西安電子科技大學
XIDIAN UNIVERSITY

概率论与数理统计



概率论与数理统计课程组

2.3 连续型随机变量的分布

离散型随机变量可能的取值是有限个或可列无限多个，它的概率分布可以用分布律来刻画，如果随机变量可能的取值充满某个区间，那么它的概率分布需要如何来刻画呢?

1、随机变量的分布函数

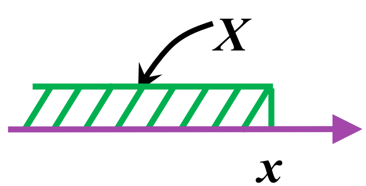
定义 设 X 是一个随机变量，称函数

$$F(x) = P(X \le x), \quad x \in \mathrm{R}$$

为随机变量 X 的分布函数。

对于随机变量$F(x)$的几何意义

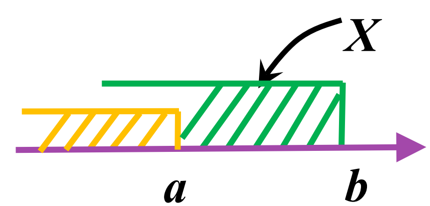
$F(x)=P(X\le x)$, $-\infty<x<+\infty$



随机变量取值落在小于等于x一侧的概率

对于任意实数a, b($a<b$)，有

$P(a<X\le b)$ $= P(X\le b) - P(X\le a)$ $= F(b) - F(a)$



注意开闭区间！

已知X分布函数，就可知X落在任意区间 $(a, b]$ 的概率

性质

1° 单调不减函数：对于任意实数$x_1<x_2$，有 $F(x_2)-F(x_1)=P(x_1<X\le x_2)\ge 0$

2° $0\le F(x)\le 1$且 $\begin{cases} F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0 \\ F(\infty)=\lim\limits_{x\to+\infty}F(x)=1 \end{cases}$

3° $F(x+0)=F(x)$且即$F(x)$为右连续函数

假想将分布函数定义改为$G(x)=P(X<x)$，则为左连续

$P\{X\le x\}$关于x右连续

$P\{X<x\}$关于x左连续

满足其上三点的$F(x)$必为某随机变量的分布函数

性质1-3是鉴别一个函数是否是某个随机变量的分布函数的充分必要条件。

离散型随机变量的分布函数

$$P(X=x_k)=p_k\ (k=1,2,3,\ldots)$$

$$F(x)=P(X\le x)=\sum_{x_k\le x}P(X=x_k)=\sum_{x_k\le x}p_k$$

$$p_i=P(X=x_i)=F(x_i+0)-F(x_i-0)=F(x_i)-F(x_i-0)\ (i=1,2,3,\ldots)$$

二、连续型随机变量的概率密度

1. 连续型随机变量概率密度的定义

定义 设 X 是随机变量，其分布函数为 $F(x)$，如果存在**非负可积函数** $f(x)$，

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

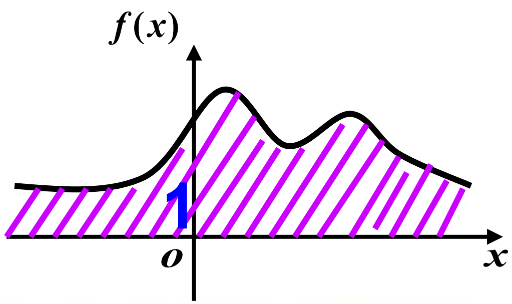
则称 X 为连续型随机变量，称F(x)为连续性分布函数，称 $f(x)$为 X 的**概率密度函数**，简称**概率密度**或**密度函数**。

连续型随机变量的性质

性质1 $f(x)\geq 0, \quad -\infty < x < +\infty$

由定义1直接得证

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



性质3 $P\{a < X \le b\} == \int_a^b f(x)\,\mathrm{d}x$

证明 $P\{a < X \le b\} = P(x \le b) - P(x \le a)$

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

同时得以下计算公式

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

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

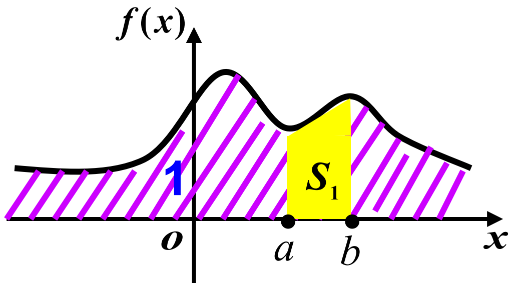
$$= \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x - \int_{-\infty}^{a} f(x)\,\mathrm{d}x$$

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

$$= \int_a^{\infty} f(x)\,\mathrm{d}x.$$

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

$$S_1 = \int_a^b f(x)\,\mathrm{d}x$$



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

证明 在 $f(x)$ 的连续点 x 处,由于 $F(x) = \int_{-\infty}^{x} f(t)dt$,

根据高等数学中所学习的变上限积分求导法则可得 $F'(x) = f(x)$.

