



大数据的统计学基础——第5周

【声明】 本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

<http://edu.dataguru.cn>

关注炼数成金企业微信



■ 提供全面的数据价值资讯，涵盖商业智能与数据分析、大数据、企业信息化、数字化技术等，各种高性价比课程信息，赶紧掏出您的手机关注吧！



基本初等函数求导公式

$$(1) \quad (C)' = 0$$

$$(2) \quad (x^\mu)' = \mu x^{\mu-1}$$

$$(3) \quad (\sin x)' = \cos x$$

$$(4) \quad (\cos x)' = -\sin x$$

$$(5) \quad (\tan x)' = \sec^2 x$$

$$(6) \quad (\cot x)' = -\csc^2 x$$

$$(7) \quad (\sec x)' = \sec x \tan x$$

$$(8) \quad (\csc x)' = -\csc x \cot x$$

$$(9) \quad (a^x)' = a^x \ln a$$

$$(10) \quad (e^x)' = e^x$$

$$(11) \quad (\log_a x)' = \frac{1}{x \ln a}$$

$$(12) \quad (\ln x)' = \frac{1}{x},$$

$$(13) \quad (\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$$

$$(14) \quad (\arccos x)' = -\frac{1}{\sqrt{1-x^2}}$$

$$(15) \quad (\arctan x)' = \frac{1}{1+x^2}$$

$$(16) \quad (\operatorname{arccot} x)' = -\frac{1}{1+x^2}$$

◆ 积分与求导互为逆运算

$$1、\int k dx = kx + c$$

$$2、\int x^a dx = \frac{x^{a+1}}{a+1} + c$$

$$3、\int \frac{1}{x} dx = \ln|x| + c$$

$$4、\int \frac{1}{1+x^2} dx = \arctan x + c$$

$$5、\int \frac{1}{\sqrt{1-x^2}} dx = \arcsin x + c$$

$$6、\int \cos x dx = \sin x + c$$

$$7、\int \sin x dx = -\cos x + c$$

$$8、\int \frac{1}{\cos^2 x} dx = \int \sec^2 x dx = \tan x + c$$

$$9、\int \frac{1}{\sin^2 x} dx = \int \csc^2 x dx = -\cot x + c$$

$$10、\int \sec x \tan x dx = \sec x + c$$

$$11、\int \csc x \cot x dx = -\csc x + c$$

$$12、\int e^x dx = e^x + c$$

$$13、\int a^x dx = \frac{a^x}{\ln a} + c$$

◆ 牛顿——莱布尼兹公式： $\int_a^b f(x)dx = F(x) \Big|_a^b = F(b) - F(a)$

◆ 其中， $F(x)$ 为 $f(x)$ 的原函数，即 $F'(x) = f(x)$

◆ 分部积分公法：

◆ 设 $u(x)$ 、 $v(x)$ 在 $[a,b]$ 上具有连续导数 $u'(x), v'(x)$ ，则

$$\int_a^b u(x)dv(x) = u(x)v(x) \Big|_a^b - \int_a^b v(x)du(x)$$

8. 甲、乙两人投篮,投中的概率分别为 0.6,0.7. 今各投 3 次. 求

(1) 两人投中次数相等的概率;

(2) 甲比乙投中次数多的概率.

- ◆ 上周的作业中, 解答这道题我们设立了两个随机变量: 用X记录甲投中的次数, Y记录乙投中的次数。我们写出了它们各自的分布律:

X	0	1	2	3
P	0.064	0.288	0.432	0.216

Y	0	1	2	3
P	0.027	0.189	0.441	0.343

- ◆ 这次我们尝试将X、Y取值的不同情况下的概率都计算出来
- ◆ 为了方便计算，我们把X与Y的分布律的数据改一下

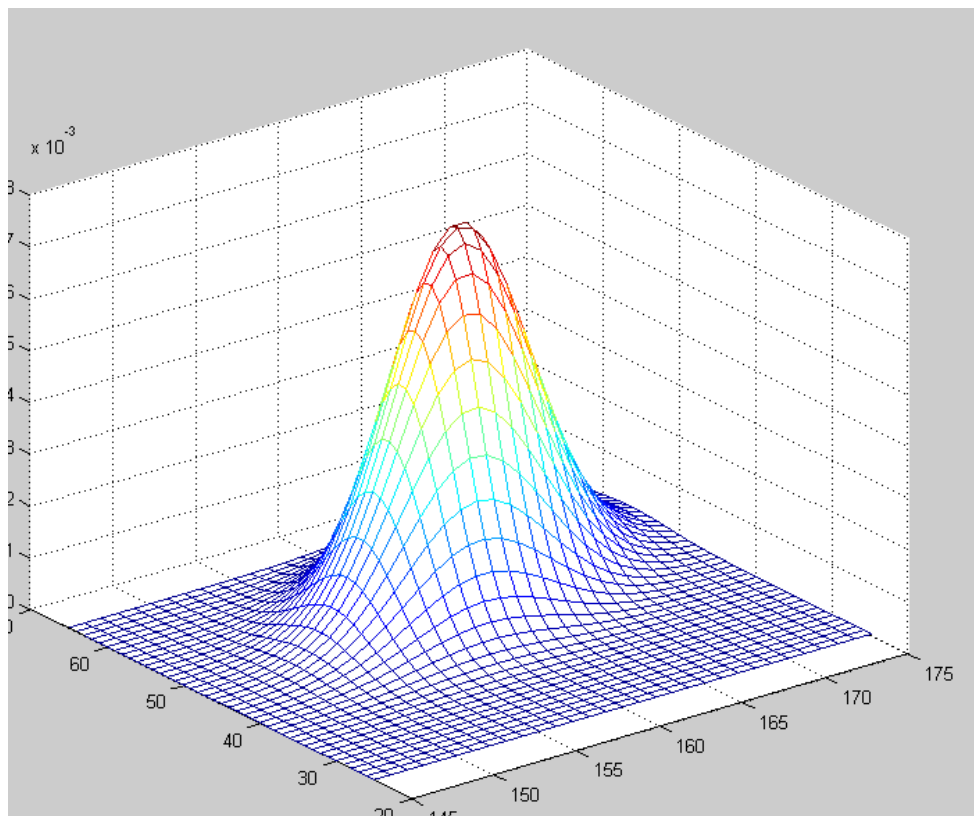
X	0	1	2	3
P	0.1	0.3	0.5	0.1
Y	0	1	2	3
P	0.1	0.2	0.5	0.2

