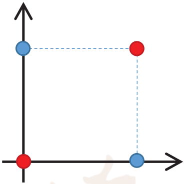
§3.2 二维随机向量的联合分布与边缘分布, §3.7 条件分布

- 定义2.1 & 2.2. 若$\boldsymbol{\xi} = (\mathbf{X}, \mathbf{Y})$ 取有限个或可列个“值”(二维向量), 则称$\boldsymbol{\xi}$ 为离散型.
- $\boldsymbol{\xi}$ 是离散型当且仅当$\mathbf{X}, \mathbf{Y}$ 都是离散型.

- 定义2.3. 设$\boldsymbol{\xi} = (\mathbf{X}, \mathbf{Y})$, 则$\mathbf{X}$ 的分布称为$\boldsymbol{\xi}$ 关于$\mathbf{X}$ 的边缘分布. 关于$\mathbf{Y}$ 的边缘分布类似.



- 例2.5. $0 \leqslant \boldsymbol{\varepsilon} \leqslant \frac{1}{4}$,

  $\mathbf{P}(\boldsymbol{\xi} = (0,0)) = \mathbf{P}(\boldsymbol{\xi} = (1,1)) = \frac{1}{4} + \boldsymbol{\varepsilon}$;

  $\mathbf{P}(\boldsymbol{\xi} = (0,1)) = \mathbf{P}(\boldsymbol{\xi} = (1,0)) = \frac{1}{4} - \boldsymbol{\varepsilon}$.

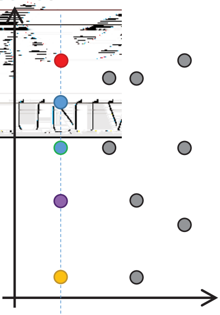
  总有, $\mathbf{X}, \mathbf{Y} \sim \mathbf{B}(1, \frac{1}{2})$.

- 给定$\mathbf{j}$, 将

  $$\mathbf{P}(\mathbf{X} = \mathbf{x}_i | \mathbf{Y} = \mathbf{y}_j), \ \mathbf{i} = 1, 2, \cdots$$

  称为在$\mathbf{Y} = \mathbf{y}_j$ 的条件下, $\mathbf{X}$ 的条件分布(列);

  $\mathbf{Y}$ 的条件分布类似. (7.3)



- 联合分布列$\Leftrightarrow$ 边缘分布列、条件分布列.

例2.2 & 2.3: 有大量粉笔, 含白、黄、红三种颜色, 比例分别为$\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$. 从中抽取$\mathbf{n}$ 支. 求: 恰好抽到$\mathbf{k}_1$ 支白, $\mathbf{k}_2$ 支黄的概 .

- $P(X=k_1, Y=k_2) = C_n^{k_1} C_{n-k_1}^{k_2} p_1^{k_1} p_2^{k_2} p_3^{n-k_1-k_2}.$
- X 的边缘分布:

$$P(X=k_1) = \sum_{k_2=0}^{n-k_1} C_n^{k_1} C_{n-k_1}^{k_2} p_1^{k_1} p_2^{k_2} p_3^{n-k_1-k_2}$$

$$= C_n^{k_1} p_1^{k_1} (1-p_1)^{n-k_1}.$$

- Y 的条件分布: 给定k_1,

$$P(Y=k_2 | X=k_1) = \frac{P(X=k_1, Y=k_2)}{P(X=k_1)} = \frac{\star\star}{**}$$

$$= C_{n-k_1}^{k_2} \left(\frac{p_2}{p_2+p_3}\right)^{k_2} \left(\frac{p_3}{p_2+p_3}\right)^{k_3}, \quad k_2 = 0, 1, \cdots, n-k_1.$$

2. 连续型情形

- 定理2.1. 若$\boldsymbol{\xi} = (\mathbf{X}, \mathbf{Y})$ 是连续型, 则$\mathbf{X}, \mathbf{Y}$ 都是连续型, 且

$$\mathbf{p}_X(\mathbf{x}) = \int_{-\infty}^{\infty} \mathbf{p}_{X,Y}(\mathbf{x}, \mathbf{y})\mathbf{dy}, \quad \mathbf{p}_Y(\mathbf{y}) = \int_{-\infty}^{\infty} \mathbf{p}_{X,Y}(\mathbf{x}, \mathbf{y})\mathbf{dx}.$$

- 称$\mathbf{p}_X(\cdot)$ 与$\mathbf{p}_Y(\cdot)$ 为$\boldsymbol{\xi}$ 的边缘密度.
- 给定$\mathbf{y}$, 满足$\mathbf{p}_Y(\mathbf{y}) > 0$. 称(关于$\mathbf{x}$的函数)

$$\mathbf{p}_{X|Y}(\mathbf{x}|\mathbf{y}) := \frac{\mathbf{p}(\mathbf{x}, \mathbf{y})}{\mathbf{p}_Y(\mathbf{y})}, \quad \mathbf{x} \in \mathbb{R}$$

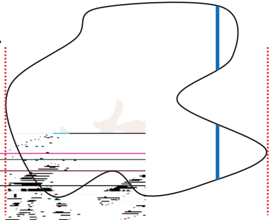
为在$\mathbf{Y} = \mathbf{y}$ 的条件下, $\mathbf{X}$ 的条件密度. (7.5)

- 联合密度$\Leftrightarrow$ 边缘密度、条件密度.

