3.5 条件分布

之前定义了条件概率，两事件A、B，若$P(A)>0$，则可考虑在A发生前提下B发生的概率：

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

对二维随机变量，也可类似分析

二维离散型随机变量的条件分布

(X, Y)联合分布律为 $P(X=x_i, Y=y_j)=p_{ij},\ i,j=1,2,\cdots$

边缘分布律为

$$P(X=x_i)=p_{i\bullet}=\sum_j p_{ij}$$

$$P(Y=y_j)=p_{\bullet j}=\sum_i p_{ij}$$

定义

若 $P(Y=y_j)=p_{\bullet j}>0$

考虑条件概率 $P(X=x_i|Y=y_j),\ i,j=1,2,\cdots$

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

称为$Y=y_j$条件下，随机变量X的条件分布律

同理，若 $P(X=x_i)=p_{i\bullet}>0$

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

称为$X=x_i$条件下，随机变量Y的条件分布律

例 盒子里装有3只黑球，4只红球，3只白球，在其中任取2球，以X表示取到黑球的数目，Y表示取到红球的只数。

求：(i) X、Y的联合分布律；(ii) X=1时Y的条件分布律；

(iii) Y=0时X的条件分布律。

解 (i) X、Y的联合分布律为

$X \backslash Y$	0	1	2
0	1/15	4/15	2/15
1	3/15	4/15	0
2	1/15	0	0

(ii) 由于 $P(X=1)=7/15$

故在X=1时Y的条件分布

$$P(Y=0|X=1)=3/7$$
$$P(Y=1|X=1)=4/7$$
$$P(Y=2|X=1)=0$$

Y	0	1	2
$P(Y=k \mid X=1)$	4/7	3/7	0

(iii)

X	0	1	2
$P(X=k \mid Y=0)$	1/5	3/5	1/5

例 一射手进行射击，击中目标的概率为 p ($0<p<1$)，射击直至击中目标两次为止，设以 X 表示首次击中目标所进行的射击次数，以 Y 表示总共进行的射击次数，试求 X 和 Y 的联合分布律和条件分布律。

要点：第m次击中第一次，第n次击中第二次，其他n-2次均未击中

解 (X, Y)的联合分布律为 $P(X=m, Y=n)=p^2q^{n-2},\ q=1-p,\quad n=2, 3, \cdots,\ \ m=1, 2, \cdots n-1$

X的边缘分布律为 $P(X=m)=\sum\limits_{n=m+1}^{\infty}P(X=m, Y=n)=\sum\limits_{n=m+1}^{\infty}p^2q^{n-2}=pq^{m-1},\quad m=1, 2, \cdots$

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

于是，对于每一 $(n=2, 3, \cdots), P(Y=n)>0$　在Y=n 条件下，X的条件分布律为

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

对于每一 $m(m=1, 2, \cdots), P(X=m)>0$　在X=m 条件下，Y的条件分布律为

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

二维连续型随机变量的条件分布

对任意x和y，均有 $P(X=x)=0$、$P(Y=y)=0$ 无法定义“条件分布函数”？ 不能用下式

$$F_{X|Y}(x|y)=P(X\le x|Y=y)=\frac{P(X\le x,Y=y)}{P(Y=y)}$$

虽然$P(Y=y)=0$，但可设$\varepsilon>0$，对于任意x，考虑条件概率 $P(X\le x|y<Y\le y+\varepsilon)$

设 $P(y<Y\le y+\varepsilon)>0$ 则 $P(X\le x|y<Y\le y+\varepsilon)=\dfrac{P(X\le x,y<Y\le y+\varepsilon)}{P(y<Y\le y+\varepsilon)}=\int_{-\infty}^{x}\left[\int_{y}^{y+\varepsilon}f(t,v)dv\right]dt\Big/\int_{y}^{y+\varepsilon}f_Y(v)dv$