Y	X	0	1	2	3
0		0.01	0.03	0.05	0.01
1		0.02	0.06	0.1	0.02
2		0.05	0.15	0.25	0.05
3		0.02	0.06	0.1	0.02

← 联合分布律

- ◆ 有了上表，再计算题目中的问题就变得简单了

- ◆ 一般，设 E 是一个随机试验，它的样本空间是 $S=\{e\}$ ，设 $X=X\{e\}$ 和 $Y=Y\{e\}$ 是定义在 S 上的随机变量，由 X 与 Y 构成的向量 (X,Y) 叫做**二维随机向量或是二维随机变量** (Two-dimensional random vector)



- ◆ 设 (X, Y) 是二维随机变量，对于任意实数 x, y ，二元函数：

$$F(x, y) = P\{(X \leq x) \cup (Y \leq y)\} = P\{X \leq x, Y \leq y\}$$

- ◆ 称为二维随机变量 (X, Y) 的联合分布函数(Joint probability distribution)

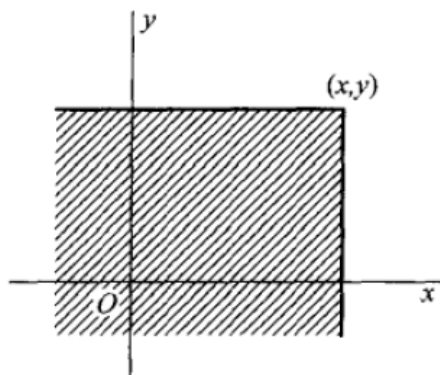


图 3-2

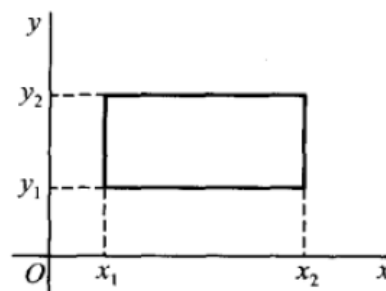


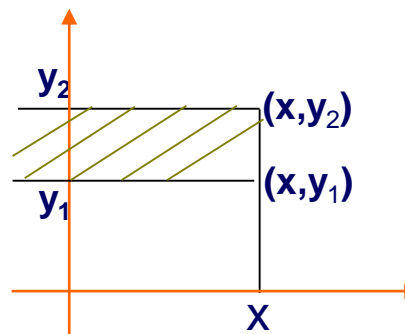
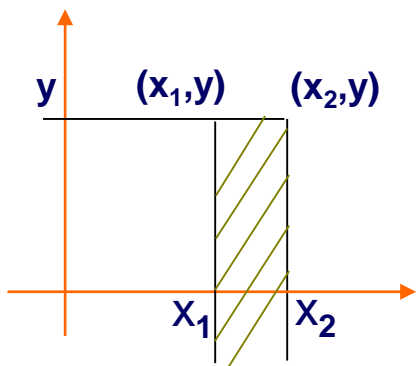
图 3-3

- ◆ 1. $F(x,y)$ 是对于 x 和 y 的不减函数，即

$$x_1 < x_2 \Rightarrow F(x_1, y) \leq F(x_2, y)$$

$$y_1 < y_2 \Rightarrow F(x, y_1) \leq F(x, y_2)$$

- ◆ 2. $0 \leq F(x,y) \leq 1$ ，且对于任意固定的 y ， $F(-\infty, y) = 0$ ；对于任意固定的 x ， $F(x, -\infty) = 0$
 $F(-\infty, -\infty) = 0$ ， $F(\infty, \infty) = 1$



