第七周 多维随机变量，独立性

7.1. 多维随机变量

多维随机变量

在同一个随机试验中，往往同时涉及多个随机变量，例如考察某地区中学生的身体素质情况，随机地选取一名学生，观察学生的身高 X，体重 Y 和肺活量 Z 等指标。随机变量 X，Y，Z 来自同一样本空间，它们的取值可能相互影响。像这样同时考虑的多个随机变量，称为多元随机变量。

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

对任意 n 个实数 $x_1, x_2, \cdots, x_n$，事件 $\{X_1 \le x_1\}, \{X_2 \le x_2\}, \cdots, \{X_n \le x_n\}$ 同时发生的概率 $F(x_1, x_2, \cdots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n)$ 称为 n 维随机变量 X 的联合分布函数。

下面我们主要讨论二维随机变量的性质，大多数二维随机变量的结果都很容易推广到 n 维的情况。

****

定理 二维联合分布函数 $F(x, y)$ 具有如下性质

(1) 单调性 $F(x, y)$ 分别对 x 和 y 是单调不减的，

当 $x_1 < x_2$ 时，$F(x_1, y) \le F(x_2, y)$；当 $y_1 < y_2$ 时，$F(x, y_1) \le F(x, y_2)$；

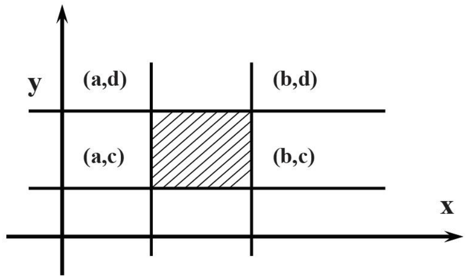
(2) 有界性 对任意 x 和 y，有 $0 \le F(x, y) \le 1$，且 $F(-\infty, y) = \lim_{x \to -\infty} F(x, y) = 0$

$$F(x, -\infty) = \lim_{y \to -\infty} F(x, y) = 0, \quad F(+\infty, +\infty) = \lim_{x, y \to +\infty} F(x, y) = 1;$$

(3) 右连续性 $F(x+0, y) = F(x, y)$，$F(x, y+0) = F(x, y)$

(4) 非负性 对任意 $a < b$，$c < d$，有 $P(a < X < b, c < Y < d)$

$$= F(b,d) - F(a,d) - F(b,c) + F(a,c) \geq 0$$



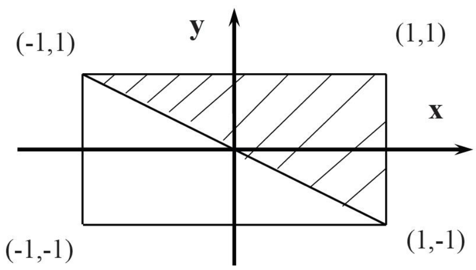
这3条性质，一元随机变量也具备。仅仅满足前3条性质，并不足以表明二元函数是某个二维随机变量的分布函数。下面看一个反例。

****

例 7.1.1　二元函数　$G(x,y) = \begin{cases} 0 & x+y<0 \\ 1 & x+y\geq 0 \end{cases}$

$G(1,1) - G(1,-1) - G(-1,1) + G(-1,-1) = -1 < 0$，

故 $G(x,y)$ 不能作为二元随机变量的分布函数。



****

二维离散型随机变量

若 (X,Y) 只取至多可列个数对，(x_i, y_j)，则称 (X,Y) 为二维离散随机变量，

$p_{ij} = P(X = x_i, Y = y_j)$ 称为 (X,Y) 的联合分布列。

$X \backslash Y$	y_1	y_2	$\cdots$	y_n	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\cdots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nn}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

其中，$p_{ij} \geq 0$，$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1$。

****

边缘分布

(X,Y)为二元离散型随机变量，其中X和Y各自的分布称为边缘分布

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ P(X=x_1) & P(X=x_2) & \cdots & P(X=x_2) & \cdots \end{pmatrix}$$ 称为X的边缘分布列

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

$$\begin{pmatrix} y_1 & y_2 & \cdots & y_n & \cdots \\ P(Y=y_1) & P(Y=y_2) & \cdots & P(Y=y_2) & \cdots \end{pmatrix}$$ 称为Y的边缘分布列

