第四节　连续型随机变量及其概率密度

一、概率密度的概念与性质

二、常见连续型随机变量的分布

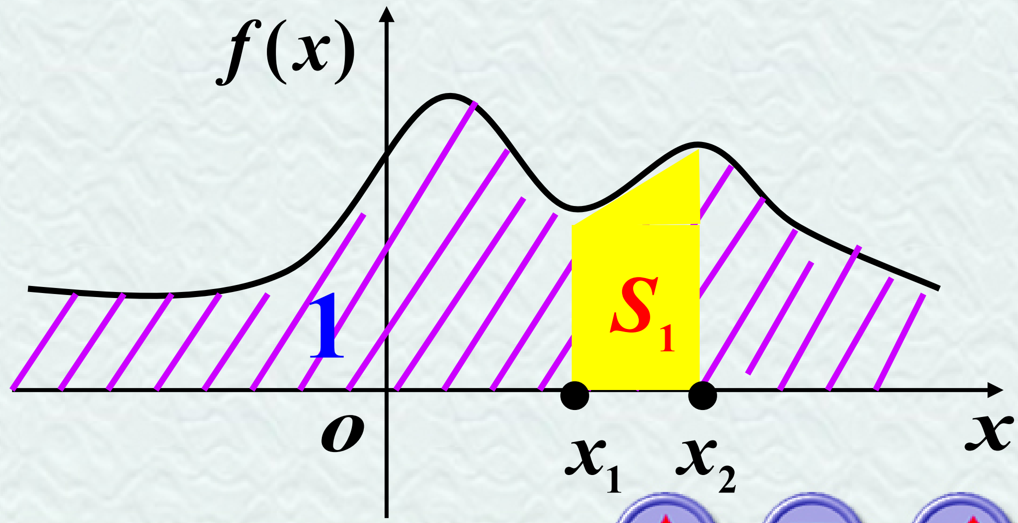
三、小结

一、概率密度的概念与性质

1.定义 如果对于随机变量 X 的分布函数 $F(x)$，存在非负函数，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 $f(x)$ 称为 X 的概率密度函数，简称概率密度.

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

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



性质 (1) $f(x) \geq 0$;

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

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

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

证明 $P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$

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

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 对于任意可能值 a ,连续型随机变量取 a 的概率等于零.即

$$P\{X=a\}=0.$$

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

由此可得

$$P\{a\le X\le b\}=P\{a<X\le b\}=P\{a\le X<b\}$$
$$=P\{a<X<b\}.$$

连续型随机变量取值落在某一区间的概率与区间的开闭无关

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

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

(1) 确定常数 k; (2) 求 X 的分布函数;

(3) 求 $P\{1<X\le \frac{7}{2}\}$.

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

得 $\int_0^3 kx\,\mathrm{d}x + \int_3^4 (2 - \frac{x}{2})\,\mathrm{d}x = 1$, 解之得 $k = \frac{1}{6}$.

(2) 由 $k = \frac{1}{6}$ 知 X 的概率密度为

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

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

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

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

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 $A = \dfrac{1}{2}, \quad B = \dfrac{1}{\pi}.$

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

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

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

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

(3) 随机变量 X 的概率密度为

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

二、常见连续型随机变量的分布

1. 均匀分布

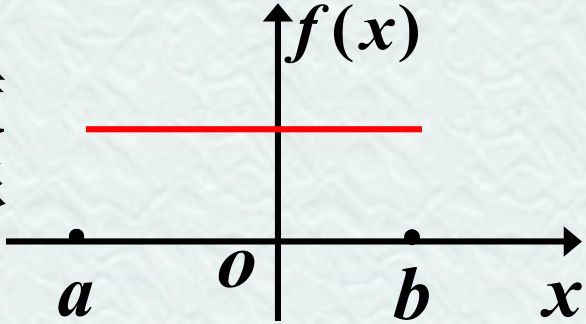
定义　设连续型随机变量 X 具有概率密度

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

则称 X 在区间 (a,b) 区间上服从均匀分布，

记为　$X\sim U(a,b)$.

