# 第三章　多维随机变量及其分布

## §3.1　多维随机变量及其联合分布

**1. $n$ 维随机变量**　如果 $X_1(\omega),X_2(\omega),\cdots,X_n(\omega)$ 是定义在同一个样本空间 $\Omega=\{\omega\}$ 上的 $n$ 个随机变量，则称 $\boldsymbol{X}(\omega)=(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))$ 为 $n$ 维随机变量，或 $n$ 元随机变量，或随机向量.

**2. 联合分布函数**　对任意的 $n$ 个实数 $x_1,x_2,\cdots,x_n$，$n$ 个事件 $X_1\leqslant x_1$，$X_2\leqslant x_2,\cdots,X_n\leqslant x_n$ 同时发生的概率

$$F(x_1,x_2,\cdots,x_n)=P(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n)$$

称为 $n$ 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的联合分布函数.

二维随机变量 $(X,Y)$ 的联合分布函数 $F(x,y)=P(X\leqslant x,Y\leqslant y)$ 具有如下四条基本性质：

(1) **单调性**　$F(x,y)$ 分别对 $x$ 或 $y$ 是单调不减的.

(2) **有界性**　对任意的 $x$ 和 $y$，有 $0\leqslant F(x,y)\leqslant 1$，且

$$F(-\infty,y)=F(x,-\infty)=0,\qquad F(+\infty,+\infty)=1.$$

(3) **右连续性**　对每个变量都是右连续的，即

$$F(x+0,y)=F(x,y),\qquad F(x,y+0)=F(x,y).$$

(4) **非负性**　对任意的 $a<b,c<d$ 有

$$P(a<X\leqslant b,c<Y\leqslant d)=F(b,d)-F(a,d)-F(b,c)+F(a,c)\geqslant 0.$$

可以证明：具有上述四条性质的二元函数 $F(x,y)$ 一定是某个二维随机变量的分布函数.注意：事件“$X\leqslant x,Y\leqslant y$”常可用平面上的**无穷直角区域**表示.

**3. 联合分布列**　如果 $(X,Y)$ 只取有限个或可列个数对 $(x_i,y_j)$，则称 $(X,Y)$ 为二维离散随机变量，称 $p_{ij}=P(X=x_i,Y=y_j)$，$i,j=1,2,\cdots$ 为 $(X,Y)$ 的联合分布列，联合分布列也可用如下表格形式表示：

| $X$ | $Y$ | | | | |
|---|---|---|---|---|---|
| | $y_1$ | $y_2$ | $\cdots$ | $y_j$ | $\cdots$ |
| $x_1$ | $p_{11}$ | $p_{12}$ | $\cdots$ | $p_{1j}$ | $\cdots$ |
| $x_2$ | $p_{21}$ | $p_{22}$ | $\cdots$ | $p_{2j}$ | $\cdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | |
| $x_i$ | $p_{i1}$ | $p_{i2}$ | $\cdots$ | $p_{ij}$ | $\cdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | |

联合分布列的基本性质：

（1）**非负性**：$p_{ij} \geqslant 0$；　（2）**正则性**：$\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} p_{ij} = 1$.

**4. 联合密度函数**　如果存在二元非负函数 $p(x,y)$，使得二维随机变量 $(X,Y)$ 的分布函数 $F(x,y)$ 可表示为

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)\,\mathrm{d}v\mathrm{d}u,$$

则称 $(X,Y)$ 为二维连续随机变量，称 $p(x,y)$ 为 $(X,Y)$ 的联合密度函数.

联合密度函数的基本性质：

（1）**非负性**：$p(x,y) \geqslant 0$；　（2）**正则性**：$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)\,\mathrm{d}x\mathrm{d}y = 1$.

