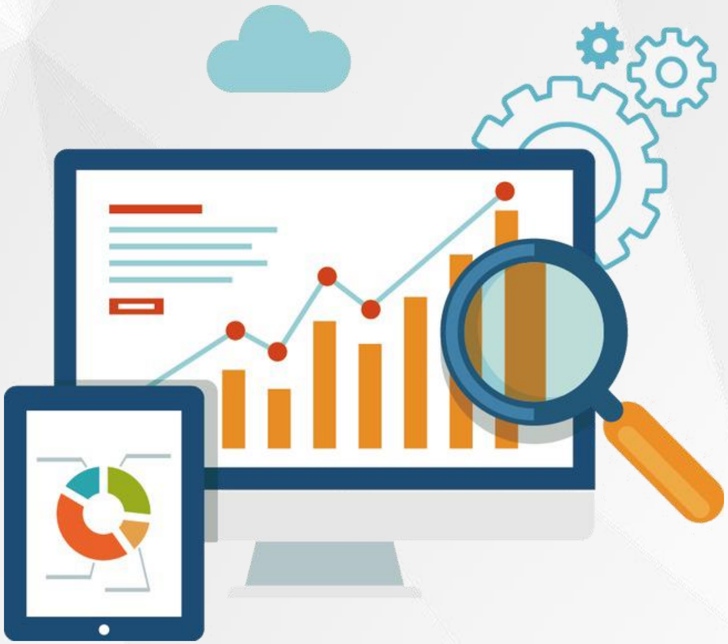
概率论与数理统计

Probability and Statistics



—概率论与数理统计教学组—

哈尔滨工程大学

第3章 多维随机变量及其分布

3.3 条件分布

学习要点

- 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

Harbin Engineering University

一、 条件分布引言

$$(X,Y) \Leftrightarrow F(x,y)$$

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

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

条件分布函数

二、二维离散型随机变量的条件分布

设(X,Y)是二维离散型随机变量，对于固定的j，若

$P\{Y=y_j\}>0$，称

$$P\{X=x_i \mid Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\bullet j}},\qquad i=1,2,\cdots$$

为$Y=y_j$条件下X的条件分布律.

X	x_1	x_2	$\cdots$	x_i	$\cdots$
$P\{X=x_i \mid Y=y_j\}$	$\frac{p_{1j}}{p_{\bullet j}}$	$\frac{p_{2j}}{p_{\bullet j}}$	$\cdots$	$\frac{p_{ij}}{p_{\bullet j}}$	$\cdots$

同理，若 $P\{X=x_i\}>0$，称

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

为 $X=x_i$ 条件下 Y 的条件分布律.

Y	y_1	y_2	$\cdots$	y_j	$\cdots$
$P\{Y=y_j \mid X=x_i\}$	$\frac{p_{i1}}{p_{i\bullet}}$	$\frac{p_{i2}}{p_{i\bullet}}$	$\cdots$	$\frac{p_{ij}}{p_{i\bullet}}$	$\cdots$

条件分布

Harbin Engineering University

注意：(1) $P\{X=x_i\,|\,Y=y_j\}\ge 0;$

(2) $\sum\limits_{i=1}^{+\infty}P\{X=x_i\,|\,Y=y_j\}=\sum\limits_{i=1}^{+\infty}\dfrac{p_{ij}}{p_{\bullet j}}=\dfrac{p_{\bullet j}}{p_{\bullet j}}=1.$

同理：(3) $P\{Y=y_j\,|\,X=x_i\}\ge 0;$

(4) $\sum\limits_{j=1}^{+\infty}P\{Y=y_j\,|\,X=x_i\}=\sum\limits_{j=1}^{+\infty}\dfrac{p_{ij}}{p_{i\bullet}}=\dfrac{p_{i\bullet}}{p_{i\bullet}}=1.$

条件分布

Harbin Engineering University

例 1 设袋中装有 2 个白球、3 个红球，现从袋中无放回地随机抽取两次，定义随机变量 X, Y 如下：

$$X=\begin{cases}1 & \text{第一次取出白球}\\ 0 & \text{第一次取出红球}\end{cases},\quad Y=\begin{cases}1 & \text{第二次取出白球}\\ 0 & \text{第二次取出红球}\end{cases}$$

(1)求 $X=0$ 条件下 Y 的条件分布律；(2)求 $Y=0$ 条件下 X 的条件分布律.