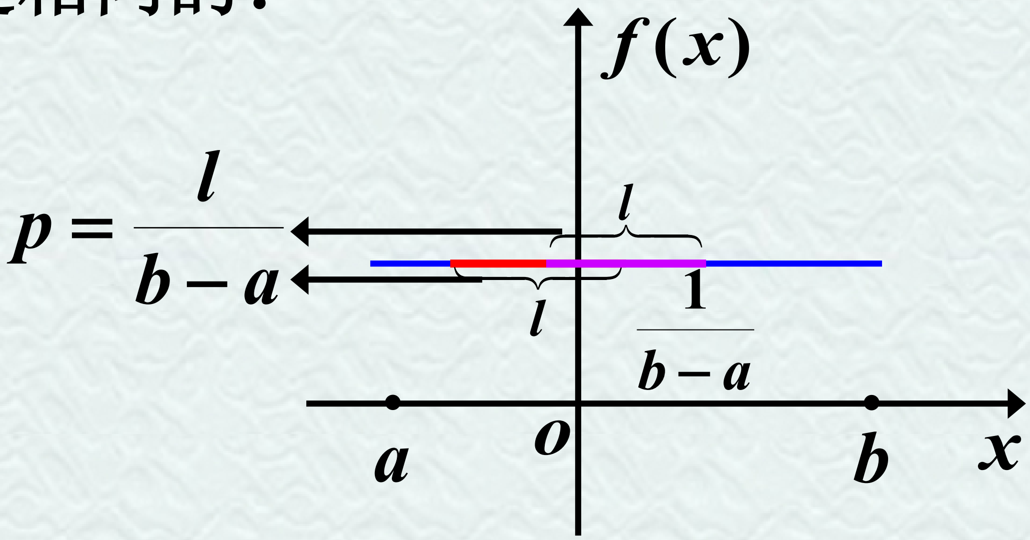
概率密度函数图形



均匀分布概率密度函数演示

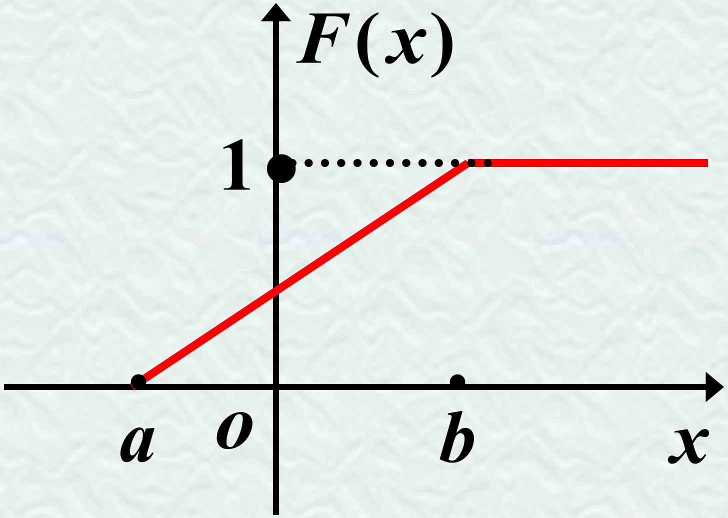
均匀分布的意义

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

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



均匀分布分布函数图形演示

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 由题意,R 的概率密度为

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

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

例4 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 的分布密度函数为

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

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

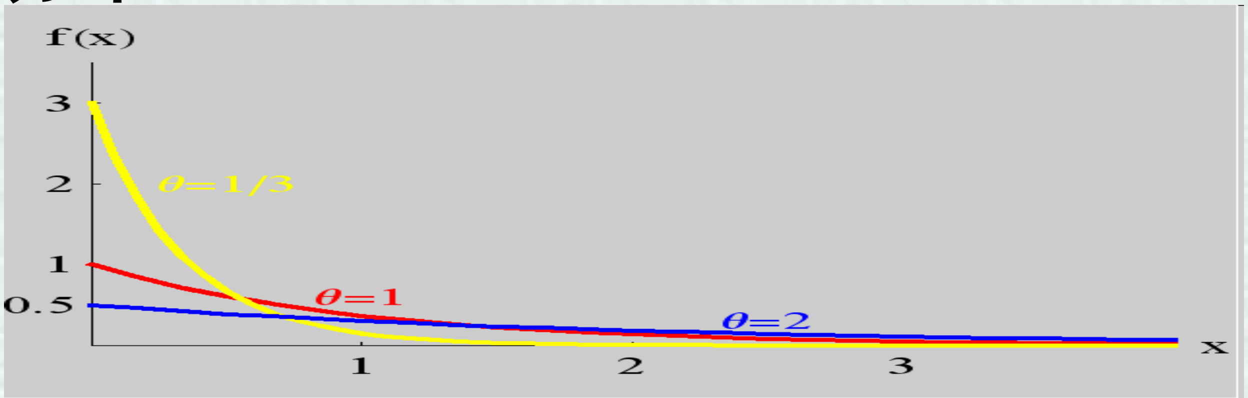
$$P\{Y\geq 2\}=\binom{3}{2}\left(\frac{2}{3}\right)^2\left(1-\frac{2}{3}\right)+\binom{3}{3}\left(\frac{2}{3}\right)^3\left(1-\frac{2}{3}\right)^0=\frac{20}{27}.$$

2. 指数分布

定义　设连续型随机变量 X 的概率密度为

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

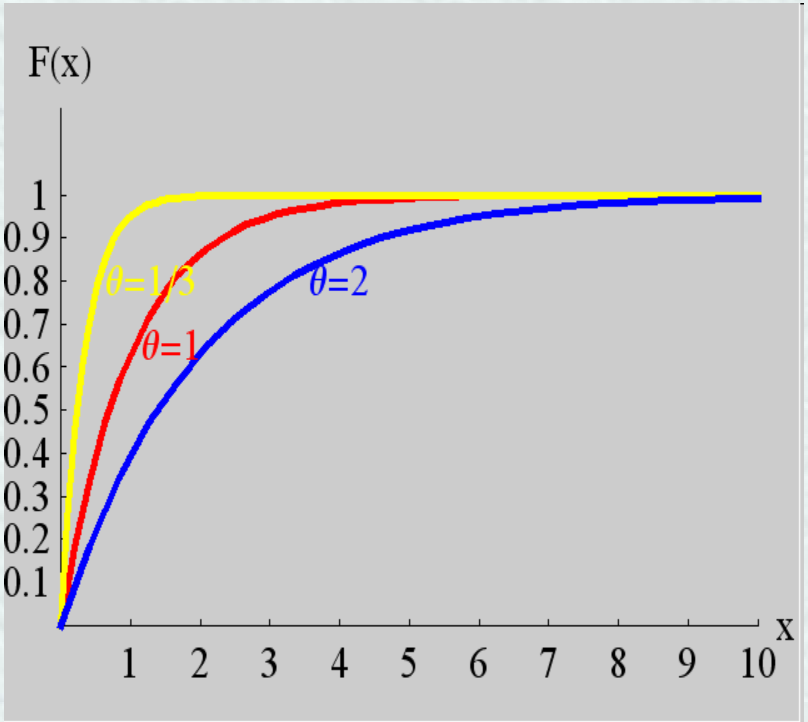
其中 $\theta>0$ 为常数,则称 X 服从参数为 θ 的指数分布.



