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. □

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

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

由此可得 $c = 12$. 进一步可得

$$P(0 < X < 1, 0 < Y < 2) = 12\int_0^1\int_0^2 e^{-(3x+4y)}dxdy = (1-e^{-3})(1-e^{-8}).$$

□

课题练习. 设二维随机变量 (X, Y) 的概率密度

$$f(x,y) = \begin{cases} x^2 + axy & 0 \le x \le 1, 0 \le y \le 2 \\ 0 & \text{其它}, \end{cases}$$

求 $P(X + Y \ge 1)$.

给定二维随机变量的联合概率密度 $f(x, y)$, 下面定义随机变量 X 和 Y 的边缘概率密度:

定义5.8. 设二维随机变量 (X, Y) 的概率密度为 $f(x, y)$, 则随机变量 X 和 Y 的边缘概率密度为

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy \qquad f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx.$$

上述的边缘概率密度可完全根据边缘分布函数 $F_X(x)$ 的定义导出, 首先可知随机变量 X 的边缘分布函数为

$$\begin{aligned} F_X(x) &= P(X \le x) = P(X \le x, Y < \infty) = F(x, +\infty) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(t,y)dtdy = \int_{-\infty}^{x}\left(\int_{-\infty}^{+\infty} f(t,y)dy\right)dt, \end{aligned}$$

由此可得随机变量 X 的边缘概率密度为

$$f_X(x) = F'_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy.$$

例5.6. 设二维随机变量 (X, Y) 的概率密度

$$f(x,y) = \begin{cases} cxy & 0 \le x \le y \le 1 \\ 0 & \text{其它}, \end{cases}$$

求 $P(X \le 1/2)$.

解. 根据概率密度的性质有

$$1 = \int_0^1\int_x^1 cxydydx = c\int_0^1 x(1-x^2)/2dx = c/8,$$

由此可解 $c = 8$. 当 $0 \le x \le 1$ 时随机变量 X 的边缘概率密度为

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_x^1 8xydy = 4x(1-x^2),$$

进一步有

$$P(X \le \frac{1}{2}) = \int_0^{\frac{1}{2}} 4x(1-x^2)dx = \frac{7}{16}.$$

□

下面定义二维连续随机变量的独立性:

定义5.9. 对任意 $x \in (-\infty, +\infty)$ 和 $y \in (-\infty, +\infty)$, 若二维连续随机变量 (X, Y) 的概率密度与边缘概率密度满足

$$f(x, y) = f_X(x) f_Y(y),$$

则称随机变量 X 和 Y 相互独立.

对连续随机变量, 上述独立性定义与基于分布函数的独立性 (定义 5.4) 等价, 即有如下定理:

定理5.4. 设二维随机变量 (X, Y) 的概率密度为 $f(x, y)$, 则有

$$F(x, y) = F_X(x)F_Y(y) \quad \Longleftrightarrow \quad f(x, y) = f_X(x)f_Y(y).$$

Proof. 首先证明充分性: 若二维连续随机变量满足 $F(x, y) = F_X(x)F_Y(y)$, 则有

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

对上式两边同时求偏导有

$$f(x, y) = f_X(x)f_Y(y).$$

其次证明必要性: 若 $f(x, y) = f_X(x)f_Y(y)$, 则有

$$\begin{aligned} F(x, y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)dudv = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)f_Y(v)dudv \\ &= \int_{-\infty}^{x} f_X(u)du \int_{-\infty}^{y} f_Y(v)dv = F_X(x)F_Y(y). \end{aligned}$$

□

例5.7. 设二维随机变量的密度函数

$$f(x, y) = \begin{cases} cxe^{-y} & 0 < x < y < +\infty \\ 0 & \text{其它}, \end{cases}$$

问 X 与 Y 是否独立.

解. 根据概率密度的性质有

$$1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)dxdy = c\int_0^{+\infty} dy \int_0^{y} xe^{-y}dx = c.$$

当 $x > 0$ 时随机变量 X 的边缘概率密度为

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy = \int_x^{+\infty} xe^{-y}dy = xe^{-x}.$$

同理当 $y>0$ 时随机变量 Y 的边缘概率密度为

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\int_0^y xe^{-y}dx=\frac{1}{2}y^2e^{-y}.$$

由此可得随机变量 X 与 Y 不独立. □

例5.8. 设随机变量 X 与 Y 相互独立, 且 X 服从 $[-1,1]$ 均匀分布, Y 服从参数为 $\lambda=2$ 的指数分布, 求 $P(X+Y\leq 1)$.

