§2.3 连续型随机变量及其分布

一 连续型随机变量及密度函数

定义1 设$F(x)$是随机变量X的分布函数，若存在非负可积函数$f(x)$，使得对任意实数x，有

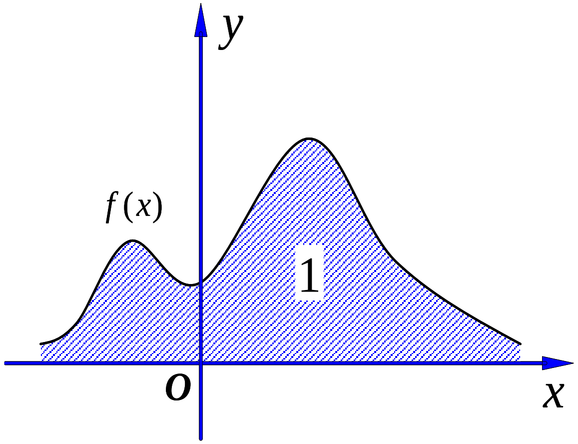
$$F(x)=\int_{-\infty}^{x} f(t)dt, \qquad x\in R.$$

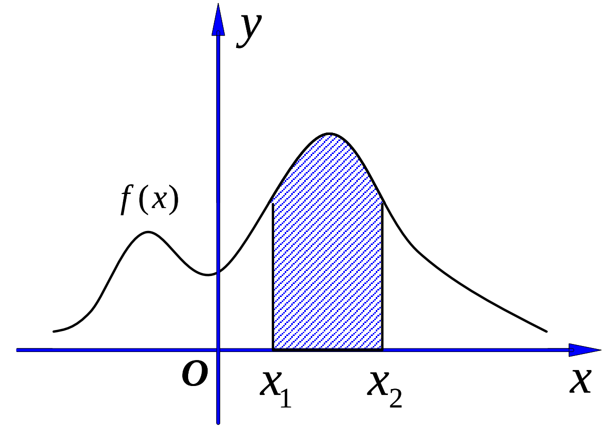
称X为连续型随机变量，称$f(x)$为X的概率密度函数，或密度函数，也称概率密度.

性质1 $f(x)\geq 0$,

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

性质3 若$f(x)$在点x处连续，则$F'(x)=f(x)$.





性质4 $P(x_1 < X \le x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\mathrm{d}x,$

性质5 $P\{X = c\} = F(c) - F(c-0) = 0.$

$$P(x_1 < X \le x_2) = P(x_1 \le X \le x_2) = P(x_1 < X < x_2)$$

$$= P(x_1 \le X < x_2) = \int_{x_1}^{x_2} f(x)\mathrm{d}x$$

显然有 $P\{X \in I\} = \int_I f(x)dx$

其中 I 是某个区间或若干个区间的并.

例1：若随机变量X的概率密度为

$$f(x)=\begin{cases} C(4x-2x^2), & 0<x<2, \\ 0, & \text{其它}. \end{cases}$$

(1)求C的值；　(2)X的分布函数；　(3)$P\{X>1\}$.

解：(1)由于 $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=1$ 有

$$C\int_0^2 (4x-2x^2)\mathrm{d}x=1$$

$$C=3/8.$$

(2)X的分布函数

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

$$=\begin{cases}\int_{-\infty}^{x}0\mathrm{d}x, & x<0,\\ \int_{-\infty}^{0}0\mathrm{d}x+\int_{0}^{x}\frac{3}{8}(4x-2x^2)\mathrm{d}x, & 0\le x<2,\\ \int_{-\infty}^{0}0\mathrm{d}x+\int_{0}^{2}\frac{3}{8}(4x-2x^2)\mathrm{d}x+\int_{2}^{x}0\mathrm{d}x, & x\ge 2.\end{cases}$$

分段讨论

$$=\begin{cases}0, & x<0,\\ \frac{3}{4}x^2-\frac{1}{4}x^3, & 0\le x<2,\\ 1, & x\ge 2.\end{cases}$$

$$P\{X>1\}=\int_{1}^{+\infty}f(x)\mathrm{d}x=\int_{1}^{2}3/8\cdot(4x-2x^2)\mathrm{d}x=1/2$$

