概率论与数理统计

第二章　随机变量及其分布

§4 连续型随机变量及其概率密度

- 概率密度的概念与性质
- 常见连续型随机变量的分布

一、概率密度的概念与性质

1.概率密度函数的定义

如果对于随机变量 X的分布函数 $F(x)$,存在非负可积函数 $f(x)$,使对于任意实数 x有

$$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t,$$

则称X为连续型随机变量 ,其中函数$f(x)$称为X的**概率密度函数**,简称**概率密度**.

连续型随机变量的分布函数是连续函数.

2.概率密度函数的性质

1° $f(x) \geq 0$;

2° $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$;

3° 对于任意实数 $x_1, x_2 (x_1 \leq x_2)$,

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\mathrm{d}x;$$

4° 若$f(x)$在点x处连续，则有$F'(x) = f(x)$.

证明　2°　$1 = F(\infty) = \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$.

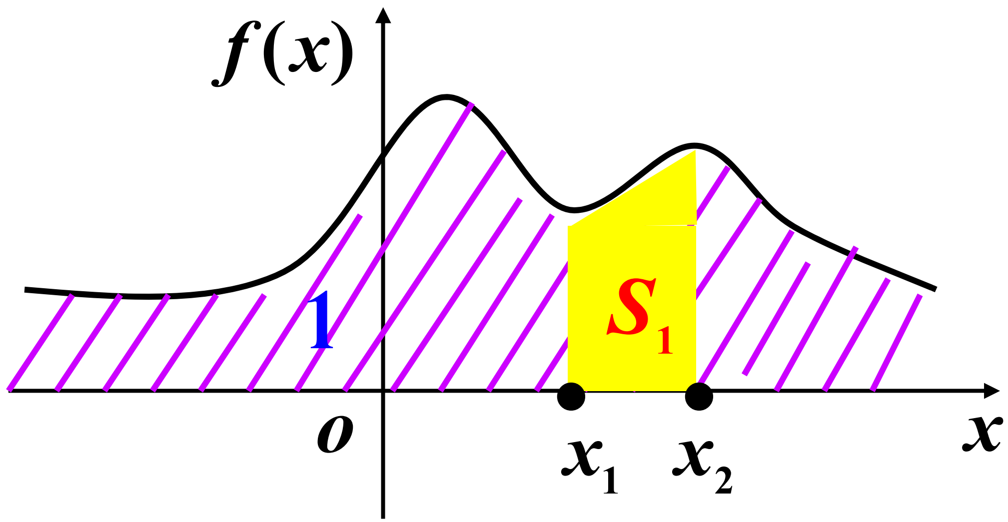
3°　$P\{x_1 < X \le x_2\} = F(x_2) - F(x_1)$

$$= \int_{-\infty}^{x_2} f(x)\,\mathrm{d}x - \int_{-\infty}^{x_1} f(x)\,\mathrm{d}x$$

$$= \int_{x_1}^{x_2} f(x)\,\mathrm{d}x .$$

$$S = \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

$$S_1 = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x$$



同时得以下计算公式

$$P\{X \le a\} = F(a) = \int_{-\infty}^{a} f(x)\,\mathrm{d}x,$$

$$P\{X > a\} = 1 - P\{X \le a\} = 1 - F(a)$$

$$= \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x - \int_{-\infty}^{a} f(x)\,\mathrm{d}x$$

$$= \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x + \int_{a}^{-\infty} f(x)\,\mathrm{d}x = \int_{a}^{\infty} f(x)\,\mathrm{d}x.$$

4° 若 $f(x)$ 在点 x 处连续，则有 $F'(x) = f(x)$.

注意

对于任意指定值 a, 连续型随机变量取 a的概率等于零. 即 $P\{X=a\}=0$.

证明 $P\{X=a\}=\lim\limits_{\Delta x\to 0}\int_a^{a+\Delta x} f(x)\,\mathrm{d}\,x=0$.

$$\boxed{\begin{aligned}P\{a\le X\le b\}&=P\{a<X\le b\}=P\{a\le X<b\}\\&=P\{a<X<b\}.\end{aligned}}$$

连续型随机变量取值落在某区间的概率与端点无关

注意

若X是连续型随机变量，$\{X=a\}$是不可能事件，则有$P\{X=a\}=0$.

若 $P\{X=a\}=0$,

则不能确定 $\{X=a\}$ 是不可能事件

连续型

若 X 为离散型随机变量,

$\{X=a\}$ 是不可能事件 $\Leftrightarrow P\{X=a\}=0$.

离散型

例1 设随机变量 X 具有概率密度

$$f(x)=\begin{cases} kx, & 0\le x<3, \\ 2-\dfrac{x}{2}, & 3\le x\le 4, \\ 0, & \text{其他}. \end{cases}$$

(1) 确定常数 k; (2) 求 X 的分布函数 ; (3) 求 $P\{1<X\le \dfrac{7}{2}\}$.

解 (1) 由 $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x=1$, 得

$$\int_0^3 kx\,\mathrm{d}x + \int_3^4 \left(2 - \frac{x}{2}\right)\mathrm{d}x = 1,$$

解得 $k = \dfrac{1}{6}$. 于是 X 的概率密度为

$$f(x) = \begin{cases} \dfrac{x}{6}, & 0 \le x < 3, \\ 2 - \dfrac{x}{2}, & 3 \le x \le 4, \\ 0, & \text{其他}. \end{cases}$$

(2) X的分布函数为

$$F(x)=\begin{cases}0, & x<0, \\ \int_0^x \frac{x}{6}\mathrm{d}x, & 0\le x<3, \\ \int_0^3 \frac{x}{6}\mathrm{d}x+\int_3^x\left(2-\frac{x}{2}\right)\mathrm{d}x, & 3\le x<4, \\ 1, & x\ge 4.\end{cases}$$

即

$$F(x)=\begin{cases}0, & x<0,\\ \dfrac{x^2}{12}, & 0\le x<3,\\ -3+2x-\dfrac{x^2}{4}, & 3\le x<4,\\ 1, & x\ge 4.\end{cases}$$

(3) $P\left\{1<X\le\dfrac{7}{2}\right\}=F\left(\dfrac{7}{2}\right)-F(1)=\dfrac{41}{48}.$

例2 设连续型随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x\le -a,\\ A+B\arcsin\dfrac{x}{a}, & -a<x\le a,\\ 1, & x>a.\end{cases}$$

求： (1) 系数 A, B 的值;

(2) $P\{-a<X<\dfrac{a}{2}\}$;

(3) 随机变量 X 的概率密度.

