第二章 随机变量

§1 离散随机变量

§2 连续随机变量

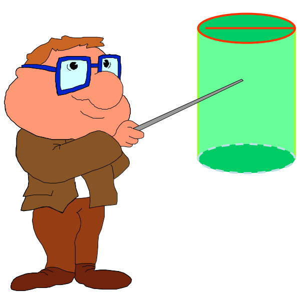
§3 随机变量的函数

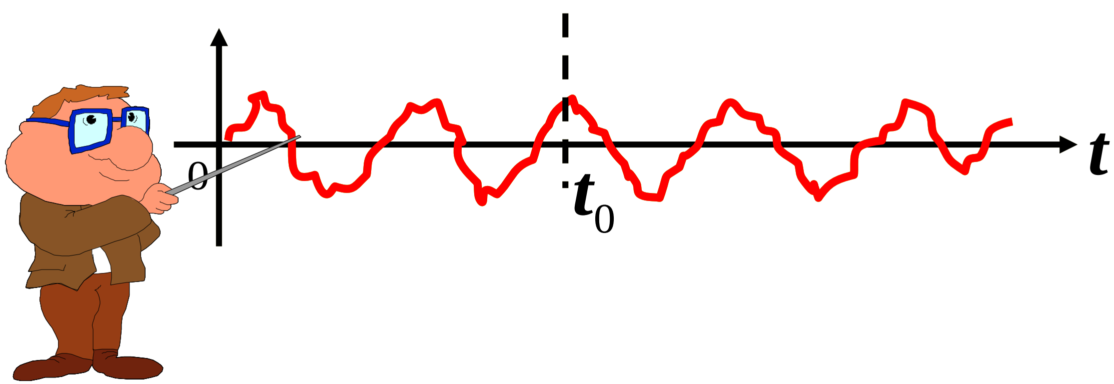
(functions of a random variable)

实际背景

- 若要得到一个圆的面积Y，总是测量其半径，半径的测量值可看作随机变量X，若$X \sim N(\mu,\sigma^2)$，则$Y=\pi X^2$的分布是什么？
- 若已知体重$W(kg)$均服从正态分布，在身高$L(m)$确定的情形下，则体质指数$BMI=W/L^2$服从什么分布？

问题：已知随机变量X的分布，$Y=g(X)$，函数$g(\cdot)$已知，求Y的分布.





问题：已知随机变量X的概率分布，
且已知Y=g(X)，求Y的概率分布。

例：已知X具有概率分布
且设$Y=X^2$，求Y的概率分布。

X	-1	0	1
p_i	0.2	0.5	0.3

解：Y的所有可能取值为0, 1

$$P(Y=0) \quad = P(X=0) = 0.5$$

$$P(Y=1) \quad = P\{(X=1)\cup(X=-1)\} = P(X=1) + P(X=-1) = 0.5$$

即找出(Y=0)的等价事件(X=0)；
(Y=1)的等价事件(X=1)或(X=-1)

(一) 离散型随机变量函数的频率函数

例 求 $Y=(X-1)^2$ 的频率函数，其中 r.v X 的频率函数为

X	-1	0	1	2
p_k	0.2	0.3	0.1	0.4

离散型— 离散型

解 $Y=(X-1)^2$ 的频率函数为

Y	4	1	0	1
p_k	0.2	0.3	0.1	0.4

0.7

设r.v X 的频率函数为

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

则 $Y=g(X)$ 的频率函数为

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_n)$	$\cdots$
p_k	p_1	p_2	$\cdots$	p_n	

有些 $g(x_k)$ 值相同

相应的概率值合并相加

例 设r.v $X \sim U(0,1)$,定义

$$Y=\begin{cases}0, & 0<X\le 0.25\\ 1, & 0.25<X\le 0.75\\ 2, & 0.75<X<1\end{cases}$$

连续型—离散型

求r.v Y 的频率函数.