指数分布密度函数图形演示

分布函数

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



指数分布分布函数图形演示

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5 设某类日光灯管的使用寿命 X 服从参数为 $\theta=2000$ 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

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

$$(1)\ P\{X > 1000\} = 1 - P\{X \le 1000\}$$

$$= 1 - F(1000)$$

$$= \mathrm{e}^{-\frac{1}{2}} \approx 0.607.$$

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

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

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

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

$$= \frac{1 - F(2000)}{1 - F(1000)}$$

$$= \mathrm{e}^{-\frac{1}{2}} \approx 0.607.$$

指数分布的重要性质 :“无记忆性”.

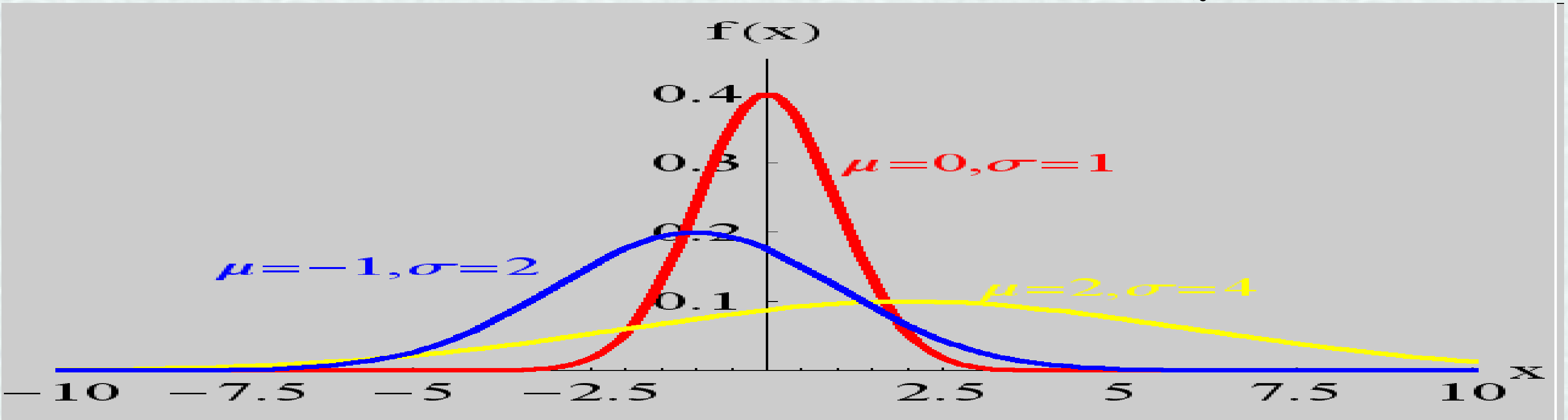
3. 正态分布(或高斯分布)