在 $F(x,y)$ 偏导数存在的点上有

$$p(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y).$$

若 $G$ 为平面上的一个区域，则有

$$P((X,Y) \in G) = \iint_G p(x,y)\,\mathrm{d}x\mathrm{d}y.$$

**5. 多项分布**　在 $n$ 次独立重复试验中，如果每次试验有 $r$ 个可能结果：$A_1, A_2, \cdots, A_r$，且每次试验中 $A_i$ 发生的概率为 $p_i = P(A_i)$，$i=1,2,\cdots,r$. $p_1+p_2+\cdots+p_r=1$. 记 $X_i$ 为 $n$ 次独立重复试验中 $A_i$ 出现的次数，$i=1,2,\cdots,r$. 则 $(X_1, X_2, \cdots, X_r)$ 服从**多项分布**，又称 $r$ **项分布**，记为 $M(n, p_1, p_2, \cdots, p_r)$，其联合分布列为

$$P(X_1=n_1, X_2=n_2, \cdots, X_r=n_r) = \frac{n!}{n_1!\ n_2!\ \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r},$$

其中 $n = n_1+n_2+\cdots+n_r$.

当 $r=2$ 时,即为二项分布.

**6. 多维超几何分布** 有 $N$ 个对象,共分 $r$ 类,其中第 $i$ 类对象有 $N_i$ 个,$N=N_1+N_2+\cdots+N_r$.从中随机取出 $n$ 个,若记 $X_i$ 为取出的 $n$ 个对象中第 $i$ 类对象的个数,$i=1,2,\cdots,r$,则$(X_1,X_2,\cdots,X_r)$服从 $r$ 维超几何分布,其联合分布列为

$$P(X_1=n_1,X_2=n_2,\cdots,X_r=n_r)=\frac{\binom{N_1}{n_2}\binom{N_2}{n_2}\cdots\binom{N_r}{n_r}}{\binom{N}{n}},$$

其中 $n_1+n_2+\cdots+n_r=n$.

**7. 多维均匀分布** 设 $D$ 为 $\mathbf{R}^n$ 中的一个有界区域,其度量(平面上为面积,空间为体积等)为 $S_D$,如果多维随机变量$(X_1,X_2,\cdots,X_n)$的联合密度函数为

$$p(x_1,x_2,\cdots,x_n)=\begin{cases}\dfrac{1}{S_D}, & (x_1,x_2,\cdots,x_n)\in D,\\ 0, & \text{其他},\end{cases}$$

则称$(X_1,X_2,\cdots,X_n)$服从 $D$ 上的多维均匀分布,记为$(X_1,X_2,\cdots,X_n)\sim U(D)$.

**8. 二元正态分布** 如果二维随机变量$(X,Y)$的联合密度函数为

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\},\quad -\infty<x,y<+\infty,$$

则称$(X,Y)$服从二元正态分布,记为$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.其中

$$E(X)=\mu_1,\quad E(Y)=\mu_2;\quad \operatorname{Var}(X)=\sigma_1^2,\quad \operatorname{Var}(Y)=\sigma_2^2;\quad -1\leqslant\rho\leqslant 1.$$

## §3.2 边际分布与随机变量的独立性

**1. 边际分布函数** 若二维随机变量$(X,Y)$的联合分布函数为$F(x,y)$,则称

$$F_X(x)=F(x,+\infty)=\lim_{y\to+\infty}F(x,y),\qquad -\infty<x<+\infty$$

为$X$的边际分布函数.称

$$F_Y(y)=F(+\infty,y)=\lim_{x\to+\infty}F(x,y),\qquad -\infty<y<+\infty$$

为$Y$的边际分布函数.

**2. 边际分布列** 若二维离散随机变量$(X,Y)$的联合分布列为$\{p_{ij}\}$,则称

$$p_{i\cdot}=\sum_{j=1}^{+\infty}p_{ij},\qquad i=1,2,\cdots$$

为$X$的边际分布列.称

$$p_{\cdot j}=\sum_{i=1}^{+\infty}p_{ij},\qquad j=1,2,\cdots$$

为$Y$的边际分布列.