解 $\because\ X \sim U(0,1)$ $\therefore$ Y的频率函数为

$$P\{Y=0\}=P\{0<X\le 0.25\}$$

$$=\int_0^{0.25} 1\cdot dx = 0.25$$

$$P\{Y=1\}=P\{0.25<X\le 0.75\}=0.50$$

$$P\{Y=2\}=P\{0.75<X<1\}$$

$$=P\{0.75<X\le 1\}=0.25$$

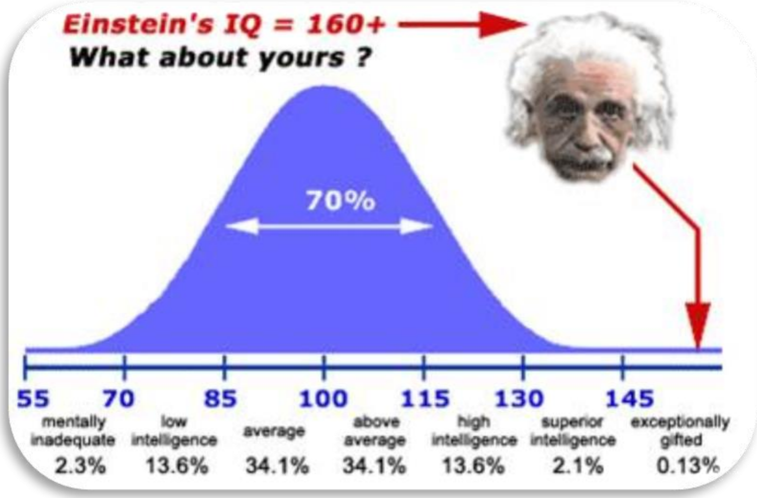
即 Y 的频率函数为

Y	0	1	2
p_k	0.25	0.50	0.25

例 (儿童智商)

设儿童智商 $X \sim N(100, 100)$, 将儿童按智商分为3类, 类标号 Y 规定如下:

$$Y = \begin{cases} 1, & X > 110 \\ 0, & 90 < X \le 110 \\ -1, & X \le 90 \end{cases}$$



求 Y 的频率函数.

Y	-1	0	1
p_k	0.16	0.68	0.16

一般，若已知X的概率分布，Y=g(X)，求Y的概率分布的过程为：

1. 若Y为离散量，则先写出Y的可能取值：$y_1, y_2, \cdots y_j, \cdots$，再找出$(Y = y_j)$的等价事件$(X \in D)$，得$P(Y = y_i) = P(X \in D)$;

✦ 关键是找出等价事件。

(二)连续型随机变量函数的分布

例 设 r.v X 的密度函数为

$$f_X(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4\\ 0, & \text{其它}\end{cases}$$

求 r.v $Y=2X+8$ 的密度函数.

解 Y 的分布函数为

$$F_Y(y)=P\{Y\le y\}=P\{2X+8\le y\}$$

$$=P\{X\le \frac{y-8}{2}\}=F_X(\frac{y-8}{2})$$

$F_X(g^{-1}(y))$

$$\therefore\quad f_Y(y)=\begin{cases}\dfrac{1}{2}\cdot\dfrac{1}{8}\cdot\dfrac{y-8}{2}, & 0<\dfrac{y-8}{2}<4\\ 0, & \text{其它}\end{cases}$$

$$=\begin{cases}\dfrac{y-8}{32}, & 8<y<16\\ 0, & \text{其它}\end{cases}$$

$$f_Y(y)=F_X'(g^{-1}(y))=(g^{-1}(y))'f_X(g^{-1}(y))$$

(二) 连续型随机变量函数的分布

问题 设 r.v X的概率密度函数为$f_X(x)$, r.v $Y=a+bX$, 求 r.v Y 的概率密度函数.

分析

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

① $b>0$时:

$$F_Y(y)=P\left\{X\le\frac{y-a}{b}\right\}=\int_{-\infty}^{\frac{y-a}{b}}f_X(x)dx$$

从而有

$$f_Y(y)=F_Y'(y)=\frac{1}{b}\cdot f_X\left(\frac{y-a}{b}\right)$$

(二)连续型随机变量函数的分布

问题 设 r.v X的概率密度函数为$f_X(x)$, r.v $Y=a+bX$, 求 r.v Y 的概率密度函数.

