概率论与数理统计

第 8 讲

随机变量函数的分布

一. 随机变量的函数

在实际中，经常遇到这样的问题，某个随机变量的分布已知，但是关心的是该随机变量的某个函数的分布。

如： 在统计物理中，气体分子运动速度的绝对值X的分布是已知的，但是我们想知道分子运动动能Y的分布，它们之间有如下关系，

$$Y=\frac{1}{2}mX^2$$

再比如： 球的直径D的分布容易测量得到，但是想知道球的体积的分布，它们有如下关系

$$V=\frac{\pi}{6}D^3$$

因此，需要求某些随机变量的函数的分布。

一. 随机变量的函数

一般地，设X、Y是两个随机变量，$y=g(x)$是一个已知函数，如果当X取值x时，Y取值为$g(x)$，则称Y是随机变量X的函数，记为$Y=g(X)$。

本节主要研究单个随机变量的函数的分布问题，即，当随机变量X的概率分布已知时，如何求它的函数$Y=g(X)$的概率分布。下面分别在X为离散型和连续型两种情况下给出$Y=g(X)$的分布。

二. 离散型随机变量函数的分布

当X为离散型随机变量时，设X的分布律为

X	x_1	x_2	$\cdots$	x_i	$\cdots$
P	p_1	p_2	$\cdots$	p_i	$\cdots$

若$Y=g(X)$也是离散型随机变量，此时Y的分布律可以表示为：

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_k)$	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

当$g(x_1)$，$g(x_2)$，$\ldots$，$g(x_i)\ldots$中有某些值相等时，则把那些相等的值分别合并，并把对应的概率相加即可。

例1.设随机变量X的分布律为

X	-1	0	1	2
P	0.2	0.3	0.1	0.4

令$Y=2X$，$Z=(X-1)^2$，求Y和Z的分布律。

解：将X，Y，Z的取值和概率列在如下的表格中

X	-1	0	1	2
$Y=2X$	-2	0	2	4
$Z=(X-1)^2$	4	1	0	1
P	0.2	0.3	0.1	0.4

二. 离散型随机变量函数的分布

X	-1	0	1	2
$Y=2X$	-2	0	2	4
$Z=(X-1)^2$	4	1	0	1
P	0.2	0.3	0.1	0.4

易知Y的取值都是不同的，因此Y的分布律为

Y	-2	0	2	4
P	0.2	0.3	0.1	0.4

而Z的取值中有两个1，则需要把它们合并，并把相应的概率相加，得Z的分布律为

Z	0	1	4
P	0.1	0.7	0.2

二. 离散型随机变量函数的分布

一般地，设X是离散型随机变量，其分布律为

$$P\{X=x_i\}=p_i, i=1,2,\cdots$$

若$Y=g(X)$可能的取值为y_k, k=1,2,.......

则Y的分布律为 $$P\{Y=y_k\}=\sum_{g(x_i)=y_k}P\{X=x_i\}=\sum_{g(x_i)=y_k}p_i$$

例2.已知随机变量Y的分布律为

$$P\{X=k\}=\frac{1}{2^k},k=1,2,\cdots$$

令 $Y=\sin(\frac{\pi}{2}X)$，求Y的分布律。

解：已知Y可能的取值为 -1，0，1

且：
$$Y=\sin(\frac{\pi}{2}X)=\begin{cases}-1, & X=4n-1 \\ 0, & X=2n \\ 1, & X=4n-3\end{cases},n=1,2,\cdots$$