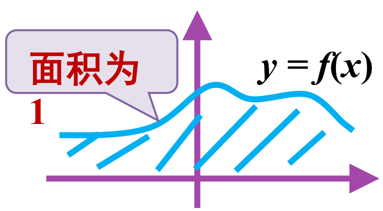
需要指出的是,

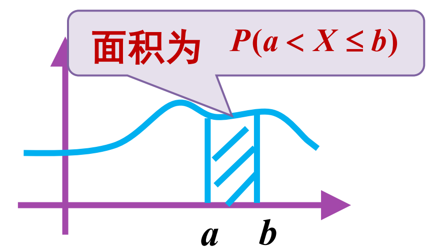
如果函数 $f(x)$ 满足上述性质 1 与性质 2，那么 $f(x)$ 一定是某个连续型随机变量的概率密度

定理 1 设连续型随机变量 X 的概率密度为 $f(x)$，分布函数为 $F(x)$，则 $F(x)$在$(-\infty,\ +\infty)$上连续。

证 由微积分学基本定理即可得证。

概率意义





概率密度与质量、电量的线密度定义类似 $P(a < X \le a + \Delta x) \approx f(x)\,\Delta x$

X落在小区间$[a, a+\Delta x]$的概率近似等于$f(x)\,\Delta x$

X落在一段区间的概率可知，那么取某一实数值的概率是？即求$P(X=a)$

定理 2 设 X 为连续型随机变量，则随机变量 X 取任一实数值 a 的概率均为零，即P (X = a) =0

证 设 X 的分布函数为 $F(x)$，$\Delta x>0$，则由$\{X=a\}\subset\{a-\Delta x<X\leq a\}$得

$$0 \leq P\{X=a\} \leq P\{a-\Delta x<X \leq a\}=F(a)-F(a-\Delta x)$$

在上述不等式中令$\Delta x\to 0$，由定理 1 得

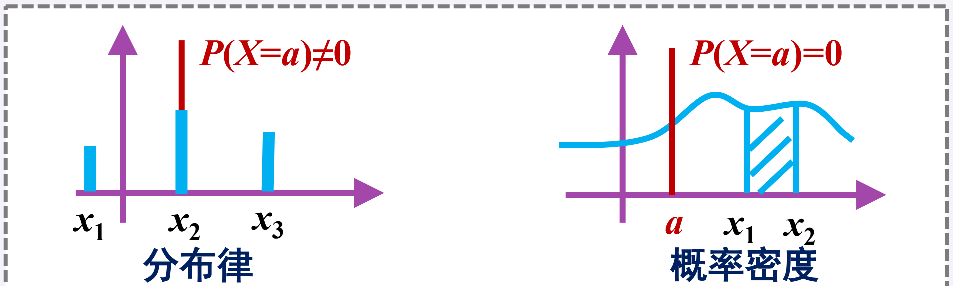
$$P\{X=a\}=0.$$

由定理2知，在计算连续型随机变量落在某一区间的概率时，可以不必区分该区间是开区间或闭区间或半闭区间，即连续积分可忽略端点处.从而

$$P\{a<X \leq b\}=P\{a \leq X<b\}=P\{a<X<b\}=P\{a \leq X \leq b\}=\int_a^b f(x) d x$$

注意

1° 连续型随机变量必有处处$P(X=a)=0$；
离散型随机变量可以有$P(X=a)\neq 0$



若 A 为离散型随机变量，

A是不可能事件$\Leftrightarrow P(A)=0$

2° **概率为0不一定是不可能事件**，例如连续型随机变量$P(X=a)=0$；但是，
若A为不可能事件，必有$P(A)=0$

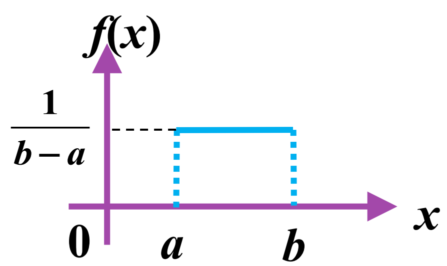
3° 由2°知，**概率为1不一定是必然事件**，例如连续型随机变量$P(X\neq a)=1$；
但是，若A为必然事件，必有$P(A)=1$