解 (1)在 $X=0$ 条件下，袋中还剩 2 个白球，2 个红球，则

$$P\{Y=0\mid X=0\}=\frac{P\{X=0,Y=0\}}{P\{X=0\}}=\frac{3/10}{3/5}=\frac{1}{2},$$

$$P\{Y=1\mid X=0\}=\frac{P\{X=0,Y=1\}}{P\{X=0\}}=\frac{3/10}{3/5}=\frac{1}{2}.$$

故所求条件分布律为：

Y	0	1
$P\{Y=y_j \mid X=0\}$	$\frac{1}{2}$	$\frac{1}{2}$

(2) **又由** $P\{X=0,Y=0\}=\frac{3}{5}\cdot\frac{2}{4}=\frac{3}{10}$， $P\{X=1,Y=0\}=\frac{2}{5}\cdot\frac{3}{4}=\frac{3}{10}$，

故 $P\{Y=0\}=\frac{3}{10}+\frac{3}{10}=\frac{3}{5}$，

$P\{X=0|Y=0\}=\frac{P\{X=0,Y=0\}}{P\{Y=0\}}=\frac{1}{2}$， $P\{X=1|Y=0\}=1-\frac{1}{2}=\frac{1}{2}$，

因此，所求的条件分布为

X	0	1
$P\{X=x_i \mid Y=0\}$	$\frac{1}{2}$	$\frac{1}{2}$

例 2 某人进行投篮练习，设其每次投篮的命中率均为 $p(0<p<1)$，投篮练习直至命中 2 次为止，以 X 表示首次命中所进行的投篮次数，以 Y 表示一共进行的投篮次数，求二维随机变量 (X,Y) 的分布律和条件分布律.

分析 每次投篮不中概率为 $q=1-p$，例如事件 $\{X=2,Y=5\}$ 表示前 5 次投篮分别为“不中，中，不中，不中，中”其概率为

$$P\{X=2,Y=5\}=q\cdot p\cdot q\cdot q\cdot p.$$

同理

$$P\{X=m,Y=n\}=q\cdots q\cdot p\cdot q\cdots q\cdot p=p^2q^{n-2},$$

$$(n=2,3,\cdots,m=1,2,\cdots,n-1)$$

即为 (X,Y) 的联合分布律，由联合分布律可求出边缘分布律，再把它们相除即得条件分布律.

例 2 某人进行投篮练习，设其每次投篮的命中率均为 $p(0<p<1)$，投篮练习直至命中 2 次为止，以 X 表示首次命中所进行的投篮次数，以 Y 表示一共进行的投篮次数，求二维随机变量 (X,Y) 的分布律和条件分布律.

解 **方法 1** (X,Y) 的联合分布律为 $P\{X=m,Y=n\}=p^2q^{n-2}$,

其中 $q=1-p$, $\quad n=2,3,\cdots;\quad m=1,2,\cdots,n-1.$

X 的边缘分布律为：

$$P\{X=m\}=\sum_{n=m+1}^{+\infty}P\{X=m,Y=n\}=\sum_{n=m+1}^{+\infty}p^2q^{n-2}$$

$$=p^2\frac{q^{m-1}}{1-q}=p^2\frac{q^{m-1}}{p}=pq^{m-1},\quad (m=1,2,\cdots)$$

例 2 某人进行投篮练习，设其每次投篮的命中率均为 $p(0<p<1)$，投篮练习直至命中 2 次为止，以 X 表示首次命中所进行的投篮次数，以 Y 表示一共进行的投篮次数，求二维随机变量 (X,Y) 的分布律和条件分布律.

解　方法 1 (X,Y) 的联合分布律为 $P\{X=m,Y=n\}=p^2q^{n-2}$,

其中 $q=1-p$,　$n=2,3,\cdots;$　$m=1,2,\cdots,n-1.$

Y 的边缘分布律为:

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

$$=(n-1)p^2q^{n-2},\quad (n=2,3,\cdots)$$

例 2 某人进行投篮练习，设其每次投篮的命中率均为 $p(0<p<1)$，投篮练习直至命中 2 次为止，以 X 表示首次命中所进行的投篮次数，以 Y 表示一共进行的投篮次数，求二维随机变量 (X,Y) 的分布律和条件分布律.

