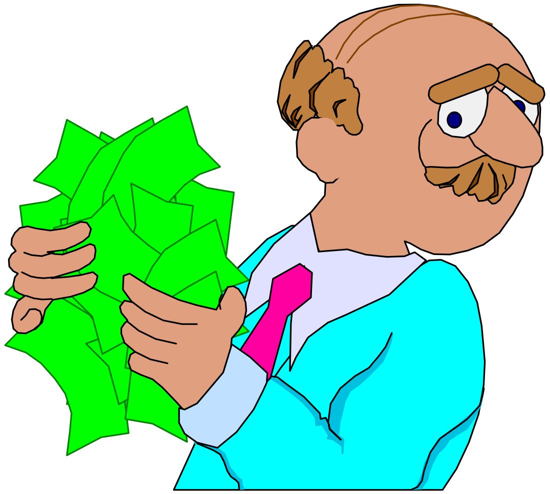
数理统计

1.2:随机变量及其分布

1.随机变量的定义:

如果对一切 $\omega \in \Omega$,有实值函数 $\xi = \xi(\omega) \in R$ 与之对应,则称实值函数 $\xi = \xi(\omega)$ （$\omega \in \Omega$）为一个随机变量.

注意：任意 $x \in R$，$\{\xi(\omega) \le x\}$为一个事件，



直观地说：随机变量 就是用来表示随机试验结果的变量.一般记为:ξ, η . . . . 或 X，Y，. . . .

随机变量的分类：

按随机变量可能取值范围

1） 离散型随机变量（有限或可列个值）

2) 连续型随机变量（某一区间内）

按随机变量的维数

1） 一维随机变量

2) 多维随机变量

离散型随机变量
及其分布列

2.定义：离散型随机变量 X 的概率分布

随机变量 ***X*** 的取值为至多可列个，则 **X** 称为离散型随机变量.

称 $P(X=x_k)=p_k, \quad k=1,2,\cdots$ 为离散型随机变量 ***X*** 的概率分布或分布律.

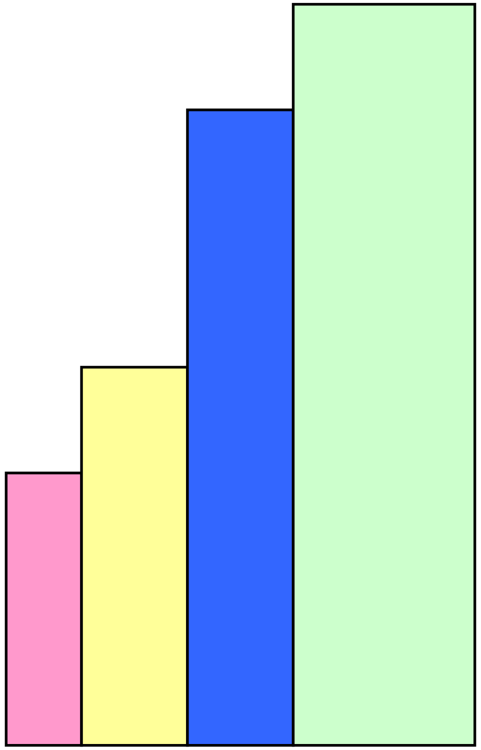
分布律一般用表格形式来表示：

X	x_1	x_2	$\cdots$	x_n	$\cdots$
p_k	p_1	p_2	$\cdots$	p_n	$\cdots$

性质：

①非负性： $p_k \geq 0, \quad k = 1,2,\cdots.$

②规范性： $\sum_k p_k = 1.$



例 1. 袋中有 2 个白球，3 个黑球，每次从中任取一球不放回，直到取到白球为止，求取球次数的概率分布。

设 Y 为取到白球时的取球次数，$Y=1,2,3,4$，则

$$P(Y=1)=\frac{2}{5}=0.4, \qquad P(Y=2)=\frac{3}{5}\cdot\frac{2}{4}=0.3,$$

$$P(Y=3)=\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{2}{3}=0.2,$$

$$P(Y=4)=\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{1}{3}\cdot\frac{2}{2}=0.1,$$

Y	1	2	3	4
P	0.4	0.3	0.2	0.1

显然，$\sum_{m=1}^{4} P(Y=m)=0.4+0.3+0.2+0.1=1$

3.定义：随机变量的分布函数：

设 ξ 是一个随机变量，对于任意的实数 $x \in R$，则概率 $P\{\omega \mid \xi(\omega) \le x\} = P\{\xi \le x\}$ 与 **x** 有关,它是 **x** 的一个函数.我们称:

$$\boldsymbol{F_\xi(x) = P\{\omega \mid \xi(\omega) \le x\} = P\{\xi \le x\}},$$

为 ξ 的分布函数，可简记为 $\boldsymbol{F(x)}$。

注意：

任意一个随机变量都存在唯一的分布函数，但同样的分布函数，可以对应不同的随机变量.

分布函数的性质

（1） 非降性：$F(x)$是单调非降函数，

即对任意的$x_1 < x_2$，有$F(x_1) \le F(x_2)$；

(2) 规范性：对任意$x \in R$,有$0 \le F(x) \le 1$，且

$F(-\infty)$=0，$F(+\infty)$=1；

(3) 右连续：即对任意的$x_0 \in R$,有

$$\lim_{x \to x_0^+} F(x) = F(x_0)\text{，即 } F(x_0+0) = F(x_0)$$

$$\lim_{x \to x_0^-} F(x) = F(x_0 -) \le F(x_0)\text{。}$$

例 2. 掷一枚硬币，把出现正面与反面分别记为 ＋1 与－1，这样可以把掷硬币的结果用随机变量ξ来表示，显然对均匀硬币 $P(\xi=\pm 1)=\dfrac{1}{2}$，

