§2.4 随机变量函数的分布

定义域为样本空间Ω值域为R的映射称作随机变量。

设X是随机变量，Y=g(X)是X的函数。那么

Y=g(X) 可以看作是函数g: R→R与X：Ω→R

的复合。即$\mathrm{Y}:\Omega\to\mathrm{R}$定义为 $\mathrm{Y}(\omega)=g(X(\omega))$.

通过上面的分析，随机变量的函数仍然可以看作是一个随机变量。

一般来说,若 X是离散的，则Y =g(X)也是离散的,

X连续型随机变量且g是连续函数，则Y=g(X)

也是连续型。本节中，我们将讨论如何由 X 的分布情况来推知 $\mathrm{Y}=\mathrm{g}(\mathrm{X})$ 的分布情况。

离散型

例2.16 X有概率分布 $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0.2 & 0.3 & 0.3 & 0.2 \end{pmatrix}$

$Y = X^2, Z = \dfrac{X^3+1}{2}$ 求Y,Z的概率分布。

解： Y的取值为0,1,4,分别求Y取这些值的概率:

P(Y=0)=P(X=0)=0.3. P(Y=1)=P(X=-1 or 1)=0.5.

P(Y=4)=P(X=2)=0.2 $Y \sim \begin{pmatrix} 0 & 1 & 4 \\ 0.3 & 0.5 & 0.2 \end{pmatrix}$

类似地,可得Z的全部可能取值为z_k： 0,0.5,1,4.5。

然后求 $P(Z=z_k)$ $Z \sim \begin{pmatrix} 0 & 0.5 & 1 & 4.5 \\ 0.2 & 0.3 & 0.3 & 0.2 \end{pmatrix}$

X离散型，求Y = g(X)的分布律的一般方法为：

设X 的分布律为 $p_k = P(X = x_k), k = 1, 2, \cdots$

则随机变量Y = g(X)的分布律为：

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

如果有若干个 $g(x_k)$ 相等，将他们合并，并将相应的概率相加（必要时可重新排序）。

连续型

例2.17 设随机变量X的密度函数为

$$f_X(x)=\begin{cases}2x, & 0<x<1\\ 0, & else\end{cases}$$

Y=-2lnX,求Y的密度函数.

解: 当X取值在(0,1)内时,Y的值域为 $(0,+\infty)$

Y的分布函数为: $F_Y(y)=P(Y\le y)=P(-2\ln X\le y)$

当$y>0$, $F_Y(y)$ $=P(X\ge e^{-\frac{y}{2}})=\int_{x\ge e^{-\frac{y}{2}}} f_X(x)dx$

$$=\int_{e^{-\frac{y}{2}}}^{1}2xdx=1-e^{-y}$$

当$y\le 0$,

$$\mathrm{F_Y}(y)=\int_{e^{-\frac{y}{2}}}^{+\infty}0dx=0.$$

$$\Rightarrow f_Y(y)=F_Y'(y)=\begin{cases}e^{-y}, y>0\\ 0, \quad y\le 0\end{cases}$$

由此可以得到：在已知连续型随机变量X 的分布情况下求 $Y=g(X)$ 的密度函数的方法为：

(1)　确定Y 的取值范围R(Y);

(2)　求出当 $y\in R(Y)$ 时Y 的分布函数 $F_Y(y)$ ：

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

其中 $G(y)$是满足 $g(X)\le y$ 的X 的取值范围;

(3)　求出当 $y\in R(Y)$ 时Y 的密度函数 $f_Y(y)$:

$$f_Y(y)=F_Y{}'(y),\ \ y\in R(Y);$$

(4) 总结Y 的密 度函数 $f_Y(y)=\begin{cases} F'_Y(y), & y\in R(Y), \\ 0, & else。\end{cases}$

例2.19 设随机变量X的密度函数为

$$f_X(x)=\begin{cases}\frac{1}{4}|x|, & -2\le x\le 0\\ x, & 0\le x\le 1\\ 0, & \text{其他。}\end{cases}$$

$Y=X^2$,求Y的密度函数.