二. 离散型随机变量函数的分布

$$P\{X=k\}=\frac{1}{2^k},k=1,2,\cdots \qquad Y=\sin(\frac{\pi}{2}X)=\begin{cases}-1, & X=4n-1\\ 0, & X=2n\\ 1, & X=4n-3\end{cases},n=1,2,\cdots$$

$$P\{Y=-1\}=\sum_{n=1}^{\infty}P\{X=4n-1\}=\sum_{n=1}^{\infty}\frac{1}{2^{4n-1}}=\frac{1/8}{1-1/16}=\frac{2}{15}$$

$$P\{Y=0\}=\sum_{n=1}^{\infty}P\{X=2n\}=\sum_{n=1}^{\infty}\frac{1}{2^{2n}}=\frac{1/4}{1-1/4}=\frac{1}{3}$$

$$P\{Y=1\}=1-P\{Y=-1\}-P\{Y=0\}=\frac{8}{15}$$

$\therefore Y$的分布律为：

Y	-1	0	1
P	2/15	1/3	8/15

1. $Y=g(X)$为离散型随机变量

当$Y=g(X)$为离散型随机变量时，只需求Y的分布律，此时，问题的实质是将Y取某个值的概率转化为X属于某范围的概率。

例5.设随机变量X的概率密度为 $f(x)=\begin{cases}2x, & 0<x<1\\ 0, & \text{其他}\end{cases}$

令 $Y=\begin{cases}1, & X\le\frac{1}{2}\\ 0, & X>\frac{1}{2}\end{cases}$ 求Y的分布。

$$f(x)=\begin{cases}2x, & 0<x<1\\ 0, & \text{其他}\end{cases}\qquad Y=\begin{cases}1, & X\le \dfrac{1}{2}\\ 0, & X>\dfrac{1}{2}\end{cases}$$

解： $P\{Y=1\}=P\{X\le \frac{1}{2}\}=\int_{-\infty}^{1/2} f(x)dx=\int_0^{1/2} 2xdx=\frac{1}{4}$

$P\{Y=0\}=1-P\{Y=1\}=1-\frac{1}{4}=\frac{3}{4}$

所以Y的分布为

Y	0	1
P	3/4	1/4

2. $y=g(x)$为严格单调函数

当$y=g(x)$为严格单调函数时，关于$Y=g(X)$的分布有如下定理。

定理 设X为连续型随机变量，其密度函数为$f_X(x)$，且当$a<x<b$时，$f_X(x)>0$，若当$a<x<b$时，$y=g(x)$为严格单调的可导函数。则$Y=g(X)$为连续型随机变量，且其密度函数为

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

其中$\alpha=\min\{g(a), g(b)\}$, $\beta=\max\{g(a), g(b)\}$， $h(y)$是$g(x)$的反函数。

证明： 不妨设，$y=g(x)$为严格单调增函数，此时$\alpha=g(a)$, $\beta=g(b)$，则$h(y)$也为严格单调增函数，且$h'(y)>0$。意味着$Y=g(X)$仅在区间(α,β)内取值。设Y的分布函数为$F_Y(y)$。

当$y\le\alpha$时，$F_Y(y)=P\{Y\le y\}=0$

当$y\ge\beta$时，$F_Y(y)=P\{Y\le y\}=1$

当$\alpha<y<\beta$时，$F_Y(y)=P\{Y\le y\}=P\{Y\le\alpha\}+P\{\alpha<Y\le y\}=P\{\alpha<Y\le y\}$

$$=P\{\alpha<g(X)\le y\}=P\{a<X\le h(y)\}=\int_a^{h(y)} f_X(x)dx$$

三. 连续型随机变量函数的分布

即Y的分布函数为 $F_Y(y)=\begin{cases} 0, & y\le\alpha \\ \int_a^{h(y)} f_X(x)dx, & \alpha<y<\beta \\ 1, & y\ge\beta \end{cases}$

所以Y的密度函数为 $f_Y(y)=F_Y'(y)=\begin{cases} f_X\big(h(y)\big)h'(y), & \alpha<y<\beta \\ 0, & \text{其它} \end{cases}$

三. 连续型随机变量函数的分布

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

步骤：

在$f_X(x)$的非零区间(a, b)内，判断$y=g(x)$的单调性

当$x\in(a, b)$时，求y的取值范围，即$y=g(x)$的值域(α, β)

