第 5 章　多维随机向量

很多随机现象可能由两种或多种随机因素造成的, 仅用一个随机变量来描述是不够的. 例如在研究某地区儿童的身体素质时, 往往需要同时观察他们的身高、体重、肺活量等; 在研究某种矿石的成分时, 往往需要同时考察铁含量、铜含量、磷含量等. 这些随机变量之间可能存在某些关联, 需要将其看作一个整体来研究, 不能简单地对每个随机变量单独进行研究.

定义 5.1　设 $X_1 = X_1(\omega), X_2 = X_2(\omega), \cdots, X_m = X_m(\omega)$ 是定义在同一样本空间 Ω 上的 m 个随机变量, 由它们构成的向量 $(X_1, X_2, \cdots, X_m)$ 称为 m **维随机向量**, 或称 m **维随机变量**.

一维随机变量可以看作多维随机变量的一种特殊情况, 本章主要讨论二维随机向量及其分布, 同理可讨论二维以上的随机向量.

5.1 二维联合分布函数

定义 5.2　设 (X, Y) 为二维随机向量, 对任意实数 x 和 y,

$$F(x, y) = P(X \leqslant x, Y \leqslant y)$$

称为二维随机向量 (X, Y) 的 **分布函数**, 或随机变量 X 和 Y 的 **联合分布函数** (joint cumulative probability distribution function).

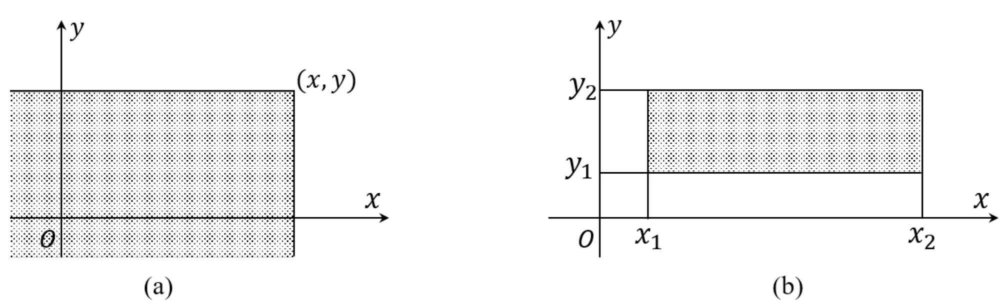

图 5.1　随机向量 (X, Y) 的分布函数 $F(x, y)$ 和概率 $P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2)$

若将 (X, Y) 看作平面上随机点的坐标, 则分布函数 $F(x, y)$ 的值表示随机向量 (X, Y) 落入以 (x, y) 为顶点的左下方无穷区域的概率, 如图 5.1(a) 所示. 随机向量 (X, Y) 落入矩形区域 $\{(x, y)\colon x_1 < x \leqslant x_2, y_1 < y \leqslant y_2\}$ 的概率如图 5.1(b) 所示, 即

$$P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1).$$

二维随机向量 (X,Y) 的分布函数 $F(x,y)$ 具有以下性质:

1) 分布函数 $F(x,y)$ 对每个变量都是单调不减的, 即对任意固定的实数 y, 当 $x_1>x_2$ 时有 $F(x_1,y)\geqslant F(x_2,y)$; 对任意固定的实数 x, 当 $y_1>y_2$ 时有 $F(x,y_1)\geqslant F(x,y_2)$.

2) 对任意实数 x 和 y, 分布函数 $F(x,y)\in[0,1]$, 而且

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

3) 分布函数 $F(x,y)$ 关于每个变量右连续, 即

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

4) 对任意实数 $x_1<x_2$ 和 $y_1<y_2$ 有

$$F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\geqslant 0.$$

任何的分布函数 $F(x,y)$ 都满足上述四条性质, 前三条性质与一维随机变量类似, 第四条性质根据图 5.1(b) 直接可证. 反之, 任何满足上面四条性质的二元函数 $F(x,y)$ 都可看成某二维随机向量的分布函数.

值得说明的是, 当二元函数 $F(x,y)$ 仅仅满足前面的三条性质时, 并不一定能成为某二维随机向量的分布函数, 例如

$$F(x,y)=\begin{cases}1 & x+y\geqslant 0,\\ 0 & x+y<0.\end{cases}$$

很容易验证 $F(x,y)$ 仅仅满足前面的三条性质, 但因为

$$F(1,1)-F(1,-1)-F(-1,1)+F(-1,-1)=-1,$$

显然不满足第四条性质, 因此不构成一个分布函数.

根据随机向量 (X,Y) 的联合分布函数, 可以研究每个随机变量的统计特征, 即将 X 和 Y 看做单独的随机变量, 通过联合分布函数来研究随机变量 X 和 Y 的分布函数 $F_X(x)$ 和 $F_Y(y)$.