（1）试求ξ的分布函数$F_1(x)$； (2) 若令$\eta=-\xi$，再求η的分布函数$F_2(x)$。

解：$$F_1(x)=\begin{cases} 0, & x\in(-\infty,-1) \\ \dfrac{1}{2}, & x\in[-1,1) \\ 1, & x\in[+1,+\infty) \end{cases}$$

注意到$P(\eta=\mp 1)=P(\xi=\pm 1)=\dfrac{1}{2}$

$\therefore \boldsymbol{F_\eta(x)=F_2(x)=F_1(x)}$。

利用分布函数计算概率：

$$P(a < \xi \le b) = F(b) - F(a)$$

$$P(\xi > b) = F(+\infty) - F(b) = 1 - F(b).$$

$$P(\xi < a) = \lim_{n\to\infty} P(\xi \le a - \frac{1}{n})$$

$$= \lim_{n\to\infty} F(a - \frac{1}{n}) = F(a-0)$$

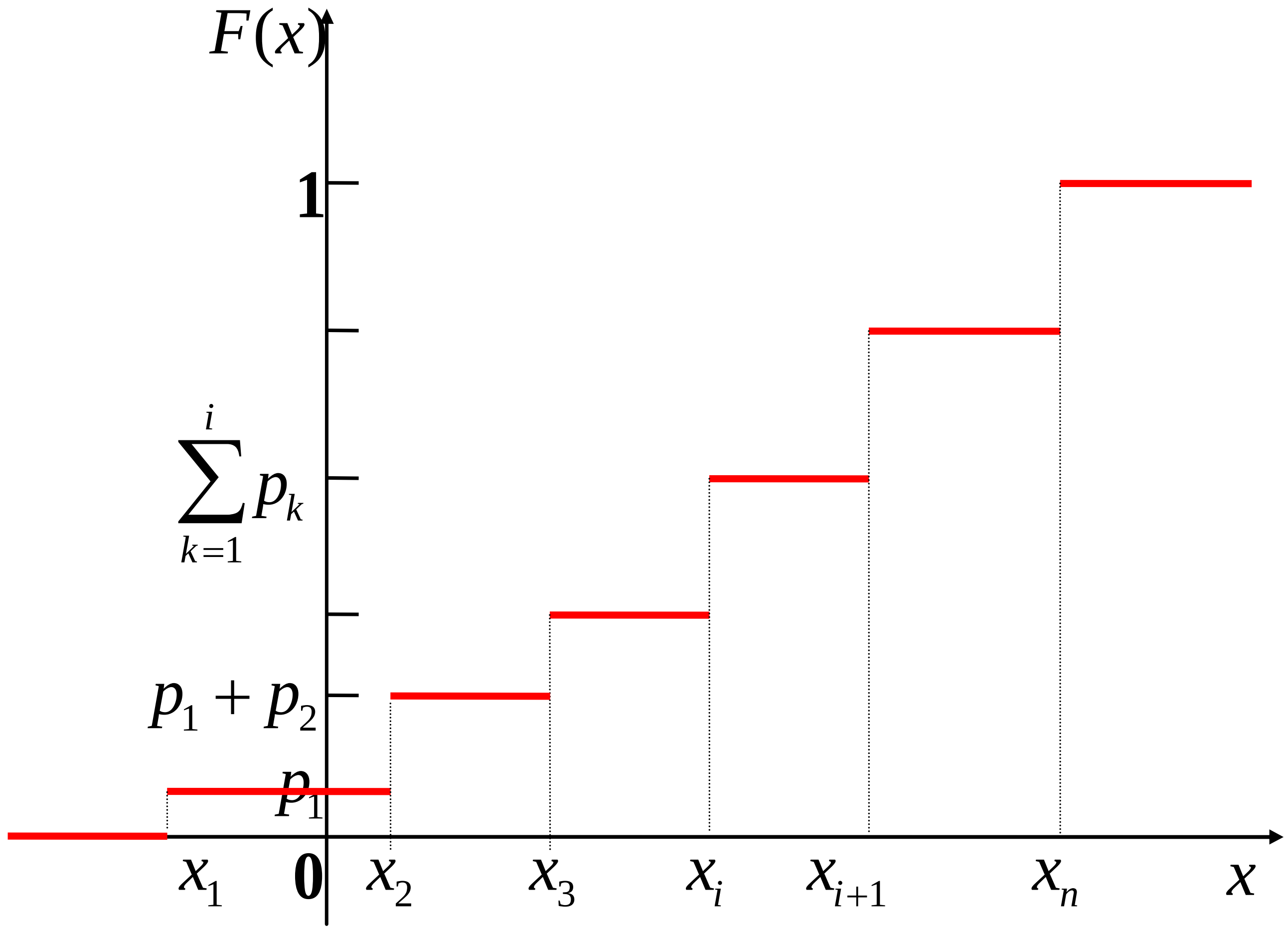
利用分布函数计算概率：

$$P(\xi = a) = P(\xi \leq a) - P(\xi < a)$$
$$= F(a) - F(a-0)$$

$$P(a \leq \xi < b) = F(b-0) - F(a-0)$$

★对离散型随机变量，已知分布律，可求它的分布函数，反之亦然。

分布函数图形特点：右连续，阶梯形



例 **3.** 随机变量 X 的分布律为

X	**-1**	**2**	**3**
p_k	**0.25**	**0.5**	**a**

求 a,及 X **的分布函数，**

并求 $P\{X \le \frac{1}{2}\}, \quad P\{\frac{3}{2} < X \le \frac{5}{2}\},$

解：a=0.25； X **的分布函数为**

$$F(x) = \begin{cases} 0, & x < -1 \\ 0.25, & -1 \le x < 2 \\ 0.75, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

$$P\{X \le \frac{1}{2}\} = P\{X = -1\} = 0.25,$$

$$P\{\frac{3}{2} < X \le \frac{5}{2}\} = P\{X = 2\} = 0.5,$$

或利用分布函数求 $P\{X \le \frac{1}{2}\} = F(\frac{1}{2}) = 0.25$，

$$P\{\frac{3}{2} < X \le \frac{5}{2}\} = F(\frac{5}{2}) - F(\frac{3}{2}) = 0.75 - 0.25 = 0.5$$

