例：圆盘直径$d \sim N\left(20, 0.15^2\right)$，面积$S = \pi\left(\frac{d}{2}\right)^2$，$S \sim ?$

统计工作的工具

已知X的分布，求$Y = g(X)$的分布

但我们并不很关心面积这样的实际问题，更关心正态分布的导出分布。

1.有实际背景的分布（研究对象分布）

$$\begin{cases} \begin{cases} X \sim B(1, p) \\ X \sim B(n, p) \\ X \sim P(\lambda) \\ X \sim G(p) \end{cases} \\ \begin{cases} X \sim U(a, b) \\ X \sim e(\lambda) \\ X \sim N\left(\mu, \sigma^2\right) \end{cases} \end{cases}$$

2.正态分布的导出分布 $\begin{cases} \chi^2(n)\text{分布} \\ t\text{分布} \\ F\text{分布} \end{cases}$

（没有实际背景，由正态分布导出）

由$X \sim N\left(\mu, \sigma^2\right)$，通过求$Y = g(X)$分布导出

四. 随机变量函数的分布

1.离散型随机变量函数分布求法

2.连续型随机变量函数分布求法

1.离散型：已知 X 的分布列 $P(X=x_i)=p_i \quad i=1,2\ldots$，求 $Y=g(X)$ 的分布列。

$$\begin{array}{ccccc} Y & & X & & P \\ \begin{pmatrix} g(x_1)\\ g(x_2)\\ \vdots \\ g(x_n)\\ \vdots\end{pmatrix} & \xleftrightarrow{\text{一对一}} & \begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\\ \vdots\end{pmatrix} & \xleftrightarrow{\text{一对一}} & \begin{pmatrix}p_1\\ p_2\\ \vdots\\ p_n\\ \vdots\end{pmatrix}\end{array}$$

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3\ldots \\ p_1 & p_2 & p_3\ldots \end{pmatrix} \qquad Y=g(X)\sim\begin{pmatrix} g(x_1)=g(x_2) & & g(x_3)\ldots & \\ p_1 \quad + \quad p_2 & & p_3 & \ldots \end{pmatrix}$$

其中$g(x_i)$有相等数值，只在分布列中用一个数值表示，其余相加。

例1.已知$X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{pmatrix}$，求$(1)$ $Y = 2X + 1$. (2) $Y = X^2$

解：$P(Y = 2X + 1) = P(Y = 2 \times (-1) + 1) = P(Y = -1) = 0.1$

$$(1) Y = 2X + 1 \sim \begin{pmatrix} -1 & 1 & 3 & 5 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{pmatrix}$$

$$(2) Y = X^2 \sim \begin{pmatrix} 0 & 1 & 4 \\ 0.2 & 0.4 & 0.4 \end{pmatrix}$$

例2.已知$X \sim \begin{pmatrix} 1 & 2 & \cdots & k & \cdots \\ 1/2 & 1/2^2 & \cdots & 1/2^k & \cdots \end{pmatrix}$， 求：(1) $Y = X^2$，(2) $Y = \sin\dfrac{\pi}{2}X$.

解:(1) $X^2 \sim \begin{pmatrix} 1 & 2^2 & \cdots & k^2 & \cdots \\ 1/2 & 1/2^2 & \cdots & 1/2^k & \cdots \end{pmatrix}$. $Y = \sin\dfrac{\pi}{2}X \sim \begin{pmatrix} -1 & 0 & 1 \\ 2/15 & 5/15 & 8/15 \end{pmatrix}$

$$(2)\ P(Y=-1) = P\left(\sin\frac{\pi}{2}X = -1\right) = \sum_{k=1}^{+\infty} P(X = 4k-1) = \frac{1}{2^3} + \frac{1}{2^7} + \cdots = \frac{2}{15}$$

$$P(Y=0) = P\left(\sin\frac{\pi}{2}X = 0\right) = \sum_{k=1}^{+\infty} P(X = 2k) = \frac{1}{2^2} + \frac{1}{2^4} + \cdots = \frac{5}{15}$$

$$P(Y=1) = P\left(\sin\frac{\pi}{2}X = 1\right) = \sum_{k=1}^{+\infty} P(X = 4k-3) = \frac{1}{2^1} + \frac{1}{2^5} + \cdots = \frac{8}{15}$$