定义 5.3 设二维随机向量 (X,Y) 的联合分布函数为 $F(x,y)$,

$$F_X(x)=P(X\leqslant x)=P(X\leqslant x,y<+\infty)=F(x,+\infty)=\lim_{y\to+\infty}F(x,y)$$

称为随机变量 X 的 **边缘分布函数** (marginal distribution function). 类似可定义随机变量 Y 的边缘分布函数 $F_Y(y)=P(Y\leqslant y)=F(+\infty,y)=\lim_{x\to+\infty}F(x,y)$.

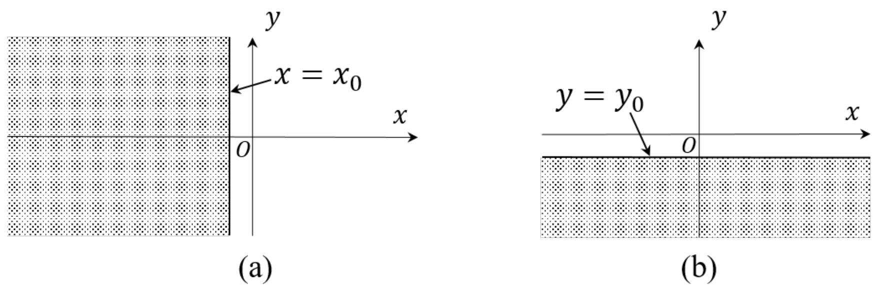

图 5.2 分布函数第四条性质的反例和边缘分布

边缘分布函数 $F_X(x_0)$ 和 $F_Y(y_0)$ 的值分别表示随机向量 (X,Y) 落入图 5.2(a) 和 5.2(b) 中阴影部分的概率. 下面来看一个例子.

例 5.1 设二维随机向量 (X,Y) 的分布函数为

$$F(x,y) = A\,(B+\arctan(x/2))\,(C+\arctan(y/3)) \qquad x,y\in(-\infty,+\infty),$$

求随机变量 X 与 Y 的边缘分布函数, 以及概率 $P(Y>3)$.

解 对任意 $x\in(-\infty,+\infty)$ 和 $y\in(-\infty,+\infty)$, 根据分布函数的性质有

$$\begin{aligned}
1 &= F(+\infty,+\infty) = A(B+\pi/2)(C+\pi/2),\\
0 &= F(x,-\infty) = A(B+\arctan(x/2))(C-\pi/2),\\
0 &= F(-\infty,y) = A(B-\pi/2)(C+\arctan(y/3)).
\end{aligned}$$

求解上述方程可得

$$A = 1/\pi^2, \quad B=\pi/2, \quad C=\pi/2.$$

从而得到 $F(x,y) = (\pi/2+\arctan x/2)(\pi/2+\arctan y/3)/\pi^2$, 进一步得到

$$F_X(x) = \lim_{y\to\infty}\frac{1}{\pi^2}\left(\frac{\pi}{2}+\arctan\frac{x}{2}\right)\left(\frac{\pi}{2}+\arctan\frac{y}{3}\right) = \frac{1}{2}+\frac{1}{\pi}\arctan\frac{x}{2},$$

同理可得

$$F_Y(y) = \frac{1}{2}+\frac{1}{\pi}\arctan\frac{y}{3}.$$

最后得到

$$P(Y>3) = 1-P(Y\leqslant 3) = 1-F_Y(3) = 1-\frac{1}{2}-\frac{1}{\pi}\arctan 1 = \frac{1}{4}.$$

表 5.1 二维随机向量的概率分布表

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_{i\cdot}=\sum_j p_{ij}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$
$p_{\cdot j}=\sum_i p_{ij}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

5.2 多维离散型随机向量

定义 5.4 若二维随机向量 (X,Y) 的取值是有限个或无限可列的, 则称 (X,Y) 为 **二维离散型随机向量**. 设离散型随机向量 (X,Y) 所有可能的取值为 $(x_i,y_j)\quad(i,j=1,2,\cdots)$,

$$p_{ij}=P(X=x_i,Y=y_j)$$

称为二维随机向量 (X,Y) 的 **联合分布列**, 简称 **分布列**.

二维随机向量分布列满足 $p_{ij}\geqslant 0$ 和 $\sum_{i,j}p_{ij}=1$.

通过随机向量 (X,Y) 的联合分布列 p_{ij}, 还可以研究每个随机变量的统计特征, 例如随机变量 X 的 **边缘分布列** 为

$$P(X=x_i)=\sum_{j=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{j=1}^{\infty}p_{ij}=p_{i\cdot},$$

以及随机变量 Y 的 **边缘分布列** 为

$$P(Y=y_j)=\sum_{i=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{i=1}^{\infty}p_{ij}=p_{\cdot j}.$$

二维随机向量的联合分布列和边缘分布列如表 5.1 所示.

