第五周讲稿

1. 方差及其性质

$$\min_C\{E((X-C)^2)\}=E((X-EX)^2)\equiv D(X)(or\,Var(X))$$，若 $EX^2<\infty$。

★ 方差的性质

性质 1： 若 EX^2 存在，则 $DX=EX^2-(EX)^2$（计算公式）.

性质 2： $D(C)=0$,且 $DX=0\Leftrightarrow P(X=EX)=1$.

性质 3： 设 $X_1,X_2,\cdots X_n$ 相互独立，则 $D(\sum_{i=1}^{n}a_iX_i)=\sum_{i=1}^{n}a_i^2DX_i$

例：（二项分布和几何分布的期望与方差）

$X\sim B(n,p)\Rightarrow EX=np,DX=npq;$

$X\sim Ge(p)\Rightarrow EX=\dfrac{1}{p},DX=\dfrac{q}{p^2}$

2. 协方差与相关系数

研究：（1）随机变量不独立时的刻画问题；

（2）最佳线性预测问题（线性回归）

$D(X+Y)=DX+DY+2E[(X-EX)(Y-EY)]$

★ 协方差与相关系数的定义

设（X,Y）是一个二维随机变量，且 X 与 Y 各自平方的数学期望均存在，则定义

$$Cov(X,Y)=E[(X-EX)(Y-EY)]$$

称为 X 与 Y 的**协方差**；定义

$$r_{X,Y}\equiv\frac{E[(X-EX)(Y-EY)]}{\sqrt{DX}\sqrt{DY}}$$

称之为 X 与 Y 的**相关系数，**称 X 与 Y 的**不相关，若** $r_{X,Y}=0(or\ Cov(X,Y)=0)$。

★ 定理（最佳线性预测） $E[X-(\hat{a}Y+\hat{b})]^2=\min_{a,b}E[X-(aY+b)]^2=DX(1-r_{X,Y}^2)$，

其中 $\hat{a},\hat{b}$ 由方程

$$\hat{a}=\frac{Cov(X,Y)}{DY}=r_{X,Y}\frac{\sqrt{DX}}{\sqrt{DY}},\quad EX=\hat{a}EY+\hat{b}$$

Remark：（1）在均方误差最小的准则下，若把预测量局限在线性函数 $L(Y)=aY+b$，则

得到最佳线性预测（线性回归），此时无需知道（X，Y）的联合分布，而只需要知道它们的前二阶矩。这套理论称为二阶矩理论，有广泛的应用，大家可以和一元回归分析作对比。

（2）预测误差与相关系数 $r_{X,Y}$ 有关，特别当 $r_{X,Y}=\pm 1$ 时，$\boldsymbol{X}$ **与** $\boldsymbol{Y}$ 完全相关，它们之间有线性关系，因此能完全准确地进行线性预测，从而预测方差为 0。而当 $r_{X,Y}=0$ 时，作线性预测完全是多余的。

（3）预测值 $\hat{X}=(\hat{a}Y+\hat{b})$ **与残差** $X-\hat{X}$ **不相关（思考）。**这是最小均方误差预测的一个重要结果。

例：给定观测值 $(x_i,y_i), i=1,2,\cdots,n$，讨论 y 关于 x 的最佳先行预测问题。

解：这是一个统计问题，把上述观测值看作总体（X，Y）的一个样本，记 X 与 Y 的均值和方差分别为 $\mu_1,\sigma_1^2;\mu_2,\sigma_2^2$，它们的相关系数 ρ，则 Y 关于 X 的最佳线性预测为

$$\mu_2+\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)$$

在统计中，参数的估计值分别为

$$\hat{\mu}_1=\bar{x},\hat{\mu}_2=\bar{y},\hat{\sigma}_1^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,\hat{\sigma}_2^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2;\hat{\rho}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\hat{\sigma}_1\hat{\sigma}_2}$$

此时最佳线性预测即为通常的一元线性回归方程

$$\hat{y}=\hat{a}x+\hat{b}$$

其中

$$\hat{a}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\hat{b}=\bar{y}-\hat{a}\bar{x}$$

★ 协方差与相关系数的性质

定理 $Cov(X,Y)$ 是对称的双线性函数，即有

(1) $Cov(X,Y)=Cov(Y,X)$.