- 定义2.5. 假设**G** 是$\mathbb{R}^2$ 中面积为**a** 的区域. 若

$$\mathbf{P}(\boldsymbol{\xi} \in \mathbf{A}) = \frac{\mathbf{A}\text{的面积}}{\mathbf{G}\text{的面积}}, \quad \forall \text{子区域}\mathbf{A},$$

则称**ξ** 服从**G** 上的均匀分布, 记为$\boldsymbol{\xi} \sim \mathbf{U}(\mathbf{G})$.



- 联合密度: $\mathbf{p}(\mathbf{x},\mathbf{y}) = \frac{1}{a}$, $(\mathbf{x},\mathbf{y}) \in \mathbf{G}$.
- 边缘密度:

$$\mathbf{p}_X(\mathbf{x}) = \frac{|\mathbf{G}_{2,x}|}{\mathbf{a}}, \quad \mathbf{x} \in \mathbf{G}_1,$$

其中, $\mathbf{G}_{2,x} := \{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in \mathbf{G}\}$, $|\mathbf{G}_{2,x}|$ 为其总长度;

$\mathbf{G}_1 = \{\mathbf{x} : |\mathbf{G}_{2,x}| > 0\}$.

- 条件密度: $\mathbf{p}_{Y|X}(\mathbf{y}|\mathbf{x}) = \frac{1}{|G_{2,x}|}$, $\mathbf{y} \in \mathbf{G}_{2,x}$.
- $\mathbf{p}_{Y|X}(\mathbf{y}|\mathbf{x})$ 就是固定**x**, 将$\mathbf{p}(\mathbf{x},\mathbf{y})$ 视为**y** 的函数归一化,

$$\mathbf{p}_{Y|X}(\mathbf{y}|\mathbf{x}) = \frac{1}{\mathbf{p}_X(\mathbf{x})}\mathbf{p}(\mathbf{x},\mathbf{y}).$$

例2.7. $\mathbf{G}$ 为由$\mathbf{y} = \mathbf{x}^2$ 和$\mathbf{y} = \mathbf{x}$ 所围成的有限区域. $\boldsymbol{\xi} \sim \mathbf{U}(\mathbf{G})$.

求: $\boldsymbol{\xi}$ 的联合密度与边缘密度.

- $\mathbf{G}$ 的面积: $\mathbf{a} = \int_0^1 \mathbf{x}\mathbf{dx} - \int_0^1 \mathbf{x}^2\mathbf{dx} = \frac{1}{6}$.
- 联合密度: $\mathbf{p}(\mathbf{x},\mathbf{y}) = 6, \quad (\mathbf{x},\mathbf{y}) \in \mathbf{G}$.
- 边缘密度:

$$\mathbf{p}_X(\mathbf{x}) = \int_{-\infty}^{\infty} \mathbf{p}(\mathbf{x},\mathbf{y})\mathbf{dy} = \int_{x^2}^{x} 6\mathbf{dy} = 6(\mathbf{x} - \mathbf{x}^2), \; 0 < \mathbf{x} < 1.$$

$$\mathbf{p}_Y(\mathbf{y}) = \int_{-\infty}^{\infty} \mathbf{p}(\mathbf{x},\mathbf{y})\mathbf{dx} = \int_{y}^{\sqrt{y}} 6\mathbf{dx} = 6(\sqrt{\mathbf{y}} - \mathbf{y}), \; 0 < \mathbf{y} < 1.$$

- 条件密度: 固定$\mathbf{y} \in (0,1)$,

$$\mathbf{p}_{X|Y}(\mathbf{x}|\mathbf{y}) = \frac{1}{\sqrt{y}-y}, \; \sqrt{\mathbf{y}} \leqslant \mathbf{x} \leqslant \mathbf{y}.$$

- 注: $\mathbf{X}$, $\mathbf{Y}$ 都取遍$(0,1)$, 但$\boldsymbol{\xi}$ 不能取遍$(0,1) \times (0,1)$.

- 定义2.6, 例2.8 & 例7.5. 若$\boldsymbol{\xi} = (\mathbf{X}, \mathbf{Y})$ 的联合密度$\mathbf{p}(\mathbf{x}, \mathbf{y})$ 有如下表达式, 则称$\boldsymbol{\xi}$ 服从二维(元)正态分布.

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}\right\},$$

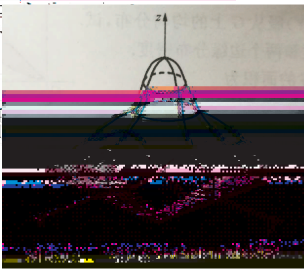
其中,

$$u = \frac{x-\mu_1}{\sigma_1}, \quad v = \frac{y-\mu_2}{\sigma_2},$$

有5个参数: $\mu_1, \mu_2 \in \mathbb{R}$,

$\sigma_1, \sigma_2 > 0$,

$\rho \in (-1, 1)$.



