

Lec-9. 边缘分布、条件分布

主讲教师：吴利苏 (wulisu@sdust.edu.cn)

主 页：wulisu.cn

目录

1. 边缘分布

- 二维离散型随机变量的边缘分布
- 二维连续型随机变量的边缘分布

2. 条件分布

- 二维离散型随机变量的条件分布
- 二维连续型随机变量的条件分布

边缘分布

目标： 已知 (X, Y) 的分布, 求 X, Y 的分布?

设 $F(x, y) = P\{X \leq x, Y \leq y\}$, 令

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= P\{X \leq x, Y \leq \infty\} \\ &= \lim_{y \rightarrow \infty} P\{X \leq x, Y \leq y\} \\ &= \lim_{y \rightarrow \infty} F(x, y) = F(x, \infty) \end{aligned}$$

设 $F(x, y) = P\{X \leq x, Y \leq y\}$, 令

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= P\{X \leq x, Y \leq \infty\} \\ &= \lim_{y \rightarrow \infty} P\{X \leq x, Y \leq y\} \\ &= \lim_{y \rightarrow \infty} F(x, y) = F(x, \infty) \end{aligned}$$

- 称 $F_X(x) = F(x, \infty)$ 为 (X, Y) 关于 X 的**边缘分布函数**.

设 $F(x, y) = P\{X \leq x, Y \leq y\}$, 令

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= P\{X \leq x, Y \leq \infty\} \\ &= \lim_{y \rightarrow \infty} P\{X \leq x, Y \leq y\} \\ &= \lim_{y \rightarrow \infty} F(x, y) = F(x, \infty) \end{aligned}$$

- 称 $F_X(x) = F(x, \infty)$ 为 (X, Y) 关于 X 的**边缘分布函数**.
- 称 $F_Y(y) = F(\infty, y)$ 为 (X, Y) 关于 Y 的**边缘分布函数**.

离散型随机变量的边缘分布律

设二维离散型随机变量 (X, Y) 的联合分布律为

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

离散型随机变量的边缘分布律

设二维离散型随机变量 (X, Y) 的联合分布律为

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

- $P\{X = x_i\} = P\{X = x_i, \bigcup_{j=1}^{\infty} \{Y = y_j\}\} = \sum_{j=1}^{\infty} p_{ij} \triangleq p_{i\bullet},$
 $i = 1, 2, \dots$ 为 (X, Y) 关于 X 的**边缘分布律**.

离散型随机变量的边缘分布律

设二维离散型随机变量 (X, Y) 的联合分布律为

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

- $P\{X = x_i\} = P\{X = x_i, \bigcup_{j=1}^{\infty} \{Y = y_j\}\} = \sum_{j=1}^{\infty} p_{ij} \triangleq p_{i\bullet},$
 $i = 1, 2, \dots$ 为 (X, Y) 关于 X 的**边缘分布律**.
- $P\{Y = y_j\} = P\{\bigcup_{i=1}^{\infty} \{X = x_i\}, Y = y_j\} = \sum_{i=1}^{\infty} p_{ij} \triangleq p_{\bullet j},$
 $j = 1, 2, \dots$ 为 (X, Y) 关于 Y 的**边缘分布律**.

$X \backslash Y$	y_1	y_2	\dots	y_j	\dots	$P\{X = x_i\}$
x_1	p_{11}	p_{12}	\dots	p_{1j}	\dots	$p_{1\bullet}$
x_2	p_{21}	p_{22}	\dots	p_{2j}	\dots	$p_{2\bullet}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_i	p_{i1}	p_{i2}	\dots	p_{ij}	\dots	$p_{i\bullet}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
$P\{Y = y_j\}$	$p_{\bullet 1}$	$p_{\bullet 2}$	\dots	$p_{\bullet j}$	\dots	1

$$P\{X = x_i\} = \sum_{j=1}^{\infty} p_{ij}, \quad i = 1, 2, \dots \quad \text{横向之和};$$

$$P\{Y = y_j\} = \sum_{i=1}^{\infty} p_{ij}, \quad j = 1, 2, \dots \quad \text{纵向之和}.$$

$\begin{array}{c} \diagdown \\ Y \end{array} \quad X$	x_1	x_2	\dots	x_i	\dots	$P\{Y = y_j\}$
y_1	p_{11}	p_{21}	\dots	p_{i1}	\dots	$p_{\bullet 1}$
y_2	p_{12}	p_{22}	\dots	p_{i2}	\dots	$p_{\bullet 2}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
y_j	p_{1j}	p_{2j}	\dots	p_{ij}	\dots	$p_{\bullet j}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
$P\{X = x_i\}$	$p_{1\bullet}$	$p_{2\bullet}$	\dots	$p_{i\bullet}$	\dots	1

$$P\{X = x_i\} = \sum_{j=1}^{\infty} p_{ij}, \quad i = 1, 2, \dots \quad \text{纵向之和};$$

$$P\{Y = y_j\} = \sum_{i=1}^{\infty} p_{ij}, \quad j = 1, 2, \dots \quad \text{横向之和}.$$