(2) 若 a, b 是两个任意常数，则 $Cov(aX,bY)=abCov(X,Y)$

(3) $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

(4) 相关系数 $r_{X,Y}$ 满足 $|r_{X,Y}|\le 1$，且 $r_{X,Y}=\pm 1 \Leftrightarrow P\left(X=\hat{a}Y+\hat{b}\right)=1$（具体的正负号，

视 $\hat{a}$ 的正负而定）

(5) $D(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i D(X_i) + 2\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i<j}}^{n} a_i a_j Cov(X_i, X_j)$

例（独立一定不相关，反之不成立）$X \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \end{pmatrix}$，则 X 与 X^2 不独立，但不相关。

★ 对随机向量 $(X_1,\cdots,X_n)$ 有协方差矩阵和相关矩阵

$\Sigma = (Cov(X_i, X_j))_{i,j\le n}$，称为协**方差矩阵；**

$R = (r_{X_i,X_j})_{i,j\le n}$，称为相关系数**矩阵，它们均为对称矩阵。（注意对角线上的元素）**

§4 条件数学期望

1. 条件数学期望

(1) 设 (X,Y) 为二维离散型随机向量，其取值为 $\{(x_i, y_j), i, j = 1,2,\cdots\}$。对一切 $y \in \{y_1, y_2, \cdots, y_n, \cdots\}$，由于 X 在 $Y=y$ 的条件下的条件分布 $\{P_{X|Y}(x_i \mid y)\}$ 是一个概率分布. X 在这个分布下的数学期望就定义为**给定 $Y=y$ 的条件下，X 的条件数学期望**（如果它存在），我们将它记为 $E(X \mid Y=y)$. 于是我们有

$$E(X \mid Y=y) = \sum_i x_i P_{X|Y}(x_i \mid y).$$

(2) 将 $E(X \mid Y=y)$ 看成 y 的函数（定义域为 $\{y_1, y_2, \cdots, y_n, \cdots\}$），记之为 $g(y)$，则**随机变量 $g(Y(\omega))$ 就定义为 X 对 Y 的条件数学期望，记为 $E(X \mid Y)$,它是一个随机变量。**

2. 条件数学期望的性质

定理 1（**全期望公式**）若随机变量 X 的数学期望 EX 存在，且 $P(Y=y_j) > 0$, j=1,2,...，则

$$E(E(X|Y)) = E(g(Y)) = \sum_j P(Y=y_j) E(X|Y=y_j) = EX$$

例子（书上小猫例子）
思考：改小猫为有判断能力的人，结果如何？

定理 2（**最佳预测**）设 X 与 Y 的平方的数学期望均存在, 则

$$E[X - E(X \mid Y)]^2 = \min_{\varphi} E[X - \varphi(Y)]^2$$

提供另一证明：先证明

$$E\{[X-\varphi(Y)]^2\}=E\{[X-E(X\mid Y)]^2\}+E\{[E(X\mid Y)-\varphi(Y)]^2\}$$

我们先考虑

$$\begin{aligned}
&E\{[X-\varphi(Y)]^2\mid Y\}=E\{[X-E(X\mid Y)+E(X\mid Y)-\varphi(Y)]^2\mid Y\}\\
&=E\{[X-E(X\mid Y)]^2+2[X-E(X\mid Y)][E(X\mid Y)-\varphi(Y)]+[E(X\mid Y)-\varphi(Y)]^2\mid Y\}\\
&=E\{[X-E(X\mid Y)]^2\mid Y\}+E\{[E(X\mid Y)-\varphi(Y)]^2\mid Y\}+2E\{[X-E(X\mid Y)][E(X\mid Y)-\varphi(Y)]\mid Y\}
\end{aligned}$$

由于

$$\begin{aligned}
&E\{[X-E(X\mid Y)][E(X\mid Y)-\varphi(Y)]\mid Y\}\\
&=[E(X\mid Y)-\varphi(Y)]E\{[X-E(X\mid Y)]\mid Y\}\\
&=[E(X\mid Y)-\varphi(Y)]\{E(X\mid Y)-E(X\mid Y)\}\\
&=0
\end{aligned}$$