**3. 边际密度函数** 若二维连续随机变量$(X,Y)$的联合密度函数为$p(x,y)$,则称

$$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)\,\mathrm{d}y,\qquad -\infty<x<+\infty$$

为$X$的边际密度函数.称

$$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)\,\mathrm{d}x,\qquad -\infty<y<+\infty$$

为$Y$的边际密度函数.

**4. 一些注意点**

- 由高维联合分布可以获得低维的边际分布,反之不一定.
- 不同的联合分布可以有相同边际分布.
- 多项分布的边际分布仍为低维的多项分布或二项分布.
- 二维正态分布的边际分布为一维正态分布.

**5. 随机变量间的独立性**

(1) 设 $n$ 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的联合分布函数为 $F(x_1,x_2,\cdots,x_n)$,且 $F_{X_i}(x_i)$ 为 $X_i$ 的边际分布函数.如果对任意 $n$ 个实数 $x_1,x_2,\cdots,x_n$,有

$$F(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}F_{X_i}(x_i),$$

则称 $X_1,X_2,\cdots,X_n$ 相互独立.否则称 $X_1,X_2,\cdots,X_n$ 不相互独立.

(2) 设 $n$ 维离散随机变量 $(X_1,X_2,\cdots,X_n)$ 的联合分布列为 $P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)$,且 $P(X_i=x_i)$ 为 $X_i$ 的边际分布列,$i=1,2,\cdots,n$.如果对其任意 $n$ 个取值 $x_1,x_2,\cdots,x_n$,有

$$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)=\prod_{i=1}^{n}P(X_i=x_i),$$

则称 $X_1,X_2,\cdots,X_n$ 相互独立.否则称 $X_1,X_2,\cdots,X_n$ 不相互独立.

(3) 设 $n$ 维连续随机变量 $(X_1,X_2,\cdots,X_n)$ 的联合密度函数为 $p(x_1,x_2,\cdots,x_n)$,且 $p_{X_i}(x_i)$ 为 $X_i$ 的边际密度函数.如果对任意 $n$ 个实数 $x_1,x_2,\cdots,x_n$,有

$$p(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}p_{X_i}(x_i),$$

则称 $X_1,X_2,\cdots,X_n$ 相互独立.否则称 $X_1,X_2,\cdots,X_n$ 不相互独立.

**6. 一些注意点**

- 多维随机变量间相互独立,必导致其部分随机变量与另一部分随机变量相互独立.
- 多维随机变量相互独立,其联合分布可由其边际分布唯一确定.
- 多维随机变量间的独立性可从定义出发加以判别,也可从实际背景加以判别.
- 多维随机变量间的独立性假设,可给理论研究和实际运用带来很多方便之处.

## §3.3 多维随机变量函数的分布

**1. 最大值、最小值分布** 设 $(X_1,X_2,\cdots,X_n)$ 是相互独立、同分布的 $n$ 维

连续随机变量,其共同的密度函数和分布函数分别为 $p(x)$ 和 $F(x)$,记

$$Y=\min\{X_1,X_2,\cdots,X_n\},\qquad Z=\max\{X_1,X_2,\cdots,X_n\}.$$

则

$$F_Y(y)=1-[1-F(y)]^n;\qquad p_Y(y)=n[1-F(y)]^{n-1}p(y);$$

$$F_Z(z)=[F(z)]^n;\qquad p_Z(z)=n[F(z)]^{n-1}p(z).$$

**2. 变量变换法** 若变换$\begin{cases}u=g_1(x,y),\\v=g_2(x,y)\end{cases}$存在唯一的反函数$\begin{cases}x=x(u,v),\\y=y(u,v),\end{cases}$且变换的雅可比行列式

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}=\left(\frac{\partial(u,v)}{\partial(x,y)}\right)^{-1}=\left(\begin{vmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{vmatrix}\right)^{-1}\neq 0.$$