解 **方法 1** (X,Y) 的联合分布律为 $P\{X=m,Y=n\}=p^2q^{n-2}$,

其中 $q=1-p$, $\quad n=2,3,\cdots;\quad m=1,2,\cdots,n-1.$

X 的边缘分布律为：$P\{X=m\}=pq^{m-1},(m=1,2,\cdots)$,

Y 的边缘分布律为：$P\{Y=n\}=(n-1)p^2q^{n-2},(n=2,3,\cdots)$,

给定 $n(n=2,3,\cdots)$，在 $Y=n$ 的条件下，X 的条件分布律为：

$$P\{X=m\mid Y=n\}=\frac{p^2q^{n-2}}{(n-1)p^2q^{n-2}}=\frac{1}{n-1},(m=1,2,\cdots,n-1).$$

例 2 某人进行投篮练习，设其每次投篮的命中率均为 $p(0<p<1)$，投篮练习直至命中 2 次为止，以 X 表示首次命中所进行的投篮次数，以 Y 表示一共进行的投篮次数，求二维随机变量 (X,Y) 的分布律和条件分布律.

解 方法 1 (X,Y) 的联合分布律为 $P\{X=m,Y=n\}=p^2q^{n-2}$,

其中 $q=1-p$, $\quad n=2,3,\cdots;\quad m=1,2,\cdots,n-1.$

X 的边缘分布律为：$P\{X=m\}=pq^{m-1},(m=1,2,\cdots)$,

Y 的边缘分布律为：$P\{Y=n\}=(n-1)p^2q^{n-2},(n=2,3,\cdots)$,

给定 $m(m=1,2,\cdots)$，在 $X=m$ 条件下，Y 的条件分布律为：

$$P\{Y=n\mid X=m\}=\frac{p^2q^{n-2}}{pq^{m-1}}=pq^{n-m-1},(n=m+1,m+2,\cdots).$$

例 2 某人进行投篮练习，设其每次投篮的命中率均为 $p(0<p<1)$，投篮练习直至命中 2 次为止，以 X 表示首次命中所进行的投篮次数，以 Y 表示一共进行的投篮次数，求二维随机变量 (X,Y) 的分布律和条件分布律.

解 方法 2 (X,Y) 的联合分布律为 $P\{X=m,Y=n\}=p^2q^{n-2}$，

其中 $q=1-p, n=2,3,\cdots; m=1,2,\cdots,n-1.$ 此时不求边缘分布律.

在 $Y=n$ 条件下，X 的所有取值可能为 $1,2,\cdots,n-1$，且 X 取以上每个值的可能是相同的，因此

$$P\{X=m\mid Y=n\}=\frac{1}{n-1},\ (m=1,2,\cdots,n-1);$$

例 2 某人进行投篮练习，设其每次投篮的命中率均为 $p(0<p<1)$，投篮练习直至命中 2 次为止，以 X 表示首次命中所进行的投篮次数，以 Y 表示一共进行的投篮次数，求二维随机变量 (X,Y) 的分布律和条件分布律.

解 方法 2 (X,Y) 的联合分布律为 $P\{X=m,Y=n\}=p^2q^{n-2}$，

其中 $q=1-p, n=2,3,\cdots; m=1,2,\cdots,n-1$. 此时不求边缘分布律.

以 $\{Y=5\mid X=2\}$ 为例，其表示在 $X=2$ 条件下 $Y=5$，即前两次投篮为“不中，中”条件下，第 3，4，5 次投篮为“不中，不中，中”的概率为 $P\{Y=5\mid X=2\}=q\cdot q\cdot p$. 同理可得

$$P\{Y=n\mid X=m\}=q\cdot q\cdots q\cdot p=pq^{n-m-1},\ (n=m+1,m+2,\cdots).$$

三、二维连续型随机变量的条件分布

设(X,Y)为二维连续型随机变量，对于给定的$x\in\mathrm{R}$及$\forall\varepsilon>0$，有$P\{y<Y\le y+\varepsilon\}>0$，若极限

$$\lim_{\varepsilon\to 0^+} P\left\{X\le x \mid y<Y\le y+\varepsilon\right\}=\lim_{\varepsilon\to 0^+}\frac{P\left\{X\le x,\ \ y<Y\le y+\varepsilon\right\}}{P\left\{y<Y\le y+\varepsilon\right\}}$$

存在，则称此极限为**在条件$Y=y$下X的条件分布函数.**

记为 $F_{X|Y}(x,y)$

类似地，可定义**在条件$X=x$下Y的条件分布函数**，记为$F_{Y|X}(y\mid x)$.

