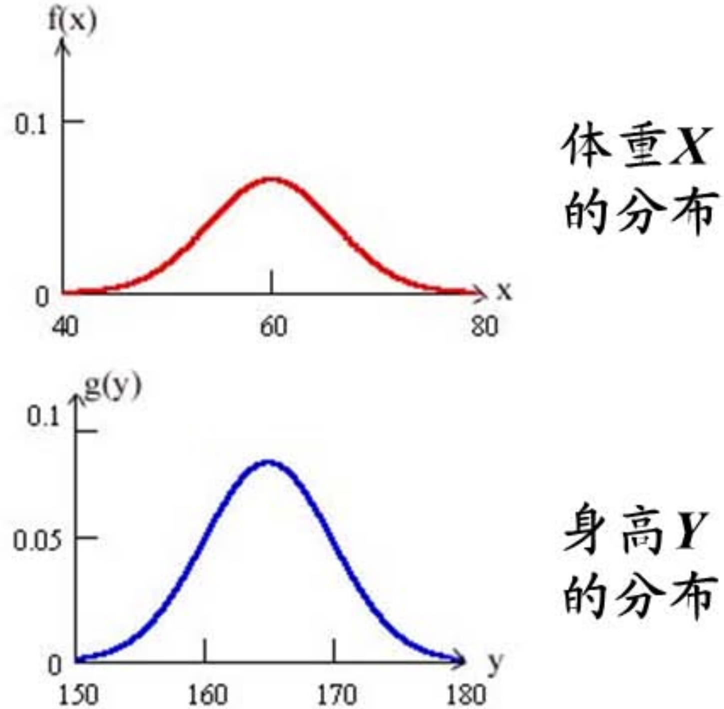
第三章 联合分布

背景例子：考虑某大学的全体学生，从其中随机抽取一个学生，分别以 X 和 Y 表示其体重和身高. 则 X 和 Y 都是r.v.，它们都有一定的概率分布.



现在若限制 $1.7<Y<1.8$(米), 在这个条件下去求 X 的条件分布，这就意味着要从该校的学生中把身高在1.7米和1.8米之间的那些人都挑出来，然后在挑出的学生中求其体重的分布.

这个分布与不加这个条件时的分布会很不一样.

如：在条件分布中体重取大值的概率会显著增加.

条件分布

回顾条件概率

$$P(A \mid B)=\frac{P(AB)}{P(B)} \qquad (P(B)>0)$$

设(X,Y)为二维r.v，$\forall\ y\in R^1$ 考虑条件概率

$$P\{X\le x \mid Y=y\} \qquad (x\in R^1)$$

这可视为在 $\{Y=y\}$ 发生的条件下 r.v X 的概率分布 —— 条件分布

问 能否由条件概率定义计算

$$P\{X\le x \mid Y=y\}=\frac{P\{X\le x, Y=y\}}{P\{Y=y\}} \ ?$$

二维离散型随机变量的条件频率函数

设 (X,Y) 的频率函数为

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

考虑在 $\{Y=y_j\}$ 已发生的条件下, $\{X=x_i\}$ 发生的条件概率

$$P\{X=x_i \mid Y=y_j\} \quad (i=1,2,\cdots)$$

由条件概率公式，有

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

同理在 $\{X=x_i\}$ 已发生的条件下, $\{Y=y_j\}$ 发生的条件概率

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

二维离散型随机变量的条件频率函数

设 (X,Y) 的频率函数为

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

定义 对于固定的j，若$P\{Y=y_j\}=p_{\cdot j}>0$，则称

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

为在 $Y=y_j$ 的条件下，r.v X 的**条件(conditional)频率函数.**

对于固定的i，若$P\{X=x_i\}=p_{i\cdot}>0$，则称

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

为在 $X=x_i$ 的条件下，r.v Y 的**条件(conditional)频率函数.**

例 设 r.v X 从1,2,3,4四个数中等可能取值,又设 r.v Y 从$1\sim X$ 中等可能取值. 问当第二次取到数字 3 时第一次取四个数字的可能性各是多少?

