§3.3　条件分布

当一个随机向量的部分分量的取值确定以后，我们考虑另外一些分量的分布问题，这就产生了条件分布的概念.

例如：对二维随机变量(X,Y)—(身高，体重)，当身高(X)确定后，我们研究体重(Y)的分布.

随机变量X的取值确定包括多种情况：$X=x$，$X\in(a,b]$，$X\in(-\infty,a)$等. 本节主要研究二维随机变量（X,Y)—当$X=x$，或$Y=y$时，另外一个分量的分布问题.

例如：

$X = 1.7$m时，Y的分布，就是身高为1.7m的这些人的体重的分布.

$Y = 60kg$时，X的分布，就是体重为$60kg$的这些人的身高的分布.

一　二维离散型随机变量的条件分布

定义1　设离散型随机变量(X,Y)的联合概率分布为

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

关于X和Y的边缘概率分布为

$$P(X=x_i)=\sum_j p_{ij}\triangleq p_{i\cdot},\qquad i=1,2,\cdots$$

$$P\{Y=y_j\}=\sum_i p_{ij}\triangleq p_{\cdot j},\qquad j=1,2,\cdots,$$

若对某一固定的i，若$P(X=x_i)=p_{i\cdot}>0$，则称

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

为在$X = x_i$的条件下，随机变量Y的条件分布.

若对某一固定的 j，若$P(Y = y_j) = p_{\cdot j} > 0$，则称

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

为在$Y = y_j$的条件下，随机变量X的条件分布.

例1 把三个球等可能地放入编号为 1,2,3 的三个盒子中，每盒可容球数无限. 记X为落入1号盒的球数，Y为落入2号盒的球数. 求

(1)在$Y = 0$ 的条件下，X的分布律;

(2)在$X = 2$ 的条件下，Y的分布律.

解：求联合分布：先求$P(X=i,Y=j)$

$$i=0,1,2,3;\quad j=0,\cdots,3-i.$$

每个球投到盒子中相当于进行一次贝努力试验，记A表示“球落入1号盒子中”，则

$P(A)=1/3.$ $X=i$ 相当于在三次贝努力试验中事件A发生i 次，故

$$P(X=i)=C_3^i(1/3)^i(2/3)^{3-i}.$$

类似得 $P(Y=j|X=i)=C_{3-i}^j(1/2)^j(1/2)^{3-i-j}.$

所以

$$P(X=i,Y=j)=P(X=i)P(Y=j|X=i)$$

$$=C_3^i(1/3)^i(2/3)^{3-i}\cdot C_{3-i}^j(1/2)^j(1/2)^{3-i-j}.$$

联合分布与边缘分布如下表所示

$X \backslash Y$	**0**	**1**	**2**	**3**	$p_{i\cdot}$
0	1/27	1/9	1/9	1/27	8/27
1	1/9	2/9	1/9	0	4/9
2	1/9	1/9	0	0	2/9
3	1/27	0	0	0	1/27
$p_{\cdot j}$	8/27	4/9	2/9	1/27	1

(1) $$P(X=i \mid Y=0)=\frac{P(X=i, Y=0)}{P(Y=0)}$$

$$=\frac{P(X=i, Y=0)}{8/27}, \qquad i=0,1,2,3$$

将表中第一列数据代入得条件分布

X	0	1	2	3
$P(X=i \mid Y=0)$	1/8	3/8	3/8	1/8

(2) 当$X=2$时，Y只可能取0与1. 将表中第三行数据代入下式

$$P(Y=j \mid X=2)=\frac{P(X=2, Y=j)}{2/9}, \qquad j=0,1$$

得Y的条件分布

Y	0	1
$P(Y=j \mid X=2)$	$1/2$	$1/2$

例2 一射手进行射击训练，设击中目标的概率为 $p(0<p<1)$,射击到击中目标两次为止. 设以X表示首次击中目标所需要的射击次数，以Y表示总的射击次数. 试求X和Y的联合分布律及条件分布律.

解：由题意知 X 取 m 且 Y 取 n 时,有

$$P\{X=m, Y=n\}=p \cdot p \cdot(1-p) \cdot(1-p) \cdots(1-p)$$

$$=p^2 q^{n-2},$$

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

此即为X和Y的联合分布律.

