§3.3 条件分布与条件密度

对于两个事件，我们可以讨论条件概率；同样，对于两个随机变量，我们也可以讨论条件分布。

一般说来，二维随机变量（X,Y）中的两个一维随机变量 X, Y 的取值是互相影响的。因此，我们引入如下条件分布的概念

离散型

设（X,Y）是二维离散型随机变量，联合分布律为

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

且对某个y_j有 $P(Y = y_j) > 0$ ，由条件概率公式有

$$P(X = x_i \mid Y = y_j) = p_{ij} / p_{.j}.$$

由此，我们定义了如下条件分布：

定义3.7 设(X,Y)为二维离散型随机变量， 若对固定的 $j(j=1,2,\mathrm{L}\ )$，有 $P(Y=y_j)>,0$称

$$P(X = x_i \mid Y = y_j) = p_{ij} / p_{.j}, i = 1,2,\mathrm{L}$$

为在 $Y=y_j$ 条件下X 的条件分布律。

类似地，若对固定的 $i(i=1,2,\mathrm{L}\ )$有 $P(X=x_i)>0$ 称

$$P(Y = y_j \mid X = x_i) = \frac{p_{ij}}{p_{i\cdot}}, \quad j = 1,2,\mathrm{L}$$

为在 $X=x_i$ 条件下Y 的条件分布律。

由以上定义易知，条件分布律有如下性质：

1) $$0 \le P(X = x_i \mid Y = y_j) \le 1,$$

$$0 \le P(Y = y_j \mid X = x_i) \le 1;$$

2) $$\sum_{i=1}^{\infty} \frac{p_{ij}}{p_{\cdot j}} = \frac{1}{p_{\cdot j}} \sum_{i=1}^{\infty} p_{ij} = 1,$$

$$\sum_{i=1}^{\infty} \frac{p_{ij}}{p_{i\cdot}} = \frac{1}{p_{i\cdot}} \sum_{j=1}^{\infty} p_{ij} = 1.$$

注：上面两条性质表明离散型随机变量的条件分布律仍然是分布律。

离散型随机变量的条件分布函数

设(X,Y)是离散型，P(Y=y)>0,那么

$$\boxed{F_{X|Y}(x\mid y)=P(X\le x\mid Y=y),\ x\in R}$$

为Y=y条件下X的条件分布函数.

可以证明：条件分布函数满足分布函数的四个特征，因此仍然是一个分布函数，

例3.10设(X,Y)的联合分布律如下：

X\Y	1	2
1	1/6	1/6
2	1/3	0
3	1/6	1/6

(1)求Y=1时X的条件分布

(2)求条件概率 $P(Y=1|X\neq 1)$

(3)求条件分布函数 $F_{X|Y}(\frac{5}{2}|Y=1)$

X\Y	1	2	$p_{i\bullet}$
1	1/6	1/6	1/3
2	1/3	0	1/3
3	1/6	1/6	1/3
$p_{\bullet j}$	2/3	1/3	

解：$p_{\bullet 1} = P(Y=1)$

$$= \sum_i p_{i1} = 2/3$$

$$P(X=i \mid Y=1) = \frac{p_{i1}}{p_{\bullet 1}}, i = 1, 2, 3.$$

$$\boxed{X|Y=1} \sim \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$$

是个随机变量

$$(2)P(Y=1\mid X\neq 1)=\frac{P(Y=1,X\neq 1)}{P(X\neq 1)}$$

$$=\frac{P(Y=1,X=2or3)}{P(X=2or3)}=\frac{3}{4}$$

X\Y	1	2	$p_{i\bullet}$
1	1/6	1/6	1/3
2	1/3	0	1/3
3	1/6	1/6	1/3
$p_{\bullet j}$	2/3	1/3	

（**3**）由于 **X|Y=1~** $\begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$

$$F_{X|Y}(\frac{5}{2}\mid 1)=P(X\leq\frac{5}{2}\mid Y=1)$$

$$=P(X=1\mid Y=1)+P(X=2\mid Y=1)$$

$$=\frac{1}{4}+\frac{1}{2}=\frac{3}{4}$$

例3.11从1,2,3中任意取一个数,记为X,在从1到X中任意取一个数,记为Y,求Y的分布律;