分析

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

② $b<0$时:

$$F_Y(y)=P\left\{X\ge\frac{y-a}{b}\right\}=1-\int_{-\infty}^{\frac{y-a}{b}}f_X(x)dx$$

从而有

$$f_Y(y)=F_Y'(y)=-\frac{1}{b}\cdot f_X\left(\frac{y-a}{b}\right)$$

(二) 连续型随机变量函数的分布

基本流程: 求r.v $Y=g(X)$的概率密度函数.

① 求r.v Y的分布函数 $F_Y(y)=P\{Y\le y\}$

② 转化为关于r.v X的概率计算问题

需用到函数 $y=g(x)$的性质!

③ 求导 $f_Y(y)=F_Y'(y)$

问题 设r.v X 的概率密度函数为 $f_X(x)$, $y=g(x)$单调递增且处处可导, 求 r.v $Y=g(X)$的概率密度函数.

分析

$$F_Y(y)=P\{Y\le y\}=P\{g(X)\le y\}$$

$y=g(x)$单调递增

$$F_Y(y)=P\{X\le g^{-1}(y)\}$$

$$=\int_{-\infty}^{g^{-1}(y)} f_X(x)dx$$

$y=g(x)$ 处处可导

$$f_Y(y)=F_Y'(y)=f_X\left(g^{-1}(y)\right)\cdot\left[g^{-1}(y)\right]'$$

讨论

① 若 r.v X 有取值范围 (a,b)，则 $f_Y(y)$ 有定义域

$$g(a) < y < g(b)$$

② 为何要求 $y = g(x)$ 严格递增？

若不然，如何求 $g^{-1}(y)$？

③ 若 $y = g(x)$ 严格单调递减，有什么结论？

$$f_Y(y) = F_Y'(y) = -f_X\left(g^{-1}(y)\right)\left[g^{-1}(y)\right]'$$

$$= f_X\left(g^{-1}(y)\right)\left|\left[g^{-1}(y)\right]'\right|$$

定理 设r.v X的密度函数为 $f(x)$,又 $y=g(x)$是严格单调函数,其反函数 $h(y)=g^{-1}(y)$ 连续可导,则 $Y=g(X)$的密度函数为

严格单调增(或单调减)

$$f_Y(y)=\begin{cases}|h'(y)|\cdot f(h(y)), & h(y)\text{有意义}\\ 0, & \text{其它}\end{cases}$$

例 设 $X\sim U(-\frac{\pi}{2},\frac{\pi}{2})$,求 $Y=\mathrm{tg}X$ 的密度函数.

解 记 $y=\mathrm{tg}x$,则

$$h(y)=\mathrm{arctg}y,\ h'(y)=\frac{1}{1+y^2}\quad(-\infty<y<\infty)$$

$\therefore$ Y的密度函数为

$$f_X(x)=\begin{cases}\frac{1}{\pi}, & -\frac{\pi}{2}<x<\frac{\pi}{2}\\ 0, & \text{其它}\end{cases}$$

$$f_Y(y)=|h'(y)|\cdot f_X(h(y))$$

$$=\frac{1}{\pi}\cdot\frac{1}{1+y^2}\quad(-\infty<y<\infty)$$

Cauchy分布

例 设 $X \sim N(\mu,\sigma^2)$，求 $Y = aX + b$ 的密度函数，其中 $a\,(\neq 0)$，b 为常数.

解 记 $y = ax + b$，则

$$h(y) = \frac{y-b}{a},\ h'(y) = \frac{1}{a} \quad (-\infty < y < \infty)$$

$\therefore$ Y 的密度函数为

$$f_Y(y) = |h'(y)| \cdot f_X(h(y)) = \frac{1}{|a|} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(h(y)-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{|a|} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}(|a|\sigma)} e^{-\frac{[y-(a\mu+b)]^2}{2(|a|\sigma)^2}}$$

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

重要结论

正态r.v的线性函数仍是正态r.v

例 （股票价格）考虑时间 u 后股票价格S_u，已知

$S_u = S_0 e^{X_u}$，而 $X_u \sim N(u\mu, u\sigma^2)$， S_0为常数，

求S_u的密度函数 $f_S(s)$.

