

5 多维随机变量及其分布

前面章节研究一个随机变量, 即一个随机事件由一种因素决定, 然而实际问题中很多随机现象往往由两个或多个随机因素造成的, 需用多个随机变量描述. 例如: 导弹攻击点的坐标 (经度、纬度); 学生的高考成绩 (语文、数学、英语等).

定义5.1. 设 $X = X(\omega)$ 和 $Y = Y(\omega)$ 为定义在样本空间 Ω 上的随机变量, 由它们构成的向量 (X, Y) 称为二维随机变量.

二维随机向量又称为二维随机变量; 将 (X, Y) 看作一个整体, 不能分开看待; 几何上, (X, Y) 可看作平面上的随机点.

5.1 二维随机变量的分布函数

首先研究二维随机变量的分布函数:

定义5.2. 设 (X, Y) 为二维随机变量, 对任意实数 $x \in (-\infty, +\infty)$ 和 $y \in (-\infty, +\infty)$,

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

称为二维随机变量 (X, Y) 的分布函数, 或称为随机变量 X 和 Y 的联合分布函数.

二维随机变量分布函数的几何意义: $F(x, y)$ 表示点 (X, Y) 落入以 (x, y) 为右上定点的无穷矩形的概率.

对二维随机变量分布函数, 有如下性质:

性质5.1. i) 分布函数 $F(x, y)$ 对每个变量单调不减: 固定 y , 当 $x_1 > x_2$ 时有 $F(x_1, y) \geq F(x_2, y)$; 同理固定 x , 当 $y_1 > y_2$ 时有 $F(x, y_1) \geq F(x, y_2)$.

ii) 对任意 $x \in (-\infty, +\infty)$ 和 $y \in (-\infty, +\infty)$, 分布函数 $F(x, y) \in [0, 1]$, 且

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

iii) 分布函数 $F(x, y)$ 关于每个变量右连续.

根据分布函数可推导概率:

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

设随机变量 (X, Y) 的联合分布函数为 $F(x, y)$, 如果将随机变量 X 和 Y 分别看, 则其分别为随机变量. 下面根据随机变量的联合分布函数 $F(x, y)$ 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. 设二维随机变量 (X, Y) 的联合分布函数为 $F(x, y)$, 称

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

为随机变量 X 的边缘分布函数. 同理定义随机变量 Y 的边缘分布函数为:

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

例5.1. 设二维随机变量 (X, Y) 的联合分布函数为 $F(x, y) = A(B + \arctan \frac{x}{2})(C + \arctan \frac{y}{3})(x, y \in \mathbb{R})$. 求随机变量 X 与 Y 的边缘分布函数和概率 $P(Y > 3)$.

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

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

同理可得

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

最后得到

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

□

前面讲过独立的随机事件 A 和 B 有 $P(AB) = P(A)P(B)$, 下面介绍随机变量的独立性:

定义5.4. 设 X, Y 为二维随机变量, 对任意 $x \in (-\infty, +\infty)$ 和 $y \in (-\infty, +\infty)$, 若事件 $X \leq x$ 和 $Y \leq y$ 相互独立, 即

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) \quad \Leftrightarrow \quad F(x, y) = F_X(x)F_Y(y),$$

则称随机变量 X 与 Y 相互独立.

定理5.1. 设随机变量 X 与 Y 相互独立, 则 $f(X)$ 与 $g(Y)$ 也相互独立, 其中 $f(x)$ 和 $g(y)$ 是连续或分段连续函数.

例如: 若随机变量 X 与 Y 相互独立, 则 X^2 与 Y^3 相互独立, $\sin X$ 与 $\cos Y$ 相互独立.

5.2 二维离散型随机变量

定义5.5. 若二维随机变量 (X, Y) 的取值是有限个或无限可列的, 称 (X, Y) 为二维离散型随机变量.

设离散型随机变量 (X, Y) 的取值分别为 $(x_i, y_j), i = 1, 2, \dots, j = 1, 2, \dots$, 则称

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

为 (X, Y) 的联合分布列. 二维随机变量的联合分布列可表示为

$X \backslash Y$	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	\ddots

根据分布列的性质可知 $p_{ij} \geq 0$ 和 $\sum_{i,j} p_{ij} = 1$. 根据二维随机变量 (X, Y) 的联合分布列 p_{ij} , 得到随机变量 X 的边缘分布列

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

同理可得随机变量 Y 的边缘分布列

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

将二维随机变量的联合分布列和边缘分布表示在同一个表格中有

$X \backslash Y$	y_1	y_2	\cdots	y_j	\cdots	$p_{i\cdot}$
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
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	\cdots	$p_{\cdot j}$	\cdots	1

下面来看一个例子:

例5.2. 有三个数 1, 2, 3, 随机变量 X 表示从这三个数中随机地抽取一个数, 随机变量 Y 表示从 1 到 X 中随机抽取一个数. 求 (X, Y) 的联合分布列和边缘分布列.

