§ 2.4 随机变量函数的分布

已知X的分布（分布律，密度函数）

求解$Y = g(X)$的分布（分布律，密度函数）

一　离散型随机变量函数的分布

设X是离散型随机变量，其分布列为

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

则$Y=g(X)$也是离散型，其分布列为

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

注意：当某些$g(x_i)$相等时，应把它们适当合并.

例如：

X	-2	0	1	2
P	$1/5$	$3/10$	$2/5$	$1/10$

则$Y=3X+2$的分布列

Y	-4	2	5	7
P	$1/5$	$3/10$	$2/5$	$1/10$

则$Y=X^2+1$的分布列

Y	1	2	5
P	$3/10$	$2/5$	$3/10$

注意合并

二　连续型随机变量函数的分布

设X是连续型随机变量，已知其分布函数$F_X(x)$或密度函数$f_X(x)$，要求$Y=g(X)$的分布函数$F_Y(y)$或密度函数$f_Y(y)$.

注：在某些特殊情况下，$Y=g(X)$仍然是连续型随机变量.

求解$Y=g(X)$分布的一般方法

先求Y的分布函数

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

求导得到Y的密度函数

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

求解的关键是积分区域$\{x:g(x)\le y\}$的确定.

求解$Y=g(X)$分布的公式法

当函数$y=g(x)$是单调函数时，根据一般方法可得到Y的密度函数求解公式. 即下面的定理1.

定理1 若X是连续型随机变量，取值范围为区间(a, b)（有限或无限），密度函数为$f(x)$，$Y = g(X)$. 则

(1) 若$y = g(x)$在(a, b)上严格单调，其反函数$x = h(y)$有连续的导函数. 则$Y = g(X)$是连续型随机变量，其密度函数为

$$f_Y(y) = f_X\left[h(y)\right]\left|h'(y)\right|.$$

(2) 若$y = g(x)$在(a, b)的不相重叠的区间$I_1, I_2, \cdots$逐段严格单调，其反函数分别为$h_1(y), h_2(y), \cdots$，而且$h_1'(y), h_2'(y), \cdots$为连续函数，则$Y = g(X)$是连续型随机变量，其密度函数为

$$f_Y(y) = f_X\left[h_1(y)\right]\left|h_1'(y)\right| + f_X\left[h_2(y)\right]\left|h_2'(y)\right| + \cdots$$

证明 （1）当$y=g(x)$是严格单调函数时, 若是严格单调上升，则

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

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

若是严格严调下降，则

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

$$=P\{X\ge g^{-1}(y)\}=1-F_X(h(y))$$

无论哪种情况，求导得到Y的密度函数为

$$f_Y(y)=f_X\left[h(y)\right]\left|h'(y)\right|.$$

(2) 当$y=g(x)$是分段严格单调函数时, 对给定实数y，记$I_i=(a_i,b_i)$

$$E_i(y)=\left\{x:x\in I_i\text{且}g(x)\le y\right\}$$

$$=\begin{cases}(a_i,h_i(y)), & \text{当}y=g(x)\text{在}I_i\text{上单调上升,}\\ (h_i(y),b_i), & \text{当}y=g(x)\text{在}I_i\text{上单调下降.}\end{cases}$$

则

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

$$=\left\{X\in\sum_i E_i(y)\right\}=\sum_i\int_{E_i(y)}f(x)dx$$

无论$E_i(y)$是哪种情况, 对y求导得到Y的密度函数为

$$f_Y(y)=f_X\left[h_1(y)\right]\left|h_1'(y)\right|+f_X\left[h_2(y)\right]\left|h_2'(y)\right|+\cdots$$

【例1】若$X\sim U(0,1)$，求$Y=2X+1$的概率密度.

解: 由题意可知，X的概率密度为

$$f(x)=\begin{cases}1, & 0<x<1,\\ 0, & \text{其它}.\end{cases}$$

$$F_Y(y)=P\{Y\le y\}=P\{2X+1\le y\}$$

$$=P\{X\le (y-1)/2\}=F_X(\frac{y-1}{2}).$$

对$F_Y(y)$关于y求导

$$f_Y(y)=f_X((y-1)/2)\cdot((y-1)/2)'$$

$$=\begin{cases}1\cdot 1/2, & 0<(y-1)/2<1,\\ 0, & \text{其它}.\end{cases}=\begin{cases}1/2, & 1<y<3,\\ 0, & \text{其它}.\end{cases}$$

【例2】若$X \sim N(\mu,\sigma^2)$，求$Y = aX + b(a \neq 0)$的概率密度.

解：X的概率密度为 $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma}\exp[-\dfrac{(x-\mu)^2}{2\sigma^2}]$,

由 $y = g(x) = ax + b$ 可得 $x = \dfrac{y-b}{a} = h(y)$.

由定理可得$Y = aX + b(a \neq 0)$的概率密度为

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

$$= \frac{1}{\sqrt{2\pi}|a|\sigma}\exp\{-\frac{[y-(a\mu+b)]^2}{2(a\sigma)^2}\}, \quad \infty < y < +\infty.$$

即 $Y \sim N(a\mu + b, (a\sigma)^2)$.

