概率论与数理统计

第二章　随机变量及其分布

练习：

1、 设随机变量 X 在 [2, 5]上服从均匀分布, 现对 X 进行三次独立观测 ,试求至少有两次观测值大于3 的概率.

解　X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{3}, & 2\le x\le 5,\\ 0, & \text{其他}.\end{cases}$$

设 A 表示 "对 X 的观测值大于 3 的次数",

即 $A=\{X>3\}$.

由于 $P(A)=P\{X>3\}=\int_3^5 \frac{1}{3}\,\mathrm{d}x=\frac{2}{3}$,

设Y表示3次独立观测中观测值大于3的次数,

则 $$Y \sim b\left(3, \frac{2}{3}\right).$$

因而有

$$P\{Y \geq 2\}=\binom{3}{2}\left(\frac{2}{3}\right)^2\left(1-\frac{2}{3}\right)+\binom{3}{3}\left(\frac{2}{3}\right)^3\left(1-\frac{2}{3}\right)^0=\frac{20}{27}.$$

2、 设某城市成年男子的身高 $X \sim N(170, 6^2)$ (单位 : cm)

(1) 问应如何设计公共汽车车门的高度，使男子与车门顶碰头的几率小于 0.01 ?

(2) 若车门高为 182 cm, 求 100 个成年男子与车门顶碰头的人数不多于 2 的概率 .

[思路] 设车门高度为 l cm，那么按设计要求应有 $P\{X > l\} < 0.01$，确定 l . 第二问首先要求出100名男子中身高超过 182cm 的人数的分布律，然后用此分布律, 求其不超过 2的概率 .

解 (1) 由题设知 $X \sim N(170, 6^2)$,

$$P\{X > l\} = 1 - P\{X \le l\}$$

$$= 1 - P\left\{\frac{X - 170}{6} \le \frac{l - 170}{6}\right\}$$

$$= 1 - \Phi\left(\frac{l - 170}{6}\right) < 0.01,$$

即 $\Phi\left(\frac{l - 170}{6}\right) > 0.99$. 查表得 $\frac{l - 170}{6} > 2.33$,

故 $l > 183.98(\text{cm})$.

(2) 设任一男子身高超过 182cm 的概率为 p .

则 $$p = P\{X > 182\} = P\left\{\frac{X-170}{6} > \frac{182-170}{6}\right\}$$

$$= 1 - \Phi(2) = 0.0228 .$$

设 Y 为100 个男子中身高超过 182cm 的人数 ,

则 $Y \sim B(100, 0.0228)$, 其中

$$P\{Y = k\} = \binom{100}{k} \times 0.0228^{k} \times 0.9772^{100-k},$$

$$k = 0,1,\cdots,100.$$

所求概率为

$$P\{Y \le 2\} = P\{Y = 0\} + P\{Y = 1\} + P\{Y = 2\},$$

由于 $n = 100$ 较大，$p = 0.0228$ 较小，故可用泊松分布来计算，其中 $\lambda = np = 2.28$，

从而

$$P\{Y \le 2\} = \frac{2.28^0 e^{-2.28}}{0!} + \frac{2.28 e^{-2.28}}{1!} + \frac{2.28^2 e^{-2.28}}{2!}$$

$$= 0.6013.$$

§5 随机变量的函数的分布

- 离散型随机变量的函数的分布
- 连续型随机变量的函数的分布

一、离散型随机变量的函数的分布

设 $f(x)$ 是定义在随机变量 X 的一切可能值 x 的集合上的函数,若随机变量 Y 随着 X 的取值 x的值而取 $y = f(x)$ 的值，则称随机变量 Y 为随机变量 X 的函数，记作 $Y = f(X)$.

问题

若已知的随机变量 X 的分布，

如何来求随机变量 $Y = f(X)$的分布？

例1　设随机变量 X 具有以下分布律，

试求 $Y=(X-1)^2$ 的分布律.