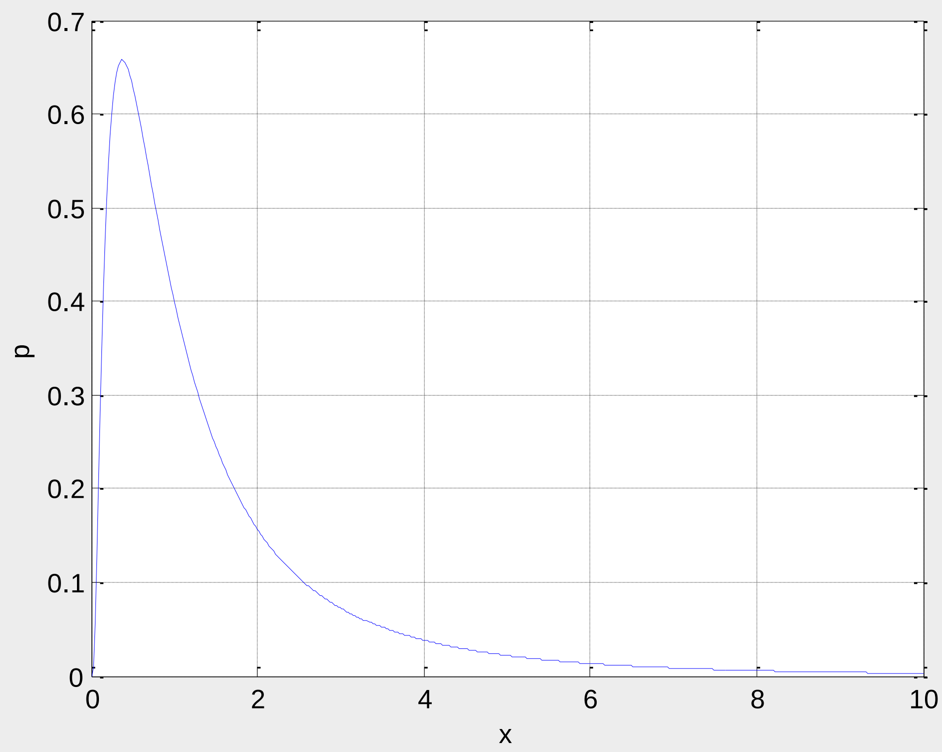
解 $h(S) = \ln\frac{S}{S_0}$， $h'(S) = \frac{1}{S}$， 故

$$f_S(s) = \frac{1}{\sqrt{2\pi u\sigma^2}\, s} e^{-\frac{(\ln\frac{s}{S_0} - u\mu)^2}{2u\sigma^2}}$$

注 $\frac{S_u}{S_0}$ 称服从参数为$(u\mu, u\sigma^2)$ 的对数正态分布，

记为 $LN(u\mu, u\sigma^2)$.

```
x = (0:0.02:10);
y = lognpdf(x,0,1);
plot(x,y); grid;
xlabel('x'); ylabel('p')
```



例 设 $X \sim U(0,1)$, 求 $Y = e^X$ 的概率密度.

问 记 $y = e^x$, 怎样确定其反函数 ?

分析 当 $y > 0$ 时，$y = e^x$ 的反函数为

$$h(y) = \ln y \quad (y > 0) \quad \times$$

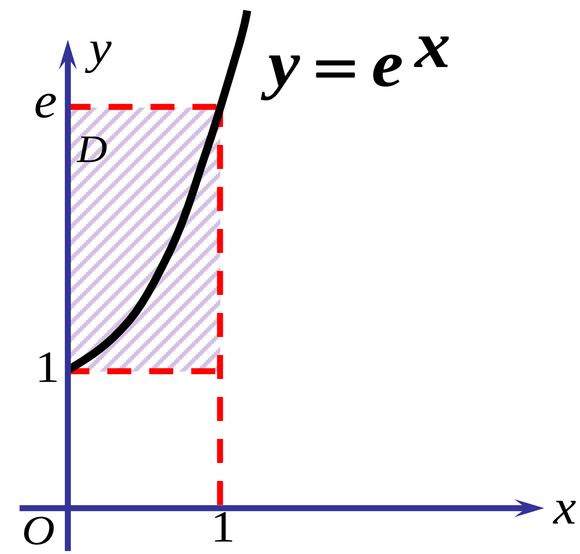
正确的分析 $\because X \sim U(0,1)$, 表明 r.v X 几乎只在 $(0,1)$ 上取值,

故 $y = e^x$ 的反函数存在的区域是

$$D: 0 < x < 1, 1 < y < e$$

其反函数为

$$h(y) = \ln y \quad (1 < y < e)$$



例 设 $X \sim U(0,1)$, 求$Y = e^X$的概率密度.