特别地有

$$Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

正态分布的标准化

【例3】若$X\sim N(\mu,\sigma^2)$，求$Y=e^X$的密度函数.

解：函数$y=e^x$在$(-\infty,+\infty)$上是严格单调增函数，其反函数$x=\ln y(y>0)$的导数为$(\ln y)'=y^{-1}$，根据公式得$Y=e^X$的密度函数为

$$f_Y(y)=f_X[h(y)]|h'(y)|=\frac{1}{\sqrt{2\pi}\sigma y}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}},y>0.$$

由于$\ln Y=X$服从正态分布，故称Y所服从的分布为对数正态分布.

【例3】若$X \sim N(0,1)$，求$Y = X^2$的密度函数$q(y)$.

解：函数$y = x^2$在$(-\infty, 0)$上单调下降，在$(0, +\infty)$上单调上升. 其反函数分别为$x = -\sqrt{y}$; $x = \sqrt{y}$，反函数的导数分别为 $(-\sqrt{y})' = -\frac{1}{2\sqrt{y}}$, $(\sqrt{y})' = \frac{1}{2\sqrt{y}}$.

根据公式得$Y = X^2$的密度函数为

$$f_Y(y) = f_X(-\sqrt{y})\left|-\frac{1}{2\sqrt{y}}\right| + f_X(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right|$$

$$= \frac{1}{\sqrt{2\pi y}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, \quad y > 0.$$

此分布是$\chi^2(1)$，是$\chi^2(n)$的特例.

【例4】若$\theta \sim U(-\frac{\pi}{2},\frac{\pi}{2})$，$\varphi = tg\theta$，试求$\varphi$的密度函数$f_Y(y)$.

解：函数$y = tgx$在$(-\frac{\pi}{2},\frac{\pi}{2})$上单调上升，其反函数为$x = tg^{-1}y$. 因为

$$(tg^{-1}y)' = \frac{1}{1+y^2}$$

根据公式得φ的密度函数为

$$f_Y(y) = f_X[h(y)]|h'(y)| = \frac{1}{\pi}\cdot\frac{1}{1+y^2}, \qquad y \in R.$$

由上式定义的分布称为柯西分布.

【例5】若$X \sim U[-1,1]$，试求$Y = X^2+1$的密度函数.

解：X的取值范围是$[-1,1]$，因而$Y = X^2+1$的取值范围是$[1,2]$，函数$y = x^2+1$在$[-1,0)$上单调下降，在$(0,1]$上单调上升.其反函数分别为

$$x = h_1(y) = -\sqrt{y-1}, \quad y \in (1,2],$$

$$x = h_2(y) = \sqrt{y-1}, \quad y \in (1,2],$$

$$h_1'(y) = -\frac{1}{2\sqrt{y-1}}; \quad h_2'(y) = \frac{1}{2\sqrt{y-1}}.$$

根据公式得$Y = X^2+1$的密度函数为

$$f_Y(y) = f_X\left[h_1(y)\right]\left|h_1'(y)\right| + f_X\left[h_2(y)\right]\left|h_2'(y)\right|$$

$$=\begin{cases} \dfrac{1}{2\sqrt{y-1}}, & y\in(1,2] \\ 0, & 其他 \end{cases}$$

【例6】若$X\sim U[-1,2]$, 试求$Y=X^2+1$的密度函数.

解：X的取值范围是$[-1,2]$，因而$Y=X^2+1$的取值范围是$[1,5]$，函数$y=x^2+1$在$[-1,0)$上单调下降，在$(0,2]$上单调上升.其反函数分别为

$$x=h_1(y)=-\sqrt{y-1},\quad y\in(1,2],$$

$$x=h_2(y)=\sqrt{y-1},\quad y\in(1,5],$$

$$h_1'(y)=-\frac{1}{2\sqrt{y-1}};\quad h_2'(y)=\frac{1}{2\sqrt{y-1}}.$$

根据公式得$Y=X^2+1$的密度函数为

$$f_Y(y)=f_X\left[h_1(y)\right]\left|h_1'(y)\right|+f_X\left[h_2(y)\right]\left|h_2'(y)\right|$$

$$=\begin{cases}\dfrac{1}{3}(y-1)^{-1/2}, & y\in(1,2], \\ \dfrac{1}{6}(y-1)^{-1/2}, & y\in(2,5], \\ 0, & \text{其他}.\end{cases}$$

若$X \sim N(0,1)$，令 $y = g(x) = \begin{cases} x, & x < 0, \\ 0, & x \geq 0. \end{cases}$

（1）求$Y = g(X)$的分布函数.

（2）判断$Y = g(X)$是离散型随机变量还是连续型随机变量，还是既不离散也不连续.

思考题？？？

设$y = g(x)$是连续函数，下列说法是否正确？

（1）若X是离散型随机变量，则$Y = g(X)$一定是离散型随机变量.

（2）若X是连续型随机变量，则$Y = g(X)$一定是连续型随机变量.

Y的分布函数为

$$F_Y(y) = P\{Y \le y\} = \begin{cases} \Phi(y), & y < 0, \\ 1, & y \ge 0. \end{cases}$$

$Y = g(X)$既不是离散型也不是连续型.

思考题答案

第一个结论正确，第二个结论不正确.