或$P\{X>1\}=1-P\{X\le 1\}=1-F(1)=1-1/2=1/2.$

课堂练习

设$X \sim p(x)$，且$p(-x) = p(x)$，$F(x)$是X的分布函数则对任意实数$a > 0$，有(　　)

①$F(-a) = 1 - \int_0^a p(x)dx$　　②$F(-a) = \frac{1}{2} - \int_0^a p(x)dx$

③$F(-a) = F(a)$　　④$F(-a) = 2F(a) - 1$.

二　几种常见的连续型随机变量的分布

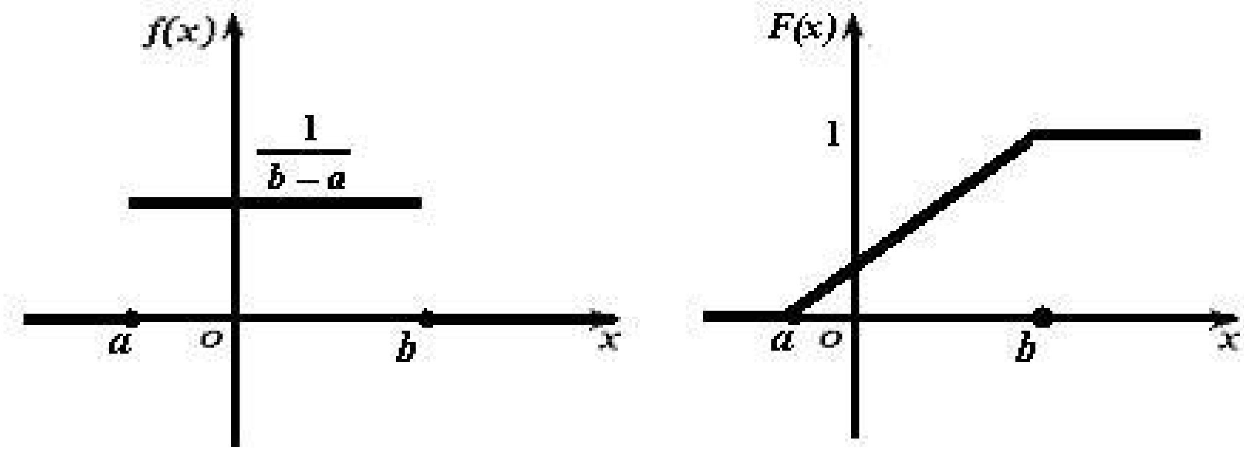
【均匀分布】如果连续型随机变量X具有下列密度函数

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

其中a,b是有限数，则称X是$[a,b]$上的均匀分布.

记作$X\sim U[a,b]$. 其分布函数为

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



特点：若$X \sim U[a,b]$，则X取值落在$U[a,b]$中的某一区域内的概率与这一区域的长度(测度)成正比，而与区域的位置无关.

误差是服从均匀分布的一个例子. 此外, 均匀分布的重要性在于：借助于均匀分布可以生成任意的分布，而在计算机上很容易生成均匀分布.

例2　设某公共汽车站从早上$7:00$开始每隔15分钟到站一辆汽车，即$7:00, 7:15, 7:30, 7:45$等时刻有汽车达到此站．如果一个乘客到达该站的时刻服从$7:00$到$7:30$之间的均匀分布．求他等待时间不超过5分钟的概率.

解：设X表示乘客到达该车站的时间，则

$X \sim U[0,30]$

$$f(x)=\begin{cases} 1/30, & 0 \le x \le 30, \\ 0, & \text{其它}. \end{cases}$$

因此所求概率为

$$P\{10 < X < 15\} + P\{25 < X < 30\}$$

$$= \int_{10}^{15} 1/30 \mathrm{d}x + \int_{25}^{30} 1/30 \mathrm{d}x = 1/3.$$

例3 设ξ在$[-1,5]$上服从均匀分布，求方程

$$x^2+2\xi x+1=0$$

有实根的概率.