解 记 $y = e^x$, 则当 $1 < y < e$ 时, 反函数是

$$h(y) = \ln y \quad (1 < y < e)$$

$\therefore$ Y的密度函数为

$$f_Y(y) = \begin{cases} |h'(y)| \cdot f_X(h(y)), & 1 < y < e \\ 0, & \text{其它} \end{cases}$$

$$= \begin{cases} \frac{1}{y} \cdot \frac{1}{1}, & 1 < y < e \\ 0, & \text{其它} \end{cases}$$

$$= \begin{cases} \frac{1}{y}, & 1 < y < e \\ 0, & \text{其它} \end{cases}$$

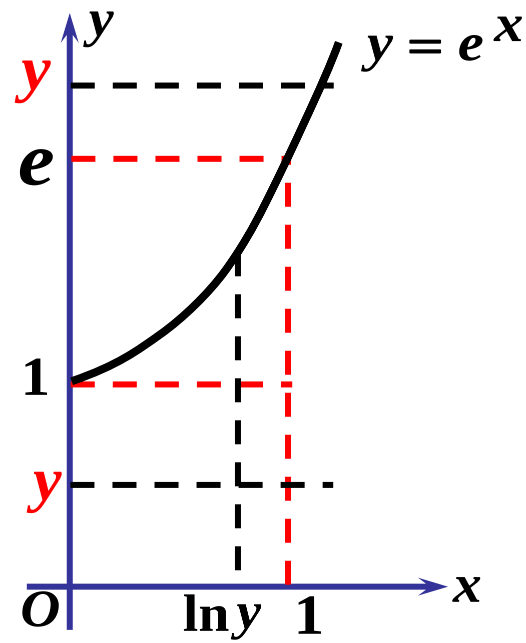
下面讨论直接计算法

例 设 $X \sim U(0,1)$, 求$Y = e^X$的概率密度.

解 Y 的分布函数为

$$F_Y(y) = P\{Y \le y\} = P\{e^X \le y\}$$

$$= \begin{cases} 1\,, & y \ge e \\ \int_0^{\ln y} 1 \cdot dx\,, & 1 < y < e \\ 0\,, & y \le 1 \end{cases}$$



$$\therefore \quad f_Y(y) = \begin{cases} F_Y'(y), & 1 < y < e \\ 0\,, & \text{其它} \end{cases}$$

$$= \begin{cases} \dfrac{1}{y}\,, & 1 < y < e \\ 0\,, & \text{其它} \end{cases}$$

问题 若 $y = g(x)$没有单调性,有什么结论？

问题 设r.v X 的概率密度函数为 $f_X(x)$，r.v $Y = X^2$ 求r.v Y 的概率密度函数.

分析 $$F_Y(y) = P\{Y \le y\} = P\{X^2 \le y\}$$

$$= P\{-\sqrt{y} \le X \le \sqrt{y}\}$$

$$= \int_{-\infty}^{\sqrt{y}} f_X(x)dx - \int_{-\infty}^{-\sqrt{y}} f_X(x)dx$$

$$f_Y(y) = F_Y'(y) = f_X(\sqrt{y})\left[\sqrt{y}\right]' - f_X(-\sqrt{y})\left[-\sqrt{y}\right]'$$

讨论 函数 $y=g(x)=x^2$ 是分段严格单调的

$$\begin{cases} y=g_1(x)=x^2, x>0 \text{ 严格递增} \\ y=g_2(x)=x^2, x<0 \text{ 严格递减} \end{cases}$$

$\Rightarrow$ $$g_1^{-1}(y)=\sqrt{y}, \quad g_2^{-1}(y)=-\sqrt{y}$$

$$f_Y(y)=F_Y'(y)$$

$$=f_X\left(g_1^{-1}(y)\right)\left[g_1^{-1}(y)\right]'-f_X\left(g_2^{-1}(y)\right)\left[g_2^{-1}(y)\right]'$$