连续型随机变量

4.连续型随机变量的概率密度

1)定义**:**如果对于随机变量 X 的分布函数 $F(x)$ 存在非负函数 $\varphi(x)$，使对于任意实数 $x \in R$，有 $F(x) = \int_{-\infty}^{x} \varphi(t)dt$，则称 X 为连续型随机变量，其中，$\varphi(x)$ 称为 X 的概率密度函数。

2)性质**:**

(1) $\varphi(x) \geq 0$**;**

(2) $\int_{-\infty}^{\infty} \varphi(x)dx = 1$**;**

3)已知概率密度求分布函数

$$P\{x_1 \le X < x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} \varphi(x)dx.$$

4)已知分布函数求概率密度：

若 $\varphi(x)$ 在点 x 处连续，则 $F'(x) = \varphi(x)$

当 $\Delta x \to 0$ 时 $P(x < X \le x + \Delta x) = \int_{x}^{x+\Delta x} \phi(x) \approx \phi(x)\Delta x \to 0$

5)结论：(1)对连续型随机变量 X， $P\{X=c\}=0$

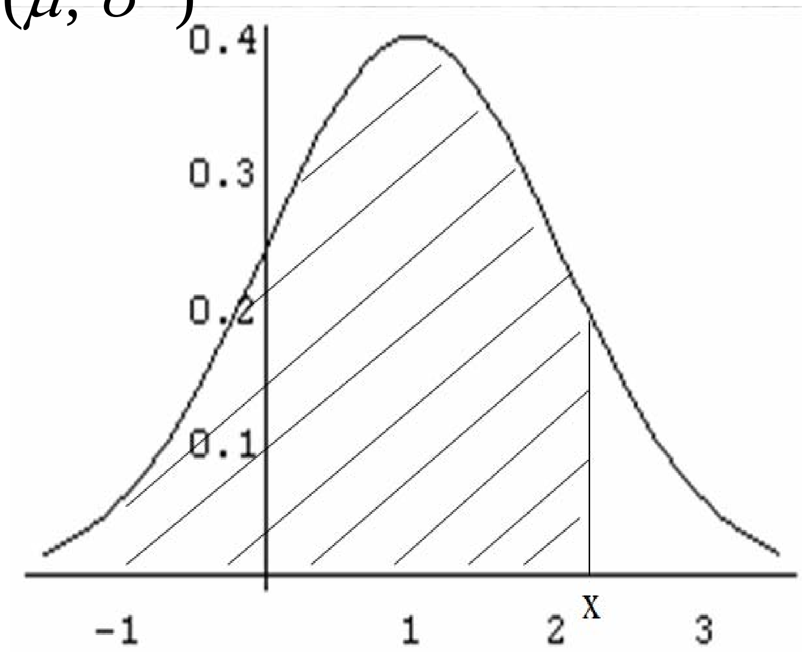
(2) $P(a\le X\le b)=P(a<X\le b)=P(a\le X<b)$
$=P(a<X<b)=F(b)-F(a)$

(3)连续型随机变量的分布函数是连续函数。

比如：随机变量ξ 的密度函数为:

$$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\quad(-\infty<x<+\infty)$$

此时我们称ξ 服从正态分布,记为 $\xi \sim N(\mu,\sigma^2)$



ξ 的密度函数如右图所示

满足 1）非负性,即 $p(x)\geq 0$；

2) 规范性,即p（x）与x轴围成的面积为 1

而 $P\{\xi \leq x\}=\int_{-\infty}^{x}p(t)dt=F_\xi(x)$就表示阴影部分的面积

例 **4.** 设随机变量 X 具有概率密度为

$$\varphi(x)=\begin{cases} ke^{-3x}, & x>0 \\ 0, & x\le 0 \end{cases},$$

试确定常数k，并求$P(X>0.1)$ 及$F(x)$。

解： **(1)** $\because \int_{-\infty}^{\infty}\varphi(x)dx=1$， $\therefore \int_{0}^{\infty}ke^{-3x}dx=1$，

$$k\cdot\left(-\frac{1}{3}\right)e^{-3x}\Big|_0^{\infty}=1, \qquad \therefore k=3,$$

即 $$\varphi(x)=\begin{cases} 3e^{-3x}, & x>0 \\ 0, & x\le 0 \end{cases}$$

(2) $P(X > 0.1) = \int_{0.1}^{\infty} \varphi(x)dx = \int_{0.1}^{\infty} 3e^{-3x}dx$

$$= -e^{-3x}\Big|_{0.1}^{\infty} = 0.7408$$

(3)当 $x \le 0$ 时， $F(x) = \int_{-\infty}^{x} 0dx = 0$

当 $x > 0$ 时， $F(x) = \int_{-\infty}^{0} 0dx + \int_{0}^{x} 3e^{-3\mathrm{t}}d\mathrm{t} = -e^{-3t}\big|_{0}^{x} = 1 - e^{-3x}$

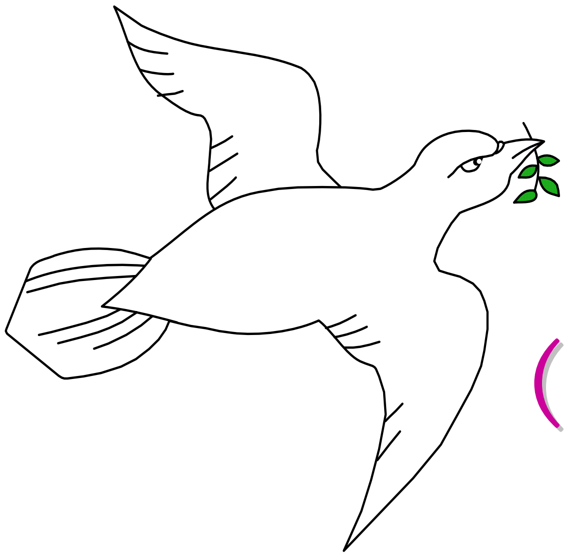
$$\therefore F(x) = \begin{cases} 1 - e^{-3x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