解 (1) 因为 X 是连续型随机变量, 所以 $F(x)$ 连续,

故有 $F(-a)=\lim\limits_{x\to -a} F(x)$,

$$F(a)=\lim_{x\to a} F(x),$$

即 $$A+B\arcsin\left(\frac{-a}{a}\right)=A-\frac{\pi}{2}B=0,$$

$$A+B\arcsin\left(\frac{a}{a}\right)=A+\frac{\pi}{2}B=1,$$

解之得 $A=\frac{1}{2},\quad B=\frac{1}{\pi}.$

所以 $$F(x)=\begin{cases}0, & x\le -a,\\ \frac{1}{2}+\frac{1}{\pi}\arcsin\frac{x}{a}, & -a<x\le a,\\ 1, & x>a.\end{cases}$$

(2) $P\{-a < X < \frac{a}{2}\} = F(\frac{a}{2}) - F(-a)$

$$= \frac{1}{2} + \frac{1}{\pi}\arcsin(\frac{a}{2a}) - 0$$

$$= \frac{1}{2} + \frac{1}{\pi} \times \frac{\pi}{6} = \frac{2}{3}.$$

(3) 随机变量 X 的概率密度为

$$f(x) = F'(x) = \begin{cases} 1/\pi\sqrt{a^2 - x^2}, & -a < x < a, \\ 0, & \text{其它}. \end{cases}$$

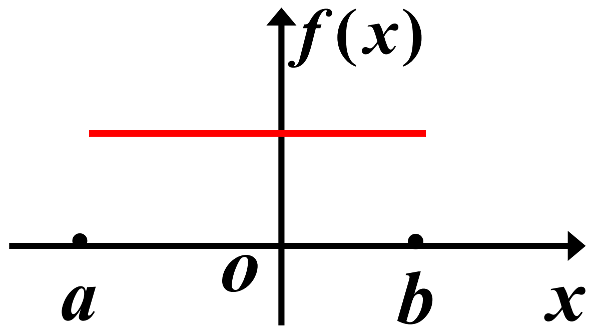
二、常见连续型随机变量及其概率分布

(一)均匀分布 若连续型随机变量 X 具有概率密度

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

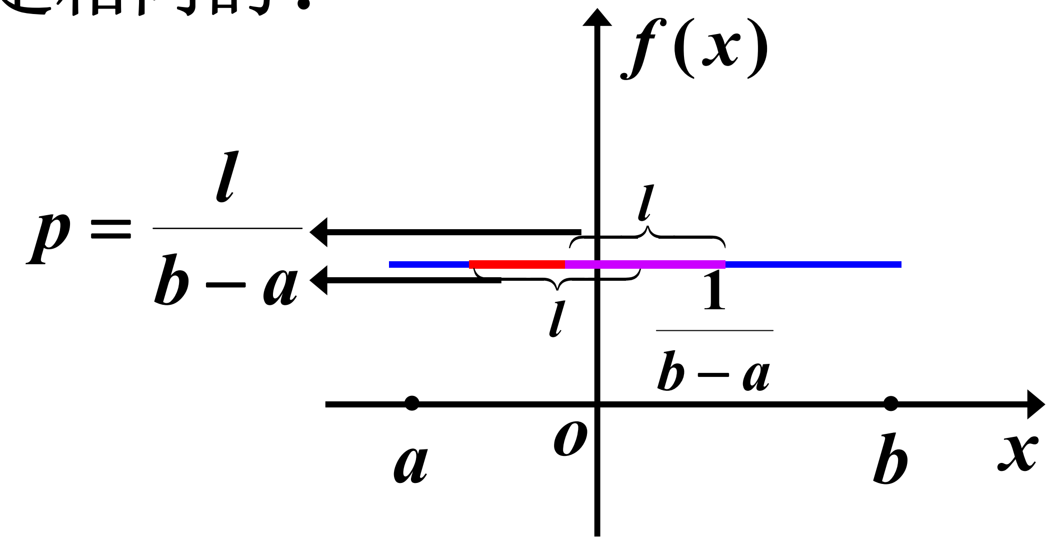
则称 X 在 (a,b) 上服从均匀分布. 记为 $X \sim U(a,b)$.

概率密度函数图形



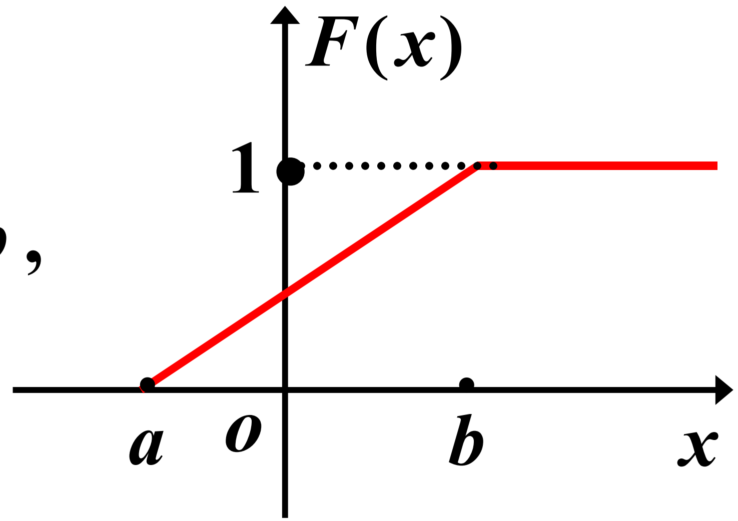
均匀分布的意义

在区间(a,b)上服从均匀分布的随机变量X，落在区间(a,b)中任意等长度的子区间内的可能性是相同的.



分布函数

$$F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\le x<b,\\ 1, & x\ge b.\end{cases}$$



例3 设电阻值 R 是一个随机变量，均匀分布在 $900\Omega \sim 1100\Omega$. 求 R 的概率密度及 R 落在 $950\Omega \sim 1050\Omega$ 的概率.