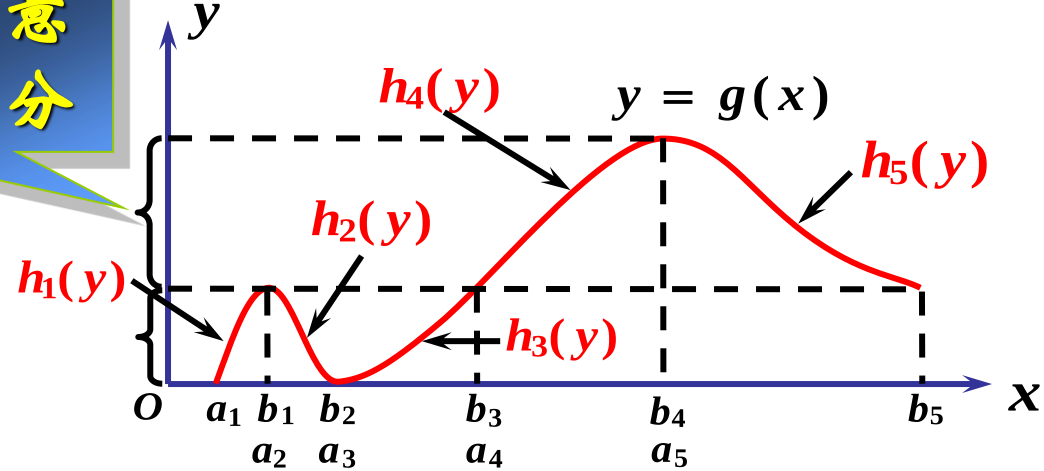
$$=\sum_{k=1}^{2} f_X\left(g_k^{-1}(y)\right)\left|\left[g_k^{-1}(y)\right]'\right|$$

推广的定理

定理 设 r.v X 的密度函数为 $f(x)$，又函数 $g(x)$ 在互不相交的区间 $(a_1,b_1),(a_2,b_2),\cdots$ 上逐段严格单调，且其反函数 $h_1(y),h_2(y),\cdots$ 均连续可导，则 $Y=g(X)$ 的密度函数为

$$f_Y(y)=\begin{cases}\sum\limits_{i=1}|h_i'(y)|\cdot f(h_i(y)), & h_1(y),h_2(y),\cdots \text{ 有意义} \\ 0, & \text{其它}\end{cases}$$

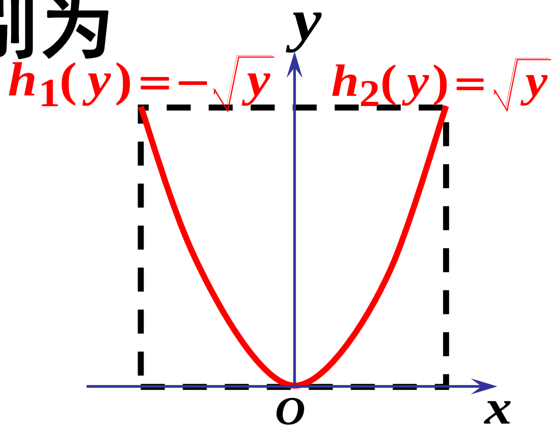
使得反函数有意义的 y 有两部分



例 设 $X \sim N(0,1)$, 求 $Y = X^2$ 的密度函数.

解 记 $g(x) = x^2$, 则 $g(x)$在 $(-\infty,0)$上严格单调减少，而在 $(0,\infty)$上严格单调增加，其反函数分别为

$$h_1(y) = -\sqrt{y}\ ,\ h_2(y) = \sqrt{y} \quad (y>0)$$



且 $$h_1'(y) = -\frac{1}{2\sqrt{y}},\quad h_2'(y) = \frac{1}{2\sqrt{y}} \quad (y>0)$$

$\therefore$ Y的密度函数为

$$f_Y(y) = \begin{cases} |h_1'(y)|\varphi(h_1(y)) + |h_2'(y)|\varphi(h_2(y)), & y>0 \\ 0\ , & y \le 0 \end{cases}$$

$$= \begin{cases} \left|-\frac{1}{2\sqrt{y}}\right| \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{y})^2}{2}} + \left|\frac{1}{2\sqrt{y}}\right| \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}}, & y>0 \\ 0\ , & y \le 0 \end{cases}$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y>0 \\ 0\ , & y \le 0 \end{cases}$$

均匀分布与其它连续分布的关系

设 r.v. X 的密度为 $f(x)$, 分布函数为$F(x)$.

