概率论与数理统计

第十六讲

条件分布与独立性（Ⅰ）

前面研究离散型随机变量 X, Y 的联合分布律时，“电游竞赛”与“昆虫产卵”两例中，联合分布均可用乘法公式来表示.

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

- **好处** 实际问题中，建模往往较方便.

  其中条件概率

$P\{Y = y_j \mid X = x_i\}$ 在此处就是条件分布律.

1. 条件分布律

一般地, 有如下定义.

设(X, Y)为离散型二维随机变量, 其分布律为

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

若对于固定的x_i, $P\{X = x_i\} > 0$, 则称条件概率

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

为在 X 为 x_i 条件下, Y 的**条件分布律**.

同理, 可定义在 Y 为y_j 条件下, X 的条件分布律.

显然，条件分布律是个分布律. 因为

$$P\{Y=y_j \mid X=x_i\} \geq 0,$$

且 $$\sum_j P\{Y=y_j \mid X=x_i\} = \frac{\sum_j p_{ij}}{p_{i\bullet}} = \frac{p_{i\bullet}}{p_{i\bullet}} = 1.$$

- 条件分布律的意义是将某变量固定时，如$X=x_i$,将原来的样本(取值)空间缩为一维点集$\{(x_i,y_1), \{(x_i,y_2), \ldots, \}$考虑其上的分布律，但因该点集原来概率之和不等于1，而等于边缘分布律$p_{i\bullet}$，故用它来规范化.

例（掷双骰子）在此例中，考虑 $P\{X=6\,|Y=2\}$及 $P\{Y\leq 3\,|\,X=9\}$.

联合分布律及边缘分布律：

$Y\backslash X$	2	3	4	5	6	7	8	9	10	11	12	
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	$\frac{6}{36}$
1	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{10}{36}$
2	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	$\frac{8}{36}$
3	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	$\frac{6}{36}$
4	0	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	0	$\frac{4}{36}$
5	0	0	0	0	0	$\frac{2}{36}$	0	0	0	0	0	$\frac{2}{36}$
	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	

解: $P\{X=6\mid Y=2\}=\dfrac{p_{6,2}}{p_{\bullet 2}}=\dfrac{\frac{2}{36}}{\frac{8}{36}}=\dfrac{1}{4}.$

$$P\{Y\le 3\mid X=9\}=\frac{p_{9,0}+p_{9,1}+p_{9,2}+p_{9,3}}{p_{9\bullet}}$$

$$=\frac{0+\frac{2}{36}+0+\frac{2}{36}}{\frac{4}{36}}=1.$$

例（昆虫产卵）在此例中，求 $P\{X=i\,|Y=j\}$.

解: $$P\{X=i\,|\,Y=j\}=\begin{cases}0, & i<j\\ \dfrac{p_{ij}}{p_{\bullet j}}=\dfrac{\dbinom{i}{j}p^j(1-p)^{i-j}\cdot e^{-\lambda}\dfrac{\lambda^i}{i!}}{e^{-\lambda p}\dfrac{(\lambda p)^j}{j!}}, & i\ge j\end{cases}$$

$$=\begin{cases}0, & i<j\\ e^{-\lambda(1-p)}\dfrac{[\lambda(1-p)]^{i-j}}{(i-j)!}, & i\ge j\end{cases}$$

表明在 $Y=j$ 的条件下，产卵数 i 与 j 的差服从 $P(\lambda(1-p))$.

如何解释？

2. 条件概率密度函数

若(X, Y)为连续型随机变量，上述的讨论不能直接套用，因为所有形如$P\{X=x, Y=y\}$的式子均为0. 为此，将条件$Y=y$放宽为

$Y \in (y-\varepsilon, y+\varepsilon]$，即考虑条件概率

$$P\{X \le x \mid y-\varepsilon < Y \le y+\varepsilon\} = \frac{P\{X \le x, y-\varepsilon < Y \le y+\varepsilon\}}{P\{y-\varepsilon < Y \le y+\varepsilon\}}$$

$$= \frac{\int_{-\infty}^{x}\int_{y-\varepsilon}^{y+\varepsilon} f(u,v)\mathrm{d}u\mathrm{d}v}{\int_{y-\varepsilon}^{y+\varepsilon} f_Y(v)\mathrm{d}v}$$

$$= \frac{2\varepsilon \int_{-\infty}^{x} f(u,\eta)\mathrm{d}u}{2\varepsilon f_Y(\xi)}.$$

令 $\varepsilon \to 0,\ \eta,\xi \to y$, 若 $f(u, v)$是连续的, 则应有

上式收敛于 $\dfrac{\int_{-\infty}^{x} f(u,y)\mathrm{d}u}{f_Y(y)}.$

- **问** 它相当于什么？ $Y=y$ 条件下X的分布函数.

于是, 下面的定义是合理的.

定义 设(X, Y)为二维随机变量，其联合密度函数为 $f(x,y)$，Y的边缘密度函数为 $f_Y(y)$，则令

$$f(x \mid y) = \begin{cases} \dfrac{f(x,y)}{f_Y(y)}, & f_Y(y) > 0 \\ 0, & \text{其他} \end{cases}$$

称为是在$Y=y$条件下，x的**条件密度**，记为$f_{X|Y}(x \mid y)$.

同理，可以定义$f_{Y|X}(y \mid x)$.

