§5 随机变量的函数分布

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	0.2	0.5	0.3

解：Y的所有可能取值为0, 1

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

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

即找出(Y=0)的等价事件(X=0)；

(Y=1)的等价事件(X=1)或(X=-1)

例如，若要测量一个圆的面积，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服从什么分布？

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=X²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，$\dfrac{d}{dx}\int_a^x f(t)dt=f(x)$

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

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。

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$\left(Y = y_j\right)$的等价

事件$(X \in D)$，得$P(Y = y_i) = P(X \in D)$；

2. 若Y为连续量，则先写出Y的概率分布函数：

$F_Y(y) = P(Y \le y)$，找出$\left(Y \le y\right)$的等价事件$(X \in D)$，

得$F_Y(y) = P(X \in D)$；再求出Y的概率密度函数$f_Y(y)$；

关键是找出等价事件。

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

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

- 例：设

X	-1	0	1
p	$1/3$	$1/3$	$1/3$

$Y=2X, Z=X^2$，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)…

(Z=1)的等价事件为$(X=1)\cup(X=-1)$

故得：

Y	-2	0	2
p	$1/3$	$1/3$	$1/3$

Z	0	1
p	$1/3$	$2/3$

例：设X的概率密度为$f(x)$, $|x| < \infty$, $Y = X^2$,

求Y的概率密度$f_Y(y)$.

解：设Y的概率分布函数为$F_Y(y)$

当 $y > 0$时，$F_Y(y) = P(Y \le y) = P(X^2 \le y) = \int_{-\sqrt{y}}^{\sqrt{y}} f(t)dt$

$$= \int_0^{\sqrt{y}} f(t)dt - \int_0^{-\sqrt{y}} f(t)dt$$

$$\Rightarrow f_Y(y) = F_Y'(y) = \begin{cases} [f(\sqrt{y}) + f(-\sqrt{y})]\dfrac{1}{2\sqrt{y}}, & y > 0 \\ 0\ , & y \le 0 \end{cases}$$

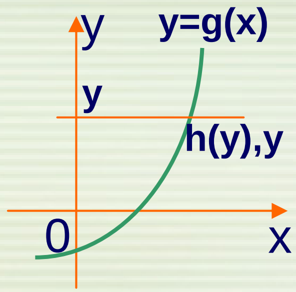
定理：设$X \sim f_X(x), -\infty < x < +\infty$，$g'(x) > 0$ (或$g'(x) < 0$)。

$Y = g(X)$, 则Y具有概率密度为：

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot \left| h'(y) \right|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$$

其中$\alpha = \min(g(-\infty), g(+\infty))$, $\beta = \max(g(-\infty), g(+\infty))$,

$h(y) = x \Leftrightarrow y = g(x)$



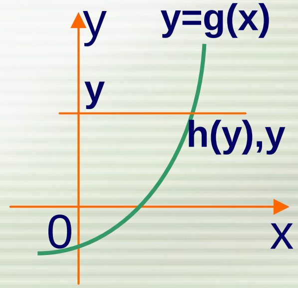
定理：设$X \sim f_X(x), -\infty < x < +\infty$，$g'(x) > 0$ (或$g'(x) < 0$)。

$Y = g(X)$, 则Y具有概率密度为：

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot |h'(y)|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$$

其中$\alpha = \min(g(-\infty), g(+\infty)), \beta = \max(g(-\infty), g(+\infty))$,

$h(y) = x \Leftrightarrow y = g(x)$



证明：不妨设$g'(x) > 0$，则$g(x)$为单调增函数，

且：$h'(y) > 0$

当 $y \le \alpha$ 时，$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le -\infty) = 0$;

当 $y \ge \beta$ 时，$F_Y(y) = 1$;

当 $\alpha < y < \beta$ 时，$F_Y(y) = P(Y \le y) = P(g(X) \le y)$

$$= P(X \le h(y)) = \int_{-\infty}^{h(y)} f_X(t)dt$$

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

同理可证：当$g'(x) < 0$时，定理为真

推论：设$X \sim f_X(x), \{x \mid f(x) \neq 0\} \subset (a,b)$,

当$a < x < b$时$g'(x) > 0$ (或$g'(x) < 0$)。

$Y = g(X)$, 则Y具有概率密度为：

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot \left|h'(y)\right|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$$

其中$\alpha = \min(g(a), g(b)), \beta = \max(g(a), g(b))$,

$$h(y) = x \Leftrightarrow y = g(x)$$

若g(x)分段单调递增或单调递减,可以利用本推论求Y的概率分布

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

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

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

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

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

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

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

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

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

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 2

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？