解 按题意, R 的概率密度为

$$f(r)=\begin{cases}\dfrac{1}{1100-900}, & 900<r<1100,\\ 0, & \text{其他}.\end{cases}$$

故有

$$P\{950<R\le 1050\}=\int_{950}^{1050}\frac{1}{200}\mathrm{d}r=0.5.$$

(二) 指数分布

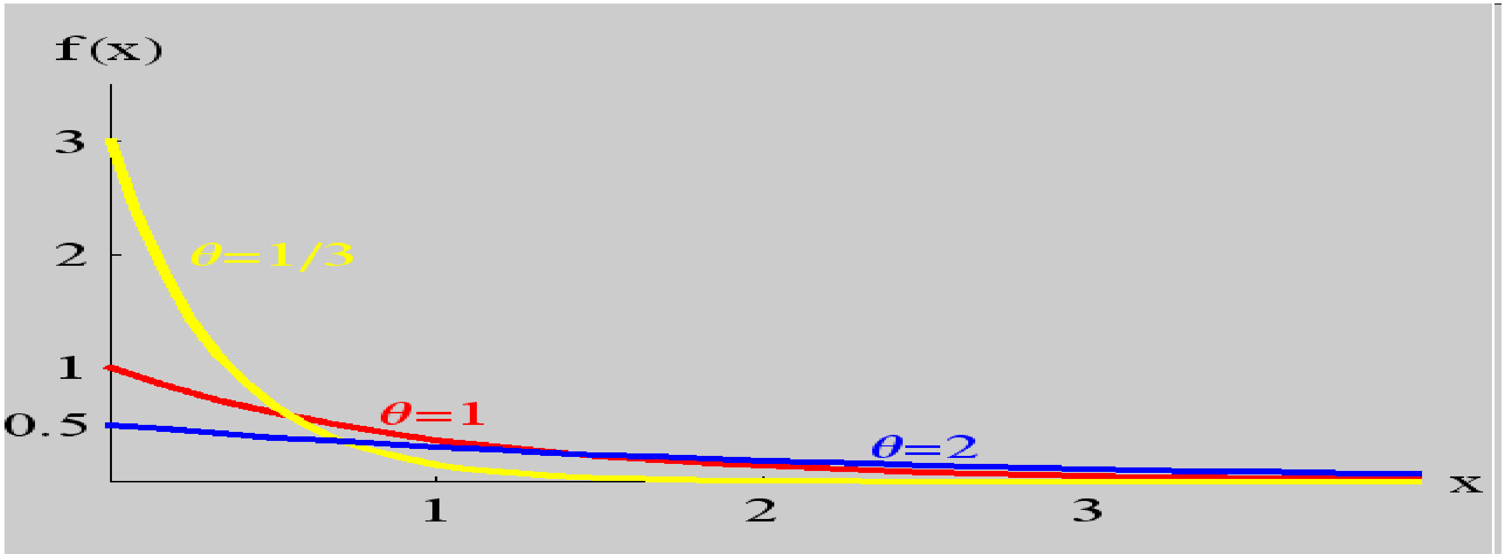
若连续型随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-x/\theta}, & x>0, \\ 0, & \text{其他},\end{cases}$$

其中$\theta>0$为常数,则称 X 服从参数为θ的指数分布 .

易知$f(x)\geq 0$, 且$\int_{-\infty}^{\infty} f(x)\mathrm{d}x=1$.

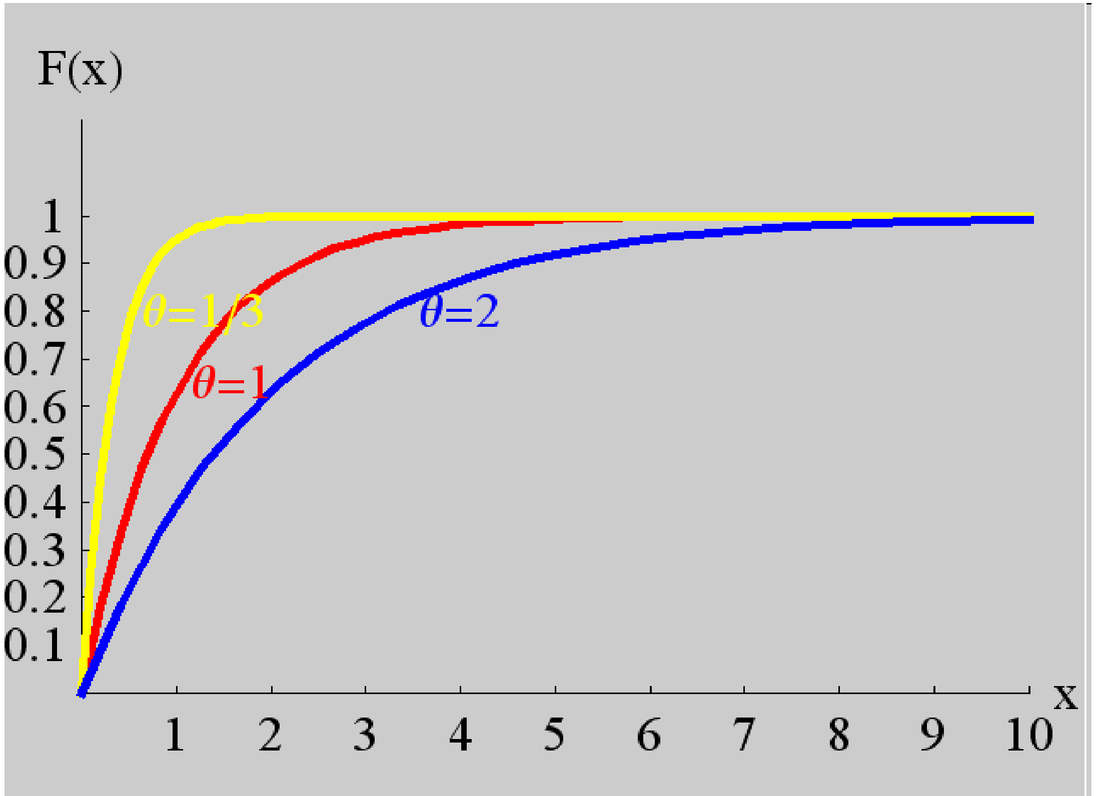
下图为 $\theta=\frac{1}{3}$, $\theta=1$, $\theta=2$ 时 $f(x)$ 的图形 .



由定义得随机变量 X 的分布函数为

$$F(x)=\begin{cases}1-\mathrm{e}^{-x/\theta}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

$\theta = \frac{1}{3}, \ \theta = 1, \ \theta = 2$时$F(x)$的图形如下：



指数分布的重要性质 :“无记忆性” .