求$y=g(x)$的反函数$x=h(y)$

对反函数$x=h(y)$关于y求导数$h'(y)$

代入公式中

例3 设随机变量X服从正态分布$N(\mu,\sigma^2)$，则X的线性函数$Y=aX+b(a\neq 0)$服从正态分布$N(a\mu+b, a^2\sigma^2)$。

证明：X的概率密度函数为

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

易知对任意 $a\neq 0$，$y=ax+b$为严格单调的可导函数，且y的取值范围为$(-\infty, +\infty)$

则有 $x=h(y)=\dfrac{y-b}{a}\quad h'(y)=\dfrac{1}{a}$

三. 连续型随机变量函数的分布

由定理知，Y的密度函数为 $f_Y(y)=f_X\left(h(y)\right)\left|h'(y)\right|,-\infty<y<\infty$

即 $$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}}\left|\frac{1}{a}\right|=\frac{1}{\sqrt{2\pi}\left|a\right|\sigma}e^{-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}},-\infty<y<\infty$$

所以有 $Y\sim N(a\mu+b,a^2\sigma^2)$

特别地，上例中若取 $a=\dfrac{1}{\sigma},b=-\dfrac{\mu}{\sigma}$　　则 $Y=aX+b=\dfrac{X-\mu}{\sigma}\sim N(0,1)$

例4.设随机变量X服从参数为λ的指数分布，试求$Y=1-e^{-\lambda X}$的分布。

解：X的概率密度为 $f_X(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$

（1）对于$y=1-e^{-\lambda x}$，当$x>0$时，严格单调可导函数；

（2）值域为$(0, 1)$；

（3）反函数为 $x=-\dfrac{\ln(1-y)}{\lambda}$

（4）反函数的导数为 $x'_y=\dfrac{1}{\lambda(1-y)}$

三. 连续型随机变量函数的分布

$$f_X(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}\qquad x=-\frac{\ln(1-y)}{\lambda}\qquad x'_y=\frac{1}{\lambda(1-y)}$$

$$f_Y(y)=\begin{cases}f_X(x)|x'_y|, & y\in(0,1)\\ 0, & \text{其它}\end{cases}=\begin{cases}\lambda e^{-\lambda(-\frac{\ln(1-y)}{\lambda})}\dfrac{1}{\lambda(1-y)}, & y\in(0,1)\\ 0, & \text{其它}\end{cases}=\begin{cases}1, & y\in(0,1)\\ 0, & \text{其它}\end{cases}$$

即： $Y=1-e^{-\lambda X}\sim U(0,1)$

推广：设$F_X(x)$为X的分布函数，严格递增，令$Y=F_X(X)$，则$Y\sim U(0,1)$。

3. $y=g(x)$为其它形式时

当$y=g(x)$不单调时，情况比较复杂，$Y=g(X)$可能是连续型随机变量，也可能不是连续型随机变量。

此时，先求Y的分布函数，$F_Y(y)=P\{Y\le y\}=P\{g(X)\le y\}$

如果Y也是连续型随机变量，再对分布函数求导数，得到其密度函数。

如果Y不是连续型随机变量，求出分布函数即可。

★关键步骤：$F_Y(y)=P\{Y\le y\}=P\{g(X)\le y\}=P\{X\in D\}$

例5.设随机变量X服从参数为1/4的指数分布，令 $Z=(X-1)^2$.求Z的概率密度。

解：易知X的密度函数为 $f_X(x)=\begin{cases}\frac{1}{4}e^{-\frac{x}{4}}, & x>0\\ 0, & x\le 0\end{cases}$

先求Z的分布函数

当$z<0$时，$F_Z(z)=P\{Z\le z\}=0$

当$z\ge 0$时，$F_Z(z)=P\{Z\le z\}=P\{(X-1)^2\le z\}=P\{1-\sqrt{z}\le X\le 1+\sqrt{z}\}=F_X(1+\sqrt{z})-F_X(1-\sqrt{z})$