一般，随机变量 X 的分布密度为

$$\varphi(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0,\end{cases}$$

$\lambda>0$，则称 X 为指数分布，记为 $e(\lambda)$。

（常用在产品的寿命）



多维随机变量

(随机向量)及其分布

5.随机向量的定义:

设 $\{\xi_i(\omega)\}\quad i=1,2,\cdots,n$ 是 n 个随机变量，则称 $\xi(\omega)=(\xi_1(\omega),\xi_2(\omega),\cdots,\xi_n(\omega))$ 为 n 维随机向量或 n 维随机变量。

6.随机向量的分布函数:

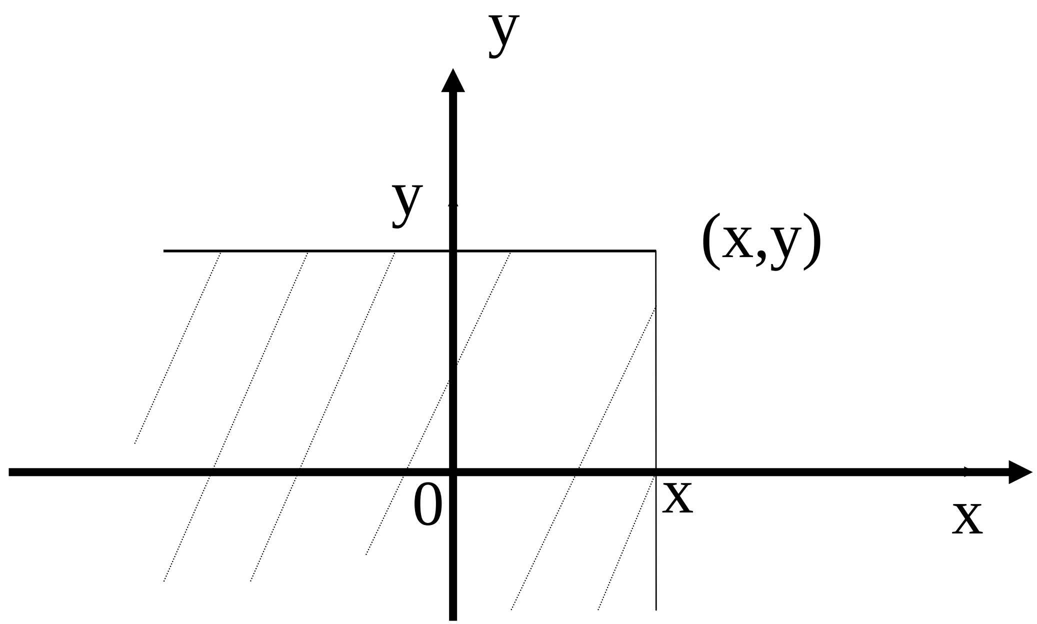
对任意实数 $x_1,x_2,\cdots,x_n$，称函数

$$F(x_1,x_2,\cdots,x_n)=P\{\xi_1\le x_1,\xi_2\le x_2,\cdots,\xi_n\le x_n\}$$

为随机向量 ξ 的（联合）分布函数。

二维随机变量(X,Y)的分布函数：

$$F(x,y)=P(X\le x,Y\le y)$$



二维分布函数$F(x,y)$的性质：

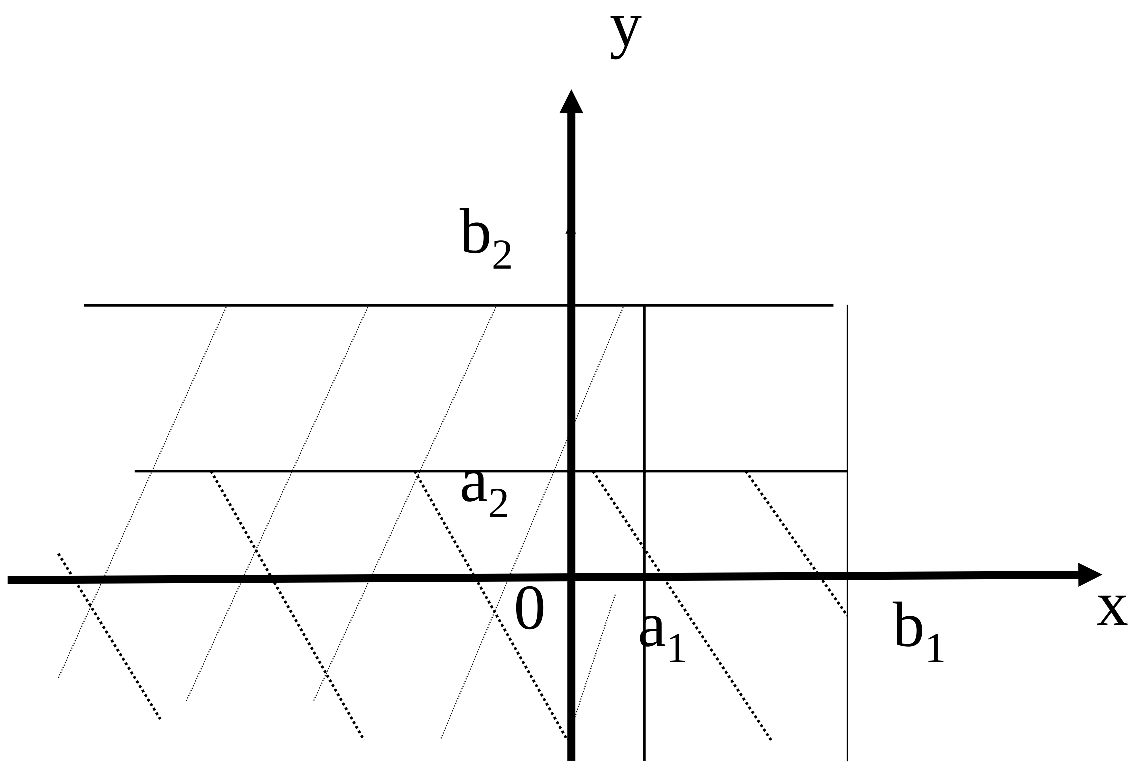
(1) $F(x,y)$是x**(**或y**)**的单调非减.