对于任意 $s, t > 0$, 有

$$P\{X > s + t \mid X > s\} = P\{X > t\}.$$

事实上

$$P\{X > s + t \mid X > s\} = \frac{P\{(X > s + t) \cap (X > s)\}}{P\{X > s\}}$$

$$= \frac{P\{X > s + t\}}{P\{X > s\}} = \frac{1 - F(s + t)}{1 - F(s)}$$

$$= \frac{\mathrm{e}^{-(s+t)/\theta}}{\mathrm{e}^{-s/\theta}} = \mathrm{e}^{-t/\theta} = P\{X > t\}.$$

如果 X 是某一元件的寿命，那么**无记忆性**表明：

已知元件已使用 s 小时，它总共能使用至少 $s+t$ 小时的条件概率与从开始使用时算起，它至少能用 t 小时的概率相等. 这就是说：

元件对它已使用过 s 小时没有记忆.

应用与背景

某些元件或设备的寿命服从指数分布.例如无线电元件的寿命、电力设备的寿命、动物的寿命等都服从指数分布.

例3　某类日光灯管的使用寿命 X 服从参数为 θ=2000的指数分布(单位:小时).

(1)任取一只这种灯管, 求能正常使用1000小时以上的概率.

(2) 有一只这种灯管已经正常使用了1000 小时以上, 求还能使用1000小时以上的概率.

解　X 的分布函数为

$$F(x)=\begin{cases}1-\mathrm{e}^{-\frac{1}{2000}x}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

$$
\begin{aligned}
(1)\ P\{X > 1000\} &= 1 - P\{X \le 1000\} \\
&= 1 - F(1000) \\
&= \mathrm{e}^{-\frac{1}{2}} \\
&= 0.607\,.
\end{aligned}
$$

(2) $P\{X > 2000 \mid X > 1000\}$

$$= \frac{P\{X > 2000, X > 1000\}}{P\{X > 1000\}}$$

$$= \frac{P\{X > 2000\}}{P\{X > 1000\}}$$

$$= \frac{1 - P\{X \le 2000\}}{1 - P\{X \le 1000\}}$$

$$= \frac{1 - F(2000)}{1 - F(1000)} \quad = \mathrm{e}^{-\frac{1}{2}} = 0.607 .$$

(三)、正态分布

正态分布的概率密度函数

若连续型随机变量 X 的概率密度为

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

其中 $\mu,\sigma(\sigma>0)$ 为常数，则称 X 服从参数为 μ,σ 的正态分布或高斯分布. 记为 $X\sim N(\mu,\sigma^2)$.

高斯资料

显然$f(x)\geq 0$,下面来证明$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.

令$(x-\mu)/\sigma = t$,得到

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-t^2/2}\mathrm{d}t,$$

记 $I = \int_{-\infty}^{\infty}e^{-t^2/2}\mathrm{d}t$, 则有 $I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(t^2+u^2)/2}\mathrm{d}t\,\mathrm{d}u$

利用极坐标将它化成累次积分, 得到

$$I^2 = \int_0^{2\pi}\int_0^{\infty} re^{-r^2/2}\mathrm{d}r\mathrm{d}\theta = 2\pi,$$

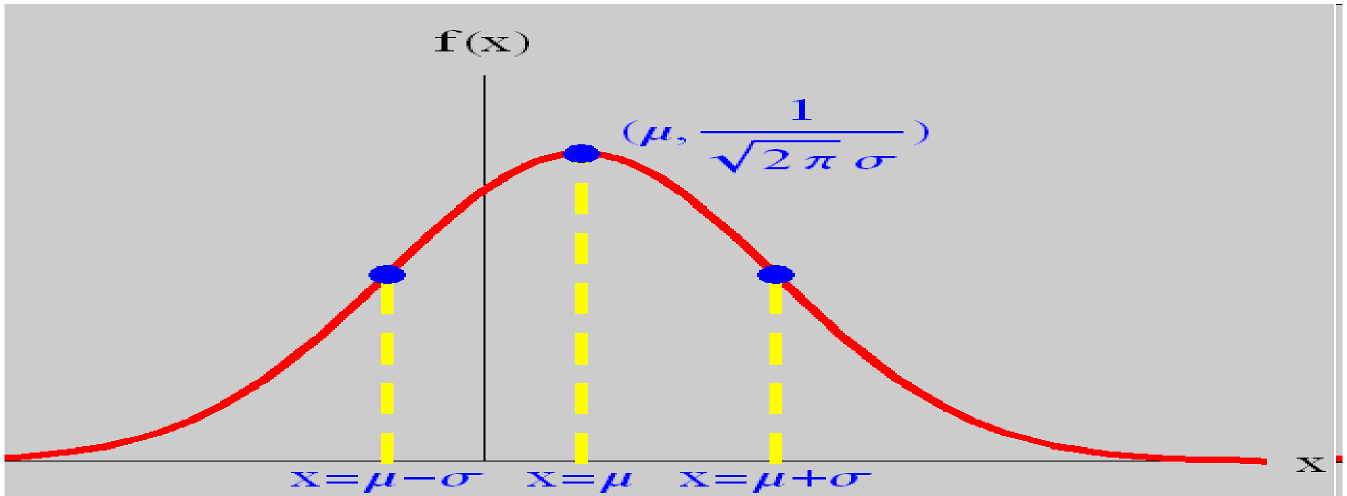
而$I > 0$, 故有$I = \sqrt{2\pi}$, 即有

$$\int_{-\infty}^{\infty} \mathrm{e}^{-t^2/2} \mathrm{d}t = \sqrt{2\pi}\,,$$

于是

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-t^2/2} \mathrm{d}t = 1\,.$$

$f(x)$的图形如图所示 .