当 $\varepsilon>0$ 时，上式为 $F_{X|Y}(x|y)=P(X\le x|y<Y\le y+\varepsilon)=\int_{-\infty}^{x}\left[\int_{y}^{y+\varepsilon}f(t,v)dv\right]dt\Big/\int_{y}^{y+\varepsilon}f_Y(v)dv$

$$\approx\frac{\varepsilon\int_{-\infty}^{x}f(t,y)dt}{\varepsilon f_Y(y)}=\int_{-\infty}^{x}\frac{f(t,y)}{f_Y(y)}dt$$

对比一维随机变量的概率密度 $F(x)=\int_{-\infty}^{x}f(t)dt$ 可见 $\dfrac{f(x,y)}{f_Y(y)}$ 是二维随机变量的条件概率密度

由定义 $f_{X|Y}(x|y)=\dfrac{\partial}{\partial x}F_{X|Y}(x|y)\Rightarrow f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$

事实上 $\dfrac{\partial}{\partial x}F_{X|Y}(x|y)=\dfrac{\partial}{\partial x}\lim\limits_{\varepsilon\to 0^+}\dfrac{P(X\le x, y<Y\le y+\varepsilon)}{P(y<Y\le y+\varepsilon)}$

$$=\frac{\partial}{\partial x}\lim_{\varepsilon\to 0^+}\frac{F(x,y+\varepsilon)-F(x,y)}{F_Y(y+\varepsilon)-F_Y(y)}$$

$$=\frac{\partial}{\partial x}\lim_{\varepsilon\to 0^+}\frac{\dfrac{F(x,y+\varepsilon)-F(x,y)}{\varepsilon}}{\dfrac{F_Y(y+\varepsilon)-F_Y(y)}{\varepsilon}}=\frac{\partial}{\partial x}\frac{\dfrac{\partial F(x,y)}{\partial y}}{\dfrac{dF_Y(y)}{dy}}$$

$$=\frac{\dfrac{\partial^2 F(x,y)}{\partial x\partial y}}{f_Y(y)}=\frac{f(x,y)}{f_Y(y)}$$

定义

若$(X$、$Y)$ 联合概率密度为$f(x, y)$，$(X$、$Y)$关于Y的边缘概率密度为$f_Y(y)$，若对于固定y，$f_Y(y)>0$，则称

$\dfrac{f(x,y)}{f_Y(y)}$ 为$Y=y$ 条件下随机变量X 的**条件概率密度**

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

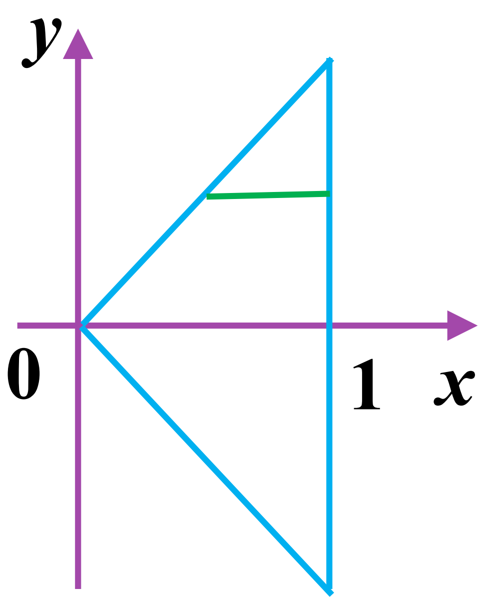
函数 $\int_{-\infty}^{x} f_{X|Y}(t|y)dt=\int_{-\infty}^{x}\dfrac{f(t,y)}{f_Y(y)}dt$

为$Y=y$ 条件下随机变量X 的**条件分布函数**，记为 $F_{X|Y}(x|y)=P(X\le x|Y=y)$

例 设二维随机变量(X,Y)在区域$\{(x,y): |y|<x<1\}$内均匀分布，求条件概率密度 $f_{X|Y}(x|y)$ 和 $P(X>\frac{2}{3}|Y=\frac{1}{2})$

