第 5 章　多维随机向量

前面讨论了一维随机变量的概率分布, 然而很多在实际问题中, 随机现象可能需要两种或两种以上的随机因素来描述, 仅仅用一个随机变量是不够的, 需要多个随机变量. 例如, 为了考察某地区儿童的身体素质时, 可以同时考虑他们的身高、体重、肺活量、视力等, 此时至少需要四个随机变量来描述. 这些随机变量之间可能存在某些关联, 因此分别对每个随机变量单独进行研究是不够的, 需要将其看作一个整体来研究, 即多维随机向量.

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

类似于一维随机变量, 我们用分布函数来研究二维随机向量的概率特性.

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**, 或称随机变量 X 和 Y 的 **联合分布函数** (joint cumulative probability distribution function).

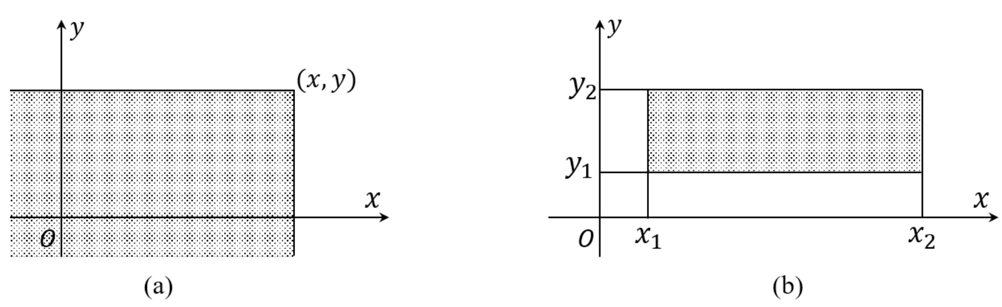

图 5.1　随机向量 (X,Y) 的分布函数 $F(x,y)$ 和概率 $P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2)$

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. 再根据图 5.1(b) 可知, 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

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

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

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

3) 分布函数 $F(x,y)$ 关于每个变量右连续, 即

$$F(x,y)=F(x+0,y)\quad \text{和}\quad F(x,y)=F(x,y+0)\ .$$

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

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

如图 5.2(a) 所示, 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 $F(x,y)$, 还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 $F(x,y)$ 来研究随机变量 X 和 Y 的分布函数 $F_X(x)$ 和 $F_Y(y)$, 即边缘分布函数.

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

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

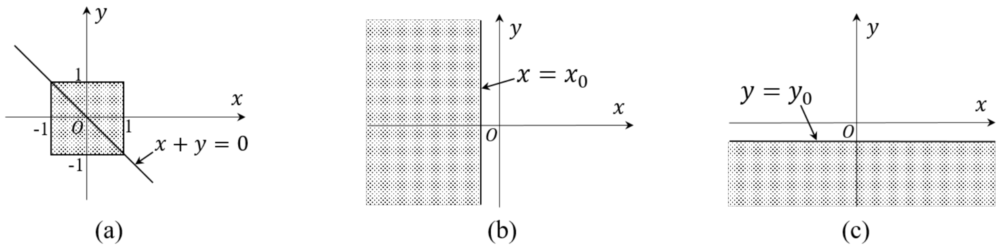

图 5.2 分布函数第四条性质的反例和边缘分布

为 (X,Y) 关于随机变量 X 的 **边缘分布函数** (marginal distribution function). 类似地定义 (X,Y) 关于随机变量 Y 的边缘分布函数

$$F_Y(y)=P(Y\leqslant y)=P(Y\leqslant y,x<+\infty)=F(+\infty,y)=\lim_{x\to+\infty}F(x,y)\ .$$

边缘分布函数 $F_X(x_0)$ 和 $F_Y(y_0)$ 所表示的概率分别如图 5.2(b) 和 5.2(c). 下面看一个例子.

例 5.1 设二维随机向量 (X,Y) 的分布函数为

$$F(x,y)=A(B+\arctan\frac{x}{2})(C+\arctan\frac{y}{3})\qquad (x,y\in\mathbb{R})\ .$$

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

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

求解上述方程可得

$$C=\frac{\pi}{2},\quad B=\frac{\pi}{2},\quad A=\frac{1}{\pi^2}\ .$$

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

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

同理可得

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

最后得到

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

5.2 二维离散型随机向量

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 $(x_i,y_j)\quad(i,j=1,2,\cdots)$, 则称

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

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具有下列性质:

$$p_{ij}\geqslant 0 \qquad \text{和} \qquad \sum_{i,j}p_{ij}=1.$$

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可通过表 5.1 来进行表示.