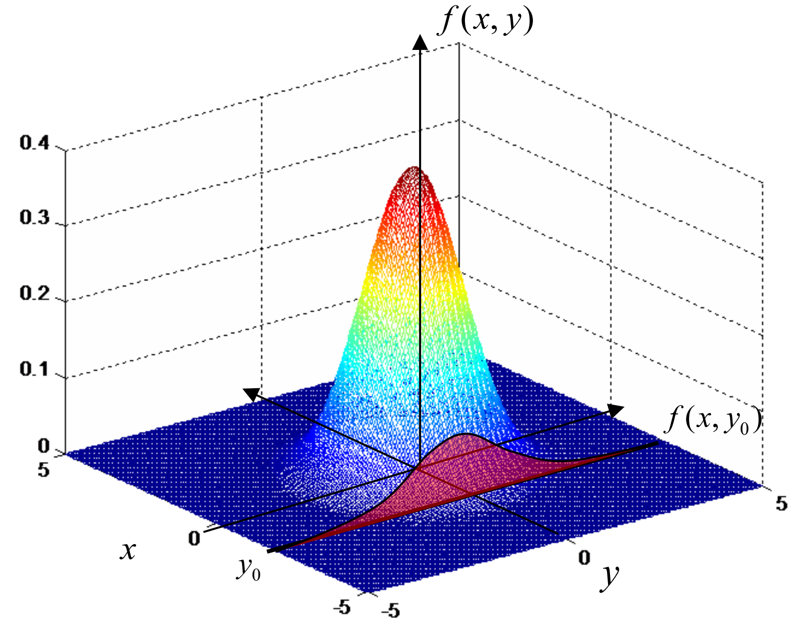
条件密度函数也是密度函数，因为

① 显然 $f_{X|Y}(x \mid y) \geq 0.$

② $$\int_{-\infty}^{\infty} f(x \mid y)dx = \int_{-\infty}^{\infty} \frac{f(x, y)\mathrm{d}x}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$$

- 它的含义是将二维分布限制在直线 $Y=y$ 上，X 的分布. 其分布关系与二维分布一致，但相差一个规范化因子 $f_Y(y)$.

条件概率密度函数的解释图示



例（上一讲用过的例子）设X, Y的联合密度函数为

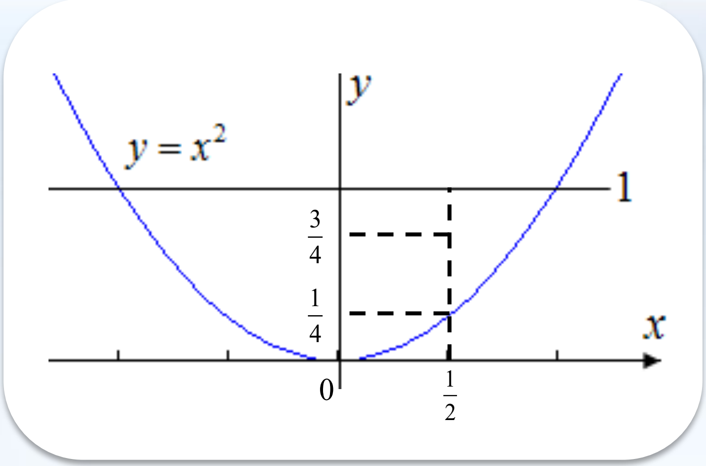
$$f(x,y)=\begin{cases}\dfrac{21}{4}x^2y, & x^2\le y\le 1\\ 0, & \text{其他}\end{cases}$$

求 $P\{Y\ge\frac{1}{4}\mid X=\frac{1}{2}\}$ 及 $P\{Y\ge\frac{3}{4}\mid X=\frac{1}{2}\}$.

解: 我们已算出 $f_X(x)=\begin{cases}\dfrac{21}{8}x^2(1-x^4), & |x|\le 1\\ 0, & \text{其他}\end{cases}$

故对于$|x|\le 1$, $\quad f_{Y|X}(y|x)=\begin{cases}\dfrac{\frac{21}{4}x^2y}{\frac{21}{8}x^2(1-x^4)}=\dfrac{2y}{1-x^4}, & x^2\le y\le 1\\ 0, & \text{其他}\end{cases}$

于是 $f_{Y|X}(y|\frac{1}{2})=\begin{cases}\frac{32}{15}y, & \frac{1}{4}\le y\le 1\\ 0, & \text{其他}\end{cases}$



故 $P\{Y\ge\frac{1}{4}\mid X=\frac{1}{2}\}=1,$

$$P\{Y\ge\frac{3}{4}\mid X=\frac{1}{2}\}=\int_{\frac{3}{4}}^{+\infty}f_{Y|X}(y|\frac{1}{2})\mathrm{d}y=\int_{\frac{3}{4}}^{1}\frac{32}{15}y\mathrm{d}y=\frac{7}{15}.$$

- 在利用条件密度进行计算时，要特别注意仔细定积分限.

注 条件密度函数只能用于条件为 $X=x\ (Y=y)$ 的情况，若条件为 $X\le x$ 等等，则直接用条件概率公式即可。如前一例中，若求 $P\{Y\ge\frac{1}{4}\,\Big|\,|X|\le\frac{1}{2}\}$，则可用下式计算

$$P\{Y\ge\frac{1}{4}\,\Big|\,|X|\le\frac{1}{2}\}=\frac{\int_{-0.5}^{0.5}\int_{\frac{1}{4}}^{1}\frac{21}{4}x^2y\mathrm{d}y\mathrm{d}x}{\int_{-0.5}^{0.5}\frac{21}{8}x^2(1-x^4)\mathrm{d}x}=\frac{105}{109}.$$