解：方程有实数根$\longleftrightarrow 4\xi^2-4\geq 0$

即 $|\xi|\geq 1$

而ξ的密度函数为 $f(x)=\begin{cases}\frac{1}{6} & (-1\leq x\leq 5)\\ 0 & \text{其它}\end{cases}$

故所求概率为

$$P\{|\xi|\geq 1\}=\int_{-\infty}^{-1}f(x)dx+\int_{1}^{+\infty}f(x)dx=\frac{2}{3}$$

【指数分布】 如果连续型随机变量X具有下列密度函数

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$$

分布函数为

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

这里参数$\lambda>0$，这种分布称为指数分布．简记为$X\sim \exp(\lambda)$.

指数分布有重要应用，常用它作为各种寿命分布的近似，例如电子元器件的寿命, 某些动物的寿命，电话的通话时间, 随机服务系统的服务时间等.

性质：无记忆性(类似于几何分布).

设 $X\sim\exp(\lambda)$. 则$\forall s>0, t>0$

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

$$=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}=P\{X\geq t\}$$

若X表示仪器的寿命, 那么上式说明:已知此仪器已使用t小时, 它总共能工作$s+t$小时的概率等于从开始使用时算起，它至少能工作s小时的概率.

练习题

证明:对于取非负值的随机变量，只有指数分布具有上述性质.

例4 设某人到银行取款时的排队时间X(分钟)服从指数分布，其概率密度为

$$f(x)=\begin{cases}\lambda e^{-1/10\cdot x}, & x>0,\\ 0, & x\le 0.\end{cases}$$

1.试确定常数λ;

2.计算排队时间超过10分钟的概率;

3.计算排队时间在10分钟到20分钟的概率.

解：（1）$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=\int_{0}^{+\infty}\lambda e^{-1/10\cdot x}\mathrm{d}x=10\lambda=1$

$$\lambda=1/10.$$

(2) $P\{X>10\}=\int_{10}^{+\infty}1/10\cdot e^{-1/10\cdot x}\mathrm{d}x=e^{-1}$

(3) $P\{10<X<20\}=\int_{10}^{20}1/10\cdot e^{-1/10\cdot x}\mathrm{d}x=e^{-1}-e^{-2}$

例5 设连续型随机变量的分布函数为

$$F(x)=\begin{cases} A+Be^{-2x}, & x\geq 0 \\ 0, & x<0 \end{cases}$$

1. 求常数A, B; 2.求X的概率密度函数.

