第七周 多维随机变量，独立性

7.2. 常见多维随机变量举例

多项分布 $(X_1,X_2,\cdots,X_m)\sim M(n,p_1,\cdots,p_m)$

设随机试验E有m个基本事件$A_1,\cdots,A_m$，$P(A_i)=p_i>0$，$\sum_{i=1}^{m}p_i=1$。将E重复n次，以X_i表示n次试验中A_i出现的次数，则

$$P(X_1=k_1,\cdots,X_m=k_m)=\frac{n!}{k_1!k_2!\cdots k_m!}p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m}\text{，}(k_1+\cdots+k_m=n)\text{。}$$

$$C_n^{k_1}C_{n-k_1}^{k_2}C_{n-k_1-k_2}^{k_3}\cdots C_{k_m}^{k_m}=\frac{n!}{k_1!(n-k_1)!}\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\cdots\frac{k_m!}{k_m!}=\frac{n!}{k_1!k_2!\cdots k_m!}$$

$$(x_1+x_2+\cdots+x_m)^n=\sum_{k_1+\cdots+k_m=n}\frac{n!}{k_1!k_2!\cdots k_m!}x_1^{k_1}x_2^{k_2}\cdots x_m^{k_m}$$

例如，某产品分为一等品（A_1）、二等品（A_2）、三等品（A_3）、不合格品（A_4），各种档次的产品出现概率分别为p_1,p_2,p_3,p_4（$p_1+p_2+p_3+p_4=1$）。任取n件产品，各品级数目服从$M(n,p_1,p_2,p_3,p_4)$。

****

例 7.2.1 已知三项分布$(X,Y,Z)\sim M(n,p_1,p_2,p_3)$，验证其边缘分布为二项分布。

$$P(X=i)=\sum_{j+k=n-i\ (j,k\in Z^+)}P(X=i,Y=j,Z=k)$$

$$=\sum_{j=0}^{n-i}P(X=i,Y=j,Z=n-i-j)=\sum_{j=0}^{n-i}\frac{n!}{i!\cdot j!\cdot(n-i-j)!}p_1^{\,i}p_2^{\,j}p_3^{\,n-i-j}$$

$$=\frac{n!}{i!\cdot(n-i)!}p_1^{\,i}(1-p_1)^{n-i}\cdot\sum_{j=0}^{n-i}\frac{(n-i)!}{j!\cdot(n-i-j)!}\left(\frac{p_2}{1-p_1}\right)^j\left(\frac{p_3}{1-p_1}\right)^{n-i-j}$$

$$=C_n^ip_1^{\,i}(1-p_1)^{n-i}\cdot\left(\frac{p_2}{1-p_1}+\frac{p_3}{1-p_1}\right)^{n-i}$$

$$=C_n^ip_1^{\,i}(1-p_1)^{n-i}\quad\Rightarrow\quad X\sim b(n,p_1)$$

同理，可也推出 Y 和 Z 的边缘分布分别为参数是 n, p2 和 n，p3 的二项分布。

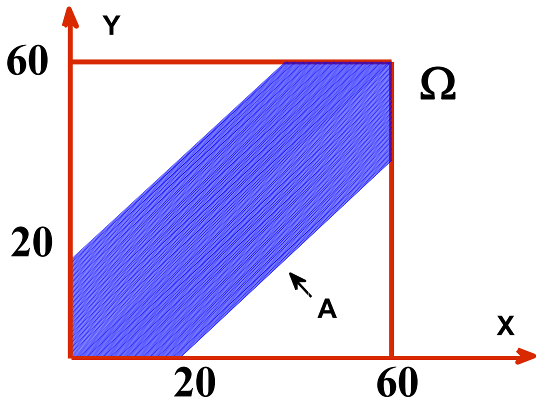
****

二维均匀分布 $(X,Y)\sim U(D)$

D 为 R^2 平面中的有界区域，其面积为 S_D， $f(x,y)=\begin{cases}1/S_D, & (x,y)\in D\\ 0, & otherwise\end{cases}$ 。

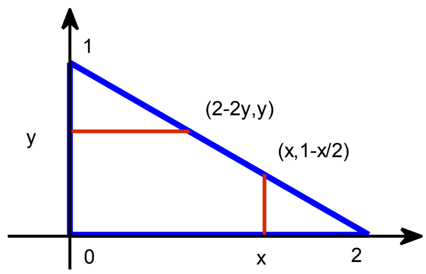
第一周古典概型的例 1.2.3 会面问题，甲、乙约定在下午 4 点至 5 点之间见面，并约定等候 20 分钟，过时即离去，求二人的会面概率。

甲、乙的到达时间 (X,Y) 服从 $0\le x,y\le 60$ 区域上的均匀分布。



****

例 7.2.2 设 (X,Y) 服从 x,y 坐标轴和 $x+2y=2$ 所围成区域内的均匀分布，求 X 和 Y 的边缘密度分布，并计算 $P(X\le 1)$。