性质:

1° 曲线关于 $x=\mu$ 对称 .这表明对于任意 $h>0$,

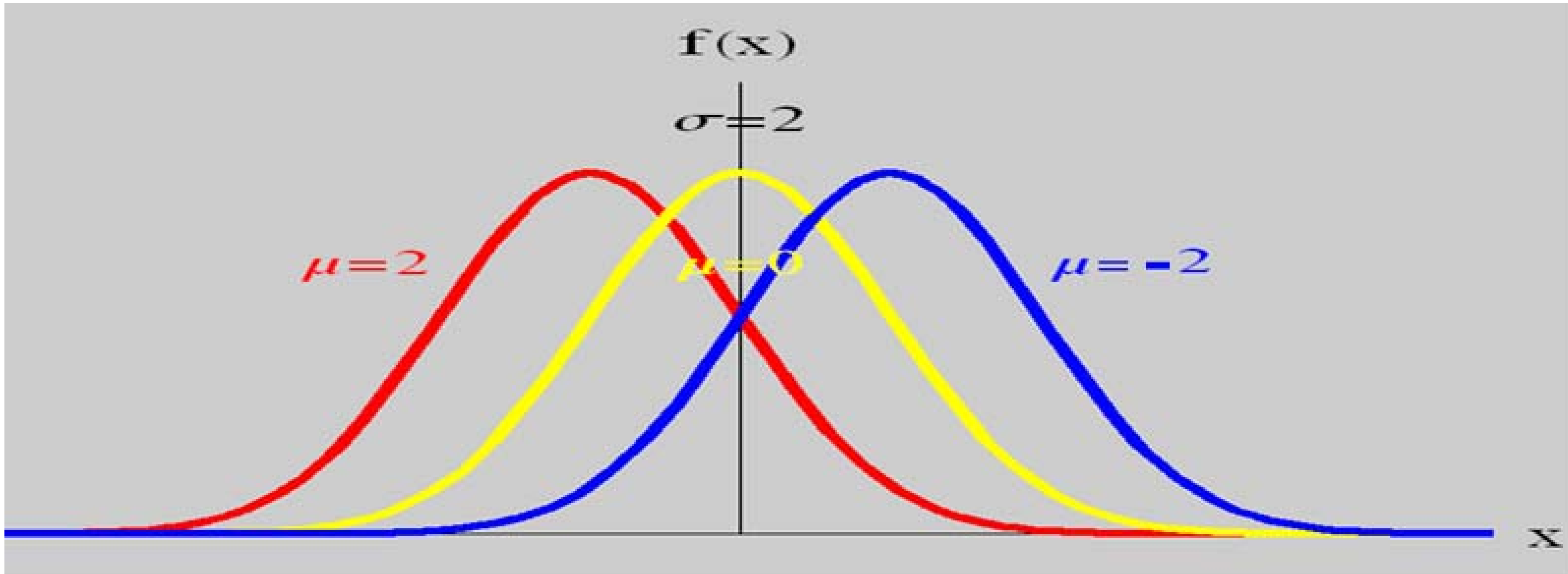
有 $P\{\mu-h<X\le\mu\}=P\{\mu<X\le\mu+h\}$.

2° 当$x=\mu$时取到最大值 $f(\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}$.

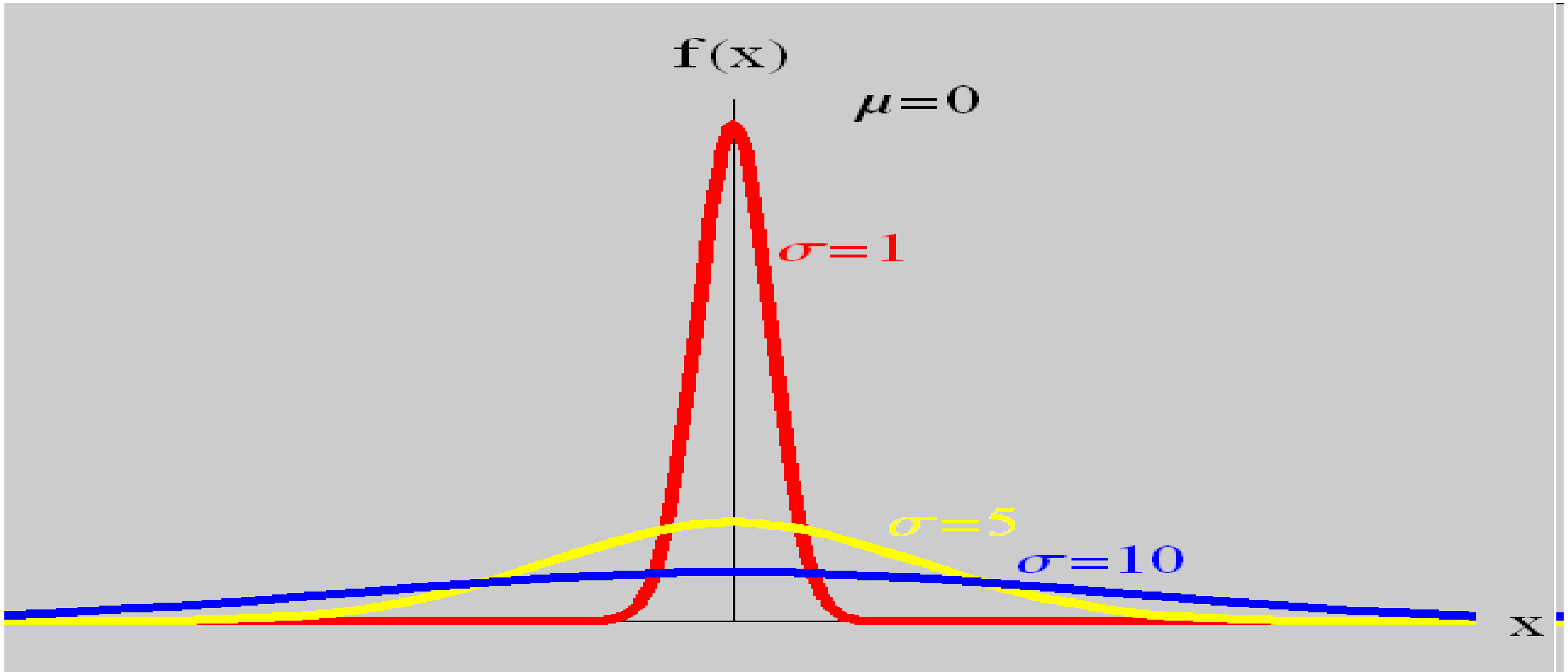
3° 在$x=\mu\pm\sigma$处曲线有拐点 ;

4° 曲线以 x 轴为渐近线 ;

5° 如果固定 σ ,改变 μ 的值，则图形沿着 Ox 轴平移，而不改变其形状，可见正态分布的概率密度曲线 $y = f(x)$ 的位置完全由参数 μ 所确定 .μ 称为位置参数.