现在求条件分布律.

$$P\{X=m \mid Y=n\}, \quad P\{Y=n \mid X=m\},$$

由于

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

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

$$m=1,2,\cdots,$$

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

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

所以当 $n=2,3,\cdots$ 时,

$$P\{X=m \mid Y=n\}=\frac{P\{X=m,Y=n\}}{P\{Y=n\}}$$

$$=\frac{p^2q^{n-2}}{(n-1)p^2q^{n-2}}=\frac{1}{n-1},$$

这里$m=1,2,\cdots,n-1.$

当 $m=1,2,\cdots$ 时

$$P\{Y=n \mid X=m\}=\frac{P\{X=m,Y=n\}}{P\{X=m\}}$$

$$=\frac{p^2q^{n-2}}{pq^{m-1}}=pq^{n-m-1},$$

这里$n=m+1,m+2,\cdots.$

二　二维连续型随机变量的条件分布

当(X,Y)是二维连续型随机变量时, $P(X=x)=0$, $P(Y=y)=0$. 所以不能直接利用条件概率去定义条件分布.

$X=x$可以理解为$x-\Delta x<X\le x$. 由于$\Delta x>0$时

$$P(Y\le y\mid x-\Delta x<X\le x)=\frac{P(x-\Delta x<X\le x,Y\le y)}{P(x-\Delta x<X\le x)}$$

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

$$=\frac{F(x,y)-F(x-\Delta x,y)}{F_X(x)-F_X(x-\Delta x)}.$$

$$\lim_{\Delta x\to 0+}\frac{[F(x,y)-F(x-\Delta x,y)]/\Delta x}{[F_X(x)-F_X(x-\Delta x)]/\Delta x}=\frac{\dfrac{\partial F(x,y)}{\partial x}}{\dfrac{dF_X(x)}{dx}}$$

$$=\frac{\int_{-\infty}^{y} f(x,v)dv}{f_X(x)}\overset{\text{def.}}{=}P(Y\le y\mid X=x)$$

$f(x,y)$连续 $\nearrow$

$f_X(x)\neq 0$,连续

定义2 若$f(x,y)$在点(x, y)连续，$f_X(x)$在点x处连续且$f_X(x)>0$，则称

$$\frac{\dfrac{\partial F(x,y)}{\partial x}}{\dfrac{dF_X(x)}{dx}} = \frac{\int_{-\infty}^{y} f(x,v)dv}{f_X(x)} = \int_{-\infty}^{y} \frac{f(x,v)}{f_X(x)}dv$$

为$X = x$时，Y的条件分布函数，记作

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

此时，Y是连续型随机变量，其密度函数

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

称为$X = x$时，Y的条件概率密度函数.

类似地有$Y=y$时，X的条件分布函数为

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

$Y=y$的条件下，X的条件概率密度函数为

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

注意：（1）$F_{Y|X}(y|x)$，$f_{Y|X}(y|x)$仅是y的函数，x是常数，对每一$f_X(x)>0$的x处，只要符合定义的条件，都能定义相应的函数. 当 $f_X(x)=0$时，可定义

$$F_{Y|X}(y|x)=0\ ,\ f_{Y|X}(y|x)=0\ .$$

（2）类似于乘法公式

$$f(x,y)=f_X(x)f_{Y|X}(y|x) \qquad f_X(x)>0,$$

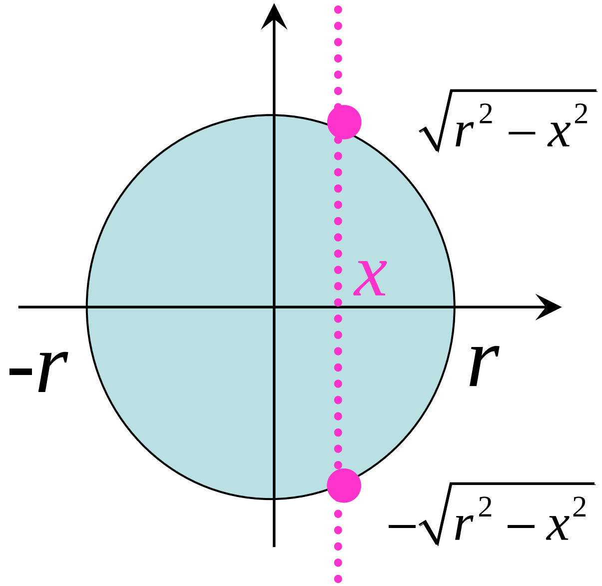
$$=f_Y(y)f_{X|Y}(x|y) \qquad f_Y(y)>0.$$