解: 易得 $R(Y)=[0,4]$

$$\forall y\in[0,4],$$

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

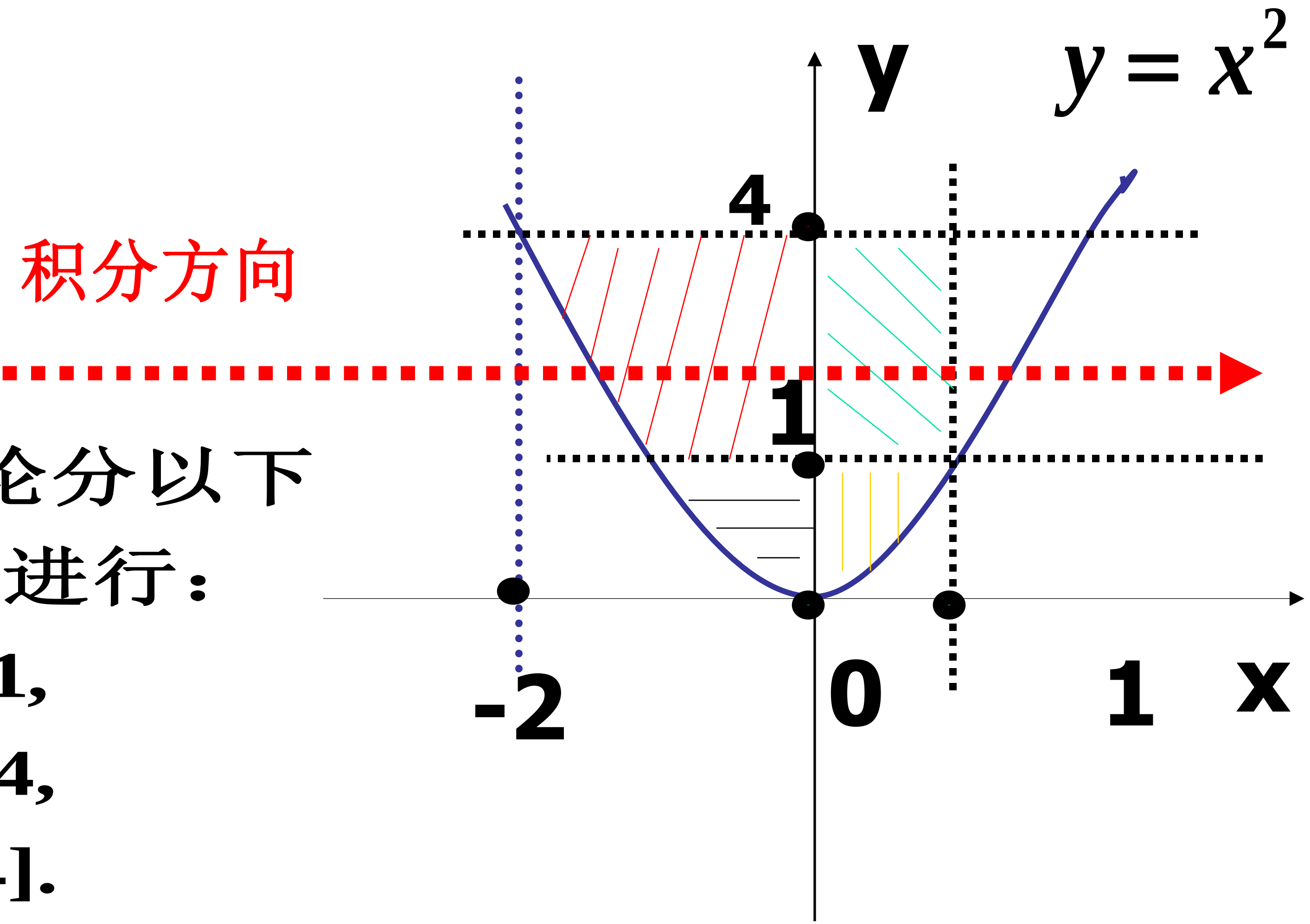
$$=\int_{-\sqrt{y}\le x\le\sqrt{y}} f_X(x)dx.$$

由于$f_X(x)$在不同的区域有不同的表示，要求出上述积分，必须把x积分区域划分成一些小的子区域使得被积函数$f_X(x)$在每个子区域上有唯一的表示,这就要求对相应的y进行讨论。

难点：　如何对y讨论？

积分区域为$y \geq x^2$，被积函数$f_X(x)$非零且有唯一表示的区域:0 ≤ x ≤ 1, 以及-2 ≤ x ≤ 0.