其中$F(x)$ 在某区间 I 上严格递增, I 的左端点处 $F=0$, 右端点处 $F=1$. I 可以是有界区间, 也可以是无界区间. 因此, $F^{-1}(x)$ 在 I 上都有定义.

① 令 $Z = F(X)$, 那么 $Z \sim U(0,1)$.

$$P\{Z \le z\} = P\{F(X) \le z\}$$

$$= P\{X \le F^{-1}(z)\}$$

$$= F(F^{-1}(z)) = z$$

均匀分布与其它连续分布的关系

② 令 $U\sim \mathrm{U}(0,1)$, $X=F^{-1}(U)$, 那么X的分布函数是$F(x)$.

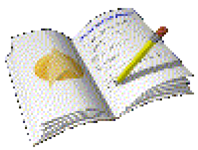
$$P\{X\le x\}=P\{F^{-1}(U)\le x\}=P\{U\le F(x)\}=F(x)$$

例： 生成给定分布的伪随机数

要生成分布函数为 $F(x)$ 的r.v., 只需将 F^{-1} 作用在均匀分布的随机数上即可.

例 为生成来自于指数分布的r.v., 可以取

$$T=-\ln V/\lambda, \qquad \text{其中 } V\sim \mathrm{U}(0,1).$$



课后作业 P49: 54、59、64

补充题1 设随机变量 X 的频率函数为

X	-2	-1	0	1	2
P	$1/5$	$1/6$	$1/5$	$1/15$	$11/30$

求 $Y=X^2$ 的频率函数.

补充题2 设随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{2x}{\pi^2} & 0<x<\pi \\ 0 & \text{其它}\end{cases}$$

求 $Y=\sin X$ 的概率密度.

3.设 $P\{X=k\}=\left(\frac{1}{2}\right)^k, k=1,2,\cdots,$ 令

$$Y=\begin{cases}1, & \text{当}X\text{取偶数时} \\ -1, & \text{当}X\text{取奇数时}.\end{cases}$$

求随机变量 X 的函数 Y 的分布律.

4．设随机变量 X 在区间 $(1，2)$ 上服从均匀分布，试求随机变量 $Y=e^{2x}$,的概率密度$f_Y(y)$.

- 例：设

X	-1	0	1
p	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$Y=2X, Z=X^2$，求Y, Z的概率分布。

解：Y的可能取值为-2, 0, 2

Z的可能取值为0, 1

$(Y=-2)$的等价事件为$(X=-1)\cdots$

$(Z=1)$的等价事件为$(X=1)\cup(X=-1)$

故得：

Y	-2	0	2
p	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Z	0	1
p	$\frac{1}{3}$	$\frac{2}{3}$

- 例：设$X \sim N(\mu,\sigma^2)$, $Y=\dfrac{X-\mu}{\sigma}$, 求Y的概率密度$f_Y(y)$

解：$y=g(x)=\dfrac{x-\mu}{\sigma}$, $g'(x)=\dfrac{1}{\sigma}>0$, $x=h(y)=\sigma y+\mu$

$$f_Y(y)=\sigma f_X(\sigma y+\mu) \quad =\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} \quad \Rightarrow Y \sim N(0,1)$$

$$\boxed{一般若X \sim N(\mu,\sigma^2),\ Y=aX+b \Rightarrow Y \sim N(a\mu+b, a^2\sigma^2)}$$

例：设随机变量X具有概率密度$f_X(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4\\ 0, & 其他\end{cases}$