解. 根据均匀分布和指数分布的定义有随机变量 X 与 Y 的边缘概率密度为

$$f_X(x)=\begin{cases}\frac{1}{2} & x\in[-1,1]\\ 0 & 其它\end{cases} \quad 和 \quad f_Y(y)=\begin{cases}2e^{-2y} & y\geq 0\\ 0 & 其它\end{cases}$$

根据随机变量的独立性可得随机变量 X 与 Y 的联合概率密度

$$f(x,y)=\begin{cases}e^{-2y} & -1\leq x\leq 1, y\geq 0\\ 0 & 其它\end{cases}.$$

由此可得

$$P(X+Y\leq 1)=\int_{-1}^1 dx\int_0^{1-x} e^{-2y}dy=\frac{3}{4}+\frac{1}{4}e^{-4}.$$

□

对常见的二维随机变量, 我们这里仅仅考虑二维正态 (Gaussian) 分布. 其定义如下

定义5.10. 设 $|\rho|<1$, 令

$$\mu=(\mu_x,\mu_y)^\top=\begin{pmatrix}\mu_x\\ \mu_y\end{pmatrix} \qquad 和 \qquad \Sigma=\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}.$$

若随机变量 X 和 Y 的联合概率密度函数为

$$\begin{aligned} f(x,y) &= (2\pi)^{-2/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(\xi-\mu)^\top\Sigma^{-1}(\xi-\mu)\right) \qquad \xi=(x,y)^\top \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho}{\sigma_x\sigma_y}(x-\mu_x)(y-\mu_y)\right]\right)\end{aligned}$$

这里利用 $|\Sigma|=(1-\rho^2)\sigma_x^2\sigma_y^2$, 以及

$$\Sigma^{-1}=\frac{1}{(1-\rho^2)\sigma_x^2\sigma_y^2}\begin{pmatrix}\sigma_y^2 & -\rho\sigma_x\sigma_y\\ -\rho\sigma_x\sigma_y & \sigma_x^2\end{pmatrix}=\frac{1}{1-\rho^2}\begin{pmatrix}1/\sigma_x^2 & -\rho/\sigma_x\sigma_y\\ -\rho/\sigma_x\sigma_y & 1/\sigma_y^2\end{pmatrix},$$

则称随机变量 X 和 Y 服从参数为 μ 和 Σ 的正太分布, 记为 $(X,Y)\sim\mathcal{N}(\mu,\Sigma)$.

若 $\mu=(0,0)^\top$ 和 Σ 为二维单位阵, 则称为二维标准正态分布. 下面研究二维正态分布的性质:

定理5.5. 设二维随机变量 (X,Y) 服从正态分布 $\mathcal{N}(\mu,\Sigma)$, 其中

$$\mu=\begin{pmatrix}\mu_x\\ \mu_y\end{pmatrix} \qquad \text{和} \qquad \Sigma=\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix},$$

则有边缘分布为 $X\sim\mathcal{N}(\mu_x,\sigma_x^2)$ 和 $Y\sim\mathcal{N}(\mu_y,\sigma_y^2)$.

Proof. 根据边缘概率密度的定义和正态分布性质有

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}e^{-\frac{1}{2(1-\rho^2)}[(\frac{y-\mu_y}{\sigma_y})^2-2\rho\frac{x-\mu_x}{\sigma_x}\frac{y-\mu_y}{\sigma_y}+(\frac{x-\mu_x}{\sigma_x})^2]}dy$$

令 $t=\frac{y-\mu_y}{\sigma_y}$, 有 $dy=\sigma_y dt$, 进一步得到

$$f_X(x)=\frac{e^{-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2}}{2\pi\sqrt{1-\rho^2}\sigma_x}\int_{-\infty}^{+\infty}e^{-\frac{(t-\rho(x-\mu_x)/\sigma_x)^2}{2(1-\rho^2)}}dt=\frac{e^{-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2}}{\sqrt{2\pi}\sigma_x},$$

这里利用了正态分布 $N(\rho(x-\mu_x)/\sigma_x,1-\rho^2)$ 的密度函数满足

$$\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}e^{-\frac{(t-\rho(x-\mu_x)/\sigma_x)^2}{2(1-\rho^2)}}dt=1.$$

□

上述定理说明正太分布的边缘分布还是正太分布.

定理5.6. 若二维随机变量 $(X,Y)\sim\mathcal{N}(\mu,\Sigma)$, 则 X 与 Y 独立的充要条件为 $\Sigma=\begin{pmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{pmatrix}$.

Proof. 若随机变量 $(X,Y)\sim\mathcal{N}(\mu,\Sigma)$, 根据定理 5.5 可知

$$f_X(x)f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_x}e^{-(x-\mu_x)^2/2\sigma_x^2}\times\frac{1}{\sqrt{2\pi}\sigma_y}e^{-(y-\mu_y)^2/2\sigma_y^2}.$$

必要性证明: 当 $\Sigma=\begin{pmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{pmatrix}$ 时有

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}e^{-\frac{1}{2}[(\frac{x-\mu_x}{\sigma_x})^2+(\frac{y-\mu_y}{\sigma_y})^2]}=f_X(x)f_Y(y).$$

充分性证明: 若 X 与 Y 独立, 则对任意 $x\in(-\infty,+\infty)$ 和 $y\in(-\infty,+\infty)$ 有 $f(x,y)=f_X(x)f_Y(y)$ 成立, 即