离散型随机变量的边缘分布函数

$$F_X(x) = F(x, \infty) = \sum_{x_i \leq x} \sum_{j=1}^{\infty} p_{ij}.$$

$$F_Y(y) = F(\infty, y) = \sum_{y_j \leq y} \sum_{i=1}^{\infty} p_{ij}.$$

例

		Y	
		0	1
X	0	$\frac{16}{49}$	$\frac{12}{49}$
	1	$\frac{12}{49}$	$\frac{9}{49}$

求其边缘分布律.

例

		Y	
		0	1
X	0	$\frac{16}{49}$	$\frac{12}{49}$
	1	$\frac{12}{49}$	$\frac{9}{49}$

求其边缘分布律.

$$\text{解: } P\{X=0\} = \frac{4}{7}, P\{X=1\} = \frac{3}{7}, \\ P\{Y=0\} = \frac{4}{7}, P\{Y=1\} = \frac{3}{7}.$$



例

X \ Y	0	1
0	$\frac{16}{49}$	$\frac{12}{49}$
1	$\frac{12}{49}$	$\frac{9}{49}$

求其边缘分布律.

$$\begin{aligned} \text{解: } P\{X=0\} &= \frac{4}{7}, \quad P\{X=1\} = \frac{3}{7}, \\ P\{Y=0\} &= \frac{4}{7}, \quad P\{Y=1\} = \frac{3}{7}. \end{aligned}$$

□

注: 联合分布 \Rightarrow 边缘分布, 反之不成立.

例

一整数 N 等可能地在 $1, 2, 3, \dots, 10$ 个值中取一个值, 设 $D = D(N)$ 是能整除 N 的正整数的个数, $F = F(N)$ 是能整除 N 的素数的个数, 试写出 D 和 F 的联合分布律, 并求边缘分布律.

例

一整数 N 等可能地在 $1, 2, 3, \dots, 10$ 个值中取一个值, 设 $D = D(N)$ 是能整除 N 的正整数的个数, $F = F(N)$ 是能整除 N 的素数的个数, 试写出 D 和 F 的联合分布律, 并求边缘分布律.

解: 写出 D, F 的可能取值.

样本点 N	1	2	3	4	5	6	7	8	9	10
D	1	2	2	3	2	4	2	4	3	4
F	0	1	1	1	1	2	1	1	1	2

D 可能取值 $1, 2, 3, 4$; F 可能取值 $0, 1, 2$.

$$P\{D = 1, F = 0\} = \frac{1}{10}$$

$$P\{D = 2, F = 1\} = \frac{4}{10}$$

$$P\{D = 3, F = 1\} = \frac{2}{10}$$

$$P\{D = 4, F = 1\} = \frac{1}{10}$$

$$P\{D = 4, F = 2\} = \frac{2}{10}$$

$\begin{array}{c} \text{F} \\ \text{D} \end{array}$	0	1	2	$P\{D = i\}$
1	$\frac{1}{10}$	0	0	$\frac{1}{10}$
2	0	$\frac{4}{10}$	0	$\frac{4}{10}$
3	0	$\frac{2}{10}$	0	$\frac{2}{10}$
4	0	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$
$P\{F = j\}$	$\frac{1}{10}$	$\frac{7}{10}$	$\frac{2}{10}$	1

连续型随机变量的边缘分布

设连续型随机变量 (X, Y) 的概率密度为 $f(x, y)$,

$$F_X(x) = F(x, \infty) = \int_{-\infty}^x \left[\int_{-\infty}^{+\infty} f(x, y) dy \right] dx$$

则

- $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$ 为 (X, Y) 关于 X 的**边缘概率密度**.
- $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$ 为 (X, Y) 关于 Y 的**边缘概率密度**.

例

设 X, Y 具有联合概率密度

$$f(x, y) = \begin{cases} 6 & x^2 \leq y \leq x, \\ 0 & \text{其他.} \end{cases}$$

求边缘概率密度 $f_X(x), f_Y(y)$.

解:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy = \begin{cases} \int_{x^2}^x 6 dy & 0 \leq x \leq 1; \\ 0 & \text{其他.} \end{cases} \\ &= \begin{cases} 6(x - x^2) & 0 \leq x \leq 1; \\ 0 & \text{其他.} \end{cases} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) dx = \begin{cases} \int_y^{\sqrt{y}} 6 dx & 0 \leq y \leq 1; \\ 0 & \text{其他.} \end{cases} \\ &= \begin{cases} 6(\sqrt{y} - y) & 0 \leq y \leq 1; \\ 0 & \text{其他.} \end{cases} \end{aligned}$$

□

例

设二维正态随机变量 (X, Y) 的概率密度为

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

其中 $x, y \in \mathbb{R}$, $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 均为常数, $\sigma_1 > 0$, $\sigma_2 > 0$, $-1 < \rho < 1$. 求 (X, Y) 的边缘概率密度.

解:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\frac{(x-\mu_1)^2}{\sigma_1^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right)^2} dy \end{aligned}$$

令 $t = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right)$, 则上式

$$= \frac{1}{2\pi\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}.$$

即, $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, x \in \mathbb{R}.$

同理, $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}, y \in \mathbb{R}.$

- 二维正态分布的两个边缘分布都是一维正态分布, 并且不依赖参数 ρ .

即对于给定的 $\mu_1, \mu_2, \sigma_1, \sigma_2$, 取不同的 ρ 对应不同的二维正态分布, 而边缘分布是一样的.

- 二维正态分布的两个边缘分布都是一维正态分布, 并且不依赖参数 ρ .

即对于给定的 $\mu_1, \mu_2, \sigma_1, \sigma_2$, 取不同的 ρ 对应不同的二维正态分布, 而边缘分布是一样的.

- 联合分布 \Rightarrow 边缘分布,
边缘分布 \nRightarrow 联合分布.

问题：边缘分布均为正态分布的随机变量，其联合分布一定是二维正态分布吗？

问题：边缘分布均为正态分布的随机变量，其联合分布一定是二维正态分布吗？

答：不一定.

例 (反例)

设 (X, Y) 的联合概率密度为

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} (1 + \sin x \sin y),$$

不服从正态分布. 但 $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$

$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$ X 和 Y 服从正态分布.

条件分布

对于事件 A, B , 若 $P(A) > 0$, 可考虑条件概率

$$P(B|A) = \frac{P(AB)}{P(A)}.$$

条件分布

对于事件 A, B , 若 $P(A) > 0$, 可考虑条件概率

$$P(B|A) = \frac{P(AB)}{P(A)}.$$

下面引入随机变量的条件分布：

- 二维离散型随机变量的条件分布；
- 二维连续型随机变量的条件分布. (下节课)

二维离散型随机变量的条件分布

对于二维离散型随机变量 (X, Y) , 设其分布律为

$$P\{X = x_i, Y = y_j\} = p_{ij},$$

若 $P\{Y = y_j\} = p_{\bullet j} > 0$, 则由条件概率公式

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_{\bullet j}}$$

当 X 取遍所有可能值, 就得到了条件分布律.

设 (X, Y) 是二维离散型随机变量. 则

- 对于固定的 y_j , 若 $P\{Y = y_j\} > 0$, 则称

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_{\bullet j}}$$

为在 $Y = y_j$ 条件下, X 的**条件分布律**.

设 (X, Y) 是二维离散型随机变量. 则

- 对于固定的 y_j , 若 $P\{Y = y_j\} > 0$, 则称

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_{\bullet j}}$$

为在 $Y = y_j$ 条件下, X 的**条件分布律**.

- 对于固定的 x_i , 若 $P\{X = x_i\} > 0$, 则称

$$P\{Y = y_j | X = x_i\} = \frac{P\{X = x_i, Y = y_j\}}{P\{X = x_i\}} = \frac{p_{ij}}{p_{i\bullet}}$$

为在 $X = x_i$ 条件下, Y 的**条件分布律**.

例

盒中装有 3 只红球, 4 只黑球, 3 只白球, 不放回取 2 只球. 以 X 表示取到红球的只数, Y 表示取到黑球的只数. 求

(1) (X, Y) 的联合分布律.

(2) $X = 1$ 时 Y 的条件分布律.

解: (1) (X, Y) 的取值

$(0, 0), (0, 1), (0, 2), (1, 0), (1, 2), (1, 1), (2, 0), (2, 1), (2, 2).$

$$P\{X = i, Y = j\} = p_{ij} = \frac{C_3^i C_4^j C_3^{2-i-j}}{C_{10}^2},$$

$i, j = 1, 2, i + j \leq 2.$

X \ Y	Y		
	0	1	2
0	$\frac{1}{15}$	$\frac{4}{15}$	$\frac{2}{15}$
1	$\frac{3}{15}$	$\frac{4}{15}$	0
2	$\frac{1}{15}$	0	0

(2)

$$\begin{aligned} P\{X = 1\} &= \frac{7}{15} & P\{Y = 0|X = 1\} &= \frac{3}{7} \\ P\{Y = 1|X = 1\} &= \frac{4}{7} & P\{Y = 2|X = 1\} &= 0 \end{aligned}$$

Y	0	1	2
$P\{Y = j X = 1\}$	$\frac{3}{7}$	$\frac{4}{7}$	0

例

已知 (X, Y) 的联合分布律

X \ Y	Y		
	-1	0	1
1	a	0.2	0.2
2	0.1	0.1	b

已知 $P\{Y \leq 0 | X < 2\} = 0.5$, 求

(1) a, b 的值.