解：（1） 由分布函数的性质 $F(+\infty)=1$

即 $\lim\limits_{x\to+\infty}(A+Be^{-2x})=1$ 所以 $A=1.$

（2） 又因为$F(x)$在点$x=0$处连续, 所以

$$A+B=F(0)=\lim_{x\to 0^-}F(x)=0\ ,\ B=-1.$$

从而分布函数为 $F(x)=\begin{cases} 1-e^{-2x}, & x\geq 0 \\ 0, & x<0 \end{cases}$

密度函数为 $f(x)=F'(x)=\begin{cases} 2e^{-2x}, & x\geq 0 \\ 0, & x<0 \end{cases}$

【正态分布】如果连续型随机变量X具有下列密度函数

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

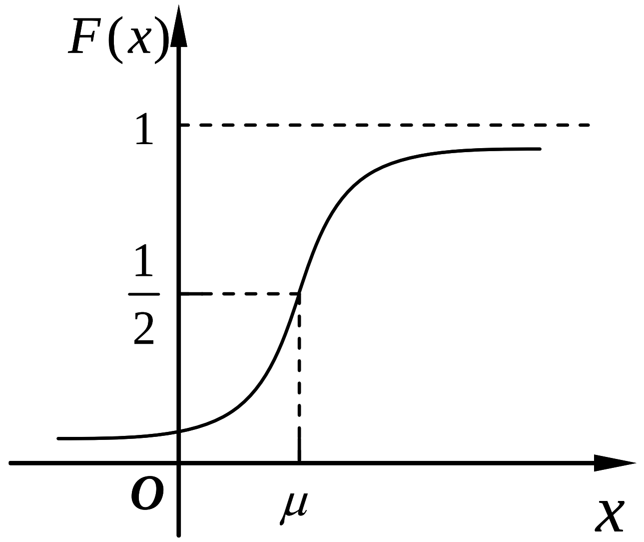
其中σ,μ 均为常数, 且$\sigma>0$，相应的分布函数为

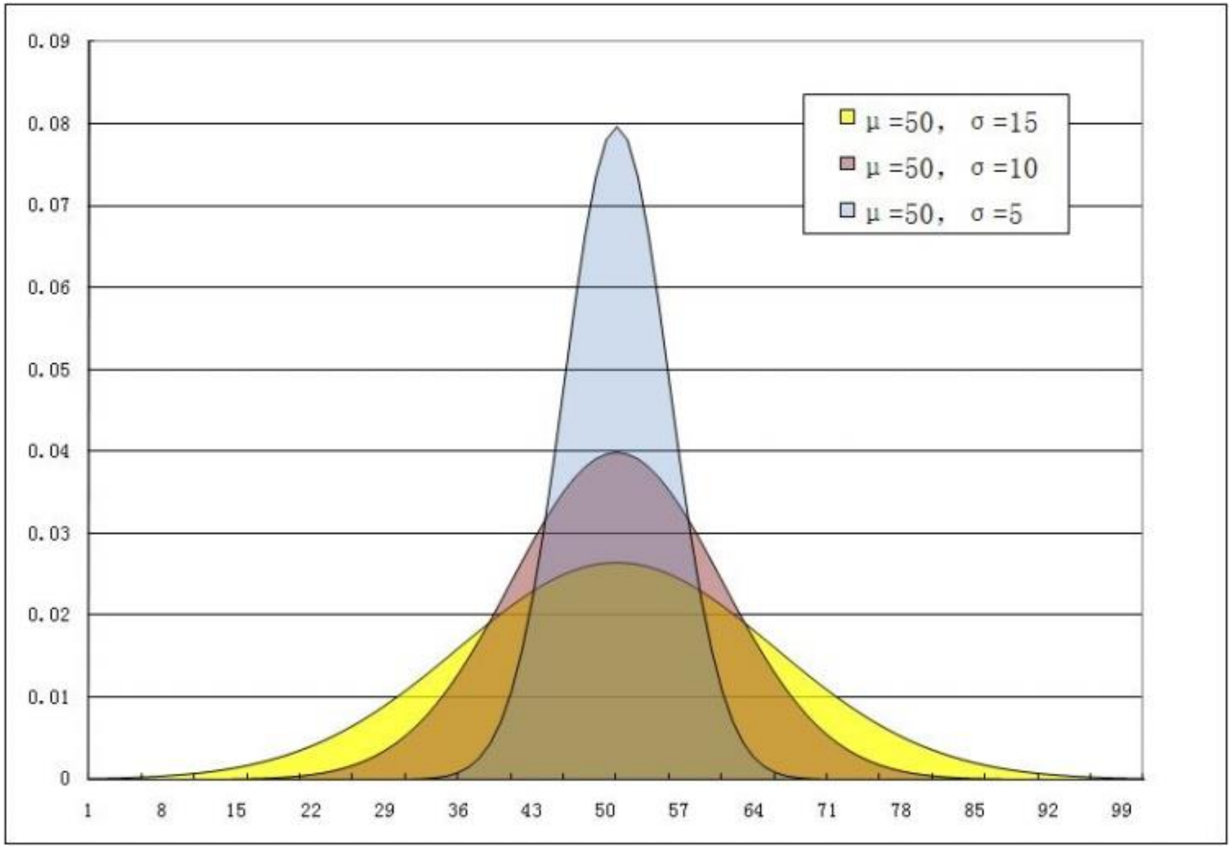
$$F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt,\qquad -\infty<x<\infty$$