X	-1	0	1	2
p	0.2	0.3	0.1	0.4

解　Y 所有可能取的值为 0, 1, 4.

$$P\{Y=0\}=P\{(X-1)^2=0\}=P\{X=1\}=0.1,$$

$$P\{Y=1\}=P\{X=0\}+P\{X=2\}=0.7,$$

$$P\{Y=4\}=P\{X=-1\}=0.2$$

即得 Y 的分布律为

Y	0	1	4
p_k	0.1	0.7	0.2

如果X是离散型随机变量，其函数$Y=g(X)$也是离散型随机变量．若X的分布律为

X	x_1	x_2	$\cdots$	x_k	$\cdots$
p_k	p_1	p_2	$\cdots$	p_k	$\cdots$

则$Y=g(X)$的分布律为

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

注意 若$g(x_k)$中有值相同的，应将相应的p_k合并．

如果设

X	-1	1	2
p_k	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

则 $Y = X^2 - 5$ 的分布律

Y	-4	-1
p	$\frac{1}{2}$	$\frac{1}{2}$

二、连续型随机变量的函数的分布

例2 设随机变量 X 具有概率密度

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

求随机变量 $Y=2X+8$ 的概率密度 .

解 分别记 X,Y 的分布函数为 $F_X(x), F_Y(y)$.

下面先来求 $F_Y(y)$.

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

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

将$F_Y(y)$关于y求导数,得$Y=2X+8$的概率密度为

$$f_Y(y)=f_X\left(\frac{y-8}{2}\right)\left(\frac{y-8}{2}\right)'$$

$$=\begin{cases}\dfrac{1}{8}\left(\dfrac{y-8}{2}\right)\cdot\dfrac{1}{2}, & 0<\dfrac{y-8}{2}<4,\\ 0, & \text{其他}\end{cases}$$

$$f_Y(y) = f_X\left(\frac{y-8}{2}\right)\left(\frac{y-8}{2}\right)'$$

$$= \begin{cases} \frac{1}{8}\left(\frac{y-8}{2}\right)\cdot\frac{1}{2}, & 0 < \frac{y-8}{2} < 4, \\ 0, & \text{其他} \end{cases}$$

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

例3 设随机变量 X 具有概率密度 $f_X(x)$，$-\infty < x < \infty$，求 $Y = X^2$ 的概率密度 .

解 分别记 X,Y 的分布函数为 $F_X(x),F_Y(y)$.

先来求 Y 的分布函数 $F_Y(y)$. 由于 $Y = X^2 \geq 0$，故当 $y \leq 0$ 时 $F_Y(y) = 0$. 当 $y > 0$ 时有

$$
\begin{aligned}
F_Y(y) &= P\{Y \leq y\} = P\{X^2 \leq y\} \\
&= P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) .
\end{aligned}
$$

将 $F_Y(y)$ 关于 y 求导数，即得 Y 的概率密度函数为

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

例如，设 $X\sim N(0,1)$，其概率密度为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\quad -\infty<x<\infty.$$

则 $Y=X^2$ 的概率密度为

$$f_Y(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2}, & y>0,\\ 0, & y\le 0.\end{cases}$$

此时称 Y 服从自由度为 1 的 χ^2 分布．

定理 设随机变量 X 具有概率密度 $f_X(x)$，$-\infty < x < \infty$，设函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$ (或恒有 $g'(x) < 0$)，则 $Y = g(X)$ 是连续型随机变量，其概率密度为

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

其中 $\alpha = \min\{g(-\infty), g(\infty)\}$，$\beta = \max\{g(-\infty), g(\infty)\}$，$h(y)$ 是 $g(x)$ 的反函数 .

证　我们只证 $g'(x)>0$ 的情况 .

此时 $g(x)$ 在 $(-\infty,\infty)$ 严格单调增加，其反函数 $h(y)$ 存在，且在 (α,β) 严格单调增加，可导. 分别记 X,Y 的分布函数为 $F_X(x),F_Y(y)$. 现在先求 Y 的分布函数 $F_Y(y)$.

因为 $Y=g(X)$ 在 (α,β) 内取值，

故当 $y\le\alpha$ 时， 有 $F_Y(y)=P\{Y\le y\}=0$;

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

当 $\alpha<y<\beta$ 时，

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

$$= P\{X \le h(y)\} = F[h(y)].$$

将$F_Y(Y)$关于y求导数，即得Y的概率密度

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

对于$g'(x) < 0$的情况可以同样地证明，此时有

$$f_Y(y) = \begin{cases} f_X[h(y)][-h'(y)], & \alpha < y < \beta, \\ 0, & \text{其他}. \end{cases}$$

合并以上两式,定理得证.

若 $f(x)$ 在有限区间 $[a,b]$ 以外等于零，则只需假设在 $[a,b]$ 上恒有 $g'(x)>0$（或恒有 $g'(x)<0$），此时，

$$\alpha=\min\{g(a),\ g(b)\},\quad \beta=\max\{g(a),\ g(b)\}.$$

例4 设随机变量 $X \sim N(\mu, \sigma^2)$. 试证明 X 的线性函数 $Y = aX + b (a \neq 0)$ 也服从正态分布.

