Lecture 10: 2-Dimensional Normal Distribution and Random Variable Function

$\boxed{\text{Definition 10.1}}$ 若二维随机变量 (X,Y) 的联合密度为

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

其中 $\mu_1,\mu_2\in\mathbb{R}$，$\sigma_1,\sigma_2>0$，$|\rho|<1$，则称 (X,Y) 服从二维正态分布。记为 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ (这里的每个参数都有具体含义)。

$\boxed{\text{Theorem 10.2}}$ 若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则其边缘分布为 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$。

证明：根据对称性，我们仅需证明 X 的边缘分布为正态分布。

$$\begin{aligned}
p_X(x)&=\int_{-\infty}^{+\infty}p(x,y)dy\\
&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}\left[\left(\frac{y-\mu_2}{\sigma_2\sqrt{1-\rho^2}}\right)^2-\frac{2\rho}{1-\rho^2}\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]}dy\\
&\overset{\text{配方}}{=}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}\left[\left(\frac{y-\mu_2}{\sigma_2\sqrt{1-\rho^2}}\right)^2-\frac{2\rho}{1-\rho^2}\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{x-\mu_1}{\sigma_1\sqrt{1-\rho^2}}\right)^2\right]+\frac{1}{2}\frac{\rho^2(x-\mu_1)^2}{\sigma_1^2(1-\rho^2)}}dy\\
&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2\sqrt{1-\rho^2}}-\frac{\rho(x-\mu_1)}{\sigma_1\sqrt{1-\rho^2}}\right)^2}dy\\
&\overset{u=\frac{y-\mu_2}{\sigma_1\sqrt{1-\rho^2}}-\frac{\rho(x-\mu_1)}{\sigma_1\sqrt{1-\rho^2}}}{=}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}u^2}d(\sigma_2\sqrt{1-\rho^2}u)\\
&=\frac{1}{2\pi\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\int_{-\infty}^{+\infty}e^{-\frac{u^2}{2}}du\\
&=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}
\end{aligned}$$

因此 $X\sim N(\mu_1,\sigma_1^2)$。■

注：二维正态分布可以用矩阵描述成如下更简洁的形式：记 $\Sigma=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$，则

$$p(x,y)=\frac{1}{2\pi\sqrt{|\Sigma|}}e^{-\frac{1}{2}(x-\mu_1,y-\mu_2)\Sigma^{-1}\binom{x-\mu_1}{y-\mu_2}}$$

二维正态分布的边缘分布都是正态分布，那么两个边缘分布都是正态分布的二维分布是否一定是二维正态分布呢？答案是否定的。考虑如下分布：令 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，$g(x)=\begin{cases}\cos x & ,|x|\le\pi\\ 0 & ,|x|>\pi\end{cases}$，可以看到 φ 就是一维标准正态分布的密度函数。令

$$p(x,y)=\varphi(x)\varphi(y)+\frac{1}{2\pi}e^{-\pi^2}g(x)g(y)$$

这显然不是联合正态密度函数，但我们逐一验证二维分布的条件和边缘分布：

- $p(x,y)\geq 0$：我们只需考虑 $|x|\leq\pi,|y|\leq\pi$ 的方形区域：

$$p(x,y)=\frac{1}{2\pi}\left(e^{-\frac{x^2+y^2}{2}}+e^{-\pi^2}\cos x\cos y\right)\geq\frac{1}{2\pi}\left(e^{-\frac{x^2+y^2}{2}}-e^{-\pi^2}\right)\geq 0$$

- $$\begin{aligned}p_X(x)&=\int_{-\infty}^{+\infty}\varphi(x)\varphi(y)dy+\frac{1}{2\pi}e^{-\pi^2}\int_{-\infty}^{+\infty}g(x)g(y)dy\\&=\varphi(x)+\frac{1}{2\pi}e^{-\pi^2}g(x)\int_{-\pi}^{\pi}\cos ydy\\&=\varphi(x)\end{aligned}$$

- $$\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}p(x,y)dy\right)dx=\int_{-\infty}^{+\infty}\varphi(x)dx=1$$

$\boxed{\text{Theorem 10.3}}$ 若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 X,Y 独立 $\Leftrightarrow\rho=0$。