3、几个重要的连续性随机变量的分布

1. 均匀分布

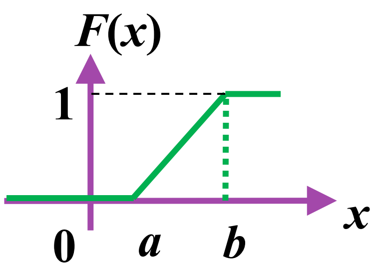
概率密度

X具有概率密度 $f(x)=\begin{cases}\dfrac{1}{b-a} & x\in(a,b)\\ 0 & \text{其他}\end{cases}$



分布函数

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



称X在区间(a, b)上服从均匀分布，

记为$X\sim U(a, b)$

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

性质： $\forall(c,d)\subset(a,b), P(c<X<d)=\dfrac{d-c}{b-a}$

证明： $P(c<X<d)=\int_c^d f(x)dx=\int_c^d \dfrac{1}{b-a}dx=\dfrac{d-c}{b-a}$

X落在(a, b) 区间中任意等长度子区间的概率相同

2. 指数分布

若连续型随机变量 X 的概率密度为

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

其中$\lambda>0$ 为常数，则称 X 服从参数为 λ的**指数分布**，记为$X\sim E(\lambda)$。

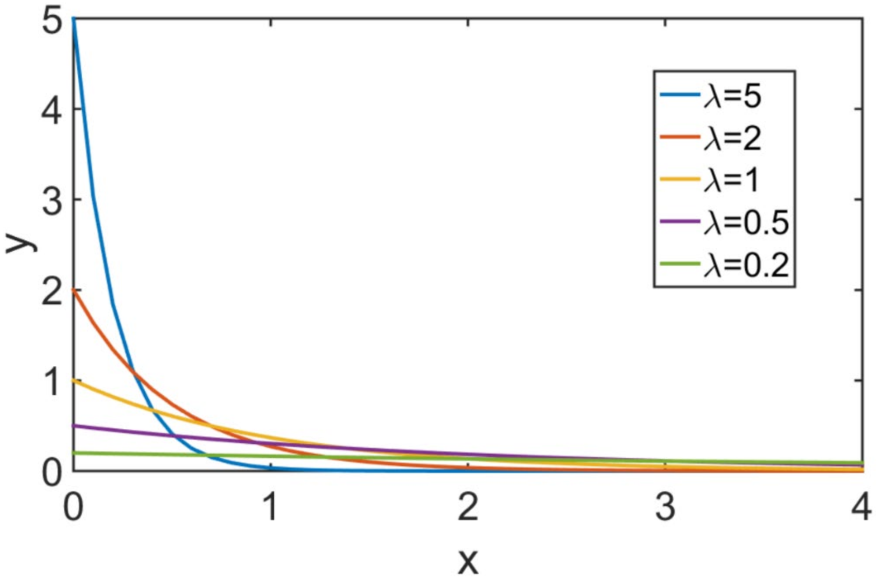
显然，$f(x)\ge 0$，且 $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=\int_{0}^{+\infty}\lambda e^{-\lambda x}\mathrm{d}x=\left[-e^{-\lambda x}\right]_{0}^{+\infty}=1$

X的分布函数为 $F(x)=\begin{cases}1-e^{-\lambda x}, x>0\\ 0, \quad x\le 0\end{cases}$

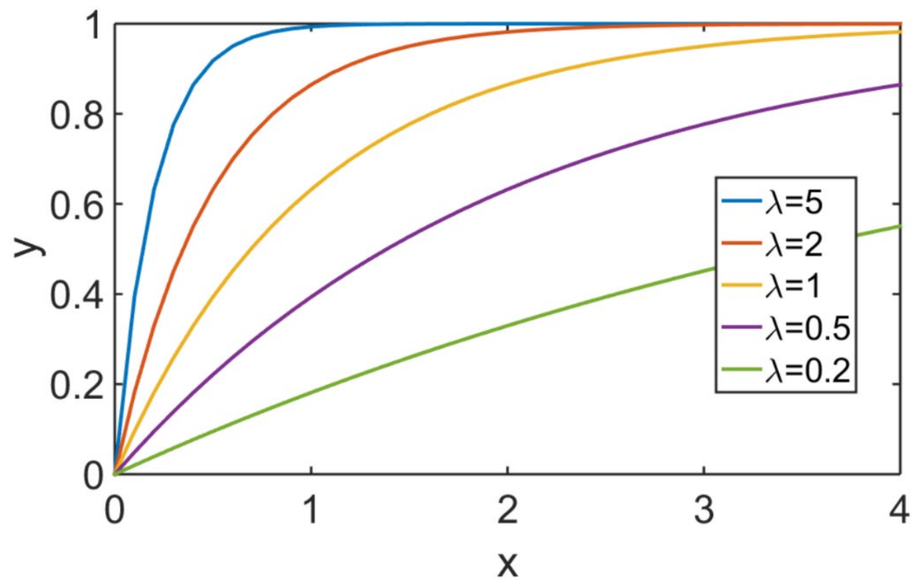
指数分布有着重要的应用，**常被用来描述寿命类随机变量的分布**。如电子元件的寿命、生物的寿命、电话的通话时间、随机服务系统的服务时间等都可以认为服从指数分布。