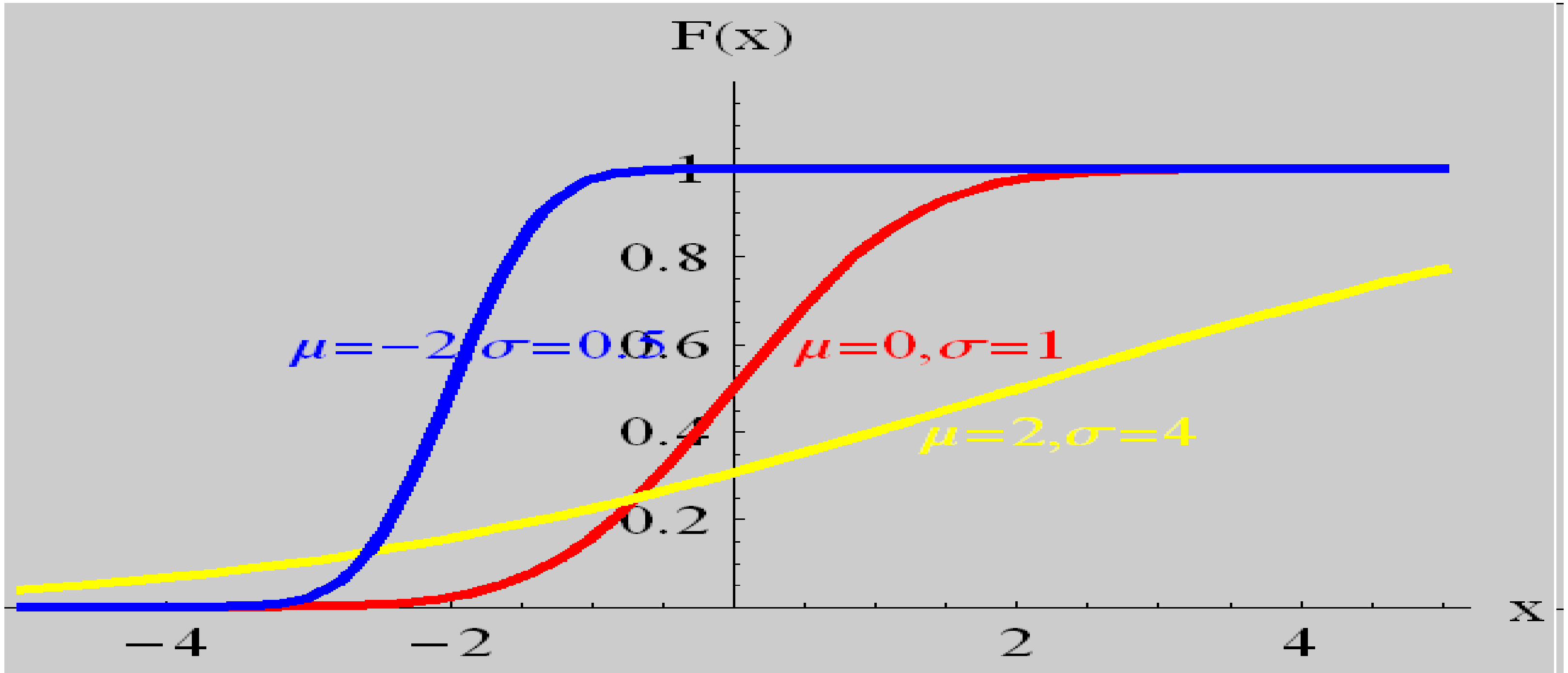


6°当固定 μ ,改变 σ 的大小时，$f(x)$ 图形的对称轴不变, 而形状在改变,σ 越小，图形越高越瘦, σ越大，图形越矮越胖.



分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(t-u)^2}{2\sigma^2}} dt$$



当$\mu=0$，$\sigma=1$时称 X 服从标准正态分布 .

其概率密度和分布函数 分别用 $\varphi(x),\Phi(x)$表示，

即有

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

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\mathrm{d}t .$$

易知

$$\Phi(-x)=1-\Phi(x)$$

$$\Phi(0)=0.5$$

性质 $\Phi(-x)=1-\Phi(x)$.

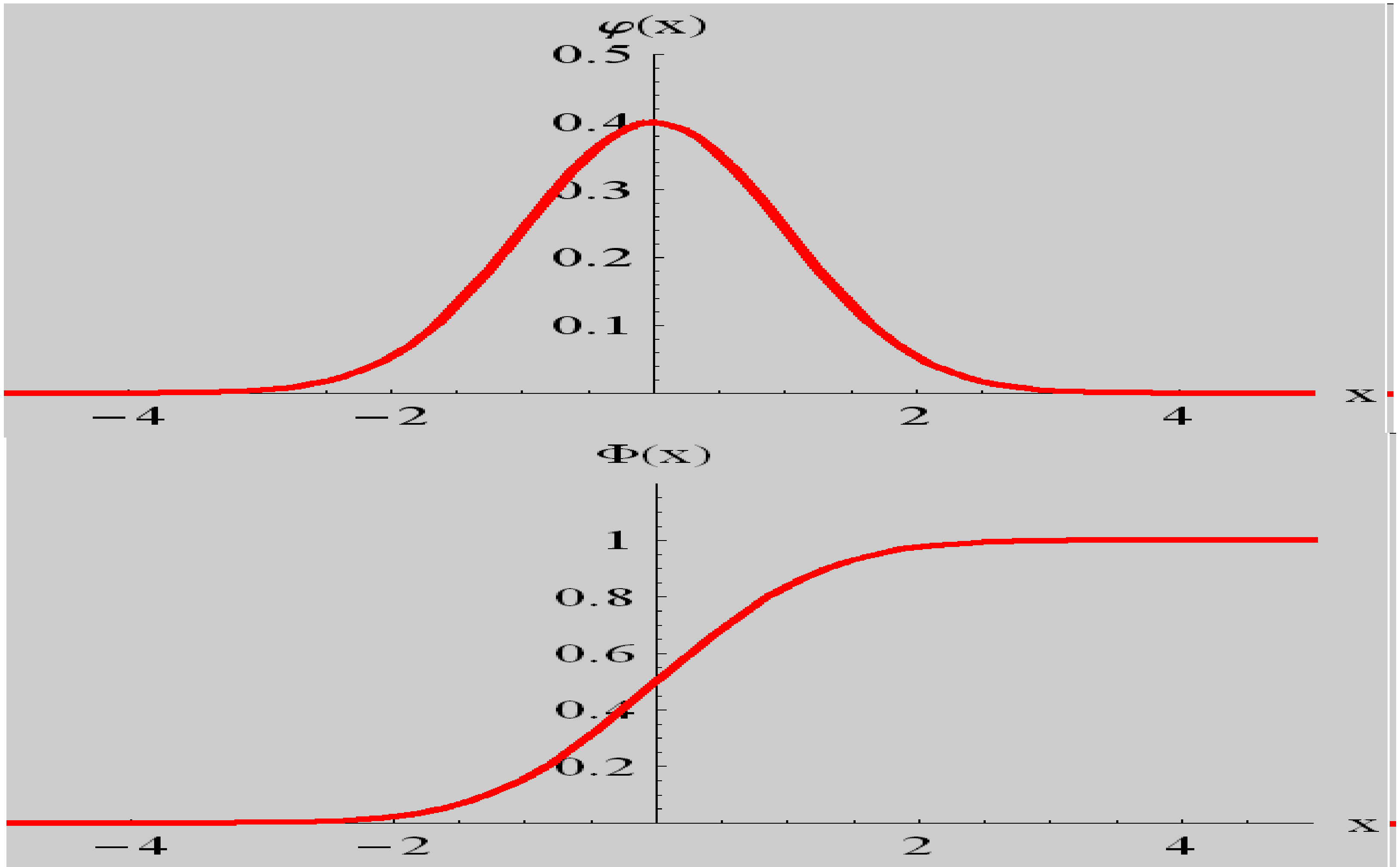
证明 $\Phi(-x)=\int_{-\infty}^{-x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x$