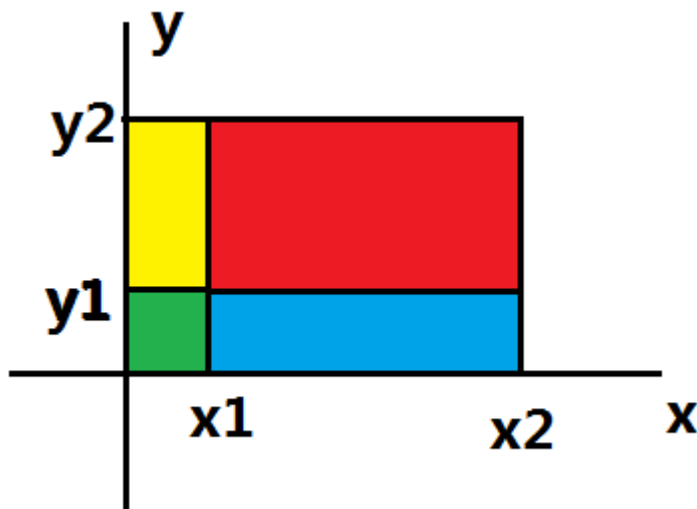
- ◆ 3. $F(x, y)$ 关于 x 右连续, 关于 y 右连续, 即

$$\lim_{\varepsilon \rightarrow 0^+} F(x + \varepsilon, y) = F(x, y)$$

$$\lim_{\varepsilon \rightarrow 0^+} F(x, y + \varepsilon) = F(x, y)$$

- ◆ 4. 对于任意 $(x_1, y_1), (x_2, y_2), x_1 < x_2, y_1 < y_2$, 下述不等式成立:

$$F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) \geq 0$$



例 3.1 设随机变量 (X, Y) 等可能地取值: $(0, 0), (0, 2), (2, 0), (2, 2)$, 求 X, Y 的联合分布函数.

解: I. $x < 0$, 或 $y < 0$ 时,

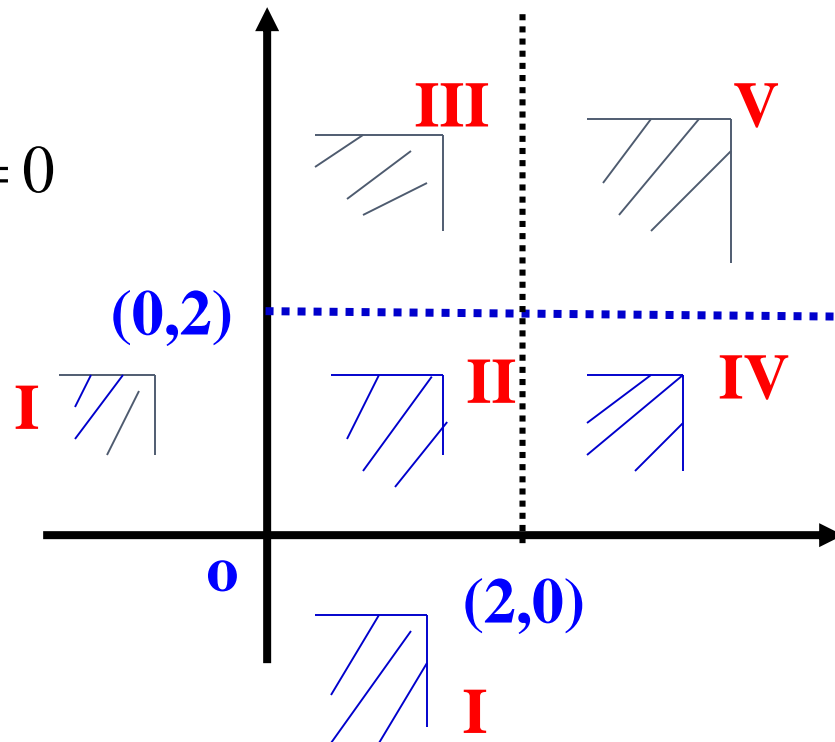
$$F(x, y) = P(X \leq x, Y \leq y) = P(\Phi) = 0$$

II. $0 \leq x < 2, 0 \leq y < 2$ 时,

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) \\ &= P\{(X, Y) = (0, 0)\} = \frac{1}{4} \end{aligned}$$

III. $0 \leq x < 2, y \geq 2$ 时.

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) \\ &= P\{(X, Y) = (0, 0)\} + P\{(X, Y) = (0, 2)\} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$



IV. $x \geq 2, 0 \leq y < 2$ 时,

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) \\ &= P\{(X, Y) = (0, 0)\} + P\{(X, Y) = (2, 0)\} \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$

V. $x \geq 2, y \geq 2$ 时,

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) \\ &= P\{(X, Y) = (0, 0)\} + P\{(X, Y) = (0, 2)\} \\ &\quad + P\{(X, Y) = (2, 0)\} + P\{(X, Y) = (2, 2)\} \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1 \end{aligned}$$

◆ 综上所述，得随机变量(X,Y)的分布函数为

$$F(x,y)=\begin{cases} 0, & x < 0 \text{ 或 } y < 0; \\ 1/4, & 0 \leq x < 2, 0 \leq y < 2; \\ 1/2, & 0 \leq x < 2, y \geq 2 \text{ 及 } x \geq 2, 0 \leq y < 2; \\ 1, & x \geq 2, y \geq 2. \end{cases}$$

- ◆ 如果二维随机变量 (X, Y) 全部可能取到的值是有限对或是可列无限对，则称 (X, Y) 为离散型的二维随机变量。
- ◆ 如果对于二维随机变量 (X, Y) 的分布函数 $F(x, y)$ ，存在非负可积函数 $f(x, y)$ 使得对于任意 x, y 有