解 由§2例,(X,Y)的频率函数及边际频率函数为

$Y \backslash X$	1	2	3	4	$p_{\cdot j}$
1	1/4	1/8	1/12	1/16	25/48
2	0	1/8	1/12	1/16	13/48
3	0	0	1/12	1/16	7/48
4	0	0	0	1/16	3/48
$p_{i\cdot}$	1/4	1/4	1/4	1/4	

故在 $Y=3$ 的条件下,X取到四个数字的概率是

$$P\{X=1 \mid Y=3\}=\frac{p_{13}}{p_{\cdot 3}}=\frac{0}{7/48}=0$$

$$P\{X=2 \mid Y=3\}=0,\quad P\{X=3 \mid Y=3\}=\frac{4}{7},\ P\{X=4 \mid Y=3\}=\frac{3}{7}.$$

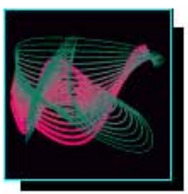
即在 $Y=3$ 的条件下,X的条件频率函数为

$X=k$	1	2	3	4
$P\{X=k \mid Y=3\}$	0	0	4/7	3/7

条件频率函数的性质

① $P\{X=x_i \mid Y=y_j\} \ge 0 \quad (i=1,2,\cdots)$

② $$\sum_{i=1}^{\infty} P\{X=x_i \mid Y=y_j\} = \sum_{i=1}^{\infty} \frac{p_{ij}}{p_{\cdot j}} = \frac{1}{p_{\cdot j}} \sum_{i=1}^{\infty} p_{ij} = \frac{1}{p_{\cdot j}} p_{\cdot j} = 1$$



这两条性质说明：

条件频率函数也是一种频率函数

二维连续型随机变量的条件概率密度

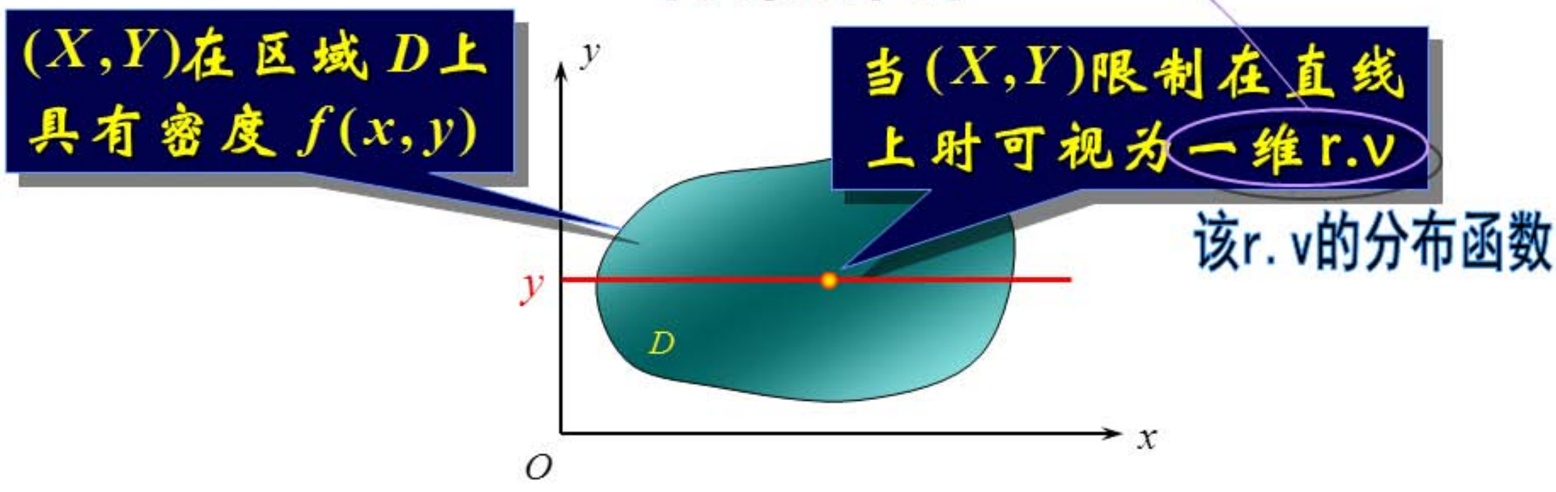
设 (X,Y) 的概率密度为

$$f(x,y)$$

考虑在 $\{Y=y\}$ 已发生的条件下, $\{X\le x\}$ 发生的条件概率

$$P\{X\le x \mid Y=y\} \quad (x\in R^1)$$

背景解释



二维连续型随机变量的条件概率密度

设 (X,Y) 的概率密度为

$$f(x,y)$$

考虑在 $\{Y=y\}$ 已发生的条件下, $\{X\le x\}$ 发生的条件概率

$$P\{X\le x\,|\,Y=y\}\qquad (x\in R^1)$$

若按条件概率公式，则有

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

对于连续型r. v

$$P\{Y=y\}=0$$

如何定义条件分布 $P\{X \le x \mid Y = y\}$?