解： X的分布律为： $X \sim \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$

当X=i时， Y的条件分布律为

$$P(Y = j \mid X = i) = \frac{1}{i}, j = 1, \cdots, i.$$

已知边缘分布律,条件分布律可写出联合

$$P(X = i, Y = j) = P(X = i).P(Y = j \mid X = i)$$

$$= \frac{1}{3}.\frac{1}{i}, i = 1,2,3, j = 1,2, i$$

X\Y	1	2	3
1	1/3	0	0
2	1/6	1/6	0
3	1/9	1/9	1/9

再利用边缘分布律与联合分布律的关系

$$p_{\bullet j} = \sum_i p_{ij}.$$

于是,**P(Y=2)=1/6+1/9=5/18**

连续型随机变量的条件分布

对于连续型随机变量 (X,Y) 不论 取何值均有 $P(X=x)=P(Y=y)=0$。因此，不能象离散型随机变量用条件概率来直接定义条件分布！尽管如此，我们可以用极限形式来定义连续型随机变量的条件分布。

定义：设（X, Y）是连续型随机变量，若对任意的 $h>0$ 总有 $P(y-h<Y$ 且对任意的 $x \in R$，极限 $\lim\limits_{h\to 0+} P(X$ 总存在，称此极限为在条件 y 下X 的条件分布函数，记为 y 。

$$F_{X|Y}(x\mid y)$$

同样可定义在条件X = x下Y 的条件分布 $F_{Y|X}(y\mid x)$。

进一步地，若连续型随机变量（X,Y）有联合密度 $f(x,y)$，则

$$
\begin{aligned}
F_{X|Y}(x\mid y) &= \lim_{h\to 0^+}\frac{P\{X\le x, y-h<Y\le y+h\}}{P\{y-h<Y\le y+h\}} \\
&= \lim_{h\to 0^+}\frac{F(x,y+h)-F(x,y-h)}{F_Y(y+h)-F_Y(y-h)} \\
&= \frac{\lim\limits_{h\to 0^+}[F(x,y+h)-F(x,y-h)]/2h}{\lim\limits_{h\to 0^+}[F_Y(y+h)-F_Y(y-h)]/2h} \\
&= \frac{\partial F(x,y)}{\partial y}\Bigg/\frac{d}{dy}F_Y(y) = \frac{\int_{-\infty}^{x} f(u,y)\,du}{f_Y(y)}
\end{aligned}
$$

即有$F_{X|Y}(x\mid y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du$。

同理，在条件X = x下 Y 的条件分布为

$$F_{Y|X}(y\mid x)=\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}dv。$$

注： 当y给定且$f_Y(y)>0$，容易验证关于x的

函数$\frac{f(x,y)}{f_Y(y)}$满足非负性与归一性，因此连

续型随机变量的条件分布仍然可以看作是

一连续型随机变量的分布函数。

定理：设（X,Y）的联合密度为 $f(x,y)$，则在条件Y = y下X 的条件分布函数及条件密度函数分别为

$$F_{X|Y}(x\,|\,y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du,\quad f_{X|Y}(x\,|\,y)=\frac{f(x,y)}{f_Y(y)};$$

在条件X = x下Y 的条件分布函数及条件密度函数分别为

$$F_{Y|X}(y\,|\,x)=\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}dv,\quad f_{Y|X}(y\,|\,x)=\frac{f(x,y)}{f_X(x)};$$