高斯资料

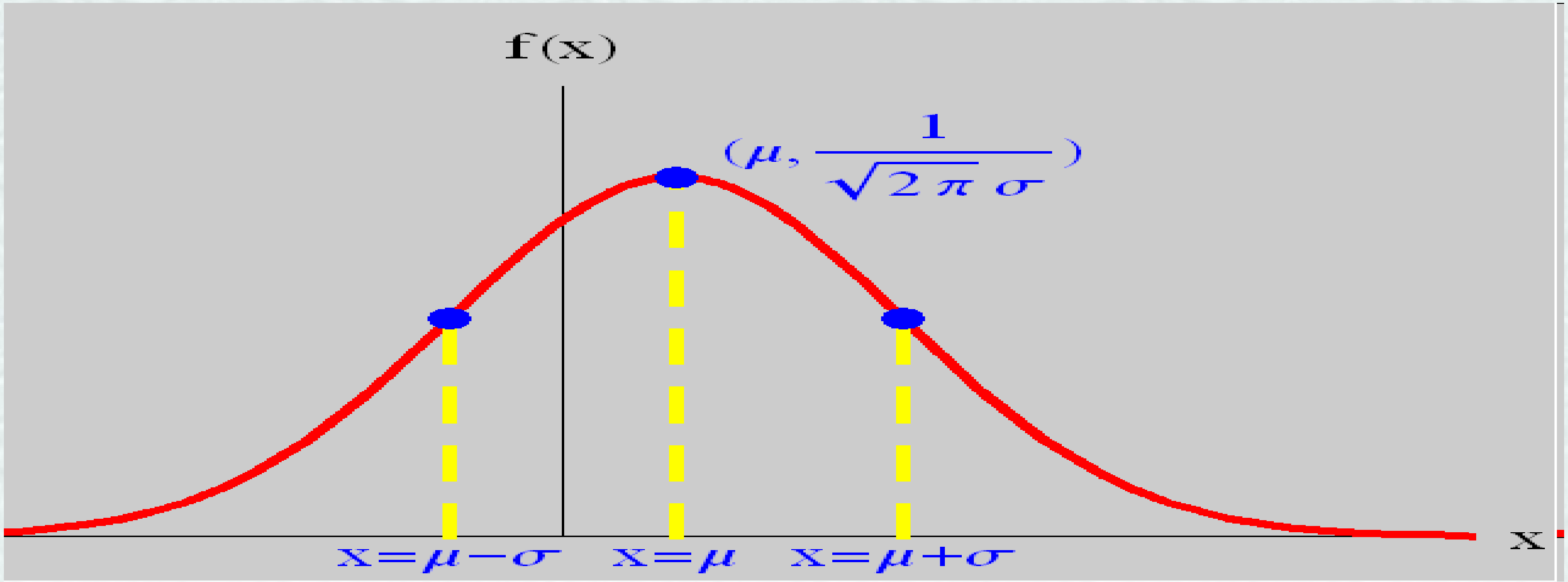
定义　设连续型随机变量 X 的概率密度为

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

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　$X\sim N(\mu,\sigma^2)$.



正态概率密度函数的几何特征



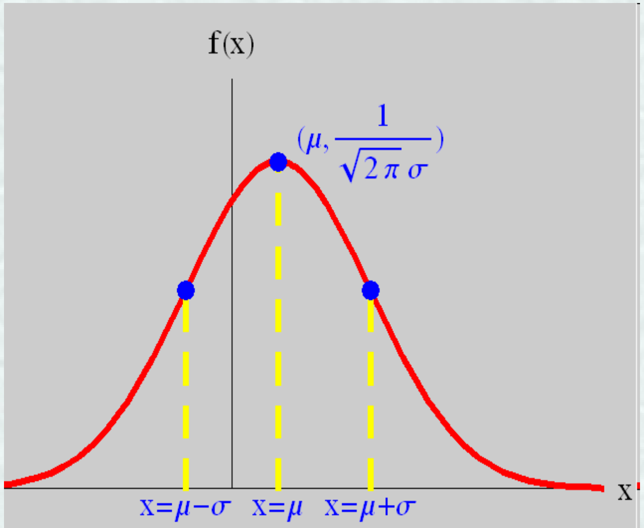
(1) 曲线关于 $x=\mu$ 对称;

(2) 当$x=\mu$时, $f(x)$取得最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$;

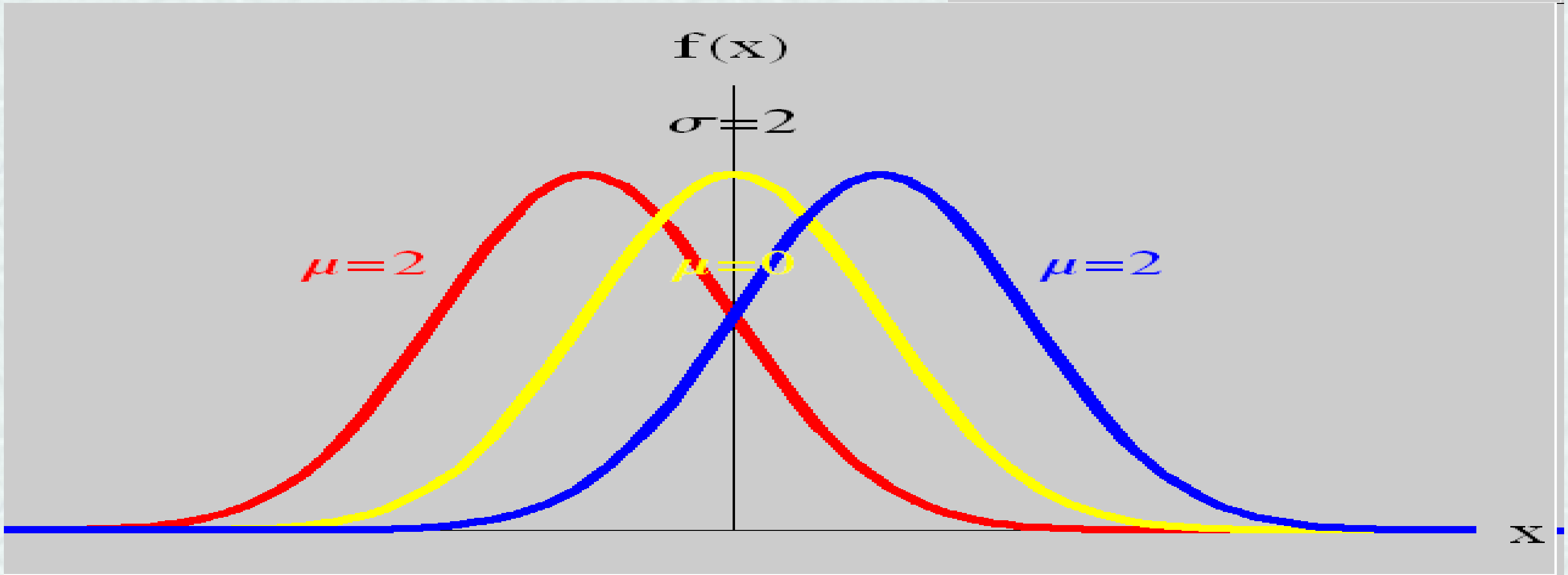
(3) 当 $x\to\pm\infty$ 时, $f(x)\to 0$;