XIDIAN UNIVERSITY

不同参数设置指数分布的概率密度



不同参数设置指数分布的分布函数



定理 指数分布具有无记忆性，即设 $X \sim E(\lambda)$，则 $\forall s, t > 0$，

$$P(X > s + t \mid X > s) = P(X > t)$$

证

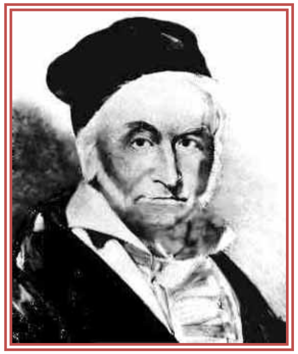
$$P(X > s + t | X > s) = \frac{P(X > s + t, X > s)}{P(X > s)}$$

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

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

3. 正态分布

高斯
Carl Friedrich Gauss



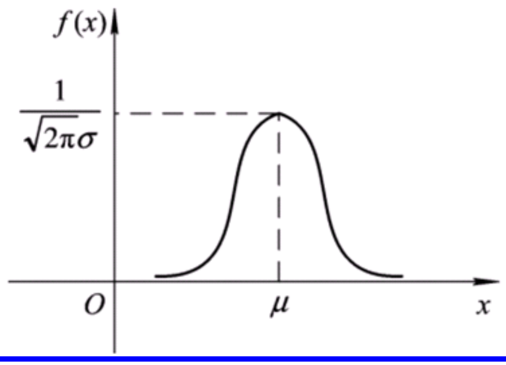
若连续型随机变量 X 的概率密度为

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

其中 μ 、 $\sigma\ (\sigma>0)$ 为常数，则称 X 服从参数为 μ 、 σ^2 的**正态分布**或**Gauss (高斯)分布**，记为 $X\sim N(\mu,\ \sigma^2)$。

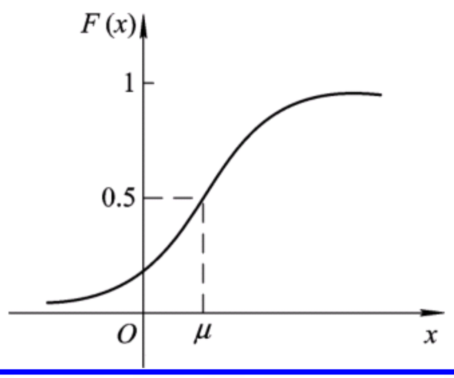
正态随机变量 X 的概率密度为：

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



正态随机变量 X 的分布函数为：

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



显然，$F(\mu)=\frac{1}{2}$。

$f(x)$的图像如图 2—5 所示，它关于 $x=\mu$ 对称，在 $x=\mu$ 处取得最大值 $f(u)=\dfrac{1}{\sqrt{2\pi}\sigma}$

若改变 μ (固定 σ)值，它将沿 x 轴平移，但其形状不变，如图 2 — 6 所示，**称 μ 为位置参数**；

若改变 σ (固定 μ)值，它的扁尖程度将改变，当 σ 越大时图形变得越扁，**当 σ 越小时图形变得越尖**，如图 2—7 所示，当 σ 变小时 X 落在 μ 附近的概率越大。

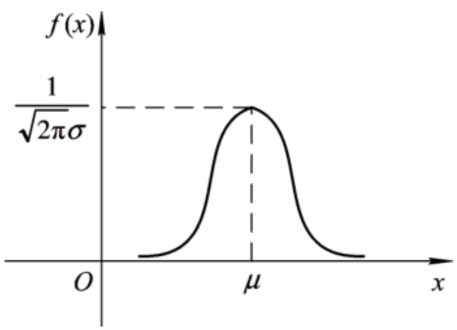


图 2-5