解 x,y 坐标轴和 $2x+y=2$ 所围成区域的面积为 1，所以

(X,Y) 的联合密度函数为 $f(x,y)=\begin{cases}1, & x+2y\le 2, x\ge 0, y\ge 0\\ 0, & 其他\end{cases}$

当 $x\le 0$ 或 $x\ge 2$ 时， $f_X(x)=0$，

当 $0<x<2$ 时， $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\int_0^{1-\frac{x}{2}} 1dy=1-\frac{x}{2}$。

当 $y\le 0$ 或 $y\ge 1$ 时， $f_Y(y)=0$，

当 $0<y<1$ 时， $f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\int_0^{2-2y} 1dx=2-2y$

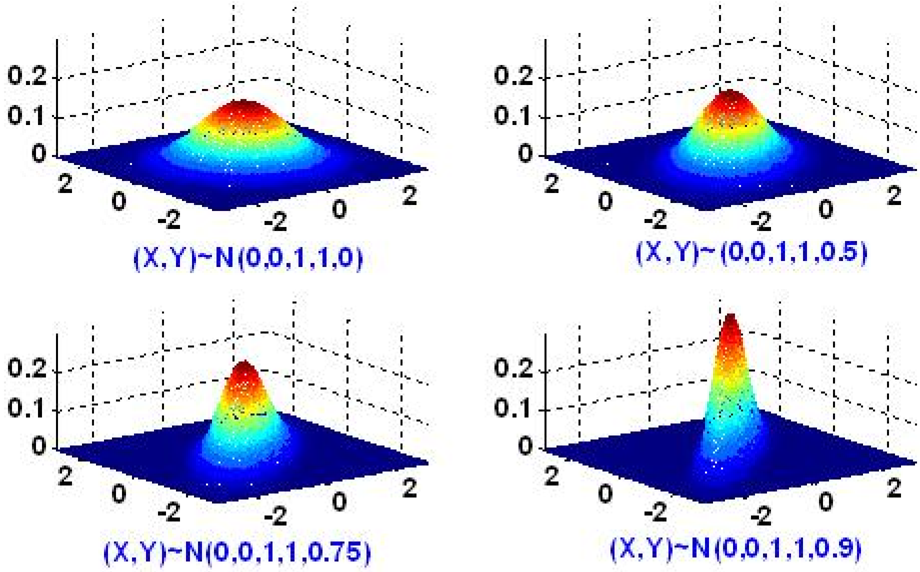
$$P(X \le 1) = \int_{-\infty}^{1} f_X(x)dx = \int_0^1 \left(1 - \frac{x}{2}\right)dx = \frac{3}{4}$$

****

二维正态分布 $(X,Y) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2\left(1-\rho^2\right)}\left[\frac{\left(x-\mu_1\right)^2}{\sigma_1^2} - 2\rho\frac{\left(x-\mu_1\right)\left(y-\mu_2\right)}{\sigma_1\sigma_2} + \frac{\left(y-\mu_2\right)^2}{\sigma_2^2}\right]\right\}$$

其中 $\mu_1, \mu_2 \in R, \sigma_1, \sigma_2 > 0, |\rho| \le 1$ 。



如图所示，为几组不同参数取值下的二维正态分布随机变量的密度函数图像。当 σ_1 和 σ_2 比较小时密度函数的取值更加集中，峰值更为明显。当 ρ 较小时，(X,Y) 的分布较均匀，而当 ρ 较大时，(X,Y) 的分布更加扁平。

****

例 7.2.3 $(X,Y) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$，计算 X 和 Y 的边缘密度函数。

解 $\quad f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2\left(1-\rho^2\right)}\left[\frac{\left(x-\mu_1\right)^2}{\sigma_1^2} - 2\rho\frac{\left(x-\mu_1\right)\left(y-\mu_2\right)}{\sigma_1\sigma_2} + \frac{\left(y-\mu_2\right)^2}{\sigma_2^2}\right]\right\}$$

$$\overset{u=\frac{x-\mu_1}{\sigma_1},\ v=\frac{y-\mu_2}{\sigma_2}}{=} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left\{\frac{-1}{2\left(1-\rho^2\right)}\left[u^2-2\rho\cdot uv+v^2\right]\right\}dv$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left\{\frac{-1}{2\left(1-\rho^2\right)}\left[\left(v-\rho u\right)^2+\left(1-\rho^2\right)u^2\right]\right\}dv$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{u^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left\{\frac{-1}{2\left(1-\rho^2\right)}\left(v-\rho u\right)^2\right\}dv$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{u^2}{2}}=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \quad\Rightarrow\quad X\sim N\left(\mu_1,\sigma_1^2\right)$$

由对称性，可知 $Y\sim N\left(\mu_2,\sigma_2^2\right)$。 二元正态分布的边缘分布是一元正态分布。

逆命题不成立，二元随机变量$\left(X,Y\right)$的边缘分布均为正态，联合分布未必是二元正态。

****