$$=\int_{x}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x-\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x,$$

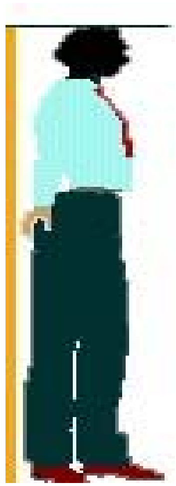
$$=1-\Phi(x).$$

标准正态分布的图形



正态分布的应用与背景

正态分布是最常见最重要的一种分布, 例如测量误差, 人的生理特征尺寸如身高、体重等；正常情况下生产的产品尺寸、直径、 长度、重量高度等都近似服从正态分布.



正态分布的计算

$$P\{X \le x\} = F(x) = \frac{1}{\sqrt{2\pi}\sigma} \boxed{\int_{-\infty}^{x} \mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t}$$

原函数不是初等函数

$$= ?$$

方法一:利用MATLAB软件包计算

方法二:转化为标准正态分布查表计算

引理 若$X \sim N(\mu,\sigma^2)$，则 $Z=\dfrac{X-\mu}{\sigma} \sim N(0,1)$.

证 $Z=\dfrac{X-\mu}{\sigma}$的分布函数为

$$P\{Z \le x\}=P\left\{\frac{X-\mu}{\sigma} \le x\right\}=P\{X \le \mu+\sigma x\}$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mu+\sigma x} \mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t,$$

令$\dfrac{t-\mu}{\sigma}=u$，得

$$P\{Z \le x\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2} du = \Phi(x)$$

由此知 $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

标准正态分布的概率密度表示为

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

则其分布函数 $F(x)$为

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad -\infty < x < \infty.$$

对于任意区间 $(x_1, x_2]$,有

$$P\{x_1 < X \le x_2\} = P\left\{\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right\}$$

$$= \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right).$$

例4　将一温度调节器放置在贮存着某种液体的容器内. 调节器整定在d °C，液体的温度X(以°C计)是一个随机变量, 且$X \sim N(d, 0.5^2)$. (1) 若$d = 90$，求X小于89°C的概率；(2) 若要求保持液体的温度至少为80 °C的概率不低于0.99，问d至少为多少？

解　(1) 所求概率为

$$P\{X < 89\} = P\left\{\frac{X-90}{0.5} < \frac{89-90}{0.5}\right\} = \Phi\left(\frac{89-90}{0.5}\right)$$

$$= \Phi(-2) \quad = 1 - \Phi(2) \quad = 1 - 0.9772 \quad = 0.0228.$$

(2) 按题意需求 d 满足

$$0.99 \le P\{X \ge 80\} = P\left\{\frac{X-d}{0.5} \ge \frac{80-d}{0.5}\right\}$$

$$=1-P\left\{\frac{X-d}{0.5} < \frac{80-d}{0.5}\right\} =1-\Phi(\frac{80-d}{0.5})$$

即 $$\Phi\left(\frac{d-80}{0.5}\right) \ge 0.99 = \Phi(2.327),$$

亦即 $$\frac{d-80}{0.5} \ge 2.327.$$

故需 $$d > 81.1635.$$

对于标准正态随机变量，我们引入上 α 分位点的定义.

设 $X \sim N(0,1)$，若 z_α 满足条件

$$P\{X > z_\alpha\} = \alpha, \quad 0 < \alpha < 1,$$

则称点 z_α 为标准正态分布的上 α 分位点.

下面列出了几个常用的 z_α 的值.

α	0.001	0.005	0.01	0.025	0.05	0.10
z_α	3.090	2.576	2.326	1.960	1.645	1.282

由 $\varphi(x)$ 图形的对称性知道 $z_{1-\alpha} = -z_\alpha$.

小结

1. 连续型随机变量 $\underbrace{F(x)}_{\text{分布函数}} = \int_{-\infty}^{x} \underbrace{f(t)}_{\text{概率密度}} \mathrm{d}t$