2.连续型：已知 X 的分布密度 $f(x)$，求 $Y=g(X)$ 的密度。

例3. 设随机变量 $X \sim U[0,2]$，求 $Y = X^2$ 的密度。

$$f(x)=\begin{cases}\dfrac{1}{2} & 0\le x\le 2\\ 0 & \text{其它}\end{cases}$$

解：由 $X \sim U[0,2]$ 及 $Y = X^2$ 知 $0 \le Y \le 4$,

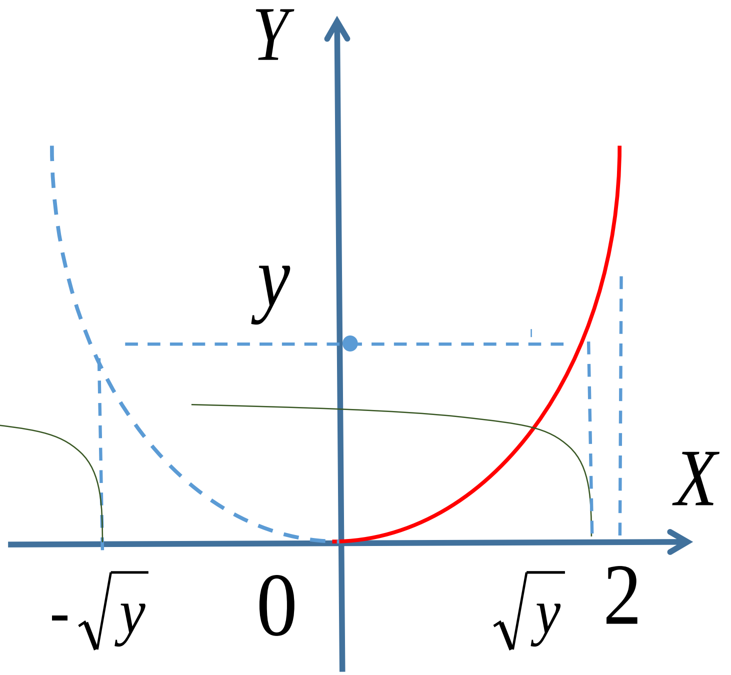
$$F_Y(y)=P(Y\le y)=P\left(X^2\le y\right)=P\left(-\sqrt{y}<X\le\sqrt{y}\right)$$

$$=P\left(X\le\sqrt{y}\right)-P\left(X\le-\sqrt{y}\right)=F_X\left(\sqrt{y}\right)-0$$

$$f_Y(y)=F_Y'(y)=F_X'\left(\sqrt{y}\right)=f_X\left(\sqrt{y}\right)\left(\sqrt{y}\right)'$$

$$=\frac{1}{2}\times\frac{1}{2\sqrt{y}}$$

$$f_Y(y)=\begin{cases}\dfrac{1}{4\sqrt{y}} & 0\le y\le 4\\ 0 & \text{其它}\end{cases}$$



I.已知X的密度$f(x)$求连续型随机变量$Y=g(X)$的一般步骤:

(i)由X的取值范围，及$Y=g(X)$确定Y的取值范围

(ii)求$F_Y(y)=P(Y\le y)=P(g(X)\le y)=P(X\in(h(y)))$

(iii)求出$F_Y(y)$的结果，再求$f_Y(y)=F_Y'(y)$。

或者直接对$F_Y'(y)=P'(X\in(h(y)))$

$$F_X'(h(y))=f_X(h(y))h(y)'$$

$$\left(F_X'=f_X\text{就如}\sin'=\cos\text{一个道理}\right)$$

例4. 设随机变量 $X\sim U[-1,2]$，求 $Y=X^2$ 的密度。

$$f(x)=\begin{cases}\dfrac{1}{3} & -1\le x\le 2\\ 0 & 其它\end{cases}$$

解：由$X\sim U[-1,2]$ 及$Y=X^2$知$0\le Y\le 4$， 当$0\le y\le 1$时，

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

$$=F_X(\sqrt{y})-F_X(-\sqrt{y})$$

$$f_Y(y)=F_Y'(y)=F_X'(\sqrt{y})-F_X'(-\sqrt{y})$$

$$=f_X(\sqrt{y})(\sqrt{y})'-f_X(-\sqrt{y})(-\sqrt{y})'$$

$$=\frac{1}{3\sqrt{y}}$$

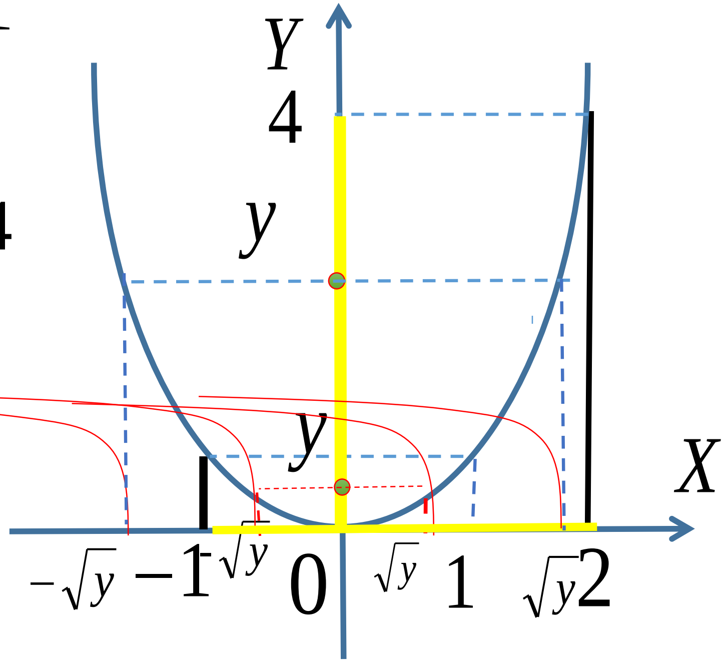
当$1\le y\le 4$时，

$$F_Y(y)=P(Y\le y)=F_X(\sqrt{y})-F_X(-\sqrt{y})$$

$$=F_X(\sqrt{y})-0$$

$$f_Y(y)=F_Y'(y)=F_X'(\sqrt{y})=\frac{1}{6\sqrt{y}}$$

$$f_Y(y)=\begin{cases}\dfrac{1}{3\sqrt{y}} & 0\le y\le 1\\ \dfrac{1}{6\sqrt{y}} & 1<y\le 4\\ 0 & 其它\end{cases}$$



例5. 设随机变量 $X \sim U(0,\pi)$，求 $Y=\sin X$ 的分布密度。

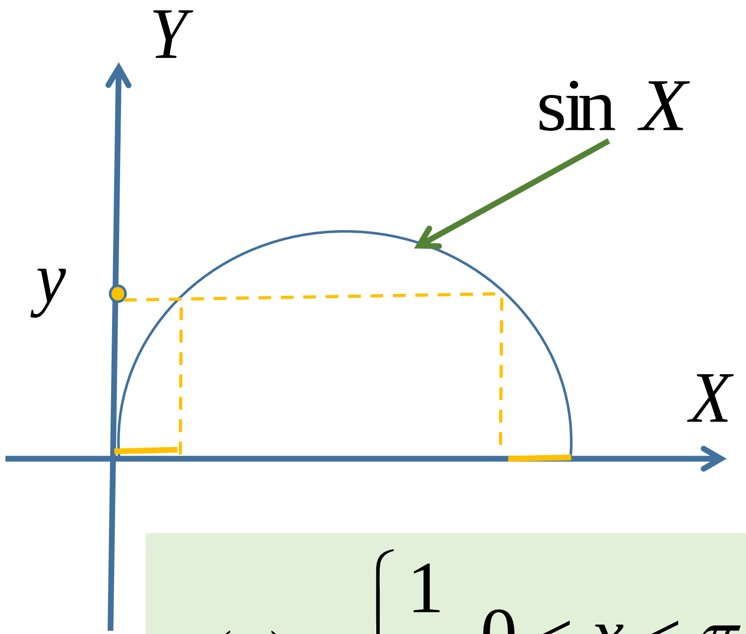
解： 由 $0<x<\pi$ 及 $Y=\sin X$ 有 $0<y<1$. 设Y的分布函数为 $F_Y(y)$,