这种分布称为正态分布

(normal distribution),

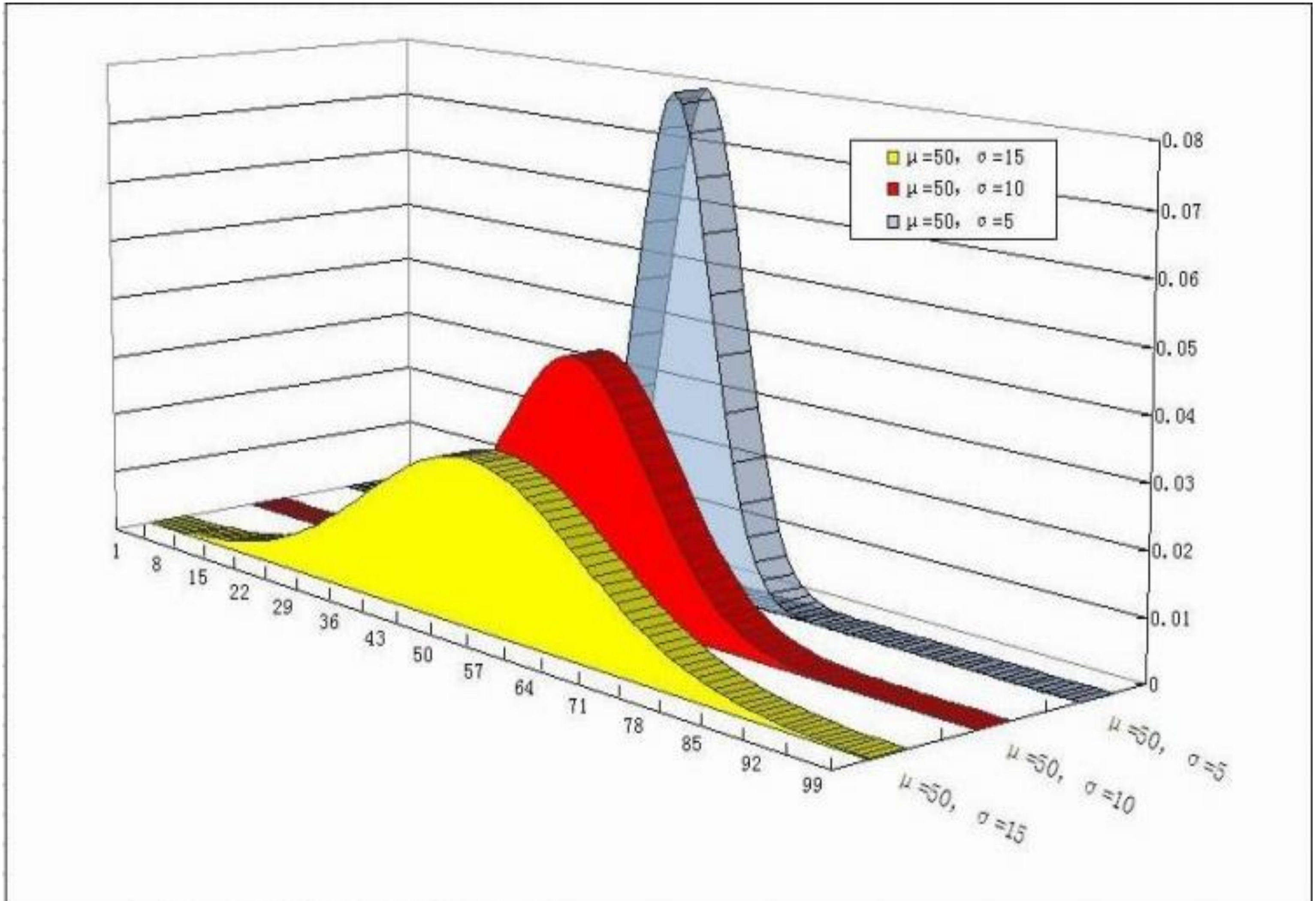
简记为$X\sim N(\mu,\sigma^2)$.





$f(x)$ 的特点：

在$x=\mu$ 处达到极大值，关于$x=\mu$对称；σ 越小, 图形越高耸，分布就越集中于μ 的附近.