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$$

- ◆ 称 (X, Y) 为连续型的二维随机变量。函数 $f(x, y)$ 称为二维随机变量 (X, Y) 的联合概率密度 (Joint probability density)。

- ◆ 对于离散型的二维随机变量 (X, Y) 的所有可能取值为 (x_i, y_j) , $i, j=1, 2, \dots$, 称 $P\{X=x_i, Y=y_j\}=p_{ij}$, $i, j=1, 2, \dots$ 为随机变量 X 和 Y 的联合分布律 (Joint distribution law)

- ◆ 性质 :

$$1^\circ \quad p_{ij} \geq 0, \quad i, j = 1, 2, \dots$$

$$2^\circ \quad \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1$$

$\begin{matrix} Y \\ X \end{matrix}$	y_1	y_2	\dots	y_j	\dots
x_1	p_{11}	p_{12}	\dots	p_{1j}	\dots
x_2	p_{21}	p_{22}	\dots	p_{2j}	\dots
\vdots	\dots		\dots		\dots
x_i	p_{i1}	p_{i2}	\dots	p_{ij}	\dots
\vdots	\dots		\dots		\dots

- ◆ 一个口袋中有三个球, 依次标有数字1, 2, 2, 从中任取一个, 不放回袋中, 再任取一个. 设每次取球时, 各球被取到的可能性相等. 以X, Y分别记第一次和第二次取到的球上标有的数字, 求X, Y的联合分布律.

解: (X, Y)的可能取值为(1, 2),(2, 1),(2, 2)

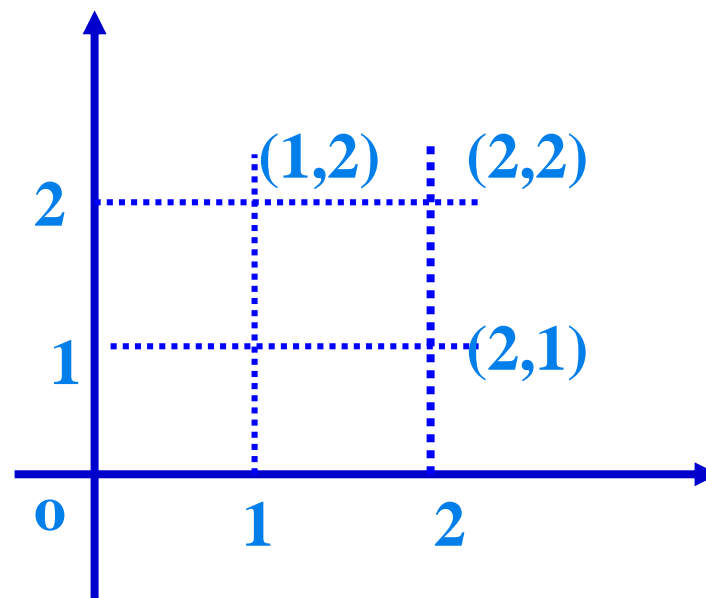
$$\begin{aligned}P\{(X, Y) = (1, 2)\} &= P\{(X = 1) \cap (Y = 2)\} \\&= P(X = 1)P(Y = 2 / X = 1) \\&= \frac{1}{3} \times \frac{2}{2} = \frac{1}{3}\end{aligned}$$

$$\begin{aligned}P\{(X, Y) = (2, 1)\} &= P(X = 2)P(Y = 1 / X = 2) \\&= \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}\end{aligned}$$

$$P\{(X, Y) = (2, 2)\} = \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}$$

◆ 故X与Y的联合分布律

Y \ X	X	
	1	2
1	0	$1/3$
2	$1/3$	$1/3$



◆ 联合概率密度的性质：

◆ 1. $f(x,y) \geq 0$

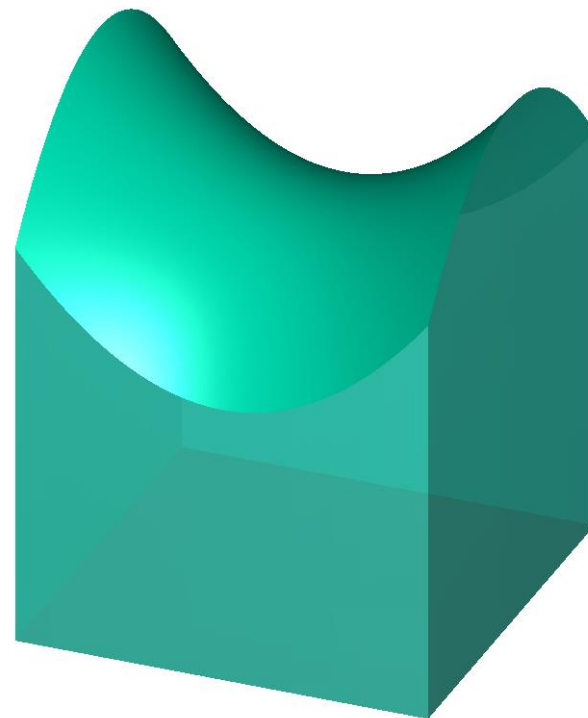
◆ 2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = F(\infty, \infty) = 1$

◆ 3. 设G是xOy平面上的区域，点 (X,Y) 落在G内的概率为

$$P\{(X,Y) \in G\} = \iint_G f(x,y) dx dy$$

◆ 4. 若点f(x,y)在点(x,y)连续，则有

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



例 2 设二维随机变量 (X, Y) 具有概率密度

$$f(x, y) = \begin{cases} 2e^{-(2x+y)}, & x > 0, y > 0, \\ 0, & \text{其他.} \end{cases}$$

(1) 求分布函数 $F(x, y)$; (2) 求概率 $P\{Y \leq X\}$.