于是

$$E\{[X-\varphi(Y)]^2\mid Y\}==E\{[X-E(X\mid Y)]^2\mid Y\}+E\{[E(X\mid Y)-\varphi(Y)]^2\mid Y\}$$

两边取数学期望，由全期望公式得到，

$$E\{[X-\varphi(Y)]^2\}=E\{[X-E(X\mid Y)]^2\}+E\{[E(X\mid Y)-\varphi(Y)]^2\}$$

从而可知道，当 $E\{[E(X\mid Y)-\varphi(Y)]^2\}=0, i.e., \varphi(Y)=E(X\mid Y)$ **时，取到最小。**

推论 1： $E\{[X-\varphi(Y)]^2\}\geq E\{[X-E(X\mid Y)]^2\}$

推论 2： $DX=E\{[X-E(X\mid Y)]^2\}+D(E(X\mid Y))$

推论 3： $DX\geq D(E(X\mid Y))$**，且等号成立当且仅当** $X=E(X\mid Y)$ **（即 X 是 Y 的一个函数）。**

定理 3　(1) 若 $a\leq X\leq b$, a, b 为常数，则 $E(X|Y)$ 存在，且 $a\leq E(X|Y)\leq b$ 。

(2) 若 $E(X_1|Y)$ 与 $E(X_2|Y)$ 存在，a,b 为常数，则 $E(aX_1+bX_2|Y)$ 也存在，且

$$E(aX_1+bX_2|Y)=aE(X_1|Y)+bE(X_2|Y)$$

定理 4　设离散型随机变量 X 与 Y 相互独立，则 $E(X|Y)=EX$.

定理 5　设（X,Y）为两维离散型随机变量，且 $g(\cdot),h(\cdot)$ 为任意两个实值函数，则

$$E(g(X)h(Y)|Y)=h(Y)E(g(X)|Y)$$

§5 随机徘徊与独立增量过程

1. 随机过程的定义

定义 1： 设 (Ω, $\mathcal{F}$, P) 是一个概率空间，一族依赖于参数 t ($t \in T$) 的随机变量 $\boldsymbol{X}$ = $\{X_t : t \in T\}$称为一个**随机过程**，其中 T 称为指标集，对 T 中的每个 t，X_t 是一个随机变量 $X_t(\omega)$。 对每个固定的基本事件 ω，{ $X_t(\omega)$: $t \in T$ }是一个定义在 T 上的实值函数，称之为随机过程 $\boldsymbol{X}$ 的一条**(样本)轨道**。

● **随机过程的分类**

按过程的指标集和状态空间分类，有

离散时间、离散状态（如随机徘徊等）；离散时间、连续状态；

连续时间、离散状态（如 Poisson 过程）；连续时间、连续状态（如 Brown 运动）

● **随机过程的有限维分布族**

对任一随机过程$\{X_t : t \in T\}$，定义：

$$F_{t_1}(x_1) \equiv P(X_{t_1} \le x_1);$$
$$F_{t_1,t_2}(x_1,x_2) \equiv P(X_{t_1} \le x_1, X_{t_2} \le x_2);$$
$$\cdots\cdots$$
$$F_{t_1,\cdots,t_k}(x_1,\cdots,x_k) \equiv P(X_{t_1} \le x_1,\cdots,X_{t_k} \le x_k);$$

为过程的**有限维分布族，它满足：**

（**a**） **对称性：若**$\{i_1,\cdots,i_j\}$**为**$\{1,\cdots,j\}$**的一个排列，则**

$$F_{t_{i_1},\cdots,t_{i_j}}(x_{i_1},\cdots,x_{i_j}) = F_{t_1,\cdots,t_j}(x_1,\cdots,x_j)$$

（b） 相容性：对任意 $i < j$，有

$$F_{t_1,\cdots,t_i,t_{i+1},\cdots,t_j}(x_1,\cdots,x_i,\infty,\cdots,\infty) = F_{t_1,\cdots,t_i}(x_1,\cdots,x_i)$$