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

根据二维随机变量 (X,Y) 的联合分布列 p_{ij}, 可以得到它们的联合分布函数

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

和边缘分布函数

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

例 5.2 假设某地区有 15% 的家庭没小孩, 20% 的家庭有一个小孩, 35% 的家庭有两个小孩, 30% 的家庭有三个小孩, 且假设每个小孩为男孩或女孩是相互独立且等可能的. 随机选择一个家庭, 用随机变量 X, Y 分别表示该家庭中男孩和女孩的个数, 求 $P(X \geqslant 1)$, $P(Y \leqslant 2)$ 和 $P(X \leqslant Y)$.

解 根据题意有 X, Y 的所有可能取值为 $\{0, 1, 2, 3\}$, 进一步有联合分布列

$$\begin{aligned} P(X=i, Y=j) &= P(\text{选择的家庭有 } i+j \text{ 个小孩, 其中 } i \text{ 个男孩和 } j \text{ 个女孩}) \\ &= P(\text{选择的家庭有 } i+j \text{ 个小孩}) P(i \text{ 个男孩和 } j \text{ 个女孩} | \text{选择的家庭有 } i+j \text{ 个小孩}) \\ &= \binom{i+j}{i} \frac{1}{2^{i+j}} P(\text{选择的家庭有 } i+j \text{ 个小孩}) \ , \end{aligned}$$

由此可得联合分布列和边缘分布列为

$X \backslash Y$	0	1	2	3	$p_{i\cdot}$
0	0.1500	0.1000	0.0875	0.0375	0.3750
1	0.1000	0.175	0.1125	0	0.3875
2	0.0875	0.1125	0	0	0.2000
3	0.0375	0	0	0	0.0375
$p_{\cdot j}$	0.3750	0.3875	0.2000	0.0375	1

最后得到

$$P(X \geqslant 1) = 0.625, \quad P(Y \leqslant 2) = 0.9625, \quad P(X \leqslant Y) = 0.6625 \ .$$

最后介绍一种常用的多维离散分布: 多项分布, 它本质上是二项分布的推广, 可用于机器学习中多分类问题. 假设试验 E 有 n 种可能的结果 $A_1, A_2, \cdots, A_n$, 每种结果发生的概率 $p_i = P(A_i)$, 则有 $p_1 + p_2 + \cdots + p_n = 1$.

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

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

$$P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) = \binom{m}{k_1, k_2, \cdots, k_n} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} \ ,$$

其中 $k_1, k_2, \cdots, k_n$ 是非负的整数且满足 $k_1 + k_2 + \cdots + k_n = m$, 则称随机向量 $(X_1, X_2, \cdots, X_n)$ 服从参数为 $m, p_1, p_2, \cdots, p_n$ 的 **多项分布** (multinomial distribution), 记为 $(X_1, X_2, \cdots, X_n) \sim M(m, p_1, p_2, \cdots, p_n)$.

很容易验证 $P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) \geqslant 0$ 以及

$$\begin{aligned}
&\sum_{k_i \geqslant 0, k_1+k_2+\cdots+k_n=m} P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) \\
&= \sum_{k_i \geqslant 0, k_1+k_2+\cdots+k_n=m} \binom{m}{k_1, k_2, \cdots, k_n} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} = (p_1 + p_2 + \cdots + p_n)^m = 1 \ .
\end{aligned}$$

当 $n = 2$ 时多项分布简化为二项分布.

引理 5.1　若多维随机向量 $(X_1, X_2, \cdots, X_n) \sim M(m, p_1, p_2, \cdots, p_n)$, 则每个随机变量 X_i 的边缘分布是二项分布 $B(m, p_i)$.

根据 X_i 的实际含义, 考虑事件 A_i 发生或不发生的伯努利试验, 则有 $X_i \sim (m, p_i)$. 另一种方法是通过多项分布的定义直接计算, 我们将其作为一个作业题.

5.3 二维连续型随机变量

定义 5.6　设二维随机向量的分布函数为 $F(x, y)$, 如果存在二元非负可积函数 $f(x, y)$ 使得对任意实数对 (x, y) 有

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv,$$

则称 (X, Y) 为 **二维连续型随机向量**, 称 $f(x, y)$ 为二维随机向量 (X, Y) 的 **密度函数**, 或称随机变量 X 和 Y 的 **联合概率密度函数**.

根据定义可知概率密度函数 $f(x, y)$ 满足如下性质:

1) 非负性: 对任意实数 x 和 y 有 $f(x, y) \geqslant 0$.

2) 规范性: $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$.
   任何满足上述两条性质的二元函数 $f(x, y)$ 可以成为某随机向量 (X, Y) 的联合密度函数.

3) 若 G 为平面上的一个区域, 则点 (X, Y) 落入 G 的概率为

$$P((X, Y) \in G) = \iint_{(x,y) \in G} f(x, y) dx dy \ ,$$

   在几何上可以看作是以 G 为底面, $z = f(x, y)$ 为顶面的柱体体积.