μ=50, σ=15
μ=50, σ=10
μ=50, σ=5
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
1
8
15
22
29
36
43
50
57
64
71
78
85
92
99
μ=50, σ=5
μ=50, σ=10
μ=50, σ=15

正态分布是概率论中最重要的分布．一方面，正态分布是最常见的分布，例如测量的误差；炮弹落点的分布, 人的生理特征的数据: 身高、体重等.

一般说来, 如果影响某一数量指标的随机因素很多，而每个因素所起的作用不太大，则这个指标服从正态分布(第五章 — 极限定理), 正态分布具有许多良好的性质, 许多分布可用正态分布来近似(二项分布, 泊松分布等), 此外一些分布(数理统计中的χ^2分布, t分布，F分布) 可由正态分布来导出.

标准正态分布

当$\mu=0,\ \sigma=1$时，称为标准正态分布，记为$X\sim N(0,1)$，相应的分布密度函数及分布函数分别记为$\varphi(x)$及$\Phi(x)$.

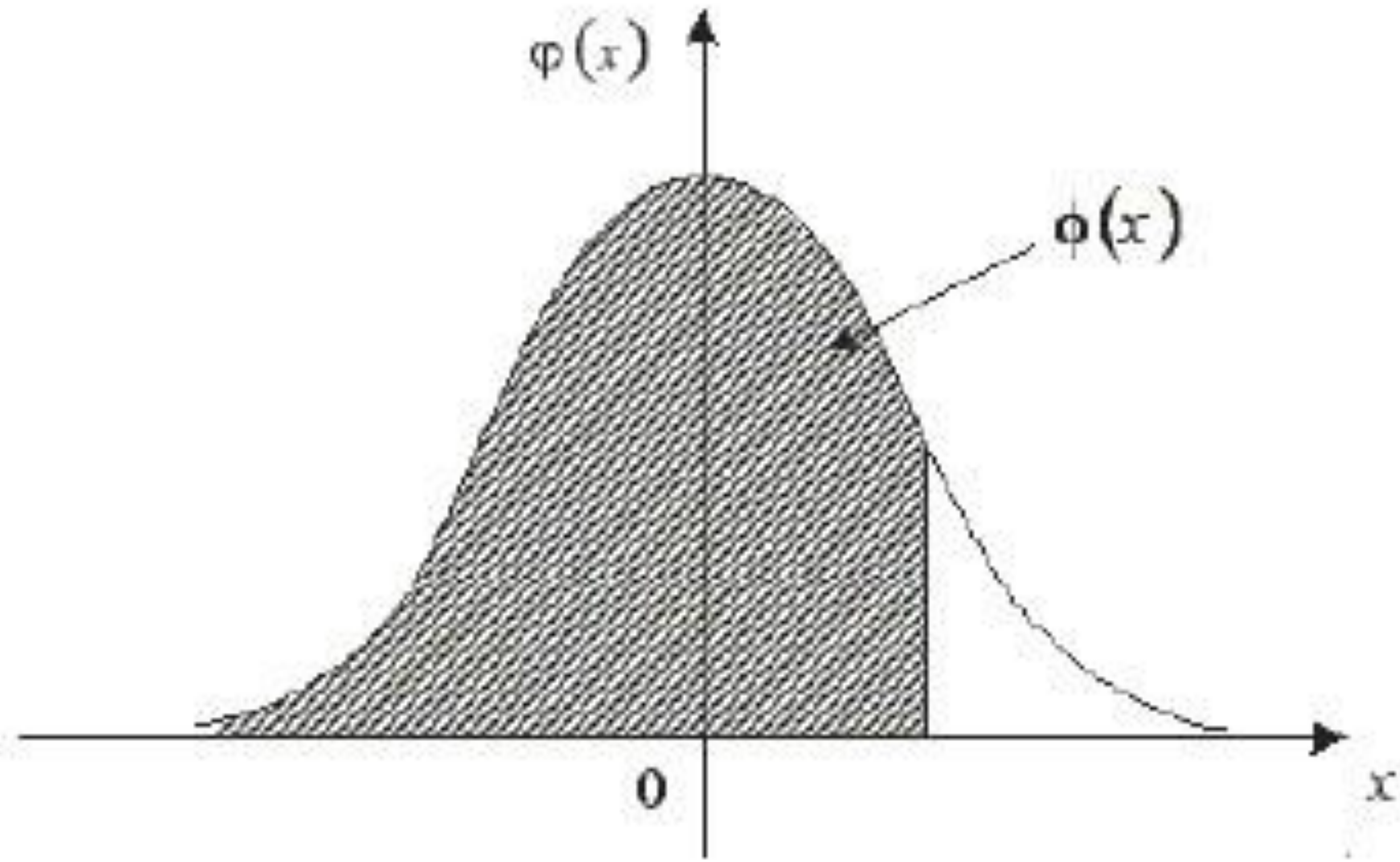
$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad \Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t.$$