(2) $\{X = 2\}$ 条件下, Y 的条件分布律.

(3) $\{X + Y = 2\}$ 条件下, X 的条件分布律.

解: (1)

$$\begin{cases} a + b = 0.4 \\ P\{Y \leq 0 | X < 2\} = 0.5 \end{cases} \Rightarrow \begin{cases} a = 0, \\ b = 0.4. \end{cases} .$$

其中

$$\begin{aligned} P\{Y \leq 0 | X < 2\} &= \frac{P\{X < 2, Y \leq 0\}}{P\{X < 2\}} \\ &= \frac{P\{X = 1, Y = -1 \text{ 或 } Y = 0\}}{P\{X = 1\}} \\ &= \frac{a + 0.2}{a + 0.4} = 0.5 \end{aligned}$$

$$(2) P\{X = 2\} = 0.6,$$

$$P\{Y = j|X = 2\} = \frac{P\{X=2, Y=j\}}{P\{X=2\}} = \begin{cases} \frac{1}{6} & j = -1; \\ \frac{1}{6} & j = 0; \\ \frac{2}{3} & j = 1. \end{cases}$$

Y	-1	0	1
$P\{Y = j X = 2\}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$

(3)

$$\begin{aligned} & P\{X + Y = 2\} \\ &= P\{X = 1, Y = 1\} + P\{X = 2, Y = 0\} \\ &= 0.3 \end{aligned}$$

$$\begin{aligned} P\{X = i | X + Y = 2\} &= \frac{P\{X + Y = 2, X = i\}}{P\{X + Y = 2\}} \\ &= \frac{P\{X = i, Y = 2 - i\}}{P\{X + Y = 2\}} \\ &= \begin{cases} \frac{2}{3} & i = 1; \\ \frac{1}{3} & i = 2. \end{cases} \end{aligned}$$

X	1	2
$P\{X = i X + Y = 2\}$	$\frac{2}{3}$	$\frac{1}{3}$

例

一射手进行射击, 击中目标概率为 $p(0 < p < 1)$, 射击直至击中目标两次为止. 设 X 表示首次击中目标所进行的射击次数, Y 表示总共射击的次数. 试求 X 和 Y 的联合分布律和条件分布律.

解: $Y = n$ 表示“第 n 次击中目标, 且前 $n - 1$ 射击中恰有一次击中”. 各次射击相互独立. $X = m$ 表示“第 m 次为首次击中目标”.

$m \leq n - 1$ 即 n 和 m 可能取值为: $n = 2, 3, \dots$;
 $m = 1, 2, \dots, n - 1$.

$$P\{X = m, Y = n\} = (1 - p)^{n-2} p^2$$

$$\begin{aligned} P\{Y = n\} &= \sum_{m=1}^{n-1} P\{X = m, Y = n\} \\ &= \sum_{m=1}^{n-1} (1 - p)^{n-2} p^2 = (n - 1) p^2 (1 - p)^{n-2} \end{aligned}$$

$$\begin{aligned}
 P\{X = m | Y = n\} &= \frac{P\{X = m, Y = n\}}{P\{Y = n\}} \\
 &= \frac{(1-p)^{n-2} p^2}{(n-1)p^2(1-p)^{n-2}} = \frac{1}{n-1}
 \end{aligned}$$

$$P\{X = m\} = \sum_{n=m+1}^{\infty} p^2(1-p)^{n-2} = p(1-p)^{m-1}.$$

$$\begin{aligned}
 P\{Y = n | X = m\} &= \frac{P\{X = m, Y = n\}}{P\{X = m\}} \\
 &= \frac{p^2(1-p)^{n-2}}{p(1-p)^{m-1}} = p(1-p)^{n-m-1}.
 \end{aligned}$$

二维连续型随机变量的条件分布

对于连续型随机变量 X, Y ,

$$P\{X = x\} = 0, \quad P\{Y = y\} = 0, \quad \forall x, y.$$

所以不能直接用条件概率引入条件分布函数.

二维连续型随机变量的条件分布

对于连续型随机变量 X, Y ,

$$P\{X = x\} = 0, \quad P\{Y = y\} = 0, \quad \forall x, y.$$

所以不能直接用条件概率引入条件分布函数.

下次课学习: 二维连续型随机变量的条件分布.