定理（Kolmogorov）设$\{F_{t_1,\cdots,t_j}(x_1,\cdots,x_j)\}$是满足对称性和相容性条件的一函数族，则存在一概率空间(Ω, $\mathcal{F}$, P)和定义于其上的随机过程$\{X_t : t \in T\}$，使 X_t 的**有限维分布族为** $\{F_{t_1,\cdots,t_j}(x_1,\cdots,x_j)\}$。

有限维分布族是研究随机过程的最重要的工具之一。

● 独立增量过程

以下我们介绍一类重要的随机过程—独立增量过程

定义 设 $\boldsymbol{X}$ = $\{X_t\}$是一个随机过程，如果它在任意 s 个互不相交的区间上的增量 $X_{m_1} - X_{n_1}$ ，$X_{m_2} - X_{n_2}$，… ，$X_{m_s} - X_{n_s}$ 都相互独立，则称随机过程 $\boldsymbol{X}$ 为一个**独立增量过程**。又如果对任意的 $n>0$，都有 $X_{m+n} - X_m$ ($n > 0$) 对一切 m 同分布，则称 $\boldsymbol{X}$ 为一个**时**

齐的独立增量过程。

结论：独立增量过程的有限维分布由其增量的分布和过程的初始分布唯一决定；时齐的独立增量过程的有限维分布由其增量的（一维）分布和过程的初始分布唯一决定。

简单证明：为方便起见，不妨考虑离散状态的随机过程，且设 $X_0 = x_0$ ， $t_1 < t_2 < \cdots < t_k$ ，**过程的有限维分布为**

$$P(X_{t_1} = x_1, \cdots, X_{t_k} = x_k) = P(X_0 = x_0, X_{t_1} - X_0 = x_1 - x_0, \cdots, X_{t_k} - X_{t_{k-1}} = x_k - x_{k-1})$$
$$= P(X_0 = x_0)P(X_{t_1} - X_0 = x_1 - x_0) \cdots P(X_{t_k} - X_{t_{k-1}} = x_k - x_{k-1})$$

若为**时齐的独立增量过程，则**

$$P(X_{t_k} - X_{t_{k-1}} = x_k - x_{k-1}) = P(X_{t_k - t_{k-1}} - X_0 = x_k - x_{k-1}) = P(X_{t_k - t_{k-1}} = x_k - x_{k-1} + x_0)$$

2. 随机徘徊及其简单性质

定义 设 $(\Omega, \mathcal{F}, P)$ 是一个概率空间, 其上的独立同分布的随机变量序列

$\{Z_n\}$ 满足

$$Z_n \sim \begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix} \qquad q \equiv 1 - p \qquad (n=1,2,\ldots).$$

令

$$X_n = X_0 + \sum_{i=1}^{n} Z_i \qquad (n=1,2,\ldots). \qquad (2.26)$$

我们称随机变量序列 $\boldsymbol{X} = \{X_n, n \geq 1\}$ 为(1-维)**简单随机徘徊**,其中 X_0 是任意一个取整数值的随机变量, 通常取 X_0 为 x. 特别, 当 $p = \dfrac{1}{2}$ 时，称之为(1-维)**对称的简单随机徘徊**。

● **简单随机徘徊的有限维分布**

定理：简单随机徘徊是时齐的独立增量过程，其一维分布为

$$P(X_n = k) = \begin{cases} C_n^{\frac{n+k-x}{2}} p^{\frac{n+k-x}{2}} (1-p)^{\frac{n-k+x}{2}}, & (n \geq |k-x|,\ n\text{与}k-x\text{同奇偶}) \\ 0, & (\text{其他情形}) \end{cases}$$

简单证明：只证一维分布的求法

由于 $Z_n \sim \begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix}$ ，将其化为标准 0-1 分布，令 $Y_n = \dfrac{Z_n + 1}{2}$ ，则 $Y_n \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$

从而 $X_n = X_0 + \sum_{i=1}^{n}(2Y_i - 1) = x + 2\sum_{i=1}^{n} Y_i - n$ ，由于 $\sum_{i=1}^{n} Y_i \sim B(n, p)$ ，故

$$P(X_n=k)=P\left(x+2\sum_{i=1}^{n}Y_i-n=k\right)=P(\sum_{i=1}^{n}Y_i=\frac{k-x+n}{2})$$，故

$$P(X_n=k)=\begin{cases}C_n^{\frac{n+k-x}{2}}p^{\frac{n+k-x}{2}}(1-p)^{\frac{n-k+x}{2}}, & (n\geq |k-x|,\ n与k-x同奇偶)\\ 0 & ,\ (其他情形)\end{cases}$$