则二维连续随机变量$(X,Y)$的函数$\begin{cases}U=g_1(X,Y)\\V=g_2(X,Y)\end{cases}$的联合密度函数为

$$p_{U,V}(u,v)=p_{X,Y}(x(u,v),y(u,v))\,|J|.$$

**3. 卷积公式** 卷积公式用于求随机变量和 $Z=X+Y$ 的分布.

**(1) 离散场合的卷积公式** $Z=X+Y$ 的分布列为

$$P(Z=z_k)=\sum_{i=1}^{+\infty}P(X=x_i,Y=z_k-x_i)$$

$$\text{或}=\sum_{j=1}^{+\infty}P(X=z_k-y_j,Y=y_j).$$

当 $X$ 与 $Y$ 独立时,

$$P(Z=z_k)=\sum_{i=1}^{+\infty}P(X=x_i)P(Y=z_k-x_i)$$

$$\text{或}=\sum_{j=1}^{+\infty}P(X=z_k-y_j)P(Y=y_j).$$

其中诸 $x_i$ 为 $X$ 的取值,诸 $y_j$ 为 $Y$ 的取值.

**(2) 连续场合的卷积公式** $Z=X+Y$ 的密度函数为

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{X,Y}(x,z-x)\,\mathrm{d}x$$

$$或 =\int_{-\infty}^{+\infty} p_{X,Y}(z-y,y)\mathrm{d}y.$$

当 $X$ 与 $Y$ 独立时,

$$p_Z(z)=\int_{-\infty}^{+\infty} p_X(x)p_Y(z-x)\mathrm{d}x$$

$$或 =\int_{-\infty}^{+\infty} p_X(z-y)p_Y(y)\mathrm{d}y.$$

**4. 积的公式** $U=XY$ 的密度函数为

$$p_U(u)=\int_{-\infty}^{+\infty} p_{X,Y}(u/v,v)\frac{1}{|v|}\mathrm{d}v$$

$$或 =\int_{-\infty}^{+\infty} p_{X,Y}(v,u/v)\frac{1}{|v|}\mathrm{d}v.$$

当 $X$ 与 $Y$ 独立时,

$$p_U(u)=\int_{-\infty}^{+\infty} p_X(u/v)p_Y(v)\frac{1}{|v|}\mathrm{d}v$$

$$或 =\int_{-\infty}^{+\infty} p_X(v)p_Y(u/v)\frac{1}{|v|}\mathrm{d}v.$$

**5. 商的公式** $U=X/Y$ 的密度函数为

$$p_U(u)=\int_{-\infty}^{+\infty} p_{X,Y}(uv,v)|v|\mathrm{d}v$$

$$或 =\int_{-\infty}^{+\infty} p_{X,Y}(v,v/u)\frac{|v|}{u^2}\mathrm{d}v.$$

当 $X$ 与 $Y$ 独立时,

$$p_U(u)=\int_{-\infty}^{+\infty} p_X(uv)p_Y(v)|v|\mathrm{d}v$$

$$或 =\int_{-\infty}^{+\infty} p_X(v)p_Y(v/u)\frac{|v|}{u^2}\mathrm{d}v.$$

**6. 分布的可加性** 若同一类分布的独立随机变量和的分布仍属于此类分布,则称此类分布具有可加性.以下一些常用分布具有可加性:

(1) 二项分布:若 $X\sim b(n,p)$, $Y\sim b(m,p)$,且 $X$ 与 $Y$ 独立,则 $Z=X+Y\sim b(n+m,p)$.注意这里两个二项分布中的参数 $p$ 要相同.

(2) 泊松分布:若 $X \sim P(\lambda_1)$,$Y \sim P(\lambda_2)$,且 $X$ 与 $Y$ 独立,则 $Z = X+Y \sim P(\lambda_1+\lambda_2)$.

(3) 正态分布:若 $X \sim N(\mu_1, \sigma_1^2)$,$Y \sim N(\mu_2, \sigma_2^2)$,且 $X$ 与 $Y$ 独立,则 $Z = X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

(4) 伽马分布:若 $X \sim Ga(\alpha_1, \lambda)$,$Y \sim Ga(\alpha_2, \lambda)$,且 $X$ 与 $Y$ 独立,则 $Z = X+Y \sim Ga(\alpha_1+\alpha_2, \lambda)$.注:这里两个伽马分布中的尺度参数 $\lambda$ 要相同.