由于

$$\int_{-\infty}^{+\infty}\varphi(t)dt=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t=\sqrt{\frac{2}{\pi}}\int_{0}^{+\infty}e^{-\frac{t^2}{2}}\mathrm{d}t$$

$$\overset{t=\sqrt{2u}}{=\!=\!=}\frac{1}{\sqrt{\pi}}\int_{0}^{+\infty}u^{\frac{1}{2}-1}e^{-u}\mathrm{d}u=\frac{\Gamma(1/2)}{\sqrt{\pi}}=1.$$

因而$\varphi(x)$是分布密度函数(见后面Γ函数的知识).

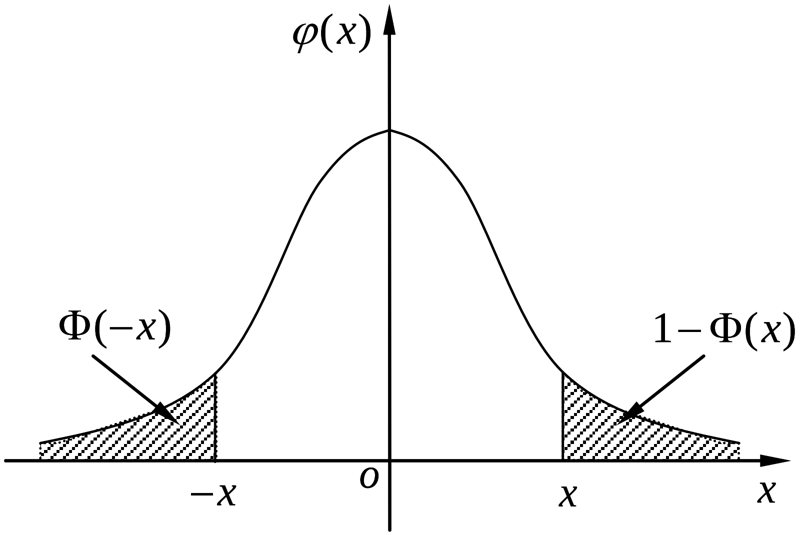

φ(x)
Φ(x)
0
x

正态分布概率的计算

标准正态分布

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

$$=\begin{cases}\text{查表}\Phi(x) & x>0\\ 1-\Phi(-x) & x<0\end{cases}$$



一般正态分布

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

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

因而若 $X\sim N(\mu,\sigma^2)$，则

$$P(a \le X < b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(|X-\mu| < k\sigma) = F(\mu + k\sigma) - F(\mu - k\sigma)$$