例3 已知 (X,Y)服从圆域 $x^2+y^2=r^2$上的均匀分布. 求 $f_{Y|X}(y|x)$, $f_{X|Y}(x|y)$.

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

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

$$= \begin{cases} \dfrac{2\sqrt{r^2 - x^2}}{\pi r^2}, & -r < x < r, \\ 0, & \text{其他}. \end{cases}$$



$\Uparrow$

边缘分布不再是均匀分布！

当 $-r < x < r$ 时

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

$$= \begin{cases} \dfrac{1}{2\sqrt{r^2 - x^2}}, & -\sqrt{r^2 - x^2} < y < \sqrt{r^2 - x^2}, \\ 0, & \text{其他}. \end{cases}$$

—这里x是常数，当$X = x$时，

$$Y \sim U(-\sqrt{r^2 - x^2}, \sqrt{r^2 - x^2}).$$

$$\Uparrow$$

$X = x$给定时，Y的条件分布是均匀分布！

同理

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

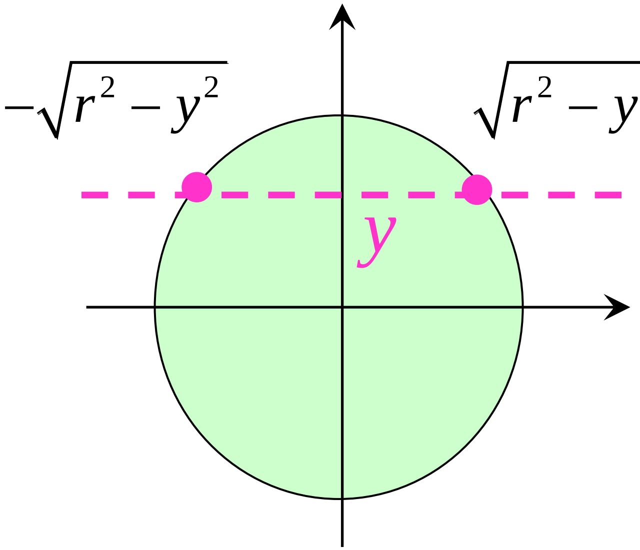
当$-r < y < r$时

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

$$=\begin{cases}\dfrac{1}{2\sqrt{r^2-y^2}}, & -\sqrt{r^2-y^2}<x<\sqrt{r^2-y^2},\\ 0, & \text{其他}.\end{cases}$$

—这里y是常数，当$Y=y$时，

$$X\sim U(-\sqrt{r^2-y^2},\sqrt{r^2-y^2}).$$



例4：已知 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. 试求

$$f_{Y|X}(y|x),\qquad f_{X|Y}(x|y).$$

解：由于二元正态分布的密度函数有下列两种形式的分解

$$f(x,y)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}e^{-\frac{\left[y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1))\right]^2}{2\sigma_2^2(1-\rho^2)}}$$

$$f(x,y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}e^{-\frac{\left[x-(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2))\right]^2}{2\sigma_1^2(1-\rho^2)}}.$$

又因为

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},\quad f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}.$$

所以

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}e^{-\frac{\left[y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1))\right]^2}{2\sigma_2^2(1-\rho^2)}}$$

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_X(x)}=\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}e^{-\frac{\left[x-(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2))\right]^2}{2\sigma_1^2(1-\rho^2)}}$$