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}\sigma_x}e^{-(x-\mu_x)^2/2\sigma_x^2}\times\frac{1}{\sqrt{2\pi}\sigma_y}e^{-(y-\mu_y)^2/2\sigma_y^2}\\ &=\frac{1}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho}{\sigma_x\sigma_y}(x-\mu_x)(y-\mu_y)\right]\right)\end{aligned}$$

令 $x=\mu_x$ 和 $y=\mu_y$, 代入上式求解可得 $\rho=0$. □

下面进一步研究多维正态 (Gaussian) 分布, 其定义如下:

定义5.11. 设向量 $\mu \in \mathbb{R}^n$ 和正定矩阵 $\Sigma \in \mathbb{R}^{n\times n}$, 若随机向量 $(X_1, X_2, \cdots, X_n)$ 的概率密度函数为

$$f(x_1, \ldots, x_n) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\xi-\mu)^\top \Sigma^{-1}(\xi-\mu)\right)$$

其中 $\xi = (x_1, \ldots, x_n)^\top$, 则称随机向量 $(X_1, X_2, \cdots, X_n)$ 服从参数为 μ 和 Σ 的多维正态分布, 记

$$(X_1, X_2, \cdots, X_n) \sim \mathcal{N}(\mu, \Sigma).$$

对多维正态分布, 有如下定理:

定理5.7. 设随机向量 $(X, Y) = (X_1, X_2, \cdots, X_n, Y_1, Y_2, \cdots, Y_m) \sim \mathcal{N}(\mu, \Sigma)$, 其中

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \mu_x = (\mu_{x_1}, \mu_{x_2}, \cdots, \mu_{x_n})^\top, \quad \mu_y = (\mu_{y_1}, \mu_{y_2}, \cdots, \mu_{y_m})^\top, \quad \Sigma = \begin{pmatrix} \sum_{xx} & \sum_{yx} \\ \sum_{xy} & \sum_{yy} \end{pmatrix},$$

则有

- 随机向量 X 和 Y 分布服从 $X \sim \mathcal{N}(\mu_x, \Sigma_{xx})$和$Y \sim \mathcal{N}(\mu_y, \Sigma_{yy})$;
- 随机向量 X 与 Y 相互独立的充要条件是 $\Sigma = \begin{pmatrix} \sum_{xx} & 0 \\ 0 & \sum_{yy} \end{pmatrix}$.

这里我们回顾正定矩阵的特征值分解, 对正定矩阵 Σ, 其特征值分解为

$$\Sigma = U^\top \Lambda U$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 为特征值构成的对角阵, U 为特征向量构成的正交矩阵. 我们有如下多维正态分布的标准正态化:

定理5.8. 若随机向量 $X = (X_1, X_2, \cdots, X_n) \sim \mathcal{N}(\mu, \Sigma)$, 且正定矩阵 Σ 的特征值分解为 $\Sigma = U^\top \Lambda U$, 则随机向量

$$Y = \Lambda^{-1/2} U(X-\mu) \sim \mathcal{N}(\mathbf{0}_n, I_n),$$

其中 $\mathbf{0}_n$ 为全为零的 n 维向量, I_n 表示 $n \times n$ 的单位阵.

Proof. 根据 $Y = \Lambda^{-1/2}U(X-\mu)$ 可得 $X = U^\top \Lambda^{1/2} Y + \mu$, 已知 X 的概率密度函数为

$$p_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^\top \Sigma^{-1}(\boldsymbol{x}-\mu)\right)$$

根据函数的概率密度公式有

$$p_Y(\boldsymbol{y}) = p_X(U^\top \Lambda^{1/2}\boldsymbol{y} + \mu)|U^\top \Lambda^{1/2}| = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{y}\right).$$

由此可得完成证明. □

下面研究多维标准正态分布的一些特征:

定理5.9. 若随机向量 $X=(X_1,X_2,\cdots,X_n)\sim\mathcal{N}(\mathbf{0}_n,I_n)$, 则其概率密度函数为 $p_X(\boldsymbol{x})$, 则有

$$\int p_X(\boldsymbol{x})d\boldsymbol{x}=1.$$

Proof. 根据概率密度的定义有

$$\int\int p_X(\boldsymbol{x})d\boldsymbol{x}=\int\int\frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{x}\right)d\boldsymbol{x}=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x_i^2/2}dx_i=1.$$

□

定理5.10. 若随机向量 $X=(X_1,X_2,\cdots,X_n)\sim\mathcal{N}(\mu,\Sigma)$, 则

$$Y=AX+b\sim\mathcal{N}(A\mu+b,A\Sigma A^\top)$$

其中 $A\in\mathbb{R}^{m\times n}$ 和 $b\in\mathbb{R}^{m\times 1}$.

证明作为练习题, 仅仅要求证明 $m=n$ 且 $|A|\neq 0$ 的情况.