求Y=X^2的概率密度。

解：分别记X, Y的分布函数为 $F_X(x), F_Y(y)$

$$F_Y(y)=P\{Y\le y\}=P\{X^2\le y\}=P\{-\sqrt{y}<X<\sqrt{y}\}$$

当 $y\le 0$时，$F_Y(y)=0$； 当 $y\ge 16$时，$F_Y(y)=1$

$f(x)$连续时，$\dfrac{d}{dx}\int_a^x f(t)dt=f(x)$

$\dfrac{d}{dx}\int_a^{u(x)} f(t)dt=f(u(x))u'(x)$

当 $0<y<16$ 时，

$$F_Y(y)=P\{0<X<\sqrt{y}\}=F_X(\sqrt{y})=\int_{-\infty}^{\sqrt{y}} f_X(t)dt$$

$$\Rightarrow f_Y(y)=\begin{cases}\dfrac{1}{2\sqrt{y}}f_X(\sqrt{y}), & 0<y<16\\ 0, & 其他\end{cases}=\begin{cases}\dfrac{1}{2\sqrt{y}}\cdot\dfrac{\sqrt{y}}{8}=\dfrac{1}{16}, & 0<y<16\\ 0, & 其他\end{cases}$$

Y在区间(0, 16)上均匀分布。

例：若$X \sim f(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4\\ 0, & 其他\end{cases}$ $Y=X^3$，求 $f_Y(y)$。

解：$y=g(x)=x^3,\ x=y^{\frac{1}{3}}=h(y)$

$g'(x)=3x^2>0,\qquad \therefore f_Y(y)=\dfrac{1}{3}y^{-\frac{2}{3}}f_X(y^{\frac{1}{3}})$

$$f_Y(y)=\begin{cases}\dfrac{1}{24}y^{-\frac{1}{3}}, & 0<y<64\\ 0\ , & 其他\end{cases}$$

例：设X服从参数为λ的指数分布，$F(x)$为X的分布函数。

(1) 求$F(x)$；

(2)设$Y=F(X)$,试证$Y \sim U(0,1)$(即均匀分布)。

解:(1) 由前知, $X \sim f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases} \Rightarrow F(x)=\begin{cases}1-e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$

(2) $Y=F(X)=\begin{cases}1-e^{-\lambda X}, & X>0\\ 0, & X\le 0\end{cases} \Rightarrow 0\le Y\le 1$

记$F_Y(y)$为Y的概率分布函数,

当$y\le 0$时，$F_Y(y)=P\{Y\le y\}=0$　　当$y\ge 1$时，$F_Y(y)=P\{Y\le y\}=1$

当$0<y<1$时，$F_Y(y)=P\{1-e^{-\lambda X}\le y\} =P\{e^{-\lambda X}\ge 1-y\}$

$$=P\left\{X\le -\frac{1}{\lambda}ln(1-y)\right\} =1-e^{-\lambda\left[-\frac{1}{\lambda}ln(1-y)\right]}=y$$

即 $F_Y(y)=\begin{cases}0, & y\le 0\\ y, & 0<y<1\\ 1, & y\ge 1\end{cases}$, $\therefore\ Y\sim U(0,1)$

复习思考题

1. 什么量被称为随机变量？它与样本空间的关系如何？
2. 满足什么条件的试验称为“n重贝努里试验”？
3. 事件A在一次试验中发生的概率为p, $0<p<1$。若在n次独立重复的试验中，A发生的总次数为X, 则X服从什么分布？并请导出：
$$P\{X=k\}=C_n^k p^k(1-k)^{n-k},\ k=0,1,\cdots,n$$
4. 什么条件下使用泊松近似公式等式较为合适？
5. 什么样的随机变量称为连续型的？
6. 若事件A为不可能事件，则$P(A)=0$, 反之成立吗？又若A为必然事件，则$P(A)=1$，反之成立吗？
7. 若连续型随机变量X在某一区间上的概率密度为0，则X落在该区间的概率为0，对吗？
8. 若随机变量X在区间(a,b)上均匀分布，则X落入(a,b)的任意一子区间(a_1,b_1)上的概率为$(b_1-a_1)/(b-a)$, 对吗？
9. 若$X\sim N(\mu,\sigma^2)$，则X的概率密度函数$f(x)$在$x=\mu$处值最大，因此X落在μ附近的概率最大，对吗？