(4) 曲线在 $x=\mu\pm\sigma$ 处有拐点;

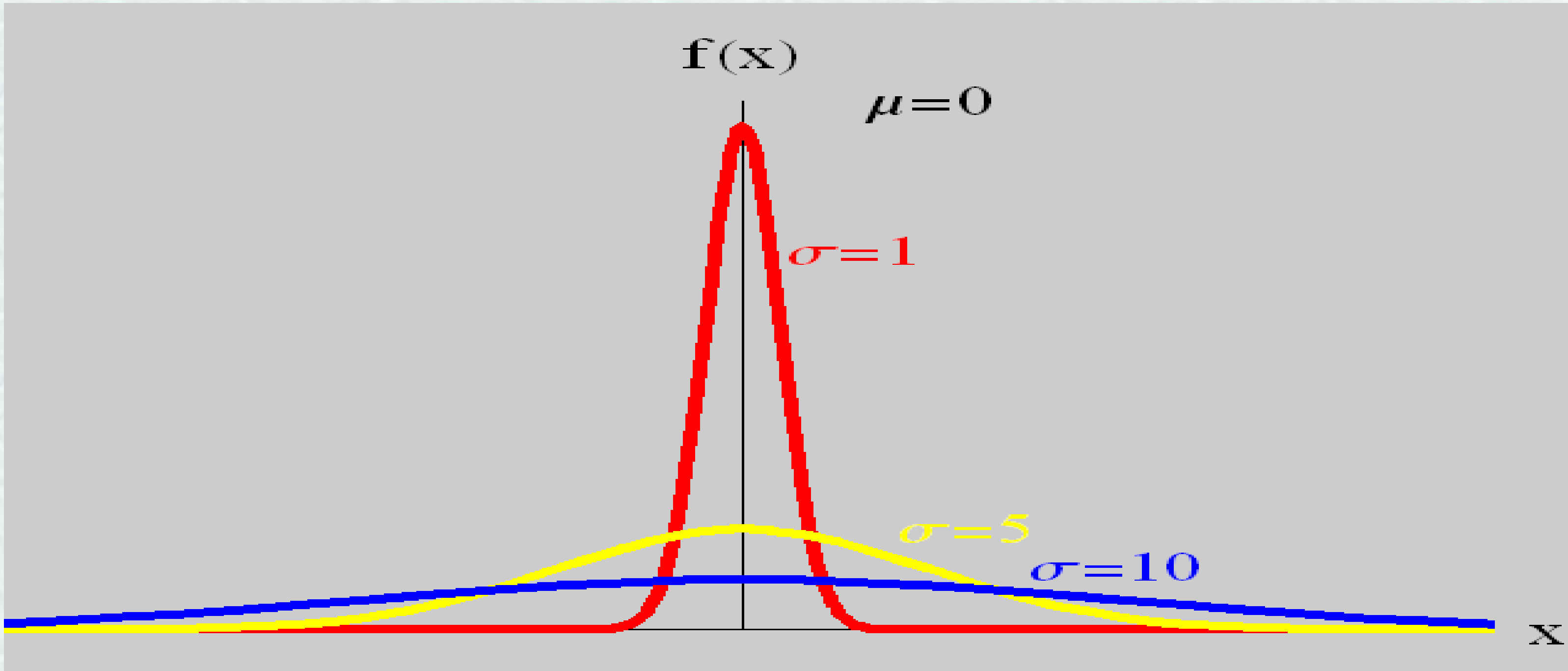
(5) 曲线以 x 轴为渐近线;



(6) 当固定 σ, 改变 μ 的大小时, $f(x)$ 图形的形状不变,只是沿着 x 轴作平移变换;