(2) $0 \le F(x,y) \le 1$，且$F(+\infty,+\infty)=1$

$F(-\infty,y)=0$，$F(x,-\infty)=0$，

(3) $F(x,y)$是x**(**或y**)**的右连续函数.

$$P(a_1 < X \le b_1, a_2 < Y \le b_2)$$
$$= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$



7. 离散型随机变量的联合分布列

若二维随机变量 (ξ,η) 的可能取值为有限个（或可列个）数对 (x_i, y_j) 时，其对应的概率：

$$p_{ij} = P\{\xi = x_i, \eta = y_j\}, \quad i, j = 1, 2, \cdots$$

满足规范性条件 $\sum_{i,\ j=1}^{+\infty} p_{ij} = 1$，则称 (ξ,η) 为二维离散型随机变量。

二维离散随机变量的联合分布表：

$X \backslash Y$	y_1	$y_2 \cdots$	$y_n \cdots$
x_1	p_{11}	$p_{12} \cdots$	$p_{1n} \cdots$
x_2	p_{21}	$p_{22} \cdots$	$p_{2n} \cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	$p_{m2} \cdots$	$p_{mn} \cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$\{p_{ij}\}$ 成为某随机变量 (ξ, η) 分布列的充要条件

仍为：**(1)**非负性：$p_{ij} \geq 0$；

(2)规范性：$\sum_{i,\ j=1}^{+\infty} p_{ij} = 1$。

8.二维连续随机变量的联合密度函数

设二维随机变量(ξ,η)的分布函数为$F(x,y)$，如果存在一个定义在$R\times R$上的二元非负可积函数$p(x,y)$，使得对$\forall\,(x,y)\in R\times R$，有

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)dudv$$

则称(ξ,η)为二维连续型随机变量，同时称$p(x,y)$为(ξ,η)的联合概率密度函数。

在$p(x,y)$的连续点有 $\dfrac{\partial^2 F(x,y)}{\partial x\partial y}=p(x,y)$

$p(x, y)$ 为密度函数的充要条件为：

(1)非负性： $p(x, y) \geq 0$；

(2)规范性： $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x, y)dxdy = 1$

★ (ξ,η) 落在平面区域 G 内的概率为：

$$P\{(\xi,\eta) \in G\} = \iint_G p(x, y)dxdy.$$

★几何上，此概率即为分布曲面之下，以区域 G 为底的曲顶柱体的体积。

例 5. 设(ξ,η)具有概率密度

$$p(x,y)=\begin{cases} ce^{-2x-3y} & , \quad x\ge 0,\ y\ge 0 \\ 0 & , \quad 其他 \end{cases}$$

试求：**(1)**常数c；（**2**）分布函数$F(x,y)$；

（**3**）$p(\xi>\eta)$。

解：（**1**）利用密度规范性得

$$\int_0^{+\infty}\int_0^{+\infty} ce^{-2x-3y}dxdy = c\int_0^{+\infty} e^{-2x}dx\int_0^{+\infty} e^{-3y}dy = 1$$

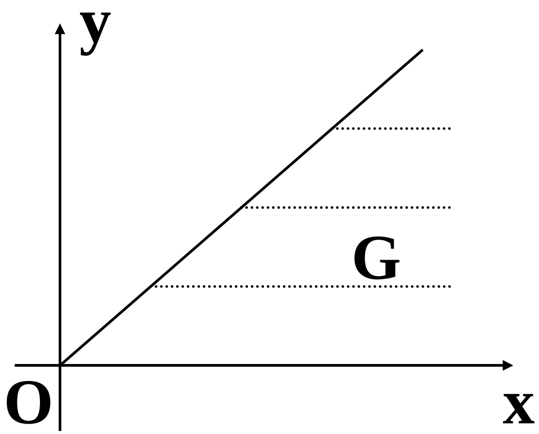
故$\boldsymbol{c=6}$。

（2） $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)dudv$

$$=\begin{cases}(1-e^{-2x})(1-e^{-3y}), & 0\le x<+\infty,\quad 0\le y<+\infty \\ 0 & \text{其他}\end{cases}$$

（3） $P(\xi>\eta)=\iint\limits_{(x,y)\in G}p(u,v)dudv$ ，

化成累次积分：



$$P(\xi>\eta)=\int_{0}^{+\infty}dx\cdot\int_{0}^{x}6e^{-2x}e^{-3y}dy$$

$$=\int_{0}^{+\infty}2e^{-2x}(1-e^{-3x})dx=\frac{3}{5}。$$

边际分布、独立性

9. 随机向量的边际分布(即随机向量的部分分量的分布)

假设二维离散随机变量(ξ,η)的概率分布为:

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

其分量ξ的分布:

$$P(\xi = x_i) = P\{\bigcup_{j=1}^{\infty}(\xi = x_i,\eta = y_j\} = \sum_{j=}^{\infty} p_{ij} = p_{i\cdot}$$

称为边际分布列;

