac{\sigma_2}{\sigma_1}(x-a_1) \tag{4.2.47}$$

值得提醒的是：这时条件数学期望$E\{\eta|\xi=x\}$是x的线性函数.

条件数学期望在**预测**问题中起重要作用. 问题这样提出：若ξ，η是相依的随机变量，我们要找ξ与η的函数关系，设这个关系是$y=h(x)$，如果$E\eta^2$及$E[h(\xi)]^2$都存在，我们的目的是找函数$h(x)$，使η与$h(\xi)$"尽可能靠近"，这里的"靠近"需要一个标准，最常用的是所谓"最小二乘法"，这时要求使

$$E[\eta-h(\xi)]^2 \tag{4.2.48}$$

达到最小. 因为

$$E[\eta - h(\xi)]^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y - h(x)]^2 p(x,y)\mathrm{d}x\mathrm{d}y$$
$$= \int_{-\infty}^{\infty} p_1(x)\left\{\int_{-\infty}^{\infty}[y - h(x)]^2 p(y|x)\mathrm{d}y\right\}\mathrm{d}x \qquad (4.2.49)$$

由 (4.2.8) 知道，当 $h(x) = E\{\eta|\xi = x\}$ 时，$\int_{-\infty}^{\infty}[y - h(x)]^2 \cdot p(y|x)\mathrm{d}y$ 达到最小，从而使 (4.2.49) 达到最小. 即当我们观察到 $\xi = x$ 时，$E\{\eta|\xi = x\}$ 是一切对 η 的估值中均方误差最小的一个.

今后我们将称 $y = E\{\eta|\xi = x\}$ 是 η 关于 ξ 的**回归**.

通常，(ξ, η) 的联合分布函数是不知道的，或者虽然知道但是却不易算出 $E\{\eta|\xi = x\}$. 假定已知 ξ 与 η 的数学期望 μ_1，μ_2，标准差 σ_1，σ_2 及相关系数 r，这时可以降低一点要求，改为求**最佳线性预测**. 也就是说，把 $h(x)$ 限定为 x 的线性函数 $L(x) = a + bx$，求 a，b 使

$$e(a,b) = E[\eta - (a + b\xi)]^2 \qquad (4.2.50)$$

达到最小.

把 $e(a, b)$ 对 a，b 求偏导数并令它们等于 0，得到

$$\begin{aligned} 2E[\eta - (a + b\xi)] &= 0 \\ 2E[(\eta - (a + b\xi))\xi] &= 0 \end{aligned} \qquad (4.2.51)$$

整理后变成

$$\begin{aligned} a + b\mu_1 &= \mu_2 \\ a\mu_1 + bE\xi^2 &= E\xi\eta \end{aligned} \qquad (4.2.52)$$

因此解得

$$a = \mu_2 - b\mu_1, \quad b = \frac{\operatorname{cov}(\xi,\eta)}{\sigma_1^2} = r \cdot \frac{\sigma_2}{\sigma_1} \qquad (4.2.53)$$

最佳线性预测为

$$L(x) = \mu_2 + r\frac{\sigma_2}{\sigma_1}(x - \mu_1) \qquad (4.2.54)$$

我们称（4.2.54）为 η 关于 ξ 的**线性回归**. 这个结果与 $E\{\eta|\xi=x\}$ 一般是不同的，但是在 (ξ,η) 是二元正态分布的场合，由（4.2.47）知两者是重合的，所以在正态分布场合，最佳预测是线性预测，这是一个十分重要的结果.

进一步，我们还可以计算最佳线性预测的均方误差.

$$
\begin{aligned}
E[\eta - L(\xi)]^2 &= E[\eta - \mu_2 - b(\xi - \mu_1)]^2 \\
&= \sigma_2^2 + b^2\sigma_1^2 - 2b\operatorname{cov}(\xi,\eta) \\
&= \sigma_2^2 - \frac{\operatorname{cov}^2(\xi,\eta)}{\sigma_1^2} = \sigma_2^2(1 - r^2) \qquad (4.2.55)
\end{aligned}
$$

因此预测误差同 η 的方差有关，也同 ξ 与 η 的相关系数有关，特别当 $|r|=1$ 时（这时 ξ 与 η 有线性关系），预测误差为 0，也就是说，可以完全准确地进行线性预测. 从这个讨论再次看出，相关系数反映了 ξ 与 η 线性联系的程度.

最佳线性预测理论中的另一个重要事实是：预测值 $\hat{\eta}=L(\xi)$ 与残差 $\eta-\hat{\eta}$ 是不相关的. 证明如下：

由（4.2.54）知

$$
\hat{\eta} = L(\xi) = \mu_2 + r\frac{\sigma_2}{\sigma_1}(\xi - \mu_1) \qquad (4.2.56)
$$

因此

$$
E\hat{\eta} = \mu_2 \qquad (4.2.57)
$$

$$
E(\eta - \hat{\eta}) = 0 \qquad (4.2.58)
$$

这样一来

$$
\begin{aligned}
\operatorname{cov}(\hat{\eta},\eta - \hat{\eta}) &= E[(\hat{\eta} - \mu_2)(\eta - \hat{\eta})] \\
&= E\left\{r\frac{\sigma_2}{\sigma_1}(\xi - \mu_1)\left[(\eta - \mu_2) - r\frac{\sigma_2}{\sigma_1}(\xi - \mu_1)\right]\right\} \\
&= r\frac{\sigma_2}{\sigma_1}\left(r\sigma_1\sigma_2 - r\frac{\sigma_2}{\sigma_1}\sigma_1^2\right) = 0 \qquad (4.2.59)
\end{aligned}
$$

这个事实可以解释为：残差中已不再包含对预测 η 有用的知识. 因此观察值 η 被分解为两个不相关的随机变量之和：

$$
\eta = \hat{\eta} + (\eta - \hat{\eta}) \qquad (4.2.60)
$$

若以 $E\{\eta|\xi\}$ 记随机变量 ξ 的如下函数：当 $\xi=x$ 时，它取值 $E\{\eta|\xi=x\}$. 这样定义的 $E\{\eta|\xi\}$ 是随机变量，对它可以求数学期望，并有下列关系式：

$$E\eta = E[E\{\eta|\xi\}] \tag{4.2.61}$$

这是条件数学期望的一个极端重要的性质，有广泛应用，在离散型及连续型随机变量的场合，读者不难对它加以证明.