$$= \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

查标准正态分布表

特别

$$P(|X-\mu| < \sigma) = 2\Phi(1) - 1 \approx 68.27\%$$

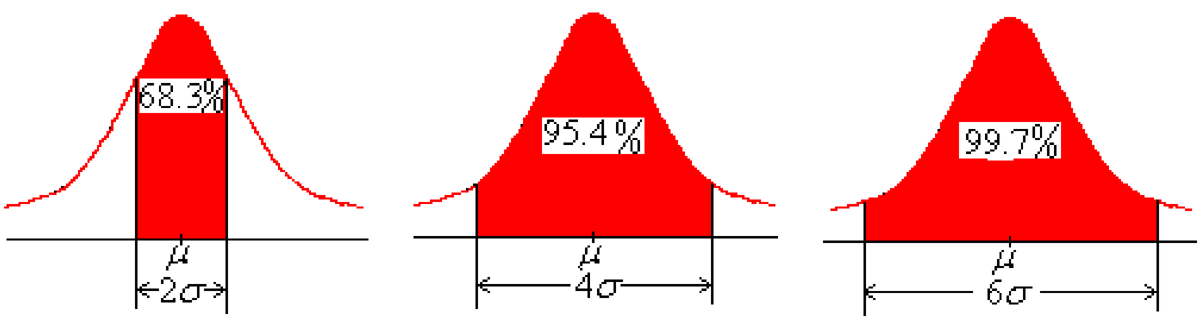
$$P(|X-\mu| < 2\sigma) = 2\Phi(2) - 1 \approx 95.45\%$$

$$P(|X-\mu| < 3\sigma) = 2\Phi(3) - 1 \approx 99.73\%$$

质量管理中的

“3σ”原则.

例：书本$P_{135} - P_{137}$



例6：若$X \sim N(3,3^2)$，试求：

1. $P\{2 < X < 5\}$； 2. $P\{X > 0\}$； 3. $P\{|X-3| > 6\}$.

解：

1. $P\{2 < X < 5\} = F(5) - F(2) = \Phi(\frac{5-3}{3}) - \Phi(\frac{2-3}{3})$

$= \Phi(\frac{2}{3}) - [1 - \Phi(\frac{1}{3})] = 0.7486 - (1 - 0.6293) = 0.3779;$

2. $P\{X > 0\} = 1 - F(0) = 1 - \Phi(\frac{0-3}{3}) = \Phi(1) = 0.8413$

3. $P\{|X-3| > 6\} = P\{X > 9\} + P\{X < -3\}$

$= 1 - P\{X \le 9\} + P\{X < -3\} = 1 - \Phi(2) + \Phi(-2)$

$= 2[1 - \Phi(2)] = 2(1 - 0.9972) = 0.0456$

练习：设$X\sim N(1.5,4)$，试计算$P\{|X|>3\}$.

例7 某零件宽度$X\sim N(0.900,\ 0.003^2)$,现规定限度是0.900 ± 0.005. 求

(1) 零件的废品率.

(2) 若要求每 100 个产品中废品不多于1个，可允许的σ最大值是多少？

$$(1)P\{|X-0.9000|\le 0.0050\}=2\Phi(\frac{0.005}{0.003})-1=90.44\%$$

$$p_{废}=100\%-90.44\%=9.56\%$$

$$(2)p_{废}\le 0.01,\qquad 1-p=P\{|X-0.900|\le 0.005\}$$

$$=2\Phi(\frac{0.005}{\sigma})-1\ge 0.99$$

查表得：$\sigma\le 0.00194$

Γ函数的知识回顾

定义 当 $r>0$ 时，积分$\int_0^{+\infty} x^{r-1}e^{-x}\mathrm{d}x$存在，记作

$$\Gamma(r)=\int_0^{+\infty} x^{r-1}e^{-x}\mathrm{d}x$$

称为(Gamma)Γ函数.

性质1 $\Gamma(r+1)=r\Gamma(r)$， 其中$r>0$.

称为递推公式. 用分部积分证明.

特别地 $\Gamma(n)=(n-1)!$， $\Gamma(1)=1$.

性质2 $\Gamma(1-r)\,\Gamma(r)=\dfrac{\pi}{\sin(\pi r)}$，

称为余元公式. 其中$0<r<1$.

特别地 $\Gamma(1/2)=2\int_0^{+\infty} e^{-t^2}\mathrm{d}t=\sqrt{\pi}$.

性质3　　与贝塔函数的关系

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

其中$\alpha>0,\quad \beta>0,$

$$B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx.$$

其他性质（略）.

斯特灵公式为：

$$n!=\sqrt{2\pi n}(n/e)^n e^{\lambda_n}$$

其中：$\dfrac{1}{12n+1}<\lambda_n<\dfrac{1}{12n}.$