解 据题意，(X, Y) 的联合概率密度为 $f(x,y)=\begin{cases}1, & |y|<x<1\\ 0, & \text{其他}\end{cases}$

Y 的边缘概率密度为 $f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\begin{cases}\int_{|y|}^{1} dx=1-|y|, & -1<y<1\\ 0, & \text{其他}\end{cases}$



于是给定 $y(-1<y<1)$，X 的条件概率密度为 $f_{X|Y}(x|y)=\begin{cases}\dfrac{1}{1-|y|}, & |y|<x<1\\ 0, & \text{其他}\end{cases}$

$$P(X>\frac{2}{3}|Y=\frac{1}{2})=\int_{2/3}^{+\infty} f_{X|Y}(x|\frac{1}{2})dx=\int_{2/3}^{1} 2dx=\frac{2}{3}$$

二维均匀分布的条件分布仍为均匀分布

例 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在区间$(x, 1)$上随机取值，求Y的概率密度$f_Y(y)$

解 设数为求Y的概率密度，就要先求(X, Y)的联合概率密度；

而根据X的边缘概率密度和Y在X给定下的条件概率密度，即可求得求(X, Y)的联合概率密度

X的边缘概率密度是 $f_X(x)=\begin{cases}1 & 0<x<1\\ 0 & \text{其他}\end{cases}$

对于任意x $(0<x<1)$时，在$X=x$条件下，Y条件概率密度为 $f_{Y|X}(y|x)=\begin{cases}\dfrac{1}{1-x} & x<y<1\\ 0 & \text{其他}\end{cases}$

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

所以Y的边缘概率密度为 $f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\displaystyle\int_0^y\frac{1}{1-x}dx=-\ln(1-y) & 0<y<1\\ 0 & \text{其他}\end{cases}$

联合分布、边缘分布和条件分布小结

二维随机变量的分布函数

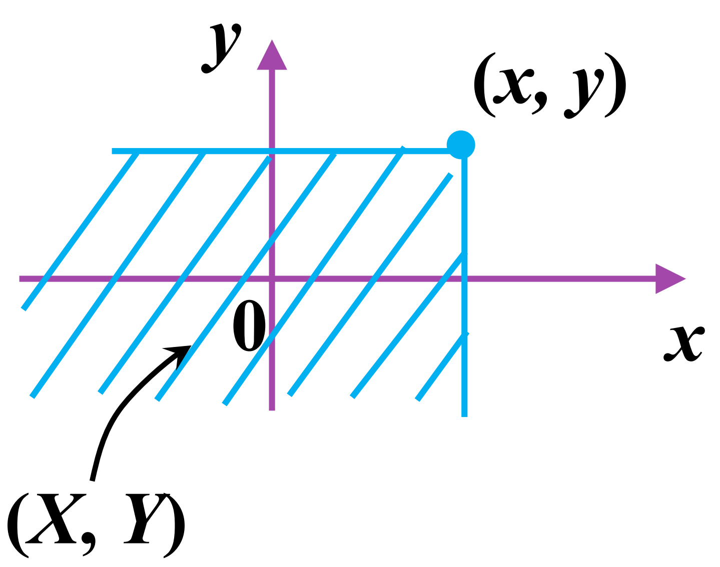
联合分布函数 $F(x,y)=P\{(X\le x)\cap(Y\le y)\}\overset{\text{记成}}{=\!=}P(X\le x, Y\le y)$ $-\infty<x<+\infty,\ -\infty<y<+\infty$