证 X 的概率密度为

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

现在 $y = g(x) = ax + b$, 由此式解得

$$x = h(y) = \frac{y-b}{a},$$

且有 $h'(y) = \frac{1}{a}$.

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), -\infty < y < \infty.$$

即

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

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

即有 $Y = aX + b \sim N(a\mu + b, (a\sigma)^2)$.

取 $a = \dfrac{1}{\sigma}$, $b = -\dfrac{\mu}{\sigma}$, 得 $Y = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

例5　设电压 $V = A\sin\Theta$，其中 A 是一个已知的正常数，相角 Θ 是一个随机变量，且有

$$\Theta \sim U\left(-\frac{\pi}{2},\frac{\pi}{2}\right),$$

求电压 V 的概率密度．

解　$v = g(\theta) = A\sin\theta$ 在 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ 上恒有

$g'(\theta) = A\cos\theta > 0$，且有反函数

$$\theta = h(v) = \arcsin\frac{v}{A}, \quad h'(v) = \frac{1}{\sqrt{A^2 - v^2}},$$

Θ的概率密度为

$$f(\theta)=\begin{cases}\dfrac{1}{\pi}, & -\dfrac{\pi}{2}<\theta<\dfrac{\pi}{2},\\ 0, & \text{其他},\end{cases}$$

$V=A\sin\Theta$的概率密度为

$$\varphi(v)=\begin{cases}\dfrac{1}{\pi}\cdot\dfrac{1}{\sqrt{A^2-v^2}}, & -A<v<A,\\ 0, & \text{其他},\end{cases}$$

注意

若$\Theta\sim U(0,\pi)$，此时$v=g(\theta)=A\sin\theta$在$(0,\pi)$上不是单调函数.

小结

1. 离散型随机变量的函数的分布

若 $Y = g(X)$ 且 X 的分布律为:

X	x_1	x_2	$\cdots$	x_k	$\cdots$
p_k	p_1	p_2	$\cdots$	p_k	$\cdots$

则 $Y = g(X)$的分布律为

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

2. 连续型随机变量的函数的分布

方法1 $F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\}$

$$= \int_{g(x)\le y} f_X(x)\,\mathrm{d}\,x, \quad (-\infty < x < +\infty),$$

再对 $F_Y(y)$ 求导得到 Y 的密度函数 $f_Y(y)$.

方法2

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

注意条件.

作业：课后习题 33、34、36、37

练习：

设某仪器上装有三只独立工作的同型号电子元件,其寿命(单位:小时)都服从同一指数分布,其中参数 $\theta = 600$,试求在仪器使用的最初200小时内,至少有一只元件损坏的概率a.

[思路] 以 $A_i(i=1,2,3)$ 分别表示三个电子元件"在使用的最初 200 小时内损坏"的事件,

于是 $a = P\{A_1 \cup A_2 \cup A_3\} = 1 - P(\overline{A_1}\ \overline{A_2}\ \overline{A_3})$

$$= 1 - P(\overline{A_1})P(\overline{A_2})P(\overline{A_3}),$$

由三个电子元件服从同一分布，

令 $p = P(A_i) \quad (i = 1,2,3)$，

由指数分布求出 p，便可得解．

解　用 $X_i \ (i = 1,2,3)$ 表示第 i 个元件的使用寿命，

由题设知 $X_i \ (i = 1,2,3)$ 的概率密度为

$$f(x) = \begin{cases} \dfrac{1}{600} e^{-\frac{x}{600}}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

从而

$$P\{X_i > 200\} = \int_{200}^{+\infty} f(x)\,\mathrm{d}\,x = \int_{200}^{+\infty} \frac{1}{600} \mathrm{e}^{-\frac{x}{600}}\,\mathrm{d}\,x = \mathrm{e}^{-\frac{1}{3}},$$

$$i = 1,2,3.$$

又 $P\{X_i > 200\} = P(\overline{A_i}) = p$,

因此所求概率为

$$a = 1 - P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})$$

$$= 1 - p^3 = 1 - (\mathrm{e}^{-\frac{1}{3}})^3$$

$$= 1 - \mathrm{e}^{-1}.$$

另解：求出单个电子器件小于200小时的概率后，在利用二项分布同样可以求解，并且要简单些。