Ch 5 多维随机向量

回顾前一次课

已知随机变量X概率密度为$f_X(x)$，求$Y=g(X)$的概率密度$f_Y(y)$

- 求解Y=g(X)的分布函数$F_Y(y)=P(Y\le y)=\int_{g(x)\le y} f_X(x)dx$
- 求解密度函数$f_Y(y)=F_Y'(y)$

随机变量X密度函数为$f_X(x)$, 函数$y=g(x)$处处可导且严格单调, 反函数$x=g^{-1}(y)=h(y)$, 则随机变量$Y=g(X)$ 的概率密度为

$$f_Y(y)=\begin{cases} f_X\big(h(y)\big)|h'(y)| & y\in(\alpha,\beta) \\ 0 & \text{其它} \end{cases}$$

$\alpha=\min\{g(-\infty),g(+\infty)\}$，$\beta=\max\{g(-\infty),g(+\infty)\}$

常用分布的随机数生成: 1) U(0,1) 均匀分布随机数
2) 离散分布随机数 3) 连续分布随机数 4) 正太分布随机数

多维随机向量

在很多实际问题中，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

为考察某个地区儿童的身体素质，可考虑他们的身高、体重、肺活量、视力等，此时至少需要四个随机变量来进行描述.

这些随机变量之间可能存在某些关联，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，即多维随机向量

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$\Omega$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**

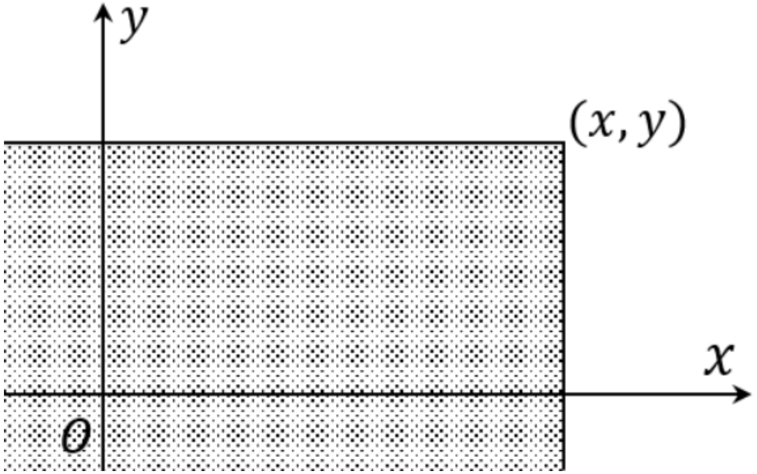
二维随机向量的分布函数

设(X,Y)为二维随机向量, 对任意实数x和y, 称

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

为**二维随机向量(X,Y)的分布函数**，或称为**随机变量X和Y的联合分布函数** (joint cumulative probability distribution function)

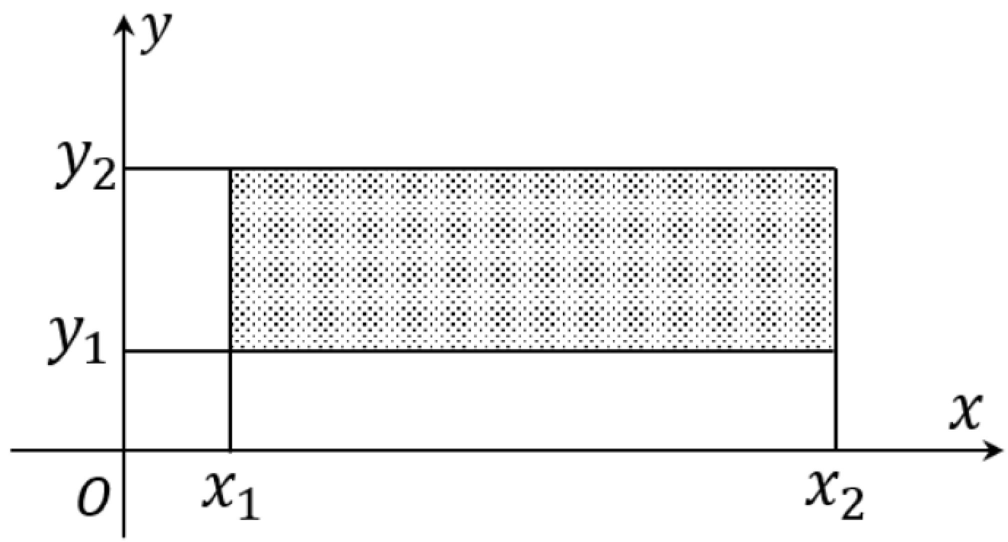
若将(X,Y)看作平面上随机点的坐标, 则分布函数$F(x,y)$在点(x,y)的值是随机向量(X,Y)落入以(x,y)为顶点的左下方无穷区域的概率