证明：$\Leftarrow$：当 $\rho=0$ 时，

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\cdot\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2}=p_X(x)p_Y(y)$$

$\Rightarrow$：当 X,Y 独立时，取 $x=\mu_1,y=\mu_2$，则

$$p(\mu_1,\mu_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}=p_X(x)p_Y(y)=\frac{1}{2\pi\sigma_1\sigma_2}$$

所以 $\sqrt{1-\rho^2}=1$，$\rho=0$。■

【习题3-5】二维随机变量 (X,Y) 的联合密度函数为

$$p(x,y)=\begin{cases}xe^{-y} & ,0<x<y\\0 & ,otherwise\end{cases}$$

求 (X,Y) 的联合分布函数。

(ATTENTION：$F(x,y)$ 的含义是落在 (x,y) 左下方矩形的概率，切勿根据 $p(x,y)$ 的非零性直接推测 $F(x,y)$ 的非零性！)

> 解：当 $x<0$ 或 $y<0$ 时，$F(x,y)=0$。
>
> 当 $0<x\leq y$ 时，我们要积的是一个梯形区域，
>
> $$\begin{aligned}F(x,y)&=\int_0^x\left(\int_u^y ue^{-v}dv\right)du\\&=\int_0^x u(e^{-u}-e^{-y})du\\&=\left(\int_0^x ue^{-u}du\right)-\frac{1}{2}x^2e^{-y}\\&=1-e^{-x}-xe^{-x}-\frac{1}{2}x^2e^{-y}\end{aligned}$$
>
> 考虑到 F 的连续性，当 $0<y<x$ 时，$F(x,y)=F(y,y)=1-e^{-y}-ye^{-y}-\frac{1}{2}y^2e^{-y}$。

2-Dimensional Random Variable Function

设 (X,Y) 为二维随机变量，$z=g(x,y)$ 为二元实函数，定义 (X,Y) 的函数 $Z=g(X,Y)$，即 Z 在 $(X,Y)=(x,y)$ 时取值 $g(x,y)$，则 Z 是一个随机变量。下面给出求 Z 的分布的方法。

$$F_Z(z)=P(Z\leq z)=P(g(x,y)\leq z)$$

对于离散情形，记 (X,Y) 的分布律为 $P(X=x_i,Y=y_j)=p_{ij},i,j=1,2,\cdots$，则 Z 也为离散型随机变量。记 Z 的可能取值为 $z_k,k=1,2,\cdots$，则

$$\begin{aligned}P(Z=z_k)=P(g(X,Y)=z_k)&=P((X,Y)\in\{(x,y)|g(x,y)=z_k\})\\&=\sum_{i,j:g(x_i,y_j)=z_k}P(X=x_i,Y=y_j)=\sum_{i,j:g(x_i,y_j)=z_k}p_{ij}\end{aligned}$$

对于连续情形，记 (X,Y) 的密度函数为 $p(x,y)$，则

$$\begin{aligned}F_Z(z)=P(g(X,Y)\le z)&=P((X,Y)\in\{(x,y)|g(x,y)\le z\})\\&=\iint_{\{(x,y):g(x,y)\le z\}}p(x,y)dxdy\end{aligned}$$

若 $F_Z(z)$ 可导，则 $p_Z(z)=F_Z'(z)$。

对于混合情形，若 X 为连续型随机变量，密度函数为 $p(x)$，Y 为离散型随机变量，分布为 $P(Y=y_j)=q_j,j=1,2,\cdots$。那么

$$\begin{aligned}F_Z(z)=P(g(X,Y)\le z)&=\sum_{j=1}^{\infty}P(g(X,Y)\le z,Y=y_j)\\&=\sum_{j=1}^{\infty}P(g(X,y_j)\le z,Y=y_j)\\&=\sum_{j=1}^{\infty}q_jP(g(X,y_j)\le z|Y=y_j)\quad\text{(条件概率)}\end{aligned}$$

若 X,Y 独立，则上式可化为

$$\sum_{j=1}^{\infty}P(g(X,y_j)\le z)=\sum_{j=1}^{\infty}q_j\int_{x:g(x,y_j)\le z}p(x)dx$$