有限维分布（当 $X_0=0$ 时）为

$$P(X_{n_1}=s_1,X_{n_2}=s_2,\ldots,X_{n_k}=s_k)=P(\sum_{i=1}^{n_1}Z_i=s_1,\sum_{i=n_1+1}^{n_2}Z_i=s_2-s_1,\ldots,\sum_{i=n_{k-1}+1}^{n_k}Z_i=s_k-s_{k-1})$$

$$=\prod_{l=1}^{k}C_{m_l}^{r_l}p^{r_l}(1-p)^{m_l-r_l}\qquad (m_l=n_l-n_{l-1},r_l=\tfrac{1}{2}(n_l-n_{l-1}+s_l-s_{l-1}))$$

- **简单随机徘徊的分布性质**

设 $X_0=x$，则有

（1）

$$EX_n^x=x+E(\sum_{i=1}^{n}Z_i)=x+nEZ_i=x+n(p-q)$$

$$D(X_n^x)=D(x+\sum_{i=1}^{n}Z_i)=nDZ_i=4npq$$

（2） $Cov(X_n^x,X_m^x)=4(n\wedge m)pq$.

- **简单随机徘徊的轨道性质（上课没讲，供学习参考）**

记 $T_y^x=min\{n\geq 0:X_n^x=y\}$ 为从 x 出发首次到达状态 y 的时刻，对整数 $c<d$，定义

$$\phi(x)=P(T_d^x<T_c^x)$$

即 $\phi(x)$ 表示从 x 出发，到达 c 之前先到达 d 的概率。由于

$$\phi(x)=p\phi(x+1)+q\phi(x-1)$$

从而

$$\phi(x+1)-\phi(x)=\frac{q}{p}\left[\phi(x)-\phi(x-1)\right],\quad c+1\leq x\leq d-1$$

$$\phi(c)=0,\quad \phi(d)=1。$$

情形 1 ： $p\neq q$

$$\phi(x)=\sum_{y=c}^{x-1}[\phi(y+1)-\phi(y)]=\phi(c+1)\frac{1-(\frac{q}{p})^{x-c}}{1-\frac{q}{p}}$$，由于 $\phi(d)=1$，知 $\phi(c+1)=\dfrac{1-\frac{q}{p}}{1-(\frac{q}{p})^{d-c}}$

所以

$$\phi(x)=P(T_d^x<T_c^x)=\frac{1-(\frac{q}{p})^{x-c}}{1-(\frac{q}{p})^{d-c}},\quad c\le x\le d, p\ne q$$

同理，

$$\psi(x)=P(T_c^x<T_d^x)=\frac{1-(\frac{p}{q})^{d-x}}{1-(\frac{p}{q})^{d-c}},\quad c\le x\le d, p\ne q \text{。}$$

情形 2 ： $p=q=1/2$，

$$\phi(x)=P(T_d^x<T_c^x)=\frac{x-c}{d-c},\quad c\le x\le d, p=q$$

$$\psi(x)=P(T_c^x<T_d^x)=\frac{d-x}{d-c},\quad c\le x\le d, p=q \text{。}$$

讨论：（1）$\phi(x)+\psi(x)=1$，即从$[c,d]$内部出发，最终一定到达边界。

（2）若 $x>c$，则

$$P(X_n^x\text{能到达}c)=P(T_c^x<\infty)=\lim_{d\to\infty}\psi(x)=\begin{cases}(\frac{q}{p})^{x-c} & p>\frac{1}{2}\\ 1 & p\le\frac{1}{2}\end{cases};$$

同理，若 $x<d$，则

$$P(X_n^x\text{能到达}d)=P(T_d^x<\infty)=\lim_{c\to-\infty}\varphi(x)=\begin{cases}1 & p\ge\frac{1}{2}\\ (\frac{p}{q})^{d-x} & p<\frac{1}{2}\end{cases}$$