§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$\xi = (X, Y)$ 取有限个或可列个“值” (二维向量), 则称ξ 为离散型.
- ξ 是离散型当且仅当X, Y 都是离散型.
- 定义2.2. 设X, Y 的可能值分别为x_i, y_j, 则称

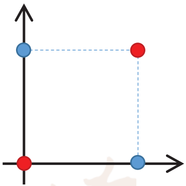
$$p_{ij} = P(X = x_i, Y = y_j), \quad i, j = 1, 2, \cdots$$

为ξ 的联合分布(列).

- 联合分布列满足: $p_{ij} \geqslant 0$, $i, j = 1, 2, \cdots$ (非负性);

$$\sum_{i,j} p_{ij} = 1 \text{ (规范性)}.$$

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



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

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

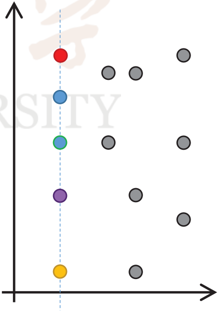
  $P(\xi=(0,1))=P(\xi=(1,0))=\frac{1}{4}-\varepsilon$.

  总有, $X,Y\sim B(1,\frac{1}{2})$.

- 给定j, 将

$$P(X=x_i|Y=y_j),\ i=1,2,\cdots$$

  称为在$Y=y_j$ 的条件下, X 的条件分布(列); 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p_1, p_2, p_3. 从中抽取n 支. 求: 恰好抽到k_1 支白, k_2 支黄的概率.

- 设恰好抽到X 支白, Y 支黄, 即求$(X, Y) = (k_1, k_2)$ 的概率.
- 可以理解为放回抽样, 连续抽取n 次.
- 所求事件包含了
$$C_n^{k_1} C_{n-k_1}^{k_2} = \frac{n!}{k_1! k_2! (n - k_1 - k_2)!}$$
个基本事件, 其中, 每一个的概率都为
$$p_1^{k_1} p_2^{k_2} p_3^{n-k_1-k_2}.$$
- 故, $\forall k_1, k_2 \geqslant 0$, $k_1 + k_2 \leqslant n$,
$$P(X = k_1, Y = k_2) = C_n^{k_1} C_{n-k_1}^{k_2} p_1^{k_1} p_2^{k_2} p_3^{n-k_1-k_2}.$$
- 称$\xi = (X, Y)$ 服从三项分布.

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

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

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

2. 连续型情形

- 定义2.4. 设$\xi = (X, Y)$. 若存在$p(x, y)$ 使得

$$P(\xi \in D) = \iint_D p(x, y)dxdy,$$

对任意开矩形 D 成立, 则称ξ 为连续型随机向量, 称$p(x, y)$ 为ξ 的联合密度(函数), 也记为$p_{X,Y}(x, y)$.

- 联合密度满足:

$$p(x, y) \geqslant 0; \quad \iint_{\mathbb{R}^2} p(x, y)dxdy = 1.$$

- ★★ 对更一般的集合D 都成立, 例如, D 是单位圆盘.

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

$$p_X(x)=\int_{-\infty}^{\infty}p_{X,Y}(x,y)dy,\quad p_Y(y)=\int_{-\infty}^{\infty}p_{X,Y}(x,y)dx.$$

- 称$p_X(\cdot)$ 与$p_Y(\cdot)$ 为ξ 的边缘密度.
- 给定y, 满足$p_Y(y)>0$. 称(关于x 的函数)

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

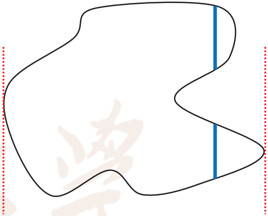
为在$Y=y$ 的条件下, X 的条件密度. (7.5)

- 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

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

则称ξ 服从G 上的均匀分布, 记为$\xi \sim U(G)$.



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

$$p_X(x) = \frac{|G_{2,x}|}{a}, \quad x \in G_1,$$

其中, $G_{2,x} := \{y : (x,y) \in G\}$, $|G_{2,x}|$ 为其总长度;
$G_1 = \{x : |G_{2,x}| > 0\}$.

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

$$p_{Y|X}(y|x) = \frac{1}{p_X(x)} p(x,y).$$

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

求: ξ 的联合密度与边缘密度.

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

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

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

- 条件密度: 固定$y\in(0,1)$,

$$p_{X|Y}(x|y)=\frac{1}{\sqrt{y}-y},\ \sqrt{y}\leqslant x\leqslant y.$$

- 注: X, Y 都取遍$(0,1)$, 但ξ 不能取遍$(0,1)\times(0,1)$.

- 定义2.6, 例2.8 & 例7.5. 若$\xi=(X,Y)$ 的联合密度$p(x,y)$ 有如下表达式, 则称ξ 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: $\mu_1,\mu_2\in\mathbb{R}$,

$\sigma_1,\sigma_2>0$,

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

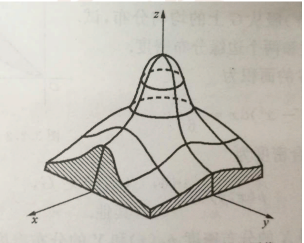


图 3.2.1 二维正态分布密度函数的图像

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

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

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

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

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

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

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

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

- 另解:

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

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, $\forall u>0$,

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

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

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

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

- 连续型向量:

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