其中，$f_X(x)$，$f_Y(y)$分别为X 与Y 的边缘密度函数。

注1：由以上公式可得：

$$f(x,y)=f_{X|Y}(x\,|\,y)f_Y(y)=f_{Y|X}(y\,|\,x)f_X(x),$$

这与条件概率的乘积公式非常类似。

注2：不管是条件密度函数，还是条件分布函数，都必须要求其分母不为 0，即

$$f_X(x)\neq 0 \qquad or \qquad f_Y(y)\neq 0$$

这是它们存在的最先决条件。即：只有在边缘密度函数不为 0 的条件下才存在相应的条件分布及条件密度函数。

例：设（X,Y）的联合密度函数为

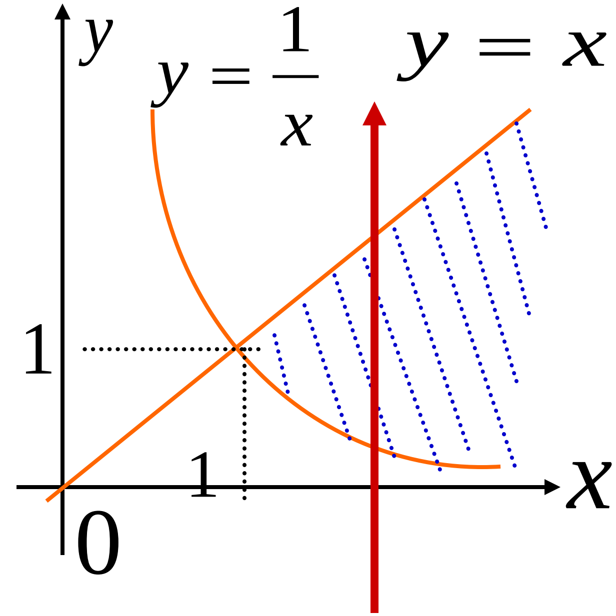
$$f(x,y)=\begin{cases}\dfrac{1}{2x^2y}, & \dfrac{1}{x}<y<x, x>0, \\ 0, & esle。\end{cases}$$

（1） 求 $f_{X|Y}(x\mid y)$，$f_{Y|X}(y\mid x)$；

（2）当事件“$X=e^2$ ”发生时，求Y的条件密度和条件分布函数及概率

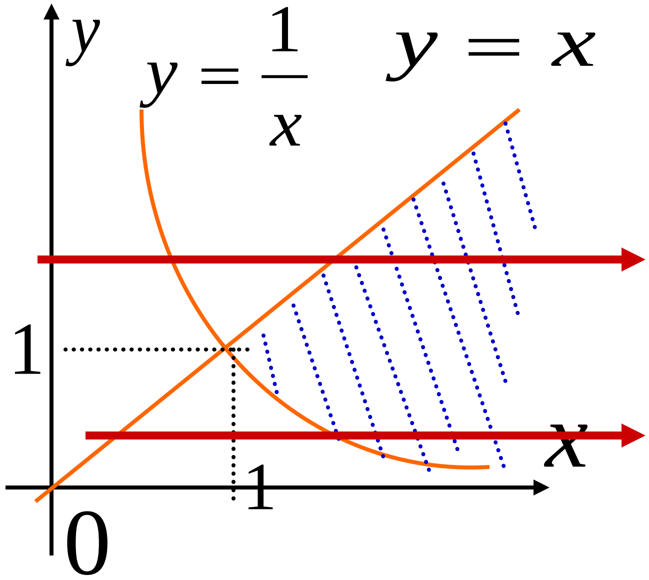
$$P(Y\le e\mid X=e^2)。$$

$$f(x,y)=\begin{cases}\dfrac{1}{2x^2y}, & \dfrac{1}{x}<y<x, x>0,\\ 0, & esle。\end{cases}$$



解1)：由边缘密度函数的定义有

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\displaystyle\int_{\frac{1}{x}}^{x}\frac{1}{2x^2y}dy=\frac{\ln x}{x^2}, & x>1,\\ 0, & x\le 1。\end{cases}$$



$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\begin{cases} 0, & y\le 0, \\ \int_{1/y}^{+\infty}\frac{1}{2x^2y}dx=\frac{1}{2}, & 0<y\le 1, \\ \int_{y}^{+\infty}\frac{1}{2x^2y}=\frac{1}{2y^2}, & 1<y。\end{cases}$$

于是，条件密度为：

1）当 $x>1$ 时，

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\begin{cases} \dfrac{1}{2x^2y}\Big/\dfrac{\ln x}{x^2}=\dfrac{1}{2y\ln x}, & \dfrac{1}{x}<y<x, \\ 0, & else。\end{cases}$$