$\forall\ \varepsilon > 0$, 考虑条件概率

$$P\{X \le x \mid y < Y \le y+\varepsilon\} = \frac{P\{X \le x, y < Y \le y+\varepsilon\}}{P\{y < Y \le y+\varepsilon\}}$$

称为条件分布

$$= \frac{\int_{-\infty}^{x}\int_{y}^{y+\varepsilon} f(u,v)\,dv\,du}{\int_{y}^{y+\varepsilon} f_Y(y)\,dy}$$

应用积分中值定理

$$= \frac{\varepsilon\int_{-\infty}^{x} f(u, y_\varepsilon)\,du}{\varepsilon f_Y(\tilde{y}_\varepsilon)}$$

$$\to \int_{-\infty}^{x} \frac{f(u,y)}{f_Y(y)}\,du \quad (\varepsilon \to 0)$$

称为条件密度

定义 设(X,Y)的概率密度为$f(x,y)$，若对于固定的y，(X,Y)关于Y的边际密度$f_Y(y)>0$，则称

$$\frac{f(x,y)}{f_Y(y)} \triangleq f_{X|Y}(x\mid y) \quad (-\infty<x<\infty)$$

为在$Y=y$的条件下，X的**条件密度(conditional density)**. 称

$$F_{X|Y}(x\mid y) \triangleq \int_{-\infty}^{x} f_{X|Y}(u\mid y)\,du \quad (-\infty<x<\infty)$$

为在$Y=y$的条件下，X的**条件分布(函数)**.

类似地，可定义

条件密度与条件概率在形式上很相似！

$$f_{Y|X}(y\mid x) \triangleq \frac{f(x,y)}{f_X(x)} \quad (-\infty<y<\infty)$$

$$F_{Y|X}(y\mid x) \triangleq \int_{-\infty}^{x} f_{Y|X}(v\mid x)\,dv \quad (-\infty<y<\infty)$$

由

$$f_{Y|X}(y\,|\,x) \triangleq \frac{f(x,y)}{f_X(x)} \quad (-\infty < y < \infty)$$

因此

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

即：联合密度可以用边际密度和条件密度表示.

两边关于 x 积分，Y 的边际密度可表示为

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y\,|\,x) f_X(x)\,\mathrm{d}\,x$$

连续情形的全概率公式

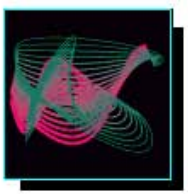
条件密度的性质

① $f_{X|Y}(x \mid y) \geq 0$

② $$\int_{-\infty}^{\infty} f_{X|Y}(u \mid y)du = \int_{-\infty}^{\infty} \frac{f(u,y)}{f_Y(y)}du$$

$$= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(u,y)du$$

$$= \frac{1}{f_Y(y)} \cdot f_Y(y) = 1$$



这两条性质说明：
条件密度也是一种密度

事件独立性与条件概率的关系

A,B 相互独立 $\Leftrightarrow P(A\mid B)=P(A), P(B\mid A)=P(B)$

r. v. 独立性与条件密度的关系

X,Y 相互独立 $\Leftrightarrow f(x,y)=f_X(x)f_Y(y) \quad (a.e)$

$$\Leftrightarrow f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}=f_X(x) \quad (a.e)$$

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

平面上的均匀分布

设G是平面上的有界区域,其面积为A.若(X,Y)的概率密度为

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

则称(X,Y)服从区域G上的均匀分布.

均匀分布的实际背景

若随机点(X,Y)在平面区域G上"等可能"取值,则(X,Y)服从G上的均匀分布.

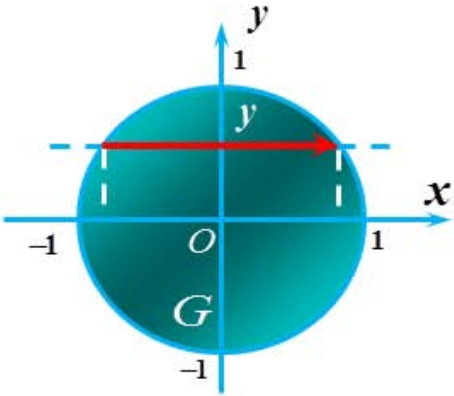
例如 设雷达的圆形屏幕半径为1,当用雷达捕捉目标时,可认为目标出现点(X,Y)在屏幕上服从圆域$G: x^2+y^2\le 1$上的均匀分布.

例 设 (X,Y) 服从圆域 $G: x^2+y^2\le 1$ 上的均匀分布. 求条件概率密度 $f_{X|Y}(x|y)$.