根据联合分布列可以得到联合分布函数

$$F(x,y)=\sum_{x_i\leqslant x,y_j\leqslant y}p_{ij},$$

和边缘分布函数分别为

$$F_X(x)=\sum_{x_i\leqslant x}p_{i\cdot}=\sum_{x_i\leqslant x}\sum_{j=1}^{+\infty}p_{ij}\quad 和\quad F_Y(y)=\sum_{y_j\leqslant y}p_{\cdot j}=\sum_{y_j\leqslant y}\sum_{i=1}^{+\infty}p_{ij}.$$

多项分布

多项分布是一种常用的多维离散分布, 本质上是二项分布的推广, 常用于机器学习的多分类问题. 设试验 E 有 m 种可能的结果 $A_1, A_2, \cdots, A_m$, 每种结果发生的概率 $p_i = P(A_i)$, 且满足 $p_1 + p_2 + \cdots + p_m = 1$.

将试验 E 独立重复地进行 n 次, 用 $X_1, X_2, \cdots, X_m$ 分别表示事件 $A_1, A_2, \cdots, A_m$ 发生的次数, 则每个随机变量 X_i 的取值为 $\{0, 1, 2, \cdots, n\}$ 且满足 $X_1 + X_2 + \cdots + X_m = n$, 则随机向量 $(X_1, X_2, \cdots, X_m)$ 服从多项分布, 其形式化定义如下:

定义 5.5 设 n 维随机向量 $(X_1, X_2, \cdots, X_m)$ 的分布列为

$$P(X_1 = k_1, X_2 = k_2, \cdots, X_m = k_m) = \binom{n}{k_1, k_2, \cdots, k_m} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m},$$

称随机向量 $(X_1, X_2, \cdots, X_m)$ 服从参数为 $n, p_1, p_2, \cdots, p_m$ 的 **多项分布** (multinomial distribution), 记为 $(X_1, X_2, \cdots, X_m) \sim M(n, p_1, p_2, \cdots, p_m)$. 这里 $k_1, \cdots, k_m$ 是非负的整数且满足 $\sum_{i=1}^m k_i = n$.

当 $n = 2$ 时多项分布退化为二项分布. 很容易验证 $P(X_1 = k_1, X_2 = k_2, \cdots, X_m = k_m) \geqslant 0$ 且

$$\begin{aligned}
&\sum_{k_i \geqslant 0, k_1 + k_2 + \cdots + k_m = n} P(X_1 = k_1, X_2 = k_2, \cdots, X_m = k_m) \\
= &\sum_{k_i \geqslant 0, k_1 + k_2 + \cdots + k_m = n} \binom{n}{k_1, k_2, \cdots, k_m} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m} = (p_1 + p_2 + \cdots + p_m)^n = 1.
\end{aligned}$$

性质 5.1 设多维随机向量 $(X_1, X_2, \cdots, X_m)$ 服从参数为 $n, p_1, p_2, \cdots, p_m$ 的多项分布, 则每个随机变量 X_i 的边缘分布为二项分布 $B(n, p_i)$.

证明 根据多项分布和边缘分布的定义有

$$\begin{aligned}
&P(X_i = k) \\
= &\sum_{k_j \geqslant 0, \sum_{j \neq i} k_j = n-k} P(X_i = k, X_1 = k_1, \cdots, X_{i-1} = k_{i-1}, X_{i+1} = k_{i+1}, \cdots, X_m = k_m) \\
= &\frac{n!}{k!(n-k)!} p_i^k \sum_{k_j \geqslant 0, \sum_{j \neq i} k_j = n-k} \frac{(n-k)!}{\prod_{j \neq i} k_j!} \prod_{j=1, j \neq i}^m p_j^{k_j} \\
= &\frac{n!}{k!(n-k)!} p_i^k \sum_{k_j \geqslant 0, \sum_{j \neq i} k_j = n-k} \binom{n-k}{k_1, \cdots, k_{i-1}, k_{i+1}, \cdots, k_m} \prod_{j=1, j \neq i}^m p_j^{k_j} \\
= &\frac{n!}{k!(n-k)!} p_i^k (p_1 + \cdots + p_{i-1} + p_{i-1} + \cdots + p_m)^{n-k} = \binom{n}{k} p_i^k (1 - p_i)^{n-k}.
\end{aligned}$$

这里利用了 $p_1 + p_2 + \cdots + p_m = 1$, 由此完成证明.

也可根据 X_i 的实际含义, 在 n 重伯努利实验中考虑事件 A_i 发生或不发生, 则有 $X_i \sim B(n, p_i)$.