(5) $\chi^2$ 分布:若 $X \sim \chi^2(n_1)$,$Y \sim \chi^2(n_2)$,且 $X$ 与 $Y$ 独立,则 $Z = X+Y \sim \chi^2(n_1+n_2)$.

**7. 一些结论**

(1) 设 $X_1, X_2, \cdots, X_n$ 独立同分布,都服从二点分布 $b(1,p)$,则 $X_1+X_2+\cdots+X_n$ 服从二项分布 $b(n,p)$.

(2) 设 $X_1, X_2, \cdots, X_n$ 独立同分布,都服从几何分布 $Ge(p)$,则 $X_1+X_2+\cdots+X_n$ 服从负二项分布 $Nb(n,p)$.

(3) 设 $X_1, X_2, \cdots, X_n$ 独立同分布,都服从指数分布 $Exp(\lambda)$,则 $X_1+X_2+\cdots+X_n$ 服从伽马分布 $Ga(n,\lambda)$.

## § 3.4 多维随机变量的特征数

**1. 多维随机变量函数的数学期望** 若二维随机变量 $(X,Y)$ 的分布用联合分布列 $P(X=x_i, Y=y_j)$ 或用联合密度函数 $p(x,y)$ 表示,则 $Z=g(X,Y)$ 的数学期望(假设存在)为

$$E[g(X,Y)] = \begin{cases} \sum_i \sum_j g(x_i, y_j) P(X=x_i, Y=y_j), & \text{在离散场合}, \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) p(x,y) \, \mathrm{d}x \mathrm{d}y, & \text{在连续场合}. \end{cases}$$

对 $n$ 维随机变量结论是类似的.

**2. 数学期望与方差的运算性质** 以下均假定有关的期望和方差存在.

(1) $E(X_1+X_2+\cdots+X_n) = E(X_1)+E(X_2)+\cdots+E(X_n)$.

(2) 若 $X_1, X_2, \cdots, X_n$ 相互独立,则

$$E(X_1 X_2 \cdots X_n) = E(X_1) E(X_2) \cdots E(X_n).$$

$$\mathrm{Var}(X_1 \pm X_2 \pm \cdots \pm X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n).$$

**3. 协方差** 若 $E[(X-E(X))(Y-E(Y))]$ 存在，则称

$$\mathrm{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))]$$

为 $X$ 与 $Y$ 的协方差.

- 当 $\mathrm{Cov}(X,Y)>0$ 时，称 $X$ 与 $Y$ **正相关**，即 $X$ 与 $Y$ 同时增加或同时减少，
- 当 $\mathrm{Cov}(X,Y)<0$ 时，称 $X$ 与 $Y$ **负相关**，即 $X$ 增加 $Y$ 减少，或 $X$ 减少 $Y$ 增加.
- 当 $\mathrm{Cov}(X,Y)=0$ 时，称 $X$ 与 $Y$ **不相关**.

**4. 协方差的性质**

(1) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$；

(2) $\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$；

(3) 若 $X$ 与 $Y$ 相互独立，则 $\mathrm{Cov}(X,Y)=0$，反之不然；

(4) $\mathrm{Cov}(X+Y,Z)=\mathrm{Cov}(X,Z)+\mathrm{Cov}(Y,Z)$；

(5) 对任意的常数 $a$，有 $\mathrm{Cov}(X,a)=0$；

(6) 对任意的常数 $a,b$，有 $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$；

(7) 对任意二维随机变量 $(X,Y)$，有

$$\mathrm{Var}(X \pm Y)=\mathrm{Var}(X)+\mathrm{Var}(Y) \pm 2\mathrm{Cov}(X,Y).$$

对任意 $n$ 个随机变量 $X_1,X_2,\cdots,X_n$，有

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{Cov}(X_i,X_j).$$

**5. 施瓦茨不等式** 对任意二维随机变量 $(X,Y)$，若 $X$ 与 $Y$ 的方差都存在，则有

$$[\mathrm{Cov}(X,Y)]^2 \leqslant \mathrm{Var}(X)\mathrm{Var}(Y).$$

**6. 相关系数** 设 $(X,Y)$ 是一个二维随机变量，且 $\mathrm{Var}(X)>0$，$\mathrm{Var}(Y)>0$. 则称

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}=\frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

为 $X$ 与 $Y$ 的(线性)相关系数.