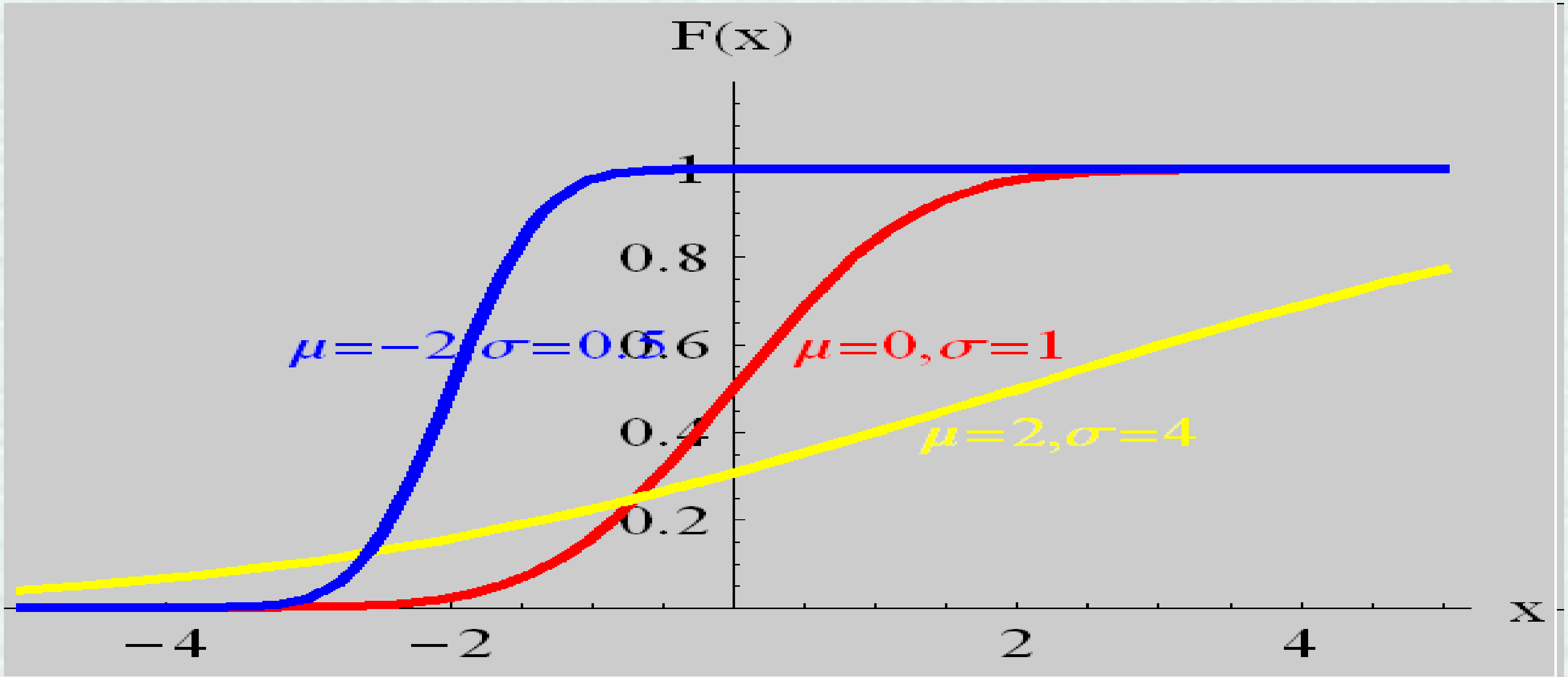
(7) 当固定 μ, 改变 σ 的大小时, $f(x)$ 图形的对称轴不变,而形状在改变, σ 越小,图形越高越瘦,σ越大,图形越矮越胖.



正态分布密度函数图形演示

正态分布的分布函数

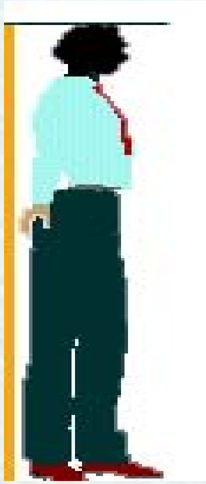
$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t$$



正态分布分布函数图形演示

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.



正态分布下的概率计算

原函数不是
初等函数

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

$$= ?$$

方法一:利用MATLAB软件包计算(演示)

方法二:转化为标准正态分布查表计算

标准正态分布

当正态分布 $N(\mu,\sigma^2)$ 中的 $\mu=0,\sigma=1$ 时,这样的正态分布称为标准正态分布,记为 $N(0,1)$.

