条件分布与随机变量的独立性

教学设计与制作：胥斌

引例1 学生成绩（数学X_1,物理X_2,外语X_3,体育X_4)

数学成绩好 ⇒ 很可能物理成绩好
可能外语成绩好
体育成绩 ？

引例2 财富指标(所得税率X_1,居住环境X_2,代步工具X_3)

豪车代步，很可能税率高，居住环境好

无自有住房，很可能税率低，无车代步

一般，要了解随机变量X与Y之间的关系，则需要研究(X,Y)关于X和Y的条件分布。

一．离散型(X,Y)的条件分布律

已知离散型(X,Y)的联合分布律p_{ij}及X的边缘分布律$p_{i\cdot}$，

则称

$$P\{Y=b_j \mid X=a_i\}=\frac{P\{X=a_i,Y=b_j\}}{P\{X=a_i\}}=\frac{p_{ij}}{p_{i\cdot}}$$

$$j=1,2,\cdots$$

为在$X=a_i$的条件下，Y的条件分布律

同理称

$$P\{X=a_i \mid Y=b_j\}=\frac{P\{X=a_i,Y=b_j\}}{P\{Y=b_j\}}=\frac{p_{ij}}{p_{\cdot j}}$$

$$i=1,2,\cdots$$

为在 $Y=b_j$ 的条件下，X 的条件分布律

例1 设甲、乙两人各进行两次射击，他们每次的命中率分别为0.8和0.6。甲先射击，且甲全部命中时乙的命中率下降10%，甲全部未命中时乙的命中率上升20%，甲命中1次时乙不受影响。令X，Y分别表示甲、乙的命中次数，

试分别求$X=0,1,2$时Y的条件分布律。

(X,Y)的联合分布律及边缘分布律如下表:

p_{ij} X \ Y	0	1	2	X
0	0.04 0.0784	0.04 0.4032	0.04 0.5184	0.04
1	0.32 0.16	0.32 0.48	0.32 0.36	0.32
2	0.64 0.2116	0.64 0.4968	0.64 0.2916	0.64
Y	0.18976	0.48768	0.32256	

代公式可得, $P\{Y=0\mid X=0\}=\dfrac{p_{00}}{p_{0\cdot}}=\dfrac{0.04\cdot 0.0784}{0.04}=0.0784$

$P\{Y=1\mid X=0\}=0.4032 \qquad P\{Y=2\mid X=0\}=0.5184$

则 $X=0$ 时**Y 的条件分布律**为

$Y\|X=0$	0	1	2
P	0.0784	0.4032	0.5184

Y的条件分布律**反映了**选手甲的成绩对选手乙的发挥有影响

$\boxed{0.5184>0.36}$

同理可得

$Y\|X=1$	0	1	2
P	0.16	0.48	0.36

$Y\|X=2$	0	1	2
P	0.2116	0.4968	0.2916

$\boxed{0.2916<0.36}$

反过来，**X的条件分布律为**

$X\mid Y=0$	0	1	2
P	0.0165	0.2698	0.7137

$X\mid Y=1$	0	1	2
P	0.033	0.315	0.652

$X\mid Y=2$	0	1	2
P	0.0643	0.3571	0.5786

X的条件分布律**反映了**由乙的成绩可以反观甲的发挥：
乙的成绩差则甲的发挥较好；
乙的成绩好则甲的发挥较差。

若比赛中，甲乙两选手**独立射击**，条件分布律如何呢?

$$p_{ij}=P\{X=i\}\cdot P\{Y=j\,|\,X=i\}=P(X=i)\cdot P(Y=j)$$

$$=C_2^i 0.8^i 0.2^{2-i}\cdot C_2^j 0.6^j 0.4^{2-j}$$

独立射击

联合分布律

X \ p_{ij} \ Y	0	1	2	X
0	0.04 0.16	0.04 0.48	0.04 0.36	0.04
1	0.32 0.16	0.32 0.48	0.32 0.36	0.32
2	0.64 0.16	0.64 0.48	0.64 0.36	0.64
Y	0.16	0.48	0.36	