解 (X,Y) 的密度及 Y 的边际密度分别为

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

$$f_Y(y)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-y^2}, & |y|<1\\ 0, & |y|\ge 1\end{cases}$$



故当 $-1<y<1$ 时有

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

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

$\because f_{X|Y}(x|y)$ 与 y 有关,

$\therefore X,Y$ 不独立.

例 设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,

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

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

则经过计算可得

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

$$\sim N(\mu_2+\rho\tfrac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2)).$$

即： **二维正态分布，给定 X 时 Y 的条件密度是一维(单变量)正态分布.**

例 湍流速度近似服从正态分布.

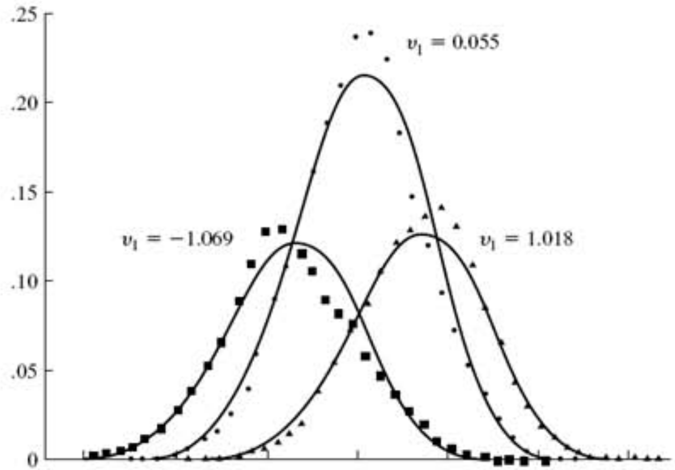


FIGURE 3.14 The conditional densities of v_2 given v_1 for selected values of v_1, where v_1 and v_2 are components of the velocity of a turbulent wind flow at different times. The solid lines are the conditional densities according to a normal fit, and the triangles and squares are empirical values determined from 409,600 observations.

v_1: t 时刻流速

v_2: $t+\tau$ 时刻流速(同一位置)

图中：给定v_1的情况下v_2的条件密度及其正态拟合

表明：v_1和v_2的联合分布不是二维正态分布.

物理解释：因为v_1和v_2的关系须遵守运动和连续性方程

再次说明：即使两个r.v.都是边际正态，它们的联合分布也不一定是正态.

例 (贝叶斯推断Bayesian Inference)

一枚新铸造的硬币沿其边转动, 最终以某概率出现正面 (未必等于1/2).

现在假设转动 n 次, 出现正面 X 次.
问硬币出现正面的机会是多少?

令 Θ 表示硬币出现正面的概率.

在搜集任何数据之前, 我们对 Θ 一无所知.

根据“同等无知原则”, 用 $\Theta\sim U[0,1]$ 表示对 Θ 的认知.

$$f_\Theta(\theta)=1,\quad 0\le\theta\le 1.$$ **(先验密度)**

问题: 观测 X 将如何改变我们对 Θ 的认知 (后验密度).

给定θ, 则$X \sim b(n,\theta)$.

$$f_{X|\Theta}(x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,1,\ldots,n.$$

Θ 是连续r.v., X是离散r.v.

二者的联合分布:

$$f_{\Theta,X}(\theta,x)=f_{X|\Theta}(x|\theta)f_{\Theta}(\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},$$
$$x=0,1,\ldots,n,\quad 0\le\theta\le 1.$$

X的边际密度:

$$f_X(x)=\int_0^1\binom{n}{x}\theta^x(1-\theta)^{n-x}\mathrm{d}\theta$$

经计算得: $$f_X(x)=\frac{1}{n+1},\quad x=0,1,\ldots,n.$$

给定 $X=x$，则 Θ 的条件密度:

$$f_{\Theta|X}(\theta|x)=\frac{f_{\Theta,X}(\theta,x)}{f_X(x)}=(n+1)\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

经计算得:

$$f_{\Theta|X}(\theta|x)=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}$$

$$\sim Beta(x+1,n-x+1)$$

(给定x时θ的后验密度)