即有 $F_Z(z)=\begin{cases}F_X(1+\sqrt{z})-F_X(1-\sqrt{z}), & z\ge 0\\ 0, & z<0\end{cases}$

$$F_Z(z)=\begin{cases}F_X(1+\sqrt{z})-F_X(1-\sqrt{z}), & z\geq 0\\ 0, & z<0\end{cases}\qquad f_X(x)=\begin{cases}\frac{1}{4}e^{-\frac{x}{4}}, & x>0\\ 0, & x\leq 0\end{cases}$$

所以Z的密度函数为 $f_Z(z)=F_Z'(z)=\begin{cases}\frac{1}{2\sqrt{z}}[f_X(1+\sqrt{z})+f_X(1-\sqrt{z})], & z\geq 0\\ 0, & z<0\end{cases}$

当$z\geq 0$时，$f_X(1+\sqrt{z})=\frac{1}{4}e^{-\frac{1+\sqrt{z}}{4}}$ 当$0\leq z<1$时，$f_X(1-\sqrt{z})=\frac{1}{4}e^{-\frac{1-\sqrt{z}}{4}}$

当$z\geq 1$时，$f_X(1-\sqrt{z})=0$

三. 连续型随机变量函数的分布

$$f_X(1+\sqrt{z})=\frac{1}{4}e^{-\frac{1+\sqrt{z}}{4}},z>0$$

$$f_X(1-\sqrt{z})=\begin{cases}\frac{1}{4}e^{-\frac{1-\sqrt{z}}{4}}, & 0\le z<1\\ 0, & z\ge 1\end{cases}$$

$$f_Z(z)=\begin{cases}0, & z<0\\ \frac{1}{2\sqrt{z}}(\frac{1}{4}e^{-\frac{1+\sqrt{z}}{4}}+\frac{1}{4}e^{-\frac{1-\sqrt{z}}{4}}), & 0\le z<1\\ \frac{1}{2\sqrt{z}}(\frac{1}{4}e^{-\frac{1+\sqrt{z}}{4}}+0), & z\ge 1\end{cases}=\begin{cases}0, & z<0\\ \frac{1}{8\sqrt{z}}[e^{-\frac{1+\sqrt{z}}{4}}+e^{-\frac{1-\sqrt{z}}{4}}], & 0\le z<1\\ \frac{1}{8\sqrt{z}}e^{-\frac{1+\sqrt{z}}{4}}, & z\ge 1\end{cases}$$

例6.在半径为R，圆心在原点O的圆周上任取一点M，设MO与x轴正向的夹角Θ~$U(-\pi,\ \pi)$，求M点与$A(-R,0)$，$B(R,0)$三点构成的三角形面积的密度函数。

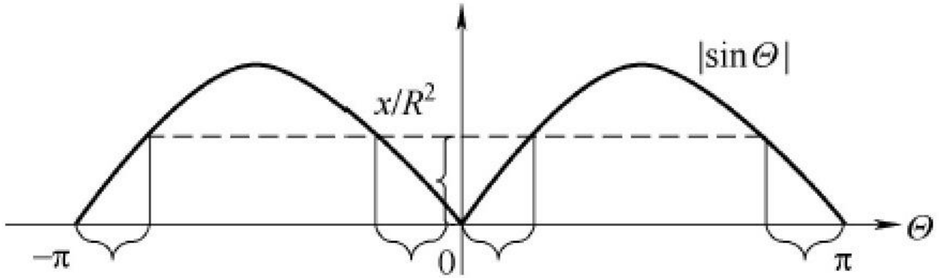
解：易知三角形的面积为 $S=R^2|\sin\Theta|,-\pi<\Theta<\pi$

先求分布函数 $F_S(x)=P\{S\le x\}=P\{R^2|\sin\theta|\le x\}=P\{|\sin\Theta|\le\frac{x}{R^2}\}$