X 的条件分布律

$X\|Y=0$	0	1	2
P	0.04	0.32	0.64

$X\|Y=1$	0	1	2
P	0.04	0.32	0.64

$X\|Y=2$	0	1	2
P	0.04	0.32	0.64

Y 的条件分布律

$Y\|X=0$	0	1	2
P	0.16	0.48	0.36

$Y\|X=1$	0	1	2
P	0.16	0.48	0.36

$Y\|X=2$	0	1	2
P	0.16	0.48	0.36

注：X,Y的条件分布律变成了X,Y的（无条件）分布律；

X与Y相互“不影响”，它们独立。

一般的，若离散型随机变量(X,Y)的联合分布律与边缘分布律满足

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \qquad i, j = 1, 2\cdots$$

则，随机变量**X与Y相互独立**。

若X与Y相互独立，则由X与Y的边缘分布律可以确定(X,Y)的联合分布律

二. 条件分布函数与条件密度

设已知连续型 (X,Y) 的联合分布函数 $F(x,y)$，联合密度函数 $f(x,y)$，边缘密度函数 $f_X(x)$ 及 $f_Y(y)$，对 $\forall \varepsilon > 0$，固定 y，考虑条件概率，

$$P(X \le x \mid Y = y) = P(X \le x \mid y - \varepsilon < Y \le y)$$

$$= \frac{F(x,y) - F(x,y-\varepsilon)}{F_Y(y) - F_Y(y-\varepsilon)} = \frac{P(X \le x, y-\varepsilon < Y \le y)}{P(y-\varepsilon < Y \le y)}$$

$$= \frac{[F(x,y) - F(x,y-\varepsilon)]/\varepsilon}{[F_Y(y) - F_Y(y-\varepsilon)]/\varepsilon}$$

取极限

取极限 $\lim_{\varepsilon \to 0} \dfrac{[F(x,y) - F(x, y-\varepsilon)]/\varepsilon}{[F_Y(y) - F_Y(y-\varepsilon)]/\varepsilon}$

$$= \frac{\dfrac{\partial F(x,y)}{\partial y}}{\dfrac{dF_Y(y)}{dy}} = \frac{\int_{-\infty}^{x} f(x,y)dx}{f_Y(y)}$$

$$= \boxed{\int_{-\infty}^{x} \frac{f(x,y)}{f_Y(y)} dx}$$

为 $Y = y$ 的条件下 X 的**条件分布函数**，记作 $F_{X/Y}(x|y)$

其中，$\dfrac{f(x,y)}{f_Y(y)}$ 称为 **$Y=y$ 的条件下 X 的条件密度**

记作 $f_{X|Y}(x|y)$

同理，**$X=x$ 的条件下 Y 的条件分布函数**

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

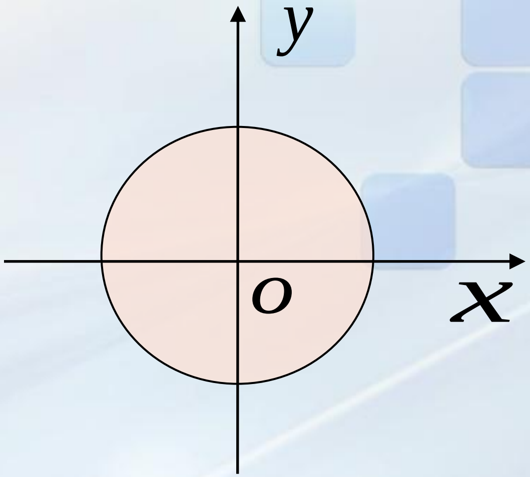
其条件密度函数为

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

例2 设(X,Y)服从单位圆上的均匀分布，概率密度为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2\le 1\\ 0, & \text{其它}\end{cases}$$

求条件密度 $f_{Y|X}(y|x)$.