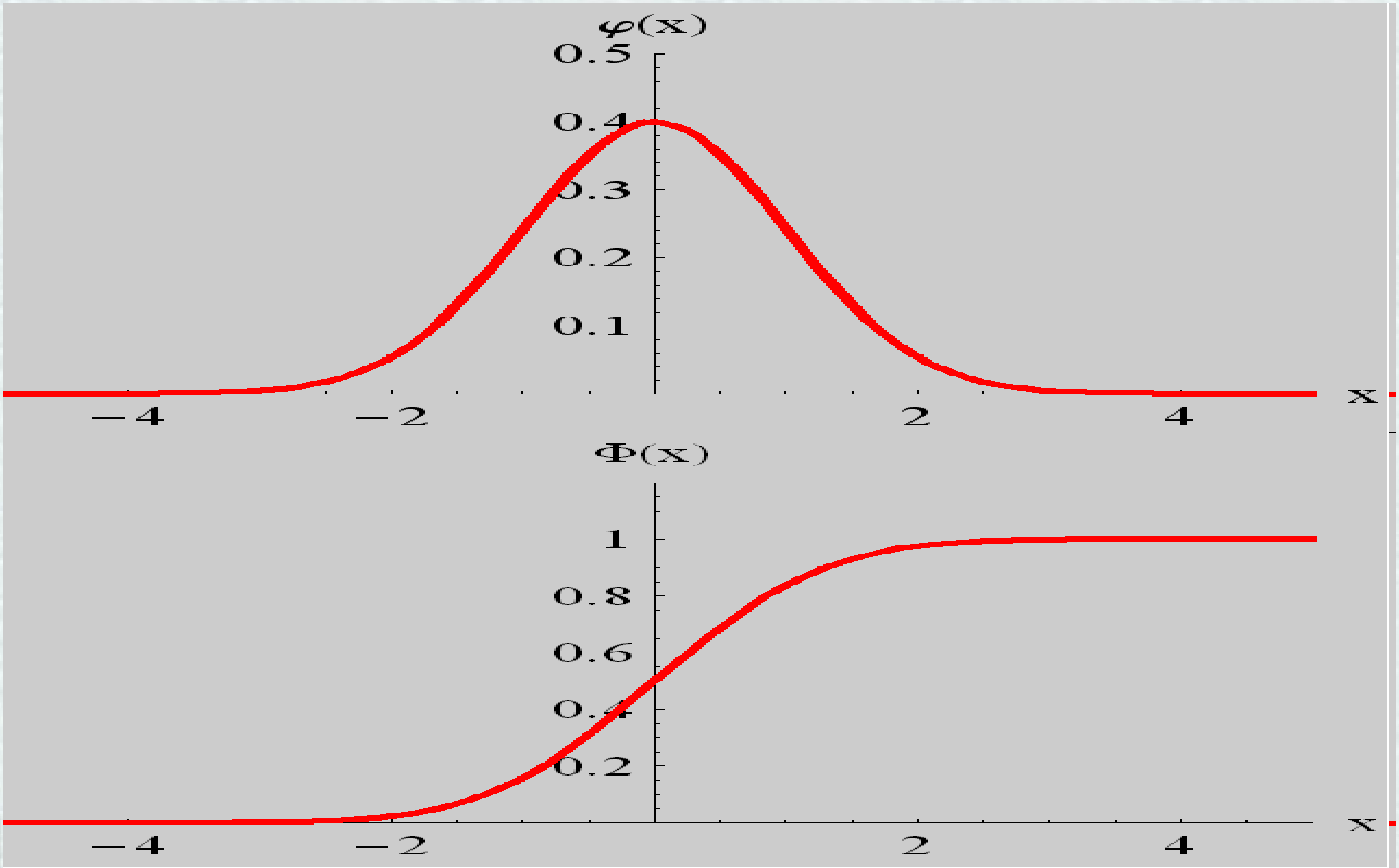
标准正态分布的概率密度表示为

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

标准正态分布的分布函数表示为

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

标准正态分布的图形



例6 已知 $X \sim N(0,1)$,求 $P\{1.25 \le X < 2\}$.

解 $P\{1.25 \le X < 2\}$

$= \Phi(2) - \Phi(1.25)$

$= 0.9772 - 0.8944$

$= 0.0828$.

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

证明 $Z=\dfrac{X-\mu}{\sigma}$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 $P\{Z \le x\}=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x} \mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u=\Phi(x),$

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

例7 已知 $X \sim N(\mu,\sigma^2)$,求 $P\{c \le X \le d\}$.

解 $$P\{c \le X \le d\} = \int_c^d \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x$$

令 $\dfrac{x-\mu}{\sigma} = u$, $$= \int_{\frac{c-\mu}{\sigma}}^{\frac{d-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{u^2}{2}} \cdot \sigma \, \mathrm{d}u$$

$$= \int_{\frac{c-\mu}{\sigma}}^{\frac{d-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u^2}{2}} \cdot \mathrm{d}u$$

$$=\int_{-\infty}^{\frac{d-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\cdot du-\int_{-\infty}^{\frac{c-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\cdot du$$

$$=\Phi\left(\frac{d-\mu}{\sigma}\right)-\Phi\left(\frac{c-\mu}{\sigma}\right).$$

因而 $P\{c\le X\le d\}=F(d)-F(c)$

$$=\Phi\left(\frac{d-\mu}{\sigma}\right)-\Phi\left(\frac{c-\mu}{\sigma}\right).$$

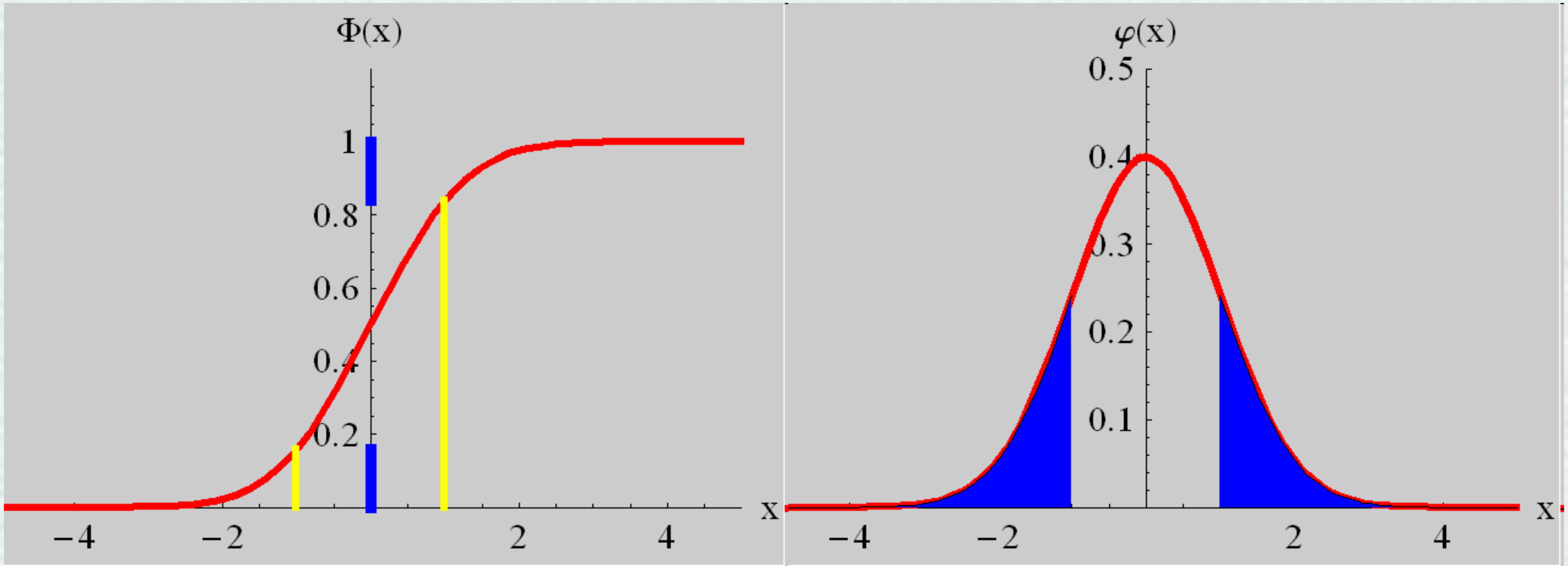
即 $P\{c\le X\le d\}=\Phi\left(\frac{d-\mu}{\sigma}\right)-\Phi\left(\frac{c-\mu}{\sigma}\right).$