当 $\frac{x}{R^2}<0$　即$x<0$时　$F_S(x)=0$

当 $\frac{x}{R^2}\ge 1$　即$x>R^2$时　$F_S(x)=1$

三. 连续型随机变量函数的分布



$$F_S(x) = P\{|\sin\Theta| \le \frac{x}{R^2}\} \qquad \Theta \sim U(-\pi, \pi)$$

当 $0 \le \frac{x}{R^2} < 1$ 即$0 \le x < R^2$时

$$F_S(x) = P\{|\sin\Theta| \le \frac{x}{R^2}\} = P\{0 \le \Theta \le \arcsin(\frac{x}{R^2})\}$$

$$+P\{\pi - \arcsin(\frac{x}{R^2}) \le \Theta \le \pi\} + P\{-\arcsin(\frac{x}{R^2}) \le \Theta \le 0\}$$

$$+P\{-\pi \le \Theta \le \pi - \arcsin(\frac{x}{R^2})\} = 4\arcsin(\frac{x}{R^2}) \Big/ 2\pi$$

故 $F_S(x)=\begin{cases} 0, & x<0 \\ \dfrac{2\arcsin(\frac{x}{R^2})}{\pi}, & 0\le x\le R^2 \\ 1, & x>R^2 \end{cases}$

所以 $f_S(x)=F_S'(x)=\begin{cases} \dfrac{2}{\pi\sqrt{R^4-x^2}}, & 0\le x\le R^2 \\ 0, & \text{其它} \end{cases}$

例7.假设一设备开机后无故障工作的时间X服从参数为1/4的指数分布。设备定时开机，出现故障自动关机，而在无故障的情况下工作3小时便关机。试求该设备每次开机无故障工作时间Y的分布函数。

解：易知X的分布函数为

$$F_X(x)=\begin{cases}1-e^{-\frac{x}{4}}, & x>0\\ 0, & x\le 0\end{cases}\qquad Y=\begin{cases}X, & X<3\\ 3, & X\ge 3\end{cases}$$

易知样本空间S可以表示为$\{X<3\}\cup\{X\ge 3\}$，所以对任意的实数y

$$\{Y\le y\}\subset\left\{\{X<3\}\bigcup\{X\ge 3\}\right\}$$

例7.假设一设备开机后无故障工作的时间X服从参数为1/4的指数分布。设备定时开机，出现故障自动关机，而在无故障的情况下工作3小时便关机。试求该设备每次开机无故障工作时间Y的分布函数。

$$\therefore F_Y(y)=P\{Y\le y\}=P\left\{\{Y\le y\}\cap\left\{\{X<3\}\cup\{X\ge 3\}\right\}\right\}$$

$$=P\left\{\left\{\{Y\le y\}\cap\{X<3\}\right\}\cup\left\{\{Y\le y\}\cap\{X\ge 3\}\right\}\right\}$$

$$=P\left\{\left\{Y\le y,X<3\right\}\cup\left\{Y\le y,X\ge 3\right\}\right\}$$

用符号“，”代替“$\cap$”

$$=P\left\{Y\le y,X<3\right\}+P\left\{Y\le y,X\ge 3\right\}$$

$$=P\left\{X\le y,X<3\right\}+P\left\{Y\le y,X\ge 3\right\}$$

三. 连续型随机变量函数的分布

$$F_X(x)=\begin{cases}1-e^{-\frac{x}{4}}, & x>0\\ 0, & x\le 0\end{cases}\quad Y=\begin{cases}X, & X<3\\ 3, & X\ge 3\end{cases}\quad F_Y(y)=P\{X\le y, X<3\}+P\{Y\le y, X\ge 3\}$$

当$y<0$时，$F_Y(y)=P\{X\le y\}+0=0$　　当$0\le y<3$时，$F_Y(y)=P\{X\le y\}+0=1-e^{-y/4}$

当$y\ge 3$时，$F_Y(y)=P\{X\le 3\}+P\{X\ge 3\}=1$

所以Y的分布函数为 $F_Y(y)=\begin{cases}0, & y<0\\ 1-e^{-y/4}, & 0\le y<3\\ 1, & y\ge 3\end{cases}$　　易知$F_Y(y)$在$y=3$时不连续。

作业： 50,53,57,59

第 8 讲

谢谢观看