如图：



对y的讨论分以下几种情况进行：

1)$0 \leq y \leq 1$,

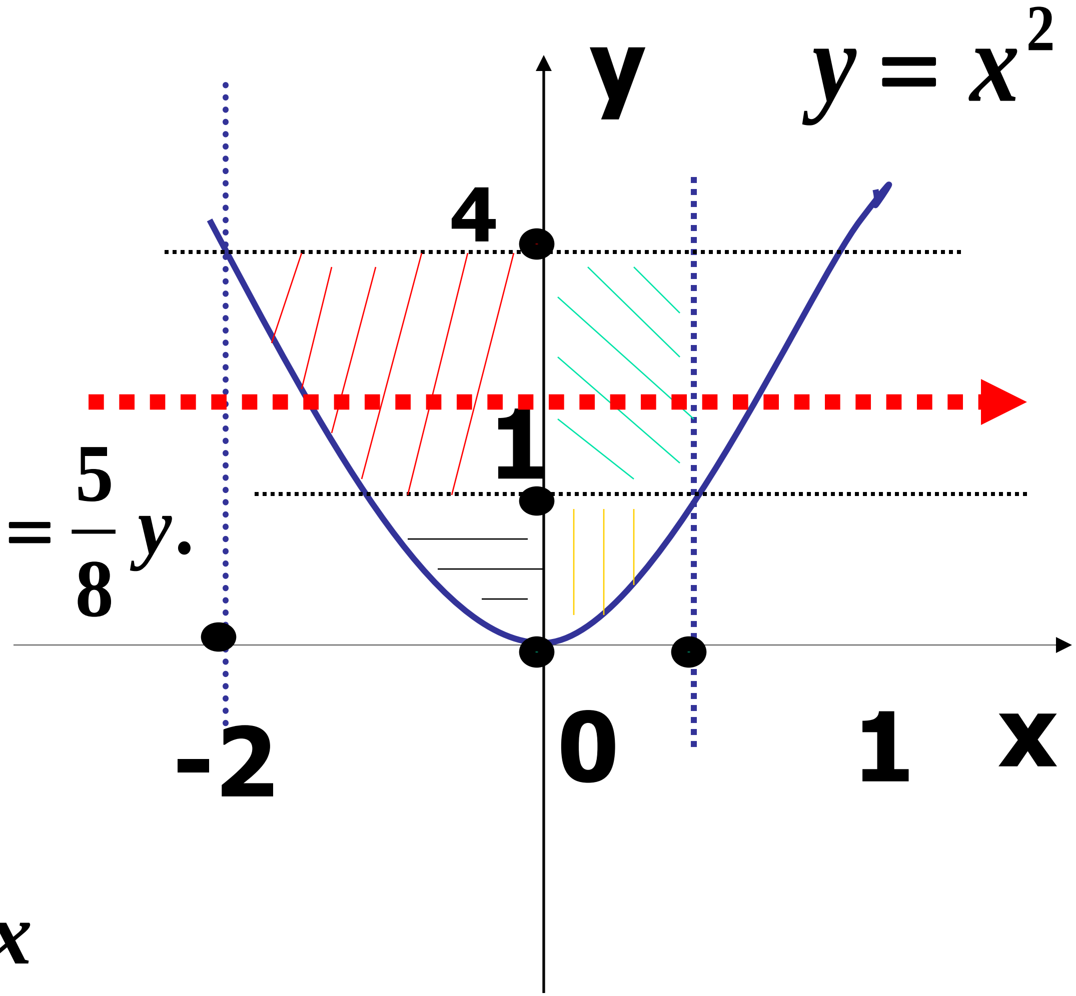
2)$1 \leq y \leq 4$,

3)$y \notin [0, 4]$.

当$0 \le y \le 1$,

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

$$= \int_{-\sqrt{y}}^{0} \frac{1}{4}(-x)dx + \int_{0}^{\sqrt{y}} xdx = \frac{5}{8}y.$$



当$1 \le y \le 4$,

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

$$= \int_{-\sqrt{y}}^{0} \frac{1}{4}(-x)dx + \int_{0}^{1} xdx + \int_{1}^{\sqrt{y}} 0dx = \frac{1}{8}y + \frac{1}{2}$$

因此

$$f_Y(y)=F_Y'(y)=\begin{cases}\frac{5}{8},0\le y\le 1\\ \frac{1}{8},1< y\le 4\\ 0, \quad else\end{cases}$$

特别地，当函数g严格单调时，有

定理：若连续型随机变量 X 的密度函数为 $f_X(x)$，函数 $Y=g(X)$ 严格单调且其反函数 $g^{-1}(Y)$ 有连续导数，则Y 也是连续型随机变量，且其密度函数为

$$f_Y(y)=\begin{cases} f_X(g^{-1}(y))\cdot|(g^{-1}(y))'|, & \alpha<y<\beta, \\ 0, & else。\end{cases}$$

其中 (α,β) 为函数 g 的值域。 （证明：略）

2、线性变换与平方变换

对于更为特殊的函数，他们有各自特殊的公式，如对于线性变换及平方变换，我们有：

定理：设随机变量X的密度函数为$f_X(x)$，则

1）若$Y=aX+b(a\neq 0)$ （a,b为常数），则

$$f_Y(y)=\frac{1}{|a|}f_X(\frac{y-b}{a}),\qquad y\in R(Y);$$

2）若 $Y=X^2$，则

$$f_Y(y)=\frac{1}{2\sqrt{y}}(f_X(\sqrt{y})+f_X(-\sqrt{y})),\quad y\in R(Y)。$$

证明：略。

例 设随机变量X在（0，1）上服从均匀分布，求Y=-2lnX的概率密度。

在区间（0，1）上，函数lnx<0,故

$$y=-2\ln x>0, \qquad y'=-\frac{2}{x}$$

于是y=-2lnx在（0，1）严格单调减，有反函数 $g^{-1}(y)=e^{-\frac{y}{2}}$,利用公式得

$$f_Y(y)=\begin{cases} f_X(e^{-\frac{y}{2}})\left|(e^{-\frac{y}{2}})'\right|, 0<e^{-\frac{y}{2}}<1, \\ 0, \qquad else. \end{cases}$$

即

$$f_Y(y)=\begin{cases} \frac{1}{2}e^{-\frac{y}{2}}, & y>0, \\ 0, & else. \end{cases}$$