$F(-\infty,-\infty)=0,\ F(+\infty,+\infty)=1$

对任意固定y, $F(-\infty, y)=0$；对任意固定x, $F(x,-\infty)=0$

$F(x, y)$关于x, y是不减函数、右连续

若 $x_1<x_2,\ y_1<y_2$，则 $F(x_2,y_2)-F(x_2,y_1)+F(x_1,y_1)-F(x_1,y_2)\ge 0$



边缘分布函数 $F_X(x)=F(x,+\infty)=P(X\le x)$ $F_Y(y)=F(+\infty,y)=P(Y\le y)$

条件分布函数 $F_{X|Y}(x|y)=P(X\le x|Y=y)$ 为$Y=y$ 条件下随机变量X 的条件分布函数

$F_{Y|X}(y|x)=P(Y\le y|X=x)$ 为$X=x$ 条件下随机变量Y 的条件分布函数

二维离散型随机变量

联合分布律 $P(X=x_i, Y=y_j)=p_{ij},\ i,j=1,2,\cdots$

联合分布函数 $F(x,y)=\sum\limits_{x_i\le x}\sum\limits_{y_j\le y} p_{ij}$

边缘分布律

$$P(X=x_i)=P(X=x_i, Y\le+\infty)=\sum_{j=1}^{\infty} p_{ij} \overset{\text{记为}}{=\!=} p_{i\bullet} \qquad i=1,2,\cdots$$

$$P(Y=y_j)=P(X\le+\infty, Y=y_j)=\sum_{i=1}^{\infty} p_{ij} \overset{\text{记为}}{=\!=} p_{\bullet j} \qquad j=1,2,\cdots$$

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

边缘分布函数 $F_X(x)=F(x,+\infty)=\sum\limits_{x_i\le x}\sum\limits_{j=1}^{\infty} p_{ij}$ $\quad F_Y(y)=F(+\infty,y)=\sum\limits_{y_j\le y}\sum\limits_{i=1}^{\infty} p_{ij}$

条件分布律

$$P(X=x_i \mid Y=y_j)=\frac{P(X=x_i, Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\bullet j}},\ i=1,2,\cdots$$ 为$Y=y_j$条件下，随机变量X的**条件分布律**

二维连续型随机变量

联合概率密度与联合分布函数 $F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v)dudv$

边缘概率密度 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$ $\quad f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$

边缘分布函数 $F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty}f(t,y)dy\right]dt=\int_{-\infty}^{x}f_X(t)dt$

$$F_Y(y)=F(+\infty,y)=\int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty}f(x,t)dx\right]dt=\int_{-\infty}^{y}f_Y(t)dt$$

条件概率密度 $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$ 为$Y=y$ 条件下随机变量X 的条件概率密度

条件分布函数 $F_{X|Y}(x|y)=\int_{-\infty}^{x}f_{X|Y}(t/y)dt$ 为$Y=y$ 条件下随机变量X 的条件分布函数

条件分布 = 联合分布 / 边缘分布　　联合分布 = 边缘分布 × 条件分布

二维离散型随机变量

条件分布律 $P(X=x_i|Y=y_j)=\dfrac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\dfrac{p_{ij}}{p_{\bullet j}},\ i=1,2,\cdots$ 为$Y=y_j$条件下，随机变量X的条件分布律

联合分布律 $P(X=x_i,Y=y_j)=P(Y=y_j)P(X=x_i|Y=y_j)=\quad i,j=1,2,\cdots$

二维连续型随机变量

条件概率密度 $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$ 为$Y=y$ 条件下随机变量X的条件概率密度

联合概率密度 $f(x,y)=f_Y(y)f_{X|Y}(x|y)$

本节回顾

二维离散型随机变量的条件分布律

若 $P(Y=y_j)=p_{\bullet j}>0$

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

称为 $Y=y_j$ 条件下，随机变量 X 的条件分布律

二维连续型随机变量的条件概率密度

若$(X、Y)$联合概率密度为$f(x,y)$，$(X、Y)$关于Y的边缘概率密度为$f_Y(y)$，若对于固定y，$f_Y(y)>0$，则称

$\dfrac{f(x,y)}{f_Y(y)}$ 为 $Y=y$ 条件下随机变量 X 的条件概率密度　记为 $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$