同样,η的边际分布为:

$$P(\eta = y_j) = \sum_{i=1}^{\infty} P(\xi = x_i,\eta = y_j) = \sum_{i=1}^{\infty} p_{ij} = \mathrm{p}_{\cdot j}, \quad j = 1,2,\cdots$$

例 6. 已知 10 件产品中有 3 件一等品，5 件二等品，2 件三等品。从这批产品中任取 4 件产品，求其中一等品、二等品件数各自的分布律。

解：设 X 及 Y 分别是取出的 **4** 件产品中一等品及二等品的件数，则我们有

$$P(X=i,Y=j)=\frac{C_3^i C_5^j C_2^{4-i-j}}{C_{10}^4},$$

$$i=0,1,2,3; \qquad j=0,1,2,3,4; \qquad 4-i-j=0,1,2$$

即 $\quad i=0,1,2,3; \qquad j=0,1,2,3,4; \qquad i+j=2,3,4$

即

$X \backslash Y$	0	1	2	3	4	$p_{i\bullet}$
0	0	0	$\frac{10}{210}$	$\frac{20}{210}$	$\frac{5}{210}$	$\frac{35}{210}$
1	0	$\frac{15}{210}$	$\frac{60}{210}$	$\frac{30}{210}$	0	$\frac{105}{210}$
2	$\frac{3}{210}$	$\frac{30}{210}$	$\frac{30}{210}$	0	0	$\frac{63}{210}$
3	$\frac{2}{210}$	$\frac{5}{210}$	0	0	0	$\frac{7}{210}$
$p_{\bullet j}$	$\frac{5}{210}$	$\frac{50}{210}$	$\frac{100}{210}$	$\frac{50}{210}$	$\frac{5}{210}$	1

一等品件数 X 的分布律为：

X	0	1	2	3
$p_{i\bullet}$	$\frac{35}{210}$	$\frac{105}{210}$	$\frac{63}{210}$	$\frac{7}{210}$

二等品件数 Y 的分布律为：

Y	0	1	2	3	4
$p_{\bullet j}$	$\frac{5}{210}$	$\frac{50}{210}$	$\frac{100}{210}$	$\frac{50}{210}$	$\frac{5}{210}$

二维随机变量(ξ,η)的边缘分布函数

$$F_{\xi}(x) = P(\xi \le x) = P(\xi \le x, \eta < +\infty) = F(x, +\infty)$$

$$F_{\eta}(y) = P(\eta \le y) = P(\xi < +\infty, \eta \le y) = F(+\infty, y)$$

对二维连续型随机变量

ξ 的边缘分布函数：

$$F_{\xi}(x)=F(x,+\infty)=\int_{-\infty}^{x}dx\int_{-\infty}^{+\infty}p(x,y)dy$$

ξ 的边缘概率密度：

$$p_{\xi}(x)=\int_{-\infty}^{+\infty}p(x,y)dy$$

η 的边缘分布函数：

$$F_{\eta}(y)=F(+\infty,y)=\int_{-\infty}^{y}dy\int_{-\infty}^{+\infty}p(x,y)dx$$

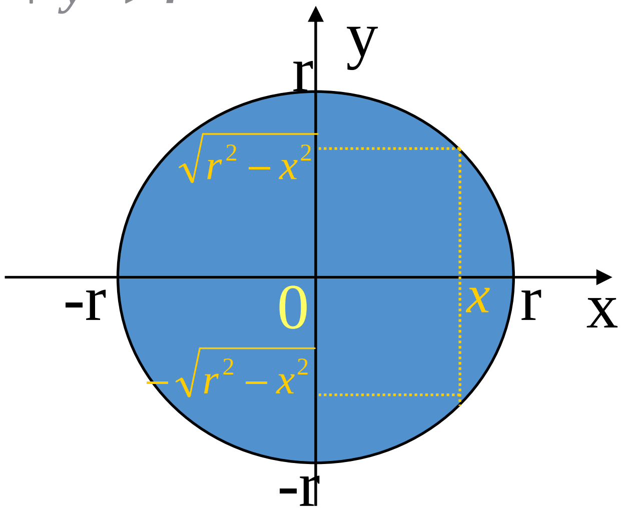
η 的边缘概率密度：

$$p_{\eta}(y)=\int_{-\infty}^{+\infty}p(x,y)dx$$

例 7. 设(X,Y)在以原点为中心，r为半径的圆域R上服从均匀分布，

$$p(x,y)=\begin{cases}\dfrac{1}{\pi r^2}, & x^2+y^2\le r^2\\ 0, & x^2+y^2>r^2\end{cases}$$

求 X 及 Y 边缘概率密度。



解：$p_X(x)=\int_{-\infty}^{\infty}p(x,y)dy$

当$|x|\le r$时，

$$p_X(x)=\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}}\frac{1}{\pi r^2}dy=\frac{2\sqrt{r^2-x^2}}{\pi r^2}$$

当$|x|>r$时，$p_X(x)=0$

$$\therefore\ p_X(x)=\begin{cases}\dfrac{2\sqrt{r^2-x^2}}{\pi r^2}, & |x|\le r\\ 0, & |x|>r\end{cases}$$

同理 $$p_Y(y)=\begin{cases}\dfrac{2\sqrt{r^2-y^2}}{\pi r^2}, & |y|\le r\\ 0, & |y|>r\end{cases}$$

10.条件分布

离散型随机变量的条件分布

$$P_{\xi|\eta}(x_i \mid y_j) = P(\xi = x_i \mid \eta = y_j)$$

$$= \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} = \frac{p_{ij}}{p_{\bullet j}}, \quad i = 1,2,\cdots$$