$$F_Y(y)=P(Y\le y)=P(\sin X\le y)$$

$$=P(0\le X\le \arcsin y)+P(\pi-\arcsin y\le X\le \pi)$$

$$=\int_0^{\arcsin y}\frac{1}{\pi}dx+\int_{\pi-\arcsin y}^{\pi}\frac{1}{\pi}dx=\frac{2\arcsin y}{\pi}$$

$$f_Y(y)=F_Y'(y)=\begin{cases}\dfrac{2}{\pi\sqrt{1-y^2}}, & 0<y<1\\ 0 & \text{其他}\end{cases}$$



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

例6. 已知随机变量 $X\sim N\left(\mu,\sigma^2\right)$，求$Y=e^X$的密度。

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$-\infty<x<+\infty$$

解：由$X\sim N\left(\mu,\sigma^2\right)$ 及$Y=e^X$ 知$y\geq 0$。

$$F_Y(y)=P(Y\leq y)=P\left(e^X\leq y\right)=P(X\leq \ln y)=F_X(\ln y)$$

$$f_Y(y)=F_Y{}'(y)=F_X{}'(\ln y)=f_X(\ln y)(\ln y)'$$

$$=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}\frac{1}{y}$$

$$f_Y(y)=\begin{cases}\dfrac{1}{y\sqrt{2\pi}\sigma}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}} & y\geq 0\\ 0 & \text{其他}\end{cases}$$

方法2. $Y=e^X$是严格单调函数， $X=\ln Y$ 可以直接带公式

$$f_Y(y)=f_X\left(g^{-1}(y)\right)\left|\left(g^{-1}(y)\right)'\right|=f_X(\ln y)\left|(\ln y)'\right|=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}\frac{1}{y}$$

II.定理： 设连续性随机变量 X 的密度函数为 $f_X(x)$，$y=g(x)$ 是一个**严格单调函数**，且具有一阶连续导数，则 $Y=g(X)$ 的密度函数为：

$$f_Y(y)=f_X\left(g^{-1}(y)\right)\left|\left(g^{-1}(y)\right)'\right|$$

III.两个重要例子

例I. 已知 $X\sim N(\mu,\sigma^2)$ 求 $Y=aX+b$ 的分布密度。

线性变换不改变正态分布的分布形式

解1：$Y=g(x)=aX+b$ 为严格单调函数，则 $X=g^{-1}(y)=(Y-b)/a$

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

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

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

解2（一般方法）： 由 $X \sim N(\mu,\sigma^2)$ 及$Y=aX+b$ 知$-\infty<y<\infty$，

$$F_Y(y)=P(Y\le y)=P(aX+b\le y)$$

$$=\begin{cases} P\left(X\le \dfrac{y-b}{a}\right) & a>0 \\ P\left(X\ge \dfrac{y-b}{a}\right) & a<0 \end{cases} =\begin{cases} F_X\left(\dfrac{y-b}{a}\right) & a>0 \\ 1-F_X\left(\dfrac{y-b}{a}\right) & a<0 \end{cases}$$

$$X \sim N(\mu,\sigma^2)$$

$$a=\frac{1}{\sigma},b=-\frac{\mu}{\sigma}$$

$$f_Y(y)=\begin{cases} f_X\left(\dfrac{y-b}{a}\right)\dfrac{1}{a} & a>0 \\ -f_X\left(\dfrac{y-b}{a}\right)\dfrac{1}{a} & a<0 \end{cases} = f_X\left(\frac{y-b}{a}\right)\frac{1}{|a|}=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}\frac{1}{|a|}$$