- 联合密度: $\mathbf{u} = \frac{x-\mu_1}{\sigma_1}$, $\mathbf{v} = \frac{y-\mu_2}{\sigma_2}$,

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{\mathbf{u}^2+\mathbf{v}^2-2\rho\mathbf{u}\mathbf{v}}{2(1-\rho^2)}\right\}.$$

- 边缘密度: $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2)^2$. 例如,

$$\begin{aligned}
p_X(x) &= \int_{-\infty}^{\infty} p(x,y)\,dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{(\mathbf{v}-2\rho\mathbf{u})^2+(1-\rho^2)\mathbf{u}^2}{2(1-\rho^2)}\right\} d\mathbf{v} \\
&= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} e^{-\frac{(1-\rho^2)u^2}{2(1-\rho^2)}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(\mathbf{v}-2\rho\mathbf{u})^2}{2(1-\rho^2)}\right\} d\mathbf{v} \\
&= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2}u^2} \sqrt{2\pi(1-\rho^2)} = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}.
\end{aligned}$$

- 联合密度: $\mathbf{u}=\frac{x-\mu_1}{\sigma_1}$, $\mathbf{v}=\frac{y-\mu_2}{\sigma_2}$,

$$\mathbf{C}\exp\left\{-\frac{\mathbf{u}^2+\mathbf{v}^2-2\rho\mathbf{u}\mathbf{v}}{2(1-\rho^2)}\right\}$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{(\mathbf{v}-2\rho\mathbf{u})^2+(1-\rho^2)\mathbf{u}^2}{2(1-\rho^2)}\right\}.$$

- 边缘密度: $\mathbf{p}_X(\mathbf{x})=\frac{1}{\sqrt{2\pi}\sigma_1}\mathbf{e}^{\frac{(1-\rho^2)u^2}{2(1-\rho^2)}}$.
- 条件密度:

$$\mathbf{p}_{Y|X}(\mathbf{y}|\mathbf{x})=\frac{\mathbf{p}(\mathbf{x},\mathbf{y})}{\mathbf{p}_X(\mathbf{x})}=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{(\mathbf{v}-2\rho\mathbf{u})^2}{2(1-\rho^2)}\right\}.$$

- 另解:

$$\mathbf{p}_{Y|X}(\mathbf{y}|\mathbf{x})=\frac{1}{\mathbf{p}_X(\mathbf{x})}\mathbf{p}(\mathbf{x},\mathbf{y})=\hat{\mathbf{C}}\exp\left\{-\frac{(\frac{y-\mu_1}{\sigma_2}-2\rho\mathbf{u})^2}{2(1-\rho^2)}\right\}.$$

例2.9. $\xi=(X,Y)$ 与$\eta=(U,V)$ 分别有联合密度

$$p(x,y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}},\quad q(u,v)=2p(u,v),\ uv>0.$$

- ξ 服从二维正态分布, $X,Y\sim N(0,1)$, $\rho=0$.
- η 不服从二维正态分布.
- 但$U,V\sim N(0,1)$. 例如, $\forall u>0$,

$$p_U(u)=\int_{-\infty}^{\infty}q(u,v)dv=\int_0^{\infty}2p(u,v)dv$$
$$=\int_0^{\infty}\frac{1}{\pi}e^{-\frac{u^2+v^2}{2}}dv=\frac{1}{\pi}e^{-\frac{u^2}{2}}\int_0^{\infty}e^{-\frac{v^2}{2}}dv$$
$$=\frac{1}{\pi}e^{-\frac{u^2}{2}}\times\frac{1}{2}\sqrt{2\pi}=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}.$$

- 注: U,V 都是正态变量, 不能推出(U,V) 是二维正态向量.

- 定义2.7. 称$\mathbf{F}(\mathbf{x},\mathbf{y}) = \mathbf{P}(\mathbf{X} \leqslant \mathbf{x}, \mathbf{Y} \leqslant \mathbf{y})$ 为$(\mathbf{X},\mathbf{Y})$ 的联合分布函数，也记为$\mathbf{F}_{X,Y}(\mathbf{x},\mathbf{y})$.
- 联合分布函数的性质：“单调”、“规范”、右连续，

$$\mathbf{F}(\mathbf{x}_2,\mathbf{y}_2) - \mathbf{F}(\mathbf{x}_1,\mathbf{y}_2) - \mathbf{F}(\mathbf{x}_2,\mathbf{y}_1) + \mathbf{F}(\mathbf{x}_1,\mathbf{y}_1) \geqslant 0;$$

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

- 连续型向量:

$$\mathbf{F}(\mathbf{x},\mathbf{y}) = \int_{-\infty}^{x}\int_{-\infty}^{y} \mathbf{p}(\mathbf{u},\mathbf{v})\mathbf{dvdu} \Rightarrow \mathbf{p}(\mathbf{x},\mathbf{y}) = \frac{\partial^2 \mathbf{F}}{\partial \mathbf{x}\partial \mathbf{y}}(\mathbf{x},\mathbf{y}).$$