解 (1)
$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(x, y) dx dy$$
$$= \begin{cases} \int_0^y \int_0^x 2e^{-(2x+y)} dx dy, & x > 0, y > 0, \\ 0, & \text{其他.} \end{cases}$$

即有
$$F(x, y) = \begin{cases} (1 - e^{-2x})(1 - e^{-y}), & x > 0, y > 0, \\ 0, & \text{其他.} \end{cases}$$

(2) 将 (X, Y) 看作是平面上随机点的坐标. 即有

$$\{Y \leq X\} = \{(X, Y) \in G\},$$

其中 G 为 xOy 平面上直线 $y=x$ 及其下方的部分, 如图 3-4. 于是

$$\begin{aligned} P\{Y \leq X\} &= P\{(X, Y) \in G\} = \iint_G f(x, y) dx dy \\ &= \int_0^\infty \int_y^\infty 2e^{-(2x+y)} dx dy = \frac{1}{3}. \quad \square \end{aligned}$$

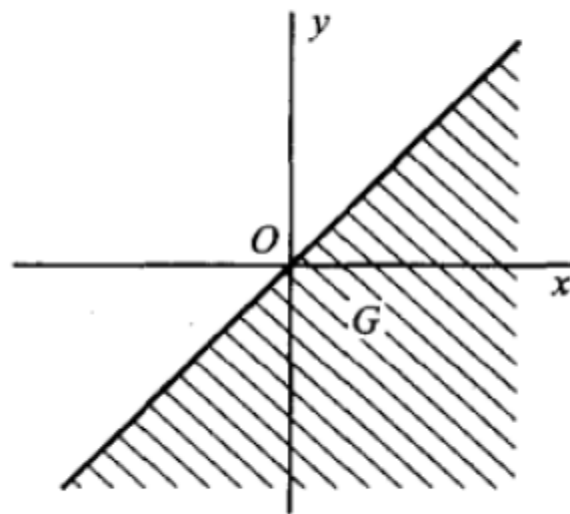


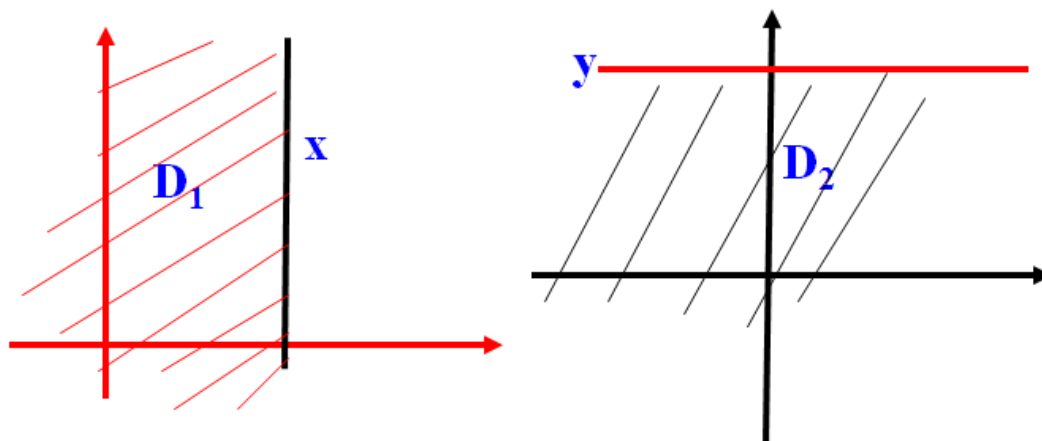
图 3-4

- ◆ 二维随机变量可以推广到多维的情况：
- ◆ 设 E 是一个随机试验，它的样本空间是 $S=\{e\}$ ，设 $X_1=X_1\{e\}$ ， $X_2=X_2\{e\}$ ，.....， $X_n=X_n\{e\}$ 是定义在 S 上的随机变量，由 X_i 构成的向量 (X_1, X_2, \dots, X_n) 叫做**多维随机向量或是多维随机变量**（Multidimensional random vector）
- ◆ 对于任意 x_1, x_2, \dots, x_n ，函数 $F(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}$ 称为 n 维随机变量的分布函数

- ◆ 在多维随机变量中，将 X ， Y 各自的分布称为边缘分布函数 (Marginal distribution)，分别记为 F_X ， F_Y

$$F_X = P\{X \leq x\} = P\{X \leq x, Y < \infty\} = F(x, \infty)$$

$$F_Y = P\{Y \leq y\} = P\{X < \infty, Y \leq y\} = F(\infty, y)$$



设 (X, Y) 是二维离散型随机变量,其(联合)分布律为:

