# 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

**$n$维随机变量(随机向量)**
若$X_1(\omega), X_2(\omega), ..., X_n(\omega)$是定义在同一样本空间$\Omega = \{\omega\}$上的随机变量，则称
$X(\omega) = (X_1(\omega), X_2(\omega), ..., X_n(\omega))$
为$n$维随机变量或随机向量。

> 关键:定义在同一样本空间上

**联合分布函数**
对任意$n$个实数,$n$个事件同时发生的概率$F(x_1, x_2, ..., x_n) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n)$称为$n$维随机变量的联合分布函数。

> **基本性质**:
> (1) 单调性:
> $x_1 < x_2, F(x_1, y) \leq F(x_2, y)$
> $y_1 < y_2, F(x, y_1) \leq F(x, y_2)$
> (2) 有界性:
> $\forall x, y, 0 \leq F(x, y) \leq 1$
> $F(-\infty, y) = 0$
> $F(x, -\infty) = 0$
> $F(+\infty, +\infty) = 1$
> (3) 右连续性：对每个变量都是右连续的
> (4) 非负性:
> $\forall a < b, c < d$
> $P(a < X \leq b, c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c) \geq 0$

**联合分布列**
二维离散分布：$p_{ij} = P(X = x_i, Y = y_j)$

> 基本性质:
> (1) 非负性：$p_{ij} \geq 0$
> (2) 正则性：$\sum_i \sum_j p_{ij} = 1$

**联合密度函数**

二维联合密度函数：$p(x,y)$

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

在$F(x,y)$偏导数存在的点上有$p(x,y)=\frac{\partial^2}{\partial x\partial y}F(x,y)$

> 基本性质：
> (1) 非负性：$p(x,y)\geq 0$
> (2) 正则性：$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}p(x,y)dydx=1$

**多项分布** $M(n,p_1,p_2,\dots,n_r)$

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

其中$n=n_1+n_2+\dots+n_r$

**多维超几何分布**

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

**多维均匀分布** $U(D)$

$$p(x_1,x_2,\dots,x_n)=\begin{cases}\frac{1}{S_D},(x_1,x_2,\dots,x_n)\in D\\ 0,其他\end{cases}$$

**二元正态分布** $N(\mu_1,\mu_2,\ \sigma_1^2,\sigma_2^2,\rho)$

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

---

# 3.2 边际分布与随机变量的独立性

**边际分布函数**：$F_X(x)=F(x,+\infty),F_Y(y)=F(+\infty,y)$

> 二维指数分布的边际函数是一维指数分布

**边际分布列**
$\sum_j P(X=x_i, Y=y_j) = P(X=x_i)$
$\sum_i P(X=x_i, Y=y_j) = P(Y=y_j)$

**边际密度函数**
$p_X(x) = \int_{-\infty}^{+\infty} p(x,y)dy$
$p_Y(y) = \int_{-\infty}^{+\infty} p(x,y)dx$

**多维变量的独立性**
$F(x,y) = F_X(x)F_Y(y)$
$p_{ij} = p_i p_j$
$p(x,y) = p_X(x)p_Y(y)$
满足三者之一，则称$X, Y$独立

---

# 3.3 多维随机变量函数的分布

**泊松分布可加性**
$X \sim P(\lambda_1), Y \sim P(\lambda_2), Z = X+Y \sim P(\lambda_1+\lambda_2)$

**离散场合的卷积公式**

$$P(Z=k) = \sum_{i=0}^{k} P(X=i)P(Y=k-i)$$

**二项分布可加性**
$X \sim b(n,p), Y \sim b(m,p), Z = X+Y \sim b(n+m,p)$

> 二项分布可以分为多个独立的伯努利分布

**最大最小值分布**：从不同分布函数，同分布函数，同概率密度函数，指数分布四个角度

**连续场合的卷积公式**

$$p_Z(z) = \int_{-\infty}^{+\infty} p_X(z-y)p_Y(y)dy = \int_{-\infty}^{+\infty} p_X(x)p_Y(z-x)dx$$

**正态分布的可加性**

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2), Z = X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

> n个相互独立的正态变量的组合仍然是正态变量

**伽马分布的可加性**
$X \sim Ga(\alpha_1, \lambda_), Y \sim Ga(\alpha_2, \lambda), Z = X + Y \sim Ga(\alpha_1 + \alpha_2, \lambda)$

> $n$个相互独立同分布的标准正态分布服从自由度为$n$的卡方分布

寻求二维连续随机变量函数的分布的方法:
**变量变换法 增补变量法**

---

# 3.4 多维随机变量的特征数

**期望**:

$$E(Z) = \begin{cases} \sum_i \sum_j g(x_i, y_j) p_{ij}, \text{离散} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) p(x, y) dy dx, \text{连续} \end{cases}$$

> 期望的性质:
> (1) $E(X + Y) = E(X) + E(Y)$
> (2) $X, Y$独立, $E(XY) = E(X)E(Y)$
> 方差的性质:
> (1) $X, Y$独立, $Var(X \pm Y) = Var(X) + Var(Y)$

**协方差**: $Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$

> 协方差的性质:
> (1) $Cov(X, Y) = E(XY) - E(X)E(Y)$
> (2) $X, Y$独立,$Cov(X, Y) = 0$
> (3) $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$
> (4) $Cov(X, Y) = Cov(Y, X)$
> (5) $Cov(X, a) = 0$
> (6) $Cov(aX, bY) = abCov(X, Y)$
> (7) $Cov(X + Y, Z) = Cov(X) + Cov(Y)$

> 独立则不相关，反之不然

**相关系数**: $Corr(X,Y)=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$

> 前提：$Var(X)>0,Var(Y)>0$

**施瓦茨(Schawrz)不等式**

$$[Cov(X,Y)]^2\leq\sigma_X^2\sigma_Y^2$$

> 前提：$X,Y$方差存在

> 相关系数的性质：
> $(1)\ |Corr(X,Y)|\leq 1$
> $(2)\ |Corr(X,Y)|=1$的充要条件$X,Y$间几乎处处有线性关系

> 二维正态分布场合，不相关和独立等价

**方差-协方差矩阵**

> 性质：对称，非负定

---

# 3.5 条件分布与条件期望

**条件分布列**：$p_{i|j}=\frac{p_{ij}}{p_j}$
**条件分布函数**：$F(x|y_i)=\sum_{x_i\leq x}p_{i|j}$

**条件密度函数**：$F(x|y)=\int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}du$
**条件分布函数**：$p(x|y)=\frac{p(x,y)}{p_Y(y)}$

**全概率公式**：$p_Y(y)=\int_{-\infty}^{+\infty}p_X(x)p(y|x)dx$
**贝叶斯公式**：$P(x|y)=\frac{p_X(x)p(y|x)}{\int_{-\infty}^{+\infty}p_X(x)p(y|x)dx}$

**条件数学期望**：

$$E(X|Y=y)=\begin{cases}\sum_i x_iP(x=x_i|Y=y)\text{离散型}\\ \int_{-\infty}^{+\infty}xp(x|y)dx,\text{连续型}\end{cases}$$

**重期望公式**: $E(X)=E(E(X|Y))$.

> 前提：$E(X)$存在