例8 证明 $\Phi(-x)=1-\Phi(x).$

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

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

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



例9 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 $d^{\circ}C$, 液体的温度 X(以$^{\circ}C$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 的概率.

(2) 若要求保持液体的温度至少为 $80^{\circ}C$ 的概率不低于 0.99, 问d至少为多少?

解 (1) 所求概率为

$$P\{X < 89\} = \Phi\left(\frac{89-90}{0.5}\right) = \Phi(-2) = 1 - \Phi(2)$$

$$= 1 - 0.9772 = 0.0228.$$

(2) $P\{X > 80\} \geq 0.99$

$$\Rightarrow 1 - P\{X \leq 80\} \geq 0.99$$

$$\Rightarrow 1 - F(80) \geq 0.99$$

$$\Rightarrow 1 - \Phi\left(\frac{80-d}{0.5}\right) \geq 0.99$$

$$\Rightarrow \Phi\left(\frac{80-d}{0.5}\right) \leq 1 - 0.99 = 0.01,$$

即 $\dfrac{80-d}{0.5} \leq -2.327 \Rightarrow d \geq 81.1635.$

三、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

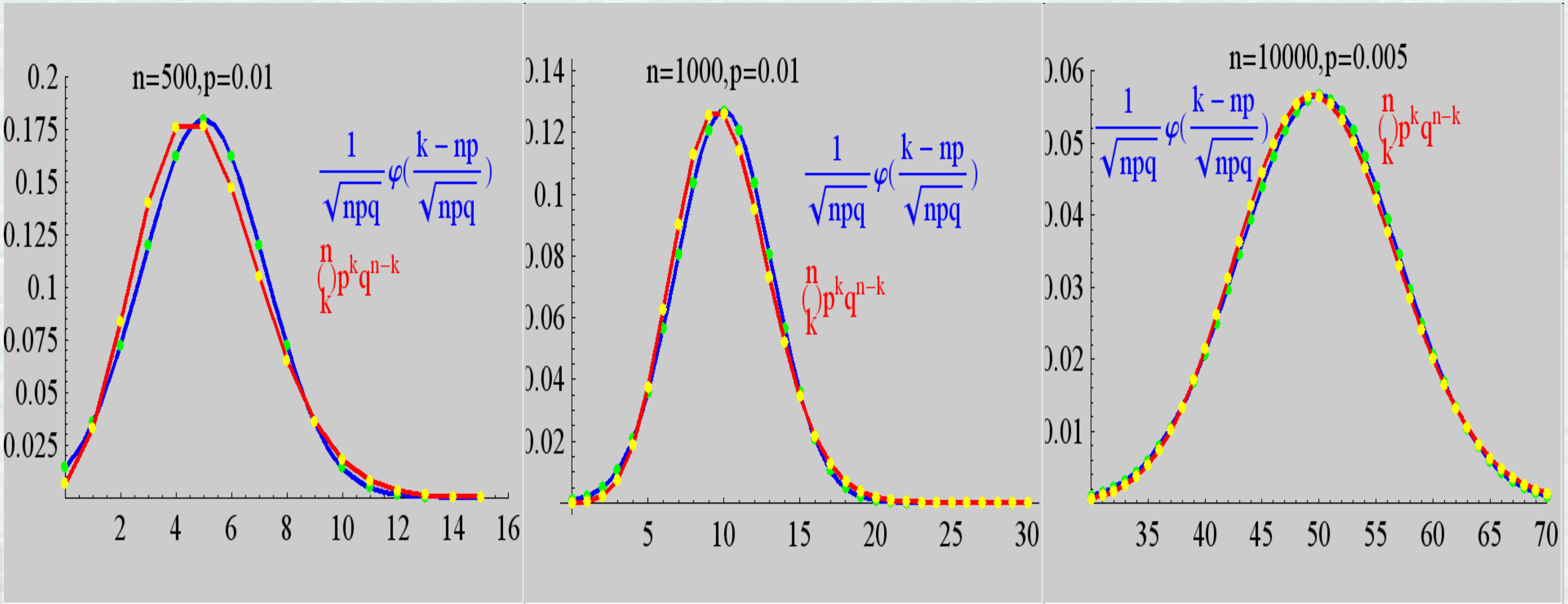
- 均匀分布
- 正态分布(或高斯分布)
- 指数分布

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, 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,炮弹的弹落点的分布等, 都服从或近似服从正态分布.可以说,正态分布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.

另一方面,有些分布(如二项分布、泊松分布)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)