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

则关于X的边缘分布律为:

$$p_{i\cdot} = P\{X = x_i\} = \sum_j p_{ij} \quad (i = 1, 2, \dots)$$

关于Y的边缘分布律为:

$$p_{\cdot j} = P\{Y = y_j\} = \sum_i p_{ij} \quad (j = 1, 2, \dots)$$

- ◆ 边缘分布律具有一维分布律的性质
- ◆ 联合分布律唯一决定边缘分布律. 具体求法是将联合分布律写成表格形式, 然后各行分别相加得关于X的分布律; 各列相加得Y的分布律

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

例子

- ◆ 已知(X, Y)的联合分布律如图
- ◆ 求X与Y的(边缘)分布律.
- ◆ 将联合分布表中的各行概率分别相加得X的分布律:

$$P(X = -1) = 0 + \frac{3}{30} + \frac{2}{30} = \frac{1}{6}$$

$$P(X = 2) = \frac{3}{30} + \frac{6}{30} + \frac{6}{30} = \frac{3}{6}$$

$$P(X = 3) = \frac{2}{30} + \frac{6}{30} + \frac{2}{30} = \frac{2}{6}$$

即：

$$\begin{pmatrix} X & -1 & 2 & 3 \\ p_{i.} & \frac{1}{6} & \frac{3}{6} & \frac{2}{6} \end{pmatrix}$$

$x \backslash y$	-1	2	3
-1	0	$\frac{3}{30}$	$\frac{2}{30}$
2	$\frac{3}{30}$	$\frac{6}{30}$	$\frac{6}{30}$
3	$\frac{2}{30}$	$\frac{6}{30}$	$\frac{2}{30}$

同理可得

$$\begin{pmatrix} Y & -1 & 2 & 3 \\ p_{.j} & \frac{1}{6} & \frac{3}{6} & \frac{2}{6} \end{pmatrix}$$

- ◆ 对于连续型随机变量(X,Y)，它的联合概率密度为 $f(x,y)$ ，则关于X和关于Y的边缘概率密度 (Marginal probability density) 如下：

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

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

- ◆ 边缘分布函数与边缘概率密度的关系：

$$F_X(x) = F(x, +\infty) = \int_{-\infty}^x \left[\int_{-\infty}^{+\infty} f(u, v) dv \right] du$$

$$= \int_{-\infty}^x f_X(u) du$$

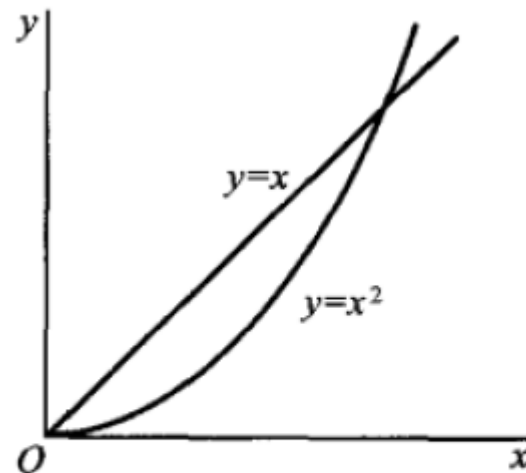
$$F_Y(y) = F(+\infty, y) = \int_{-\infty}^y \left[\int_{-\infty}^{+\infty} f(u, v) du \right] dv$$

$$= \int_{-\infty}^y f_Y(v) dv$$

例 2 设随机变量 X 和 Y 具有联合概率密度
(图 3-5)

$$f(x, y) = \begin{cases} 6, & x^2 \leq y \leq x, \\ 0, & \text{其他.} \end{cases}$$

求边缘概率密度 $f_X(x), f_Y(y)$.



$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \begin{cases} \int_{x^2}^x 6 dy = 6(x - x^2), & 0 \leq x \leq 1, \\ 0, & \text{其他.} \end{cases}$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \begin{cases} \int_y^{\sqrt{y}} 6 dx = 6(\sqrt{y} - y), & 0 \leq y \leq 1, \\ 0, & \text{其他.} \end{cases} \end{aligned}$$

- ◆ 考虑一大群人，从中随机挑选一个人，记此人的身高和体重分别为 X 和 Y ，则 X 和 Y 是随机变量。
- ◆ X 和 Y 有相应的分布函数 $F_X(x)$ 和 $F_Y(y)$ 。
- ◆ 如果已知 X 的取值是160cm，则此时 Y 也有一个分布函数 $F^*_Y(y)$ 。
- ◆ 问题： $F_Y(y)$ 与 $F^*_Y(y)$ 相等吗？在 X 未知的情况下，抽到的人体重小于60kg的概率和在已知 $X=160\text{cm}$ 的情况下，抽到抽到的人体重小于60kg的概率相等吗？