2. 常见连续型随机变量的分布

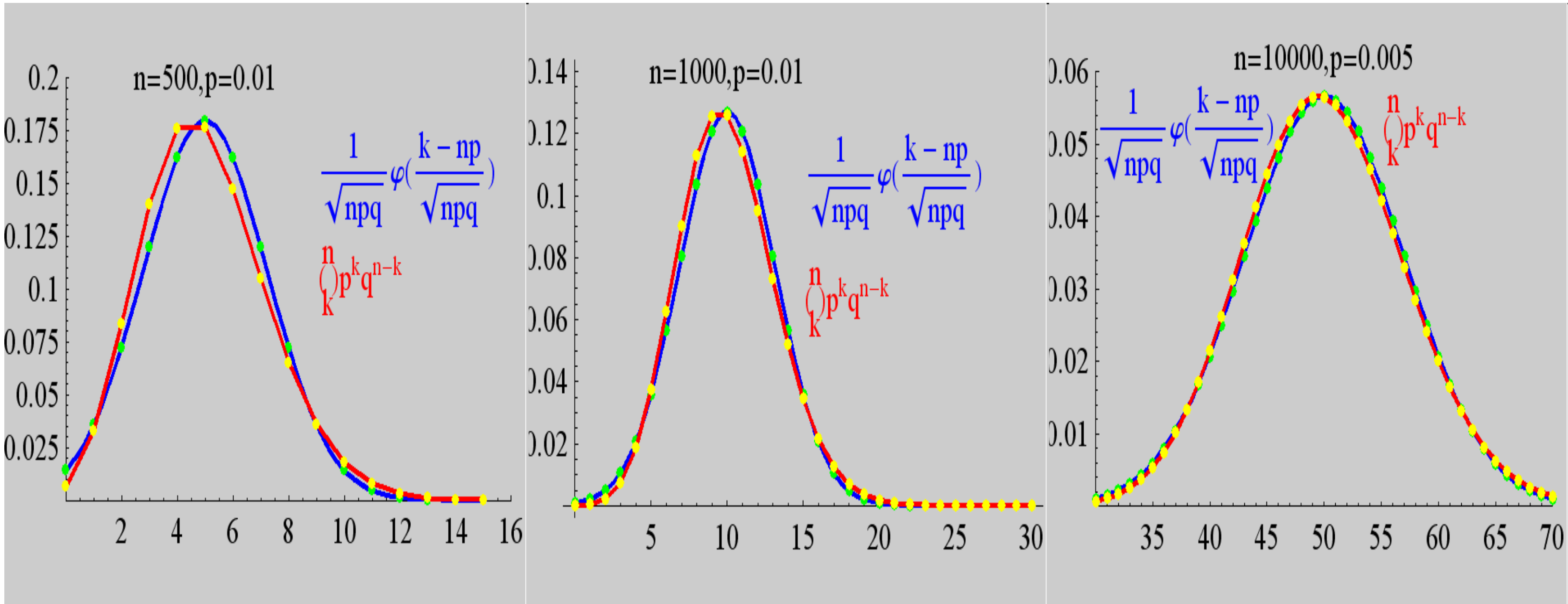
$$\begin{cases} \text{均匀分布} \\ \text{正态分布(或高斯分布)} \\ \text{指数分布} \end{cases}$$

3. 正态分布是概率论中最重要的分布

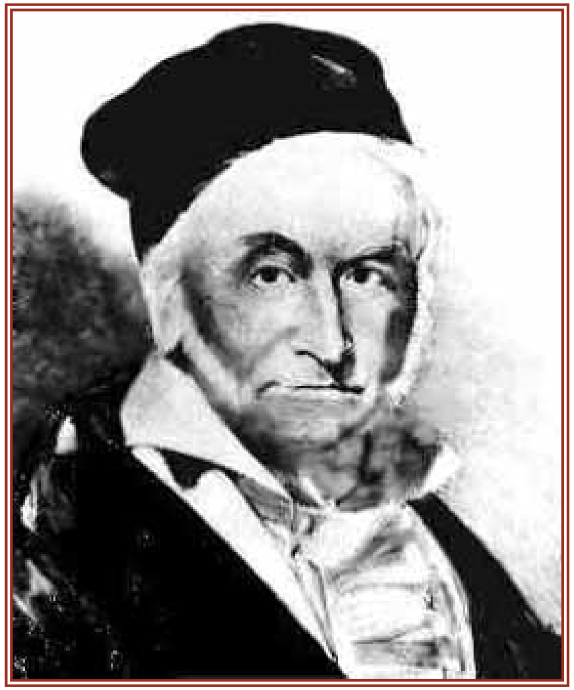
正态分布有极其广泛的实际背景，是自然界和社会现象中最为常见的一种分布，一个变量如果受到大量微小的、独立的随机因素的影响，那么这个变量一般是一个正态随机变量.

二项分布、泊松分布等的极限分布是正态分布．所以，无论在实践中，还是在理论上，正态分布是概率论中最重要的一种分布.

二项分布向正态分布的转换



高斯资料



Carl Friedrich Gauss

Born: 30 Apr. 1777 in Brunswick, Duchy of Brunswick (now Germany)

Died: 23 Feb. 1855 in Göttingen, Hanover (now Germany)

作业：课后习题 19、21、23、26、27、29、31

练习：

已知随机变量 X 的概率密度为

$$f(x)=A\mathrm{e}^{-|x|},\quad -\infty<x<+\infty.$$

(1) 求系数 A；

(2) 求 X 的分布函数 $F(x)$；

(3) 求 $Y=X^2$ 的概率密度．

解 (1) 由概率密度的性质，有

$$1=\int_{-\infty}^{+\infty}f(x)\,\mathrm{d}x=\int_{-\infty}^{+\infty}A\mathrm{e}^{-|x|}\,\mathrm{d}x=2\int_{0}^{+\infty}A\mathrm{e}^{-x}\,\mathrm{d}x$$

$$=2A,$$

故 $A=\dfrac{1}{2}.$

(2) $F(x)=\int_{-\infty}^{x}\frac{1}{2}\mathrm{e}^{-|x|}\,\mathrm{d}x$，

当 $x<0$ 时，有 $F(x)=\frac{1}{2}\int_{-\infty}^{x}\mathrm{e}^{x}\,\mathrm{d}x=\frac{1}{2}\mathrm{e}^{x}$；

当 $x\geq 0$ 时，有 $F(x)=\frac{1}{2}[\int_{-\infty}^{0}\mathrm{e}^{x}\,\mathrm{d}x+\int_{0}^{x}\mathrm{e}^{-x}\,\mathrm{d}x]=1-\frac{1}{2}\mathrm{e}^{-x}$；

所以 X 的分布函数为

$$F(x)=\begin{cases}\frac{1}{2}\mathrm{e}^{x}, & x<0,\\ 1-\frac{1}{2}\mathrm{e}^{-x}, & x\geq 0.\end{cases}$$

(3) 由于 $Y = X^2 \geq 0$,

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

当 $y > 0$ 时,有

$$F_Y(y) = P\{Y \leq y\} = P\{X^2 \leq y\}$$

$$= P\{-\sqrt{y} \leq X \leq \sqrt{y}\}$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} \mathrm{e}^{-|x|} \,\mathrm{d}x = 2\int_0^{\sqrt{y}} \frac{1}{2} \mathrm{e}^{-x} \,\mathrm{d}x,$$

由于 $F'_Y(y) = f_Y(y)$,

故当 $y>0$ 时，有

$$\frac{\mathrm{d}}{\mathrm{d}y}F_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}\left[\int_0^{\sqrt{y}}\mathrm{e}^{-x}\,\mathrm{d}x\right]$$

$$=\mathrm{e}^{-\sqrt{y}}\cdot\frac{1}{2\sqrt{y}},$$

从而，Y 的概率密度为

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