条件分布

设(X,Y)为二维连续型随机变量，分布函数为$F(x,y)$，概率密度函数为$f(x,y)$且连续，则

$$F_{X|Y}(x\mid y)=\lim_{\varepsilon\to0^+}\frac{P\{X\le x,\ y<Y\le y+\varepsilon\}}{P\{y<Y\le y+\varepsilon\}}$$

$$=\lim_{\varepsilon\to0^+}\frac{\int_{-\infty}^{x}\left[\int_{y}^{y+\varepsilon}f(x,y)dy\right]dx}{\int_{y}^{y+\varepsilon}f_Y(y)dy}$$ (下步用积分中值定理)

$$=\lim_{\varepsilon\to0^+}\frac{\varepsilon\int_{-\infty}^{x}f(x,\xi)dx}{\varepsilon f_Y(\eta)}=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}dx$$

在$Y=y$条件下，X的条件概率密度为 $f_{X|Y}(x\mid y)=\dfrac{f(x,y)}{f_Y(y)}$

在$X=x$条件下，Y的条件概率密度为 $f_{Y|X}(y\mid x)=\dfrac{f(x,y)}{f_X(x)}$

例 3 设(X,Y)的概率密度为 $f(x,y)=\begin{cases} xe^{-x(1+y)}, & x>0, y>0, \\ 0, & \text{其他}. \end{cases}$

(1) 求 $f_{Y|X}(y|x)$; (2) 求 $P(Y>1 | X=3)$.

解 (1) 由题意可知

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\begin{cases} \int_0^{+\infty} xe^{-x(1+y)}dy, & x>0, \\ 0, & x\le 0, \end{cases}=\begin{cases} e^{-x}, & x>0, \\ 0, & x\le 0. \end{cases}$$

例 3 设(X,Y)的概率密度为 $f(x,y)=\begin{cases} xe^{-x(1+y)}, & x>0, y>0, \\ 0, & \text{其他}. \end{cases}$

(1) 求 $f_{Y|X}(y|x)$;　　(2) 求 $P(Y>1|X=3)$.

解 (1) $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\begin{cases} e^{-x}, & x>0, \\ 0, & x\le 0. \end{cases}$

因此当 $x>0$ 时，有

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\begin{cases} \dfrac{xe^{-x(1+y)}}{e^{-x}}, & y>0 \\ 0, & y\le 0 \end{cases}=\begin{cases} xe^{-xy}, & y>0 \\ 0, & y\le 0 \end{cases}.$$

(2) 当 $X=3$ 时，有

$$P\{Y>1|X=3\}=\int_1^{+\infty} f_{Y|X}(y|3)dy=\int_1^{+\infty} 3e^{-3y}dy=e^{-3}.$$

二维均匀分布

设G是一个平面区域，A为区域G的面积，若二维连续型随机变量(X,Y)具有概率密度

$$f(x,y)=\begin{cases}\dfrac{1}{A}, & (x,y)\in G,\\ 0, & \text{其他}.\end{cases}$$

则称(X,Y)在G上服从**均匀分布.**

注意： 若(X,Y)在G上服从均匀分布，则(X,Y)落在G的任何一个子区域的概率只与子区域的面积成正比，与子区间的形状、位置无关。

例 4 设二维随机变量(X,Y)在圆域$x^2+y^2\le 1$上服从均匀分布，当$-1<y<1$时，求条件概率密度$f_{X|Y}(x|y)$.

解 由题意 (X,Y)的概率密度为

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

则Y的边缘概率密度为

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\begin{cases}\displaystyle\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{1}{\pi}dx=\frac{2}{\pi}\sqrt{1-y^2}, & -1<y<1,\\ 0, & \text{其他}.\end{cases}$$

于是当$-1<y<1$时有

$$f_{X|Y}(x|y)=\begin{cases}\dfrac{f(x,y)}{f_Y(y)}=\dfrac{1/\pi}{(2/\pi)\sqrt{1-y^2}}=\dfrac{1}{2\sqrt{1-y^2}}, & -\sqrt{1-y^2}\le x\le\sqrt{1-y^2},\\ 0, & \text{其他}.\end{cases}$$

例 5 设数 X 在区间$(0,1)$上随机取值，当观察到 $X = x\ (0 < x < 1)$时，数Y 在$(x,1)$上随机取值，求Y 的概率密度 $f_Y(y)$.