所以

$$f_{X|Y}(x|y) \sim N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2));$$

$$f_{Y|X}(y|x) \sim N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)).$$

$$\Uparrow$$

二元正态分布的条件分布仍然是正态分布！

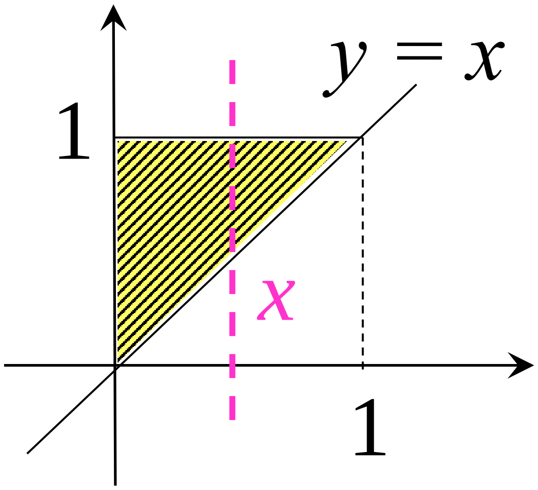
练习题： 设(X,Y)的密度函数为

$$f(x,y) = \begin{cases} 8xy, & 0 \le x \le y, 0 \le y \le 1 \\ 0, & \text{其他} \end{cases},$$

试求： $f_{Y|X}(y|x)$， $f_{X|Y}(x|y)$.

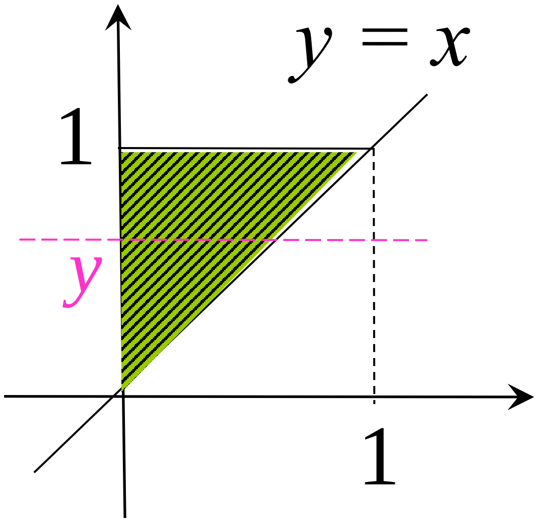
解： $f_X(x)=\begin{cases}\int_x^1 8xydy, & 0\le x\le 1,\\ 0, & \text{其他}.\end{cases}$

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



$$f_Y(y)=\begin{cases}\int_0^y 8xydx, & 0\le y\le 1\\ 0, & \text{其他}\end{cases}$$

$$=\begin{cases}4y^3, & 0\le y\le 1\\ 0, & \text{其他}\end{cases}$$



故

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

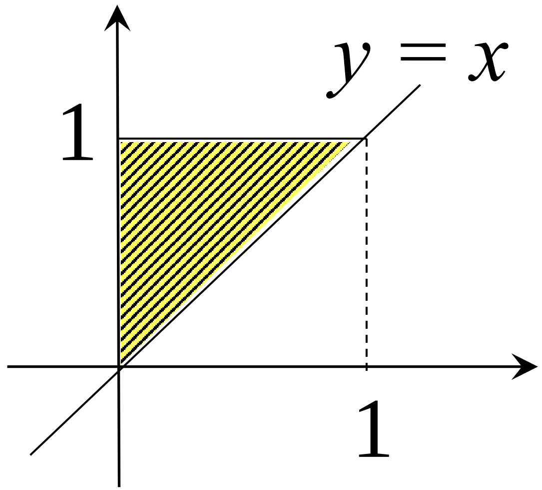
$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\begin{cases}\dfrac{2x}{y^2}, & 0\le x\le y,\ \ 0<y<1.\\ 0, & \text{其他}.\end{cases}$$

例6: 已知 $f_{Y|X}(y|x)=\begin{cases}\dfrac{2y}{1-x^2}, & x\le y\le 1,\\ 0, & \text{其他}\end{cases}$