在n次转动中观察到x次正面后，利用后验密度来量化我们对Θ的认知

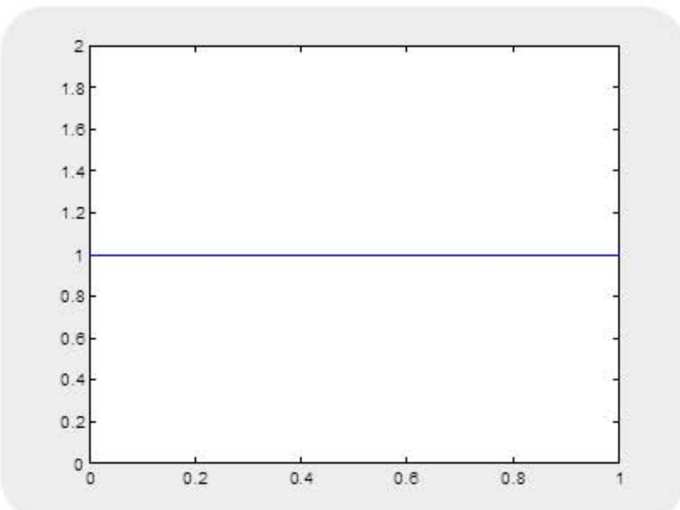
Beta密度刻画 [0,1] 区间上的r.v:

$$f(u)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}u^{a-1}(1-u)^{b-1},\quad 0\le u\le 1$$

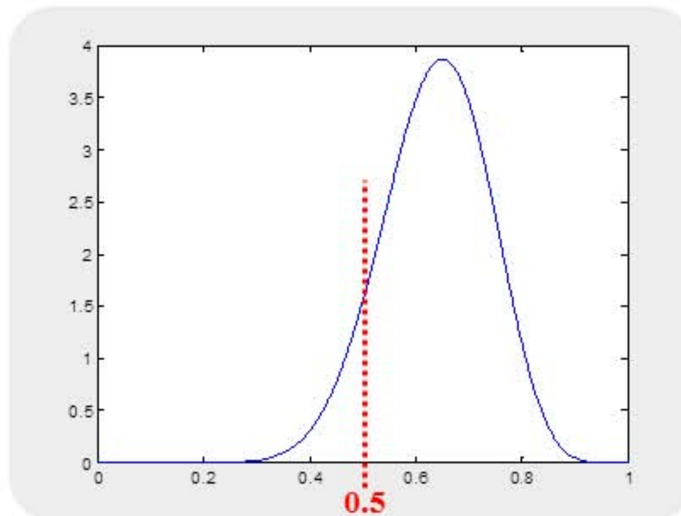
$$f_{\Theta|X}(\theta|x)=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^{x}(1-\theta)^{n-x}\quad\sim Beta(x+1,n-x+1)$$

如:转动20次硬币，则 $f_{\Theta}(\theta)=1\qquad(0\le\theta\le 1)$

若观察到13次正面，则： $f_{\Theta|X}(\theta|13)\sim Beta(14,8)$



θ 的先验密度



θ 的后验密度

小结

边际分布

随机变量的边际分布完全由它们的联合分布确定

$$F_X(x)=F(x,+\infty) \qquad F_Y(y)=F(+\infty,y)$$

离散型r.v的边际频率函数

$$P\{X=x_i\}=\sum_{j=1}^{\infty}p_{ij}\triangleq p_{i\cdot} \quad (i=1,2,\cdots)$$

$$P\{Y=y_j\}=\sum_{i=1}^{\infty}p_{ij}\triangleq p_{\cdot j} \quad (j=1,2,\cdots)$$

连续型r.v的边际分布密度

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy \quad (-\infty<x<+\infty)$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx \quad (-\infty<y<+\infty)$$

定理 若$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则

$$X\sim N(\mu_1,\sigma_1^2)\,,\quad Y\sim N(\mu_2,\sigma_2^2)$$

小结

条件分布

$$P\{X \le x \mid Y = y\} \quad (x \in R^1)$$

这可视为在 $\{Y = y\}$ 发生的条件下
r.v X 的概率分布 —— 条件分布

离散型r. v的条件频率函数

$$P\{X = x_i \mid Y = y_j\} = \frac{p_{ij}}{p_{\cdot j}}$$

$$(p_{\cdot j} > 0, i = 1, 2, \cdots)$$

$$P\{Y = y_j \mid X = x_i\} = \frac{p_{ij}}{p_{i\cdot}}$$

$$(p_{i\cdot} > 0, j = 1, 2, \cdots)$$

连续型r. v的条件概率密度

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

$$(f_Y(y) > 0, -\infty < x < \infty)$$

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

$$(f_X(x) > 0, -\infty < y < \infty)$$

课后作业

P77：1、9、10、15

思考题*

将长度为d的一根木棒任意截去一段，再将剩下的木棒任意截为两段. 求这三段木棒能构成三角形的概率.