连续型随机变量的条件分布

$$p_{\xi|\eta}(x \mid y) = \frac{p(x,y)}{p_\eta(y)}$$

称 $p_{\xi|\eta}(x \mid y)$ 为在 $\eta = y$ 条件下，连续随机变量 ξ 的条件概率密度函数。

例 **8** （二维正态分布）设随机变量(ξ,η)具有密度函数（$\sigma_1>0$，$\sigma_2>0,|\rho|\le 1$）：

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

求边际分布密度

解：$p_{\xi}(x)=\int_{-\infty}^{+\infty} p(x, y)dy$

作变换 $\dfrac{x-\mu_1}{\sigma_1}=u, \dfrac{y-\mu_2}{\sigma_2}=v$

则$p_{\xi}(x)=\dfrac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}[u^2-2\rho uv+v^2]}dv$

$$=\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty} e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}}\cdot e^{-\frac{u^2}{2}}dv$$

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

$$x\in(-\infty, +\infty)$$

即$\xi\sim N(\mu_1, \sigma_1^2)$，类似地$\eta\sim N(\mu_2, \sigma_2^2)$。

11. 随机变量的独立性

随机变量$\xi_1, \xi_2, \cdots, \xi_n$相互独立表示他们取值互不影响.用分布函数来表示:

随机变量$\xi_1, \xi_2, \cdots, \xi_n$相互独立等价于联合分布函数等于它们边际分布函数的乘积:

$$F_{\xi_1\cdots\xi_n}(x_1, x_2, \cdots x_n) = \prod_{i-1}^{n} F_{\xi_i}(x_i)$$

二维离散型随机变量的独立性

$$P(\xi = x_i, \eta = y_j) = P(\xi = x_i) \cdot P(\eta = y_j),$$

$$i = 1,2,\cdots,;\ j = 1,2,\cdots.$$

二维连续型随机变量的独立性

$$p(x, y) = p_\xi(x) \cdot p_\eta(y)$$

例 9.　设二维随机变量(ξ,η)的联合分布列为

$\xi \backslash \eta$	**1**	**2**	**3**
1	$1/6$	$1/9$	$1/18$
2	$1/3$	α	β

问其中的α，β取什么值时ξ与η独立？

解：由规范性可知$1=\frac{1}{6}+\frac{1}{9}+\frac{1}{18}+\frac{1}{3}+\alpha+\beta$，

从而$\beta=\frac{1}{3}-\alpha$.

再求边际分布

$$P(\xi=1)=\frac{1}{6}+\frac{1}{9}+\frac{1}{18}=\frac{1}{3}, \quad P(\xi=2)=1-\frac{1}{3}=\frac{2}{3},$$

$$P(\eta=1)=\frac{1}{6}+\frac{1}{3}=\frac{1}{2}, \quad P(\eta=2)=\frac{1}{9}+\alpha,$$

$$P(\eta=3)=\frac{7}{18}-\alpha$$

根据独立性： $P(\xi=1,\eta=2)=P(\xi=1)P(\eta=2)$

即 $1/9=1/3\times(\frac{1}{9}+\alpha)$ 。 $\alpha=2/9$ $\beta=1/3-\alpha=1/9$

例 10. 设 (X,Y) 在以原点为中心，r 为半径的圆域 R 上服从均匀分布，二维概率密度为：

$$p(x,y)=\begin{cases}\dfrac{1}{\pi r^2}, & x^2+y^2\le r^2\\ 0, & x^2+y^2>r^2\end{cases}$$

问：X 与 Y 是否相互独立？

解：已经求出 X,Y 的边缘分布密度分别为：

$$p_X(x)=\begin{cases}\dfrac{2\sqrt{r^2-x^2}}{\pi r^2}, & |x|\le r\\ 0, & |x|>r\end{cases}$$

$$p_Y(y)=\begin{cases}\dfrac{2\sqrt{r^2-y^2}}{\pi r^2}, & |y|\le r\\ 0, & |y|>r\end{cases}$$

显然，$p(x,y)\neq p_X(x)p_Y(y)$ 因此，X 与 Y 不相互独立.

随机变量函数
的概率分布

随机变量函数的分布

问题：如何从已知分布的随机变量出发，去求其函数的分布？

比如：测量圆轴直径 d，而关心的是截面积 A，$A=\frac{1}{4}\pi d^2$。

已知直径 d 的分布，要求截面积 A 的分布。

注意：随机变量的函数仍然是随机变量。

例 **11.** 已知随机变量 X 的分布律为

X	-1	0	1	2
p_k	0.2	0.3	0.1	0.4

求 $(1) Y=-2X$， $(2) Y=(X-1)^2$ 的分布律。

解：$(1) Y=-2X$ 是单值函数， $f(x_i)$ 各不相等，所以有

$$P(Y=2)=P(X=-1)=0.2$$

$$P(Y=0)=P(X=0)=0.3$$

$$P(Y=-2)=P(X=1)=0.1$$

$$P(Y=-4)=P(X=2)=0.4$$

Y	2	0	-2	-4
p_k	0.2	0.3	0.1	0.4

$(2) Y=(X-1)^2$ 不是单值函数,

$$P(Y=0)=P(X=1)=0.1,$$

$$P(Y=1)=P(X=0)+P(X=2)=0.3+0.4=0.7,$$

$$P(Y=4)=P(X=-1)=0.2$$

即

Y	0	1	4
p_k	0.1	0.7	0.2

注:离散型随机变量(包括多维)的函数分布都可按上例的方法来求解

例 12. 若随机变量ξ具有如下分布率:

$$P\{\xi=k\}=\frac{\lambda^k}{k!}e^{-\lambda} \quad (k=0,1,2\ldots)$$

则称ξ服从参数为λ的波松分布,记为$\xi \sim P(\lambda)$.

设 $X \sim P(\lambda_1)$， $Y \sim P(\lambda_2)$， X 与 Y 独立.