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

$$Y=aX+b=\frac{X-\mu}{\sigma}\sim N(0,1)$$

例II. 已知$X \sim N(0,1)$　求$Y=X^2$的密度。

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

$$-\infty < x < +\infty$$

解：由$X \sim N(0,1)$ 及$Y=X^2$ 知$y \ge 0$,

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

$$=P(X\le\sqrt{y})-P(X\le-\sqrt{y})=\Phi_X(\sqrt{y})-\Phi_X(-\sqrt{y})$$

$$f_Y(y)=F_Y'(y)=\Phi_X'(\sqrt{y})-\Phi_X'(-\sqrt{y})=\varphi_X(\sqrt{y})(\sqrt{y})'-\varphi_X(-\sqrt{y})(-\sqrt{y})'$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{\sqrt{y}^2}{2}}\left(\frac{1}{2\sqrt{y}}\right)-\frac{1}{\sqrt{2\pi}}e^{-\frac{(-\sqrt{y})^2}{2}}\left(-\frac{1}{2\sqrt{y}}\right)$$

标准正态的平方分布是卡方分布

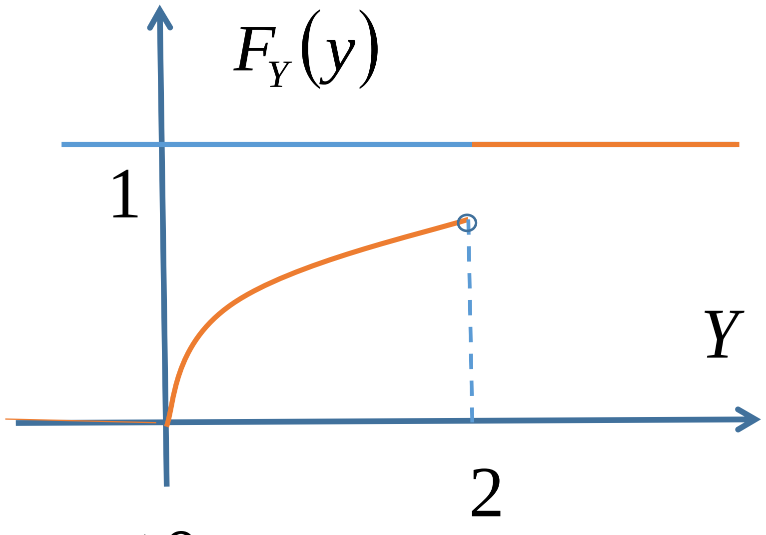
$$=\frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{y}{2}}$$

称Y服从$\chi^2(1)$分布

$$f_Y(y)=\begin{cases}\frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{y}{2}} & y\ge 0\\ 0 & y<0\end{cases}$$

例7. 设随机变量 $X \sim e(1/4)$，试求 $Y=\min(X,2)$ 的分布函数。

解：$X \sim e(1/4)$，$f(x)=\begin{cases}\frac{1}{4}e^{-\frac{x}{4}} & x>0\\ 0 & 其他\end{cases}$；　$Y=\min(X,2)=\begin{cases}X & X<2\\ 2 & X\ge 2\end{cases}$　则 $0<y\le 2$



当 $y\le 0$ 时，$F_Y(y)=0$　　当 $0<y<2$ 时，$Y=X$

$$F_Y(y)=P(Y\le y)=P(X\le y)=F_X(y)=1-e^{-\frac{y}{4}}$$

$$P(Y=2)=P(X\ge 2)$$

$$=\int_2^{+\infty}\frac{1}{4}e^{-\frac{x}{4}}dx=e^{-\frac{1}{2}}$$

于是 $F_Y(y)=\begin{cases}0 & y\le 0\\ 1-e^{-\frac{y}{4}} & 0<y<2\\ 1 & y\ge 2\end{cases}$