**7. 相关系数的性质**

(1) $-1\leqslant \mathrm{Corr}(X,Y)\leqslant 1$.

(2) $\mathrm{Corr}(X,Y)$ 与 $\mathrm{Cov}(X,Y)$ 同号,或同为零.

(3) $\mathrm{Corr}(X,Y)=\mathrm{Cov}(X^*,Y^*)$,其中 $X^*,Y^*$ 分别为 $X,Y$ 的标准化随机变量.

(4) $\mathrm{Corr}(X,Y)=\pm 1$ 的充要条件是 $X$ 与 $Y$ 间几乎处处有线性关系,即存在 $a(a\neq 0)$ 与 $b$,使得 $P(Y=aX+b)=1$.其中当 $\mathrm{Corr}(X,Y)=1$ 时,有 $a>0$;当 $\mathrm{Corr}(X,Y)=-1$ 时,有 $a<0$.

(5) 在二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 场合,不相关与独立是等价的.

**8. 随机向量的数学期望与协方差阵** 记 $n$ 维随机向量为 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$,以下假设所涉及的数学期望、方差、协方差均存在.

(1) 随机向量 $\boldsymbol{X}$ 的数学期望向量为

$$E(\boldsymbol{X})=(E(X_1),E(X_2),\cdots,E(X_n))^{\mathrm{T}}.$$

(2) 随机向量 $\boldsymbol{X}$ 的协方差阵为

$$E[(\boldsymbol{X}-E(\boldsymbol{X}))(\boldsymbol{X}-E(\boldsymbol{X}))^{\mathrm{T}}]=\begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2,X_n) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Var}(X_n) \end{pmatrix}.$$

也记以上的协方差阵为 $\mathrm{Cov}(\boldsymbol{X})$,或记成 $\boldsymbol{\Sigma}$.

(3) 随机向量 $\boldsymbol{X}$ 的协方差阵 $\mathrm{Cov}(\boldsymbol{X})=(\mathrm{Cov}(X_i,X_j))_{n\times n}$ 是一个对称的非负定矩阵.

(4) **$n$ 元正态分布** 设 $n$ 维随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的协方阵为 $\boldsymbol{\Sigma}=\mathrm{Cov}(\boldsymbol{X})$,数学期望向量为 $\boldsymbol{a}=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$.又记 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$,则由密度函数

$$p(x_1,x_2,\cdots,x_n)=p(\boldsymbol{x})=\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$$

定义的分布称为 $n$ 元正态分布,记为 $\boldsymbol{X}\sim N(\boldsymbol{a},\boldsymbol{\Sigma})$.

## §3.5 条件分布与条件期望

**1. 离散随机变量的条件分布**

（1）**条件分布列** 对一切使 $P(Y=y_j)=p_{\cdot j}=\sum\limits_{i=1}^{+\infty}p_{ij}>0$ 的 $y_j$，称

$$p_{i|j}=P(X=x_i\mid Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}},\quad i=1,2,\cdots$$

为给定 $Y=y_j$ 条件下 $X$ 的条件分布列.

同理，对一切使 $P(X=x_i)=p_{i\cdot}=\sum\limits_{j=1}^{+\infty}p_{ij}>0$ 的 $x_i$，称

$$p_{j|i}=P(Y=y_j\mid X=x_i)=\frac{P(X=x_i,Y=y_j)}{P(X=x_i)}=\frac{p_{ij}}{p_{i\cdot}},\quad j=1,2,\cdots$$

为给定 $X=x_i$ 条件下 $Y$ 的条件分布列.