解 X的边缘密度函数为

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\begin{cases}\dfrac{2}{\pi}\sqrt{1-x^2}, & |x|\le 1\\ 0, & |x|>1\end{cases}$$

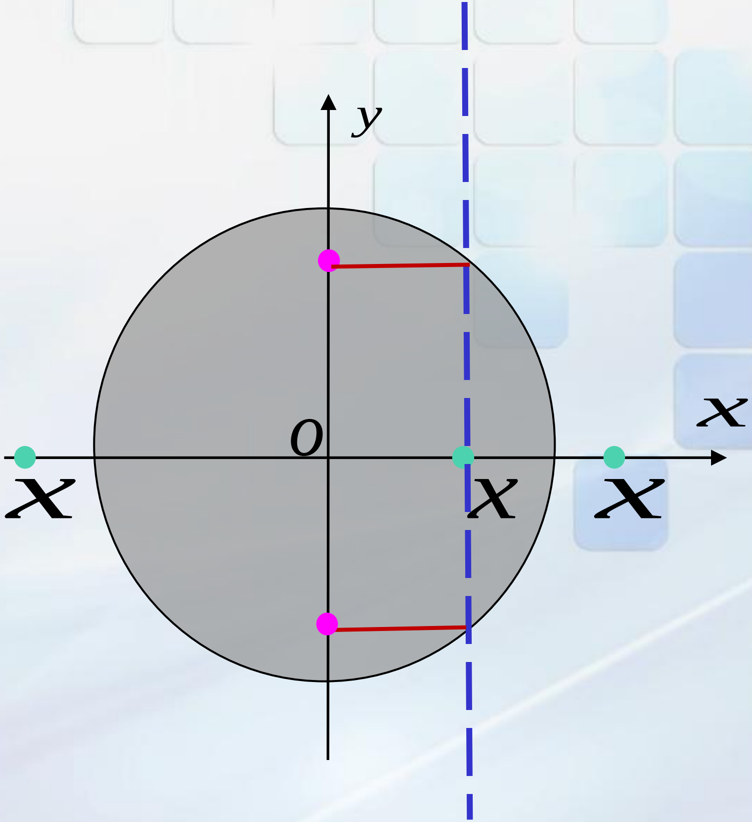
当$|x|>1$时,有 $$f_{Y|X}(y|x)=0$$

当$|x|<1$时,有

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

$$=\frac{1/\pi}{(2/\pi)\sqrt{1-x^2}}=\frac{1}{2\sqrt{1-x^2}},$$

$$-\sqrt{1-x^2}\le y\le\sqrt{1-x^2}$$



即 当 $|x|<1$ 时,有

X 作为已知变量

$$f_{Y|X}(y|x)=\begin{cases}\dfrac{1}{2\sqrt{1-x^2}}, & -\sqrt{1-x^2}\le y\le\sqrt{1-x^2}\\ 0, & y\text{ 取其它值}\end{cases}$$

X已知的条件下
Y的条件密度

这里是y的取值范围

特别的,

$$f_{Y|X}(y|x=0.5)=\begin{cases}\dfrac{1}{\sqrt{3}}, & -\sqrt{3}/2\le y\le\sqrt{3}/2\\ 0, & \text{其它}\end{cases}$$

三. 连续型(X,Y)的独立性

条件密度函数为

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

类似于乘法公式可得

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

结论 已知边缘分布和条件分布，则可确定联合分布

若X与Y“相互不影响”，**结论如何变化?**

此时，条件分布转变为无条件分布，即

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

可得，

随机变量X与Y相互独立

$$\boxed{f(x,y) = f_X(x) \cdot f_Y(y) \quad \forall (x,y) \in R^2}$$

结论 若已知X与Y相互独立，则由X与Y的边缘分布可以确定(X,Y)的联合分布

例3 设二维连续型随机变量(X,Y)的联合密度为

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

验证 X 与 Y 相互独立.

解 易求得边缘密度函数,

$$f_X(x)=\begin{cases}2x, & 0\le x\le 1\\ 0, & \text{其他}\end{cases}$$

$$f_Y(y)=\begin{cases}2y, & 0\le y\le 1\\ 0, & \text{其他}\end{cases}$$

$$f(x,y)=f_X(x)\cdot f_Y(y)$$

$$\forall (x,y)\in R^2$$

所以,X 与 Y 相互独立.

小结

- 条件分布的概念和计算
- 随机变量的独立性