解. 由题意可知随机变量 X 和 Y 的取值为 1, 2, 3: 当 $X = 1$ 时有 $Y = 1$; 当 $X = 2$ 时有 Y 等可能取 1, 2; 当 $X = 3$ 时 Y 等可能取 1, 2, 3. 从而得到

$X \backslash Y$	1	2	3	$p_{i\cdot}$
1	1/3	0	0	1/3
2	1/6	1/6	0	1/3
3	1/9	1/9	1/9	1/3
$p_{\cdot j}$	11/18	5/18	1/9	1

□

定义5.6. 对离散型随机变量 (X, Y) , 若对所有 (x_i, y_j) 有

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j), \quad \text{即} \quad p_{ij} = p_{i\cdot}p_{\cdot j}$$

称离散随机变量 X 与 Y 相互独立.

定理5.2. 对二维离散型随机变量 (X, Y) , 定义 5.4 与上述定义等价, 即对所有 (x_i, y_j) 有

$$F(x_i, y_j) = F_X(x_i)F_Y(y_j) \iff p_{i,j} = p_{i\cdot}p_{\cdot j}.$$

Proof. 首先证明必要性, 根据定义 5.4 分布函数的独立性有

$$\begin{aligned} p_{i,j} &= F(x_i, y_j) - F(x_{i-1}, y_j) - F(x_i, y_{j-1}) + F(x_{i-1}, y_{j-1}) \\ &= F_X(x_i)F_Y(y_j) - F_X(x_{i-1})F_Y(y_j) - F_X(x_i)F_Y(y_{j-1}) + F_X(x_{i-1})F_Y(y_{j-1}) \\ &= (F_X(x_i) - F_X(x_{i-1}))F_Y(y_j) - (F_X(x_i) - F_X(x_{i-1}))F_Y(y_{j-1}) \\ &= p_{i\cdot}F_Y(y_j) - p_{i\cdot}F_Y(y_{j-1}) = p_{i\cdot}p_{\cdot j}. \end{aligned}$$

其次证明充分性, 根据定义 5.6 有

$$F(x_i, y_j) = \sum_{l \leq i} \sum_{k \leq j} p_{lk} = \sum_{l \leq i} \sum_{k \leq j} p_{l\cdot}p_{\cdot k} = \sum_{l \leq i} p_{l\cdot} \sum_{k \leq j} p_{\cdot k} = F_X(x_i)F_Y(y_j).$$

□

定理5.3. 设离散随机变量 X 和 Y 独立, 则对任意集合 $A \in \mathbb{R}, B \in \mathbb{R}$, 有事件 $X \in A$ 和 $Y \in B$ 独立.

Proof. 对离散型随机变量, 不妨假设 $A = \{x_1, x_2, \dots, x_k\}$ 和 $B = \{y_1, y_2, \dots, y_l\}$, 则有

$$P(X \in A, Y \in B) = \sum_{i=1}^k \sum_{j=1}^l p_{ij} = \sum_{i=1}^k \sum_{j=1}^l p_{i\cdot}p_{\cdot j} = \sum_{i=1}^k p_{i\cdot} \sum_{j=1}^l p_{\cdot j} = P(X \in A)P(Y \in B).$$

□

例5.3. 设离散型 X, Y 独立, 求解 (X, Y) 的联合分布律为

$X \backslash Y$	y_1	y_2	y_3	$p_{i\cdot}$
x_1		1/8		
x_2	1/8			
$p_{\cdot j}$	1/6			

求解可得

$X \backslash Y$	y_1	y_2	y_3	$p_{i\cdot}$
x_1	1/24	1/8	1/12	1/4
x_2	1/8	3/8	1/4	3/4
$p_{\cdot j}$	1/6	1/2	1/3	1

例5.4. 将两个球 A, B 放入编号为 1, 2, 3 的三个盒子中, 用随机变量 X 放入 1 号盒的球数, 用随机变量 Y 表示放入 2 号盒的球数, 判断 X 和 Y 是否独立.

解. 由题意可知

$X \backslash Y$	0	1	2	$p_{i\cdot}$
0	1/9	2/9	1/9	4/9
1	2/9	2/9	0	4/9
2	1/9	0	0	1/9
$p_{\cdot j}$	4/9	4/9	1/9	

由此可知 $P(X=2, Y=2) \neq P(X=2)P(Y=2)$, 所以 X 和 Y 不独立. \square

5.3 二维连续型随机变量

定义5.7. 设二维随机变量的分布函数为 $F(x, y)$, 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) 为二维连续型随机变量, 称 $f(x, y)$ 称为二维随机变量 (X, Y) 的概率密度, 或称为随机变量 X 和 Y 的联合概率密度.

根据概率密度的定义可知概率密度函数 $f(x, y)$ 满足如下性质:

- 1) $p(x, y) \geq 0$.
- 2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$.
- 3) 若 $f(x, y)$ 在 (x, y) 连续, 则 $f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$.
- 4) 若 G 为平面上的一个区域, 则点 (X, Y) 落入 G 的概率为

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

例5.5. 设二维随机变量 (X, Y) 的概率密度为

$$f(x, y) = \begin{cases} ce^{-(3x+4y)} & x > 0, y > 0 \\ 0 & \text{其它} \end{cases}$$

求 $P(0 < X < 1, 0 < Y < 2)$.

解. 根据概率密度的性质可知

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ce^{-(3x+4y)} dx dy = \frac{c}{12},$$