- ◆ 回顾条件概率： $P(A|B) = \frac{P(AB)}{P(B)}, P(B) > 0$
- ◆ 在X未知的情况下，抽到的人体重小于60kg的概率： $P(X \text{取任意值}, Y \leq 60)$
- ◆ 已知 $X=160\text{cm}$ 的情况下，抽到抽到的人体重小于60kg的概率：

$$P(Y \leq 60|X = 160) = \frac{P(X = 160, Y \leq 60)}{P(X = 160)}$$

- ◆ 设二维随机变量 (X, Y) ，条件概率 $P(X \leq x|Y=y)$ 可以看做是在 $Y=y$ 的条件下，X的分布函数 $F_{X|Y}(x)$

- ◆ 设离散二维随机变量 (X, Y) 的分布律为

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

- ◆ 关于X和关于Y的边缘分布律分别为

$$p_{i\cdot} = P\{X = x_i\} = \sum p_{ij} \quad (i = 1, 2, \dots)$$

$$p_{\cdot j} = P\{Y = y_j\} = \sum_i p_{ij} \quad (j = 1, 2, \dots)$$

- ◆ 则事件 $\{X = x_i | Y = y_j\}$ 的概率由条件概率的定义可知：

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

- ◆ 称上式为在 $Y=y_j$ 条件下随机变量X的**条件分布律**

- ◆ 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
 - (1) X, Y 的联合分布律；
 - (2) $X=1$ 时 Y 的条件分布律；
 - (3) $Y=0$ 时 X 的条件分布律。
- ◆ 解： 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

- ◆ 由于 $P(X=1)=7/15$ ，故在 $X=1$ 的条件下， Y 的分布律为

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

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

- ◆ 同理 $P(Y=0)=1/3$ ，故在 $Y=0$ 的条件下， X 的分布律为：

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

- ◆ 对于连续型随机变量 (X, Y) , 其联合概率密度为 $f(x, y)$, (X, Y) 关于 Y 的边缘概率密度为 $f_Y(y)$ 。若对固定的 y , $f_Y(y) > 0$, 则称 $\frac{f(x, y)}{f_Y(y)}$ 为在 $Y=y$ 条件下 X 的 **条件概率密度** (**Conditional probability density**) , 记为 $f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$
- ◆ 称 $F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(t|y) dt = \int_{-\infty}^x \frac{f(t, y)}{f_Y(y)} dt$ 为在 $Y=y$ 条件下 X 的 **条件分布函数**

- ◆ 设 G 是平面上的有界区域,其面积为 A 。若二维随机变量 (X,Y) 的概率密度为

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

- ◆ 则称 (X,Y) 在 G 上服从均匀分布。现设二维随机变量 (X,Y) 在圆域 $x^2 + y^2 \leq 1$ 上服从均匀分布,求条件概率密度 $f_{X|Y}(x|y)$ 。

解 由假设随机变量 (X,Y) 具有概率密度

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

且有边缘概率密度

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

于是当 $-1 < y < 1$ 时有

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

$$f_Y(y) = \int_{-\infty}^{\infty} \frac{1}{\pi} dx = \int_{x^2+y^2 \leq 1} \frac{1}{\pi} dx = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{1}{\pi} dx = \frac{x}{\pi} \Big|_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} = \frac{2}{\pi} \sqrt{1-y^2}, -1 \leq y \leq 1$$

积分对象是x，要确定x的积分范围，此时y可以看做是一个单纯的常数，而不是自变量

最后得出的函数表达式中只含有y，是y的边缘概率密度

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

最后得到的条件概率密度是一个关于y的函数

当 $y=0$ 和 $y=\frac{1}{2}$ 时 $f_{X|Y}(x|y)$ 的图形分别如图 3-6, 图 3-7 所示.

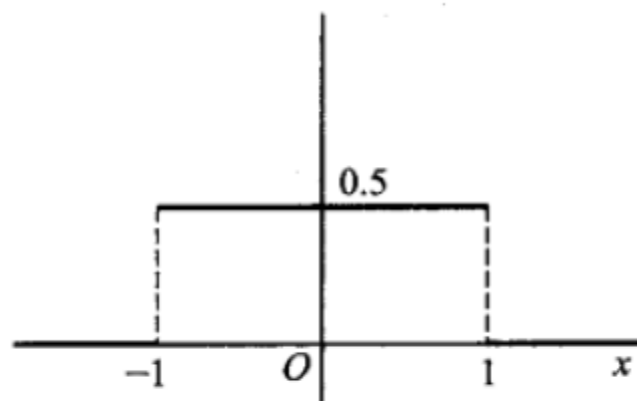


图 3-6

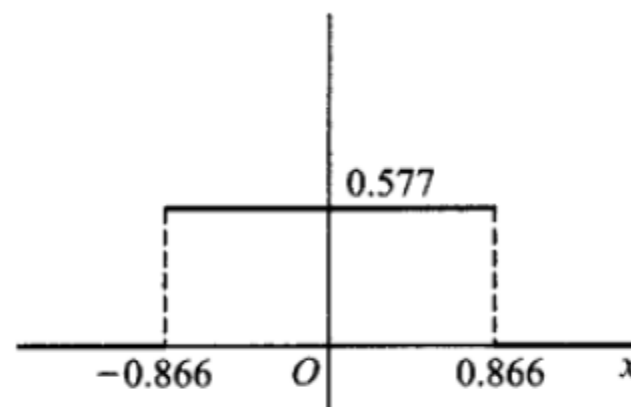
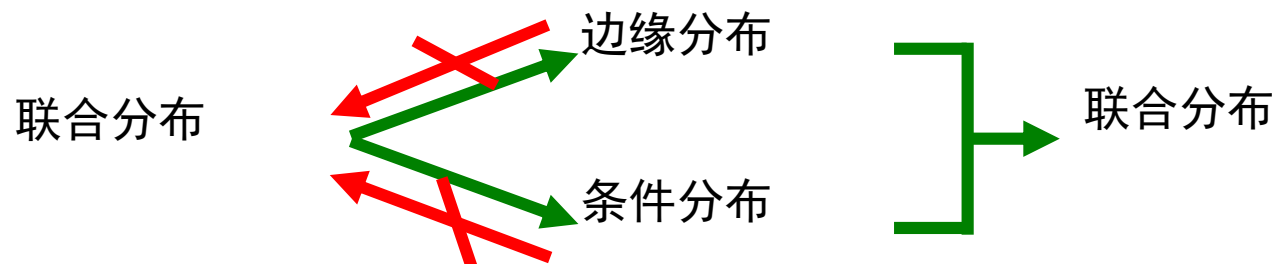