解 由题意 $f_X(x) = \begin{cases} \boxed{1, \quad 0 < x < 1,} \\ 0, \quad \text{其他.} \end{cases}$

对任意$0 < x < 1$，$f_{Y|X}(y|x) = \begin{cases} \boxed{\dfrac{1}{1-x}, \quad x < y < 1,} \\ 0, \quad \text{其他.} \end{cases}$

当$0 < x < 1$时:

$$f(x,y) = f_X(x) f_{Y|X}(y|x) = \begin{cases} \boxed{\dfrac{1}{1-x}, \quad 0 < x < y < 1,} \\ 0, \quad \text{其他.} \end{cases}$$

当$x \le 0$或$x \ge 1$时: $f(x,y) = 0$.

例 5 设数 X 在区间$(0,1)$上随机取值，当观察到 $X=x\ (0<x<1)$时，数Y 在$(x,1)$上随机取值，求Y 的概率密度 $f_Y(y)$.

解 由题意 $f_X(x)=\begin{cases}1, & 0<x<1,\\ 0, & \text{其他}.\end{cases}$

对任意$0<x<1$， $f_{Y|X}(y\mid x)=\begin{cases}\dfrac{1}{1-x}, & x<y<1,\\ 0, & \text{其他}.\end{cases}$

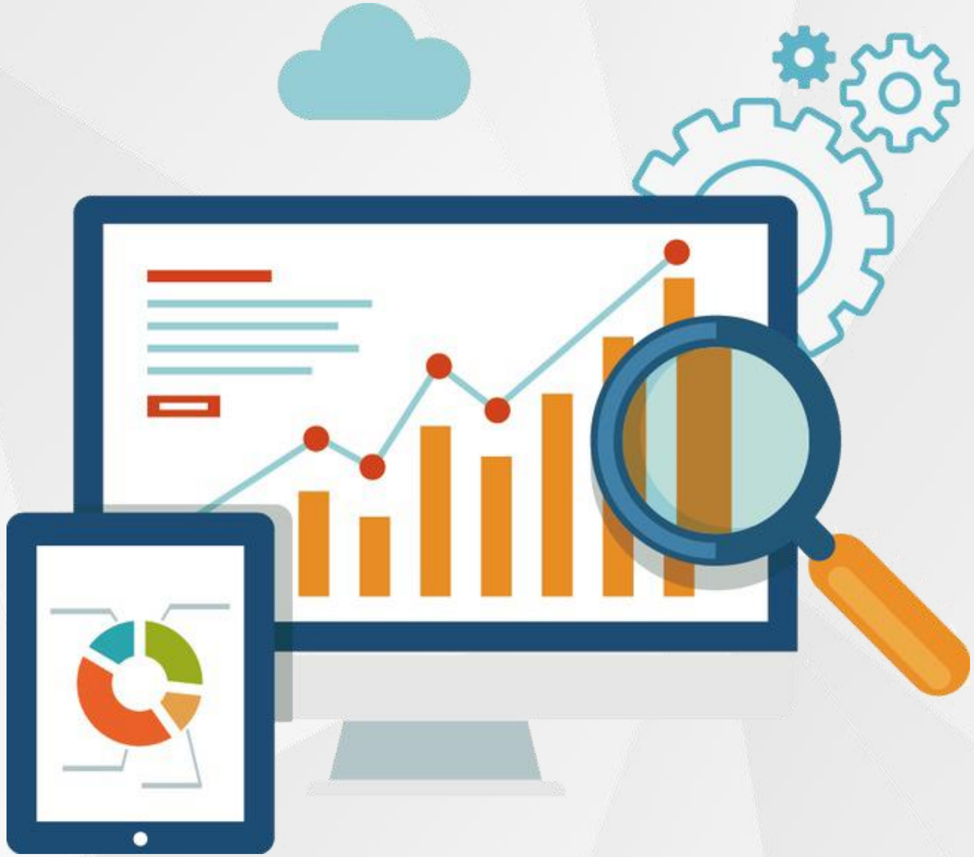
综上，(X,Y)的联合概率密度 $f(x,y)=\begin{cases}\dfrac{1}{1-x}, & 0<x<y<1,\\ 0, & \text{其他}.\end{cases}$

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\begin{cases}\displaystyle\int_0^y \frac{1}{1-x}\mathrm{d}x=-\ln(1-y), & 0<y<1,\\ 0, & \text{其他}.\end{cases}$$

哈尔滨工程大学
Harbin Engineering University

小结 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



谢谢您的观看

HARBIN ENGINEERING UNIVERSITY