二维随机向量的矩形概率

(X,Y)落入矩形区域$\{(x,y): x\in(x_1,x_2], y\in(y_1,y_2]\}$的概率

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



二维随机向量分布函数性质

- 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$F(x_1,y) \geq F(x_2,y)$
  - 固定x, 当$y_1 > y_2$时有$F(x,y_1) \geq F(x,y_2)$
- 对任意实数x和y, 分布函数$F(x,y) \in [0,1]$, 且

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

$$F(-\infty,y) = F(x,-\infty) = F(-\infty,-\infty) = 0$$

- 分布函数$F(x,y)$关于每个变量右连续

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

- 对任意实数$x_1 < x_2$和$y_1 < y_2$有

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

二维随机变量分布函数性质

任何的分布函数$F(x,y)$ 满足前面四条性质. 任何满足前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，不一定能成为某二维随机向量的分布函数

例 设随机向量(X,Y)的分布函数为

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

很容易验证$F(x,y)$ 仅仅满足前面的三条性质, 但因为

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

因此不满足第四条性质, 不构成一个分布函数

边缘分布函数

根据联合分布函数$F(x,y)$，还可研究每个随机变量的统计特征，将X和Y看做单独的随机变量，根据联合分布函数$F(x,y)$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

设二维随机变量(X,Y)的联合分布函数为$F(x,y)$，称

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

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:

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

例题

设二维随机变量(X,Y)的联合分布函数为

$$F(x,y)=A\left(B+\arctan\frac{x}{2}\right)(C+\arctan\frac{y}{3})$$

求随机变量X与Y的边缘分布函数, 以及概率$P(Y>3)$

二维离散型随机向量

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**

设离散型随机向量(X,Y)的取值分别为$\left(x_i,y_j\right)$, $i,j=1,2,\cdots$, 则称

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

为**(X,Y)的联合分布列,** 简称**分布列**

性质：$p_{ij}\geq 0$和$\sum_{i,j}p_{ij}=1$

$X \backslash Y$	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}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$

边缘分布列

根据二维随机变量(X,Y)的联合分布列$\{p_{ij}\}$,还可研究每个随机变量的统计特征,将X和Y看做单独的随机变量

随机变量X的边缘分布列

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

随机变量Y的边缘分布列

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

边缘分布列

$X \backslash Y$	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}, 可得

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

$$F_X(x) = \sum_{x_i \le x} p_{i\cdot} = \sum_{x_i \le x} \sum_{j=1}^{+\infty} p_{ij} \qquad F_Y(y) = \sum_{y_j \le j} p_{\cdot j} = \sum_{y_j \le j} \sum_{i=1}^{+\infty} p_{ij}$$

例题

设某一个地区15%的家庭没有小孩, 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．设每个家庭中小孩为男孩和女孩是等可能的且相互独立．现从该地区随机任意选择一个家庭，用随机变量X, Y分别表示该家庭中男孩和女孩的个数, 求概率$P(X \geq 1)$, $P(Y \leq 2)$和$P(X \leq Y)$

多项分布

伯努利试验：试验E有2种可能的结果A或$\bar{A}$

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$p_1 + p_2 + \cdots + p_n = 1$成立. 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，每个随机变量$X_i$取值为$\{0,1,\cdots,m\}$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

多项分布

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$k_1, k_2, \cdots, k_n$是非负整数且满足$k_1 + k_2 + \cdots + k_n = m$，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$(X_1, X_2, \cdots, X_n) \sim M(m, p_1, p_2, \cdots, p_n)$

性质

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$X_i$ $(i \in [n])$的边缘分布为二项分布$B(m, p_i)$