当 $0<y<1$ 时，

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\frac{1}{2}}=2f(x,y)=\begin{cases}\frac{1}{x^2y}, & x>\frac{1}{y},\\ 0, & else。\end{cases}$$

当 $y>1$ 时，

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\frac{1}{2y^2}}=\begin{cases}\frac{y}{x^2}, & x>y,\\ 0, & else。\end{cases}$$

2）直接由（1）的结论知，在条件$X=e^2$ 下Y 的条件密度为

$$f_{Y|X}(y|e^2)=\begin{cases}\frac{1}{2y\ln e^2}=\frac{1}{4y}, & e^{-2}<y<e^2\\ 0, & else。\end{cases},$$

在条件 $X=e^2$ 下Y 的条件分布函数为

$$F_{Y|X}(y|e^2)=\int_{-\infty}^{y} f_{Y|X}(v|e^2)dv=\begin{cases} 0, & y\le e^{-2}, \\ \frac{1}{4}(\ln y+2), & e^{-2}<y<e^2, \\ 1, & e^2<y。\end{cases}$$

于是有

$$P(Y\le e\mid X=e^2)=\int_{-\infty}^{e} f_{Y|X}(y|e^2)dy=\int_{e^{-2}}^{e}\frac{1}{4y}dy=\frac{3}{4}。$$

(又解：$P(Y\le e\mid X=e^2)=F_{Y|X}(e|e^2)=\frac{1}{4}(\ln e+2)=\frac{3}{4}$。)

例：设随机变量 X 在[0，1]上均匀分布，且对每个 $x \in (0,1)$，当 $X = x$ 时有 $Y: e(x/2)$ ，求概率 $P(X \geq Y)$。

解：由题意知，X 的密度函数为

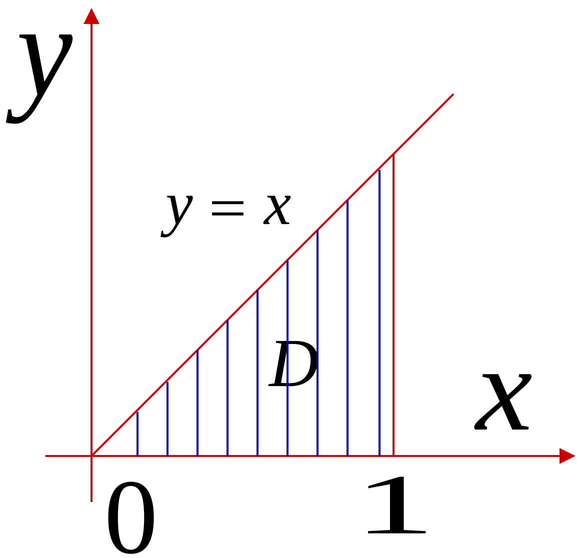
$$f_X(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & else。\end{cases}$$

且当 $x \in (0,1)$，Y 的条件密度函数为

$$f_{Y|X}(y|x) = \begin{cases} \frac{x}{2} e^{-\frac{1}{2}xy}, & y > 0, \\ 0, & else。\end{cases}$$

则（X, Y）的联合密度函数为

$$f(x,y)=f_X(x)f_{Y|X}(y|x)=\begin{cases}\dfrac{x}{2}e^{-\frac{1}{2}xy}, & x\in[0,1], y>0,\\ 0, & else。\end{cases}$$



从而，所求的概率为

$$P(X\geq Y)=\iint_D f(x,y)dxdy$$

$$=\int_0^1 dx\int_0^x \frac{x}{2}e^{-\frac{1}{2}xy}dy=1-\int_0^1 e^{-\frac{1}{2}x^2}dx$$

$$=1-\sqrt{2\pi}\int_0^1\frac{1}{\sqrt{2\pi}\times 1}e^{-\frac{x^2}{2}}dx=1-\sqrt{2\pi}(\Phi(1)-\Phi(0))\approx 0.1445。$$