求 $Z = X + Y$ 的分布。

解：Z的一切可能取值为0,1,2,3...

$$P(Z=k)=\sum_{i=0}^{k}P(X=i)P(Y=k-i)$$

$$=\sum_{i=0}^{k}\frac{\lambda_1^i}{i!}e^{-\lambda_1}\cdot\frac{\lambda_2^{k-i}}{(k-i)!}e^{-\lambda_2}=\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\lambda_1^i\lambda_2^{k-i}$$

$$=\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}C_k^i\lambda_1^i\lambda_2^{k-i}=\frac{e^{-(\lambda_1+\lambda_2)}}{k!}(\lambda_1+\lambda_2)^k$$

$$=\frac{(\lambda_1+\lambda_2)^k}{k!}e^{-(\lambda_1+\lambda_2)}$$

$$\therefore Z\sim P(\lambda_1+\lambda_2)$$

结论：独立的服从泊松分布的随机变量之和仍服从泊松分布，且参数为前两个参数之和。

连续型随机变量函数的分布

问题：如何根据 X 的密度函数 $\boldsymbol{p_X(x)}$ 寻找 $Y = f(X)$ 的分布?

求解的基本思想:

从有 **Y** 的分布函数入手,把 **Y** 的分布函数 **F(y)** 表示为 **X** 的密度函数在对应区域上的积分:

$$F_Y(y) = P(Y \le y) = P(f(X) \le y) = \int_{f(x)\le y} p_X(x)dx$$

两边再对 y 求导即得 **Y** 的密度函数 $p_Y(y)$,这个方法对多维连续型随机变量的函数也适用

例 13. 设 $X \sim N(\mu,\sigma^2)$，即

$$p_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < +\infty)$$

求 $Y=a+bX,(b\neq 0)$ 的概率密度。

解：不妨设 b>0,当 b<0 时同理可求:

$$F_Y(y)=P\{Y\le y\}=P\{f(X)\le y\}=P\{a+bX\le y\}$$

$$=P\{X\le\frac{y-a}{b}\}=\int_{-\infty}^{\frac{y-a}{b}}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$p_Y(y) = F_Y^{'}(y) = \frac{1}{\sqrt{2\pi}\sigma b} e^{-\frac{(y-a-\mu b)^2}{2\sigma^2 b^2}}$$

$$\therefore Y \sim N(a + b\mu, \sigma^2 b^2)$$

结论：服从正态分布的随机变量的线性函数仍然服从正态分布。

即：若 $X \sim N(\mu, \sigma^2)$，则 $a + bX \sim N(a + b\mu, \sigma^2 b^2)$

特别地,取 $a = -\frac{\mu}{\sigma}, b = \frac{1}{\sigma}$,则 $a + bX = \frac{X-\mu}{\sigma} \sim N(0,1)$

例 **14.** $X \sim N(\mu_1, \sigma_1^2)$， $Y \sim N(\mu_2, \sigma_2^2)$， X 与 Y 独立，

求 $Z = X + Y$ 的分布。

解:因X, Y独立,故联合密度函数:

$$p(x,y) = p_X(x)\, p_Y(y)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

$$F(z)=P\{Z\le z\}=P\{X+Y\le z\}=\iint\limits_{x+y\le z}P(x,y)dxdx$$

$$=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{z-x}\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}dy$$

$$p(z) = F'(z)$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(z-x-\mu_2)^2}{2\sigma_2^2}} dx$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} e^{-(Ax^2-2Bx+C)} dx$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} \cdot \sqrt{\frac{\pi}{A}} e^{-\frac{AC-B^2}{A}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \cdot \sqrt{1/2A}} e^{-\frac{(x-\frac{B}{A})^2}{2\cdot(\sqrt{1/2A})^2}} dx$$

$$= \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\pi\sigma}_{\mathbf{1}}\boldsymbol{\sigma}_{\mathbf{2}}} \cdot \sqrt{\frac{\boldsymbol{\pi}}{\boldsymbol{A}}} \boldsymbol{e}^{-\frac{\boldsymbol{AC-B^2}}{\boldsymbol{A}}}$$

其中

$$A=\frac{1}{2}\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right),\quad B=\frac{1}{2}\left(\frac{\mu_1}{\sigma_1^2}+\frac{z-\mu_2}{\sigma_2^2}\right),$$

$$C=\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2}+\frac{(z-\mu_2)^2}{\sigma_2^2}\right)$$

$$\therefore p_Z(z)=\frac{1}{\sqrt{2\pi}\cdot\sqrt{\sigma_1^2+\sigma_2^2}}e^{-\frac{[z-(\mu_1+\mu_2)]^2}{2(\sigma_1^2+\sigma_2^2)}}$$

$$\therefore Z\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$$

结论：两个独立的正态随机变量之和仍为正态分布。

推论：有限个独立的正态随机变量之和仍为正态分布。

即：　相互独立的 $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \cdots, n$ ，

$$\text{则} \sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

例 15. ξ与η独立，且$\xi \sim N(0,1)$， $\eta \sim N(0,1)$，

求$\zeta = \xi^2 + \eta^2$的概率密度。

解：$p(x,y) = \mathrm{p}_X(x) p_Y(y) = \dfrac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$

当$z \le 0$时， $F_\zeta(z) = 0$， $p_\zeta(z) = 0$

当$z>0$时，$$F_\zeta(z)=\iint\limits_{x^2+y^2<z}\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}dxdy$$

$$=\frac{1}{2\pi}\cdot 2\pi\cdot(-1)e^{-\frac{r^2}{2}}\Big|_0^{\sqrt{z}}$$

$$=1-e^{-\frac{z}{2}}$$

$$\therefore p_\zeta(z)=\frac{1}{2}e^{-\frac{z}{2}}$$

$$\therefore p_\zeta(z)=\begin{cases}\frac{1}{2}e^{-\frac{z}{2}}, & z>0\\ 0, & z\le 0\end{cases}$$

$\therefore \zeta\sim E(\frac{1}{2})$ 或称: $\zeta\sim\chi^2(2)$