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

****

例 7.1.2 从$1,2,3,4$中等可能地随机取一个数记为X，再从$1,2,\cdots,X$中等可能地随机取一个数记为Y。（1）写出(X,Y)的联合分布列，并计算$P(X=Y)$；（2）写出(X,Y)的边缘分布列。

解：（1）

$X\setminus Y$	1	2	3	4
1	1/4	0	0	0
2	1/8	1/8	0	0
3	1/12	1/12	1/12	0
4	1/16	1/16	1/16	1/16

$$P(X=i,Y=j)=\begin{cases}\dfrac{1}{4i}, & 1\le j\le i\le 4\\ 0, & \text{其他}\end{cases}$$

$$P(X=Y)=P(X=1,Y=1)+P(X=2,Y=2)+P(X=3,Y=3)+P(X=4,Y=4)$$

$$=\frac{1}{4}+\frac{1}{8}+\frac{1}{12}+\frac{1}{16}=\frac{25}{48}。$$

（2） 边缘分布列 $P(X=i)=\dfrac{1}{4}$，$i=1,2,3,4$

$$P(Y=1)=\sum_{i=1}^{4}p_{ij}=\frac{1}{4}+\frac{1}{8}+\frac{1}{12}+\frac{1}{16}=\frac{25}{48},$$

同理可得，$P(Y=2)=\dfrac{13}{48}$，$P(Y=3)=\dfrac{7}{48}$，$P(Y=4)=\dfrac{1}{16}$。

$X\setminus Y$	1	2	3	4	$P(X=i)$
1	1/4	0	0	0	1/4
2	1/8	1/8	0	0	1/4
3	1/12	1/12	1/12	0	1/4
4	1/16	1/16	1/16	1/16	1/4
$P(Y=j)$	25/48	13/48	7/48	1/16	1

****

例 7.1.3. 设二元随机变量(X,Y)的联合分布列为，$P(X=1,Y=1)=\dfrac{4}{9}$，

$P(X=1,Y=2)=\dfrac{2}{9}$，$P(X=2,Y=1)=\dfrac{2}{9}$，$P(X=2,Y=2)=\dfrac{1}{9}$。设 $U=\max(X,Y)$，

$V=\min(X,Y)$。

（1） 求(U,V)的联合概率分布； （2） 求U,V的期望和方差。

解：（1）

$X \backslash Y$	1	2
1	$\frac{4}{9}$	$\frac{2}{9}$
2	$\frac{2}{9}$	$\frac{1}{9}$

$\Rightarrow$

$U \backslash V$	1	2
1	$\frac{4}{9}$	0
2	$\frac{4}{9}$	$\frac{1}{9}$

（2）

$U \backslash V$	1	2
1	$\frac{4}{9}$	0
2	$\frac{4}{9}$	$\frac{1}{9}$

得到U,V的边缘分布列分别为 $U \sim \begin{pmatrix} 1 & 2 \\ \frac{4}{9} & \frac{5}{9} \end{pmatrix}$，$V \sim \begin{pmatrix} 1 & 2 \\ \frac{8}{9} & \frac{1}{9} \end{pmatrix}$

$E(U)=1\cdot\frac{4}{9}+2\cdot\frac{5}{9}=\frac{14}{9}$， $E(U^2)=1\cdot\frac{4}{9}+2^2\cdot\frac{5}{9}=\frac{24}{9}$， $Var(U)=E(U^2)-E(U)^2=\frac{20}{81}$

$E(V)=1\cdot\frac{8}{9}+2\cdot\frac{1}{9}=\frac{10}{9}$， $E(V^2)=1\cdot\frac{8}{9}+2^2\cdot\frac{1}{9}=\frac{12}{9}$， $Var(V)=E(V^2)-E(V)^2=\frac{8}{81}$

****

二维连续型随机变量

存在二元非负函数$f(x,y)$，使得二维随机变量(X,Y)的分布函数$F(x,y)$可表示为$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$，则$(X,Y)$为二维连续型随机变量，$f(x,y)$为$(X,Y)$的联合密度函数，$f(x,y)=\frac{\partial^2}{\partial x\partial y}F(x,y)$。（稍停顿）

边缘密度函数 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$；

边缘分布函数 $F_X(x) = F(x,+\infty) = \int_{-\infty}^{x}\left(\int_{-\infty}^{+\infty} f(u,v)dv\right)du = \int_{-\infty}^{x} f_X(u)du$ 。

****