$$f_X(x)=\begin{cases}4x(1-x^2), & 0\le x\le 1\\ 0, & \text{其他}\end{cases}$$

试求：$P(X+Y\ge 1)$, $P(Y<1/2)$,

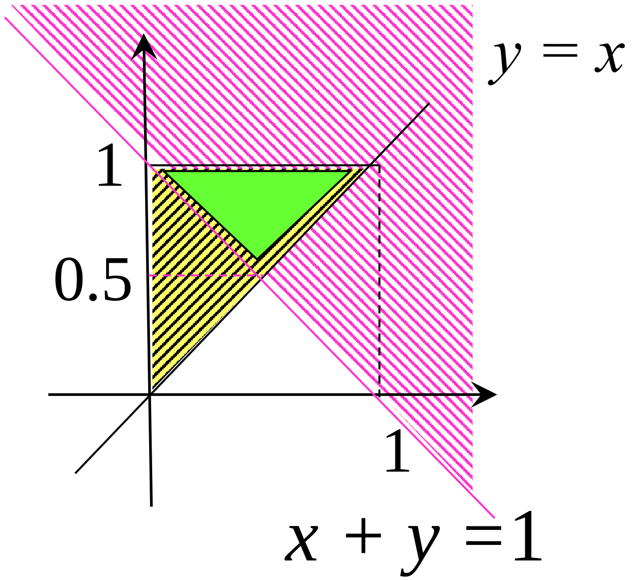
$P(Y<2/3|X=1/2)$.



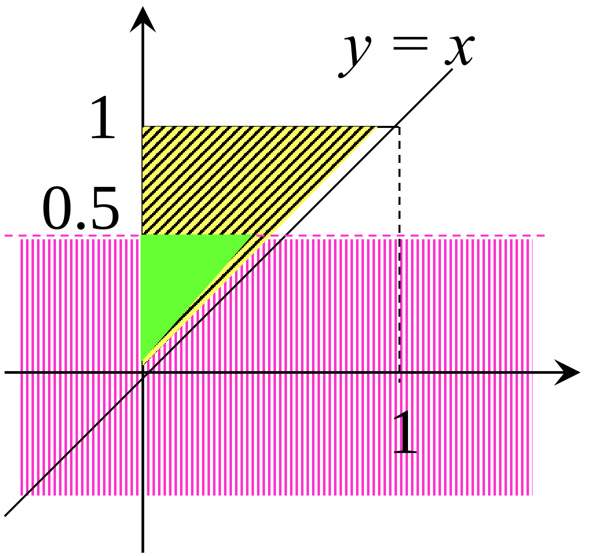
解： $f(x,y)=f_{Y|X}(y|x)f_X(x)$, 故

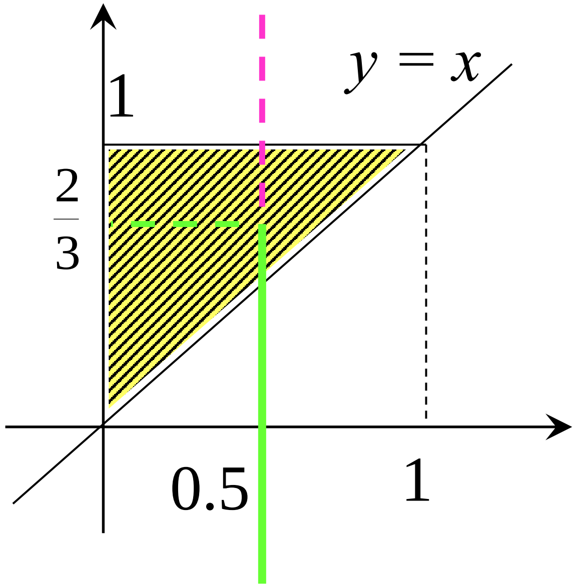
$$f(x,y)=\begin{cases}8xy, & 0\le x\le y, 0\le y\le 1.\\ 0, & \text{其他}.\end{cases}$$

$$P(X+Y\geq 1)=\int_{0.5}^{1}dy\int_{1-y}^{y}8xydx=\frac{5}{6}.$$



$$P(Y<0.5)=\int_{0}^{1/2}dy\int_{0}^{y}8xydx=\frac{1}{16}$$





$$P(Y < 2/3 \mid X = 1/2) = \int_{-\infty}^{2/3} f_{Y|X}\left(y \,\middle|\, \frac{1}{2}\right) dy$$

$$= \int_{1/2}^{2/3} \frac{2y}{1 - 0.5^2} dy = \int_{1/2}^{2/3} \frac{8y}{3} dy$$

$$= \frac{7}{27}.$$

说明

联合分布、边缘分布、条件分布的关系如下