图 3-7

- ◆ 联合分布可以唯一地确定边缘分布和条件分布



- ◆ 回顾事件的独立性：
- ◆ 对于事件A，B，若 $P(AB)=P(A)P(B)$ 成立，则称事件A，B相互独立
- ◆ 类比随机变量的独立性：
- ◆ 对于随机变量 X, Y ，记 $A=\{X \leq x\}$ ， $B=\{Y \leq y\}$ ，若对于任意的 x, y ，都有 $P(AB)=P(A)P(B)$ ，即 $P\{X \leq x, Y \leq y\}=P\{X \leq x\}P\{Y \leq y\}$ ，即 $F(x, y) = F_X(x)F_Y(y)$ 成立，称随机变量 X 和 Y 是相互独立的。
- ◆ 当 (X, Y) 是连续型随机变量时 $F(x, y) = F_X(x)F_Y(y)$ 等价于 $f(x, y) = f_X(x)f_Y(y)$ 几乎处处成立
- ◆ 当 (X, Y) 是离散型随机变量时 $F(x, y) = F_X(x)F_Y(y)$ 等价于 $P\{X = x_i, Y = y_j\} = P\{X =$

8. 甲、乙两人投篮,投中的概率分别为 0.6,0.7. 今各投 3 次. 求

(1) 两人投中次数相等的概率;

(2) 甲比乙投中次数多的概率.

- ◆ 对于开头的例子,用 X 记录甲投中的次数, Y 记录乙投中的次数。因为两人投篮是互不影响的,所以 X 与 Y 是相互独立的。可以根据公式 $P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}$ 验证 X 与 Y 的独立性。



1. 在一箱子中装有 12 只开关,其中 2 只是次品,在其中取两次,每次任取一只,考虑两种试验:(1)放回抽样;(2)不放回抽样. 我们定义随机变量 X, Y 如下:

$$X = \begin{cases} 0, & \text{若第一次取出的是正品,} \\ 1, & \text{若第一次取出的是次品;} \end{cases}$$

$$Y = \begin{cases} 0, & \text{若第二次取出的是正品,} \\ 1, & \text{若第二次取出的是次品.} \end{cases}$$

试分别就(1)、(2)两种情况,写出 X 和 Y 的联合分布律.

◆ (1) 放回抽样:

◆ (X, Y) 的所有可能取值: $(0, 0)$ $(0, 1)$ $(1, 0)$ $(1, 1)$

◆ $P(X=0, Y=0) = \frac{10}{12} * \frac{10}{12} = \frac{25}{36}$; $P(X=0, Y=1) = \frac{10}{12} * \frac{2}{12} = \frac{5}{36}$

◆ $P(X=1, Y=0) = \frac{2}{12} * \frac{10}{12} = \frac{5}{36}$; $P(X=1, Y=1) = \frac{2}{12} * \frac{2}{12} = \frac{1}{36}$

- ◆ 故放回抽样时 (X,Y) 的联合分布律为

Y \ X	0	1
0	25/36	5/36
1	5/36	1/36

- ◆ (2) 不放回抽样：

◆ $P(X=0, Y=0) = \frac{10}{12} * \frac{9}{11} = \frac{45}{66}$; $P(X=0, Y=1) = \frac{10}{12} * \frac{2}{11} = \frac{10}{66}$

◆ $P(X=1, Y=0) = \frac{2}{12} * \frac{10}{11} = \frac{10}{66}$; $P(X=1, Y=1) = \frac{2}{12} * \frac{1}{11} = \frac{1}{66}$

Y \ X	0	1
0	45/66	5/33
1	5/33	1/66

- ◆ 对于上述的X与Y，它们相互独立吗？
- ◆ (1) 放回抽样：先求出X与Y的边缘分布律

Y \ X	0	1	Y
0	25/36	5/36	5/6
1	5/36	1/36	1/6
X	5/6	1/6	

- ◆ 根据边缘分布律与联合分布律，很容易验证 $P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}$ 的正确性，所以X与Y在放回抽样下是相互独立的。

- ◆ (2) 无放回抽样 :

Y \ X	0	1	Y
0	45/66	5/33	5/6
1	5/33	1/66	1/6
X	5/6	1/6	

- ◆ 很明显 , $P\{X=0,Y=0\}=45/66 \neq 25/36 = P\{X=0\}P\{Y=0\}$
- ◆ 故在无放回抽样下 , X与Y不相互独立

- ◆ **Dataguru（炼数成金）**是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- ◆ 关于逆向收费式网络的详情，请看我们的培训网站 <http://edu.dataguru.cn>



Thanks

FAQ时间