（2）**条件分布函数** 给定 $Y=y_j$ 条件下 $X$ 的条件分布函数为

$$F(x\mid y_j)=\sum_{x_i\leqslant x}P(X=x_i\mid Y=y_j)=\sum_{x_i\leqslant x}p_{i|j},$$

给定 $X=x_i$ 条件下 $Y$ 的条件分布函数为

$$F(y\mid x_i)=\sum_{y_j\leqslant y}P(Y=y_j\mid X=x_i)=\sum_{y_j\leqslant y}p_{j|i}.$$

**2. 连续随机变量的条件分布** 对一切使 $p_Y(y)>0$ 的 $y$，给定 $Y=y$ 条件下 $X$ 的条件密度函数和条件分布函数分别为

$$p(x\mid y)=\frac{p(x,y)}{p_Y(y)},\qquad F(x\mid y)=\int_{-\infty}^{x}p(u\mid y)\,\mathrm{d}u=\int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}\mathrm{d}u.$$

类似对一切使 $p_X(x)>0$ 的 $x$，给定 $X=x$ 条件下 $Y$ 的条件密度函数和条件分布函数分别为

$$p(y\mid x)=\frac{p(x,y)}{p_X(x)},\qquad F(y\mid x)=\int_{-\infty}^{y}p(v\mid x)\,\mathrm{d}v=\int_{-\infty}^{y}\frac{p(x,v)}{p_X(x)}\mathrm{d}v.$$

**3. 连续场合的全概率公式和贝叶斯公式**

（1）全概率公式的密度函数形式：

$$p_Y(y)=\int_{-\infty}^{+\infty}p_X(x)p(y\mid x)\,\mathrm{d}x,\qquad p_X(x)=\int_{-\infty}^{+\infty}p_Y(y)p(x\mid y)\,\mathrm{d}y.$$

（2）贝叶斯公式的密度函数形式：

$$p(x\mid y)=\frac{p_X(x)p(y\mid x)}{\int_{-\infty}^{+\infty}p_X(x)p(y\mid x)\,\mathrm{d}x},\qquad p(y\mid x)=\frac{p_Y(y)p(x\mid y)}{\int_{-\infty}^{+\infty}p_Y(y)p(x\mid y)\,\mathrm{d}y}.$$

**4. 条件数学期望** 条件分布的数学期望(若存在)称为条件期望,其定义如下:

$$E(X\mid Y=y)=\begin{cases}\sum\limits_i x_iP(X=x_i\mid Y=y), & (X,Y)\text{ 为二维离散随机变量},\\ \int_{-\infty}^{+\infty}xp(x\mid y)\,\mathrm{d}x, & (X,Y)\text{ 为二维连续随机变量},\end{cases}$$

$$E(Y\mid X=x)=\begin{cases}\sum\limits_j y_jP(Y=y_j\mid X=x), & (X,Y)\text{ 为二维离散随机变量},\\ \int_{-\infty}^{+\infty}yp(y\mid x)\,\mathrm{d}y, & (X,Y)\text{ 为二维连续随机变量}.\end{cases}$$

- 条件期望具有数学期望的一切性质.
- 条件期望 $E(X\mid Y=y)$ 是 $y$ 的函数,记为 $g(y)$,它是另一个随机变量 $g(Y)=E(X\mid Y)$ 的取值.$E(X\mid Y)$ 与 $E(X\mid Y=y)$ 虽都称为条件期望,但含义不同.前者是特定的随机变量,后者是其取值.

**5. 重期望公式** 设$(X,Y)$是二维随机变量,若 $E(X)$ 存在,则 $E(X)=E[E(X\mid Y)]$.注意:$E[E(X|Y)]$中里面的期望是用条件分布 $p(x|y)$ 计算的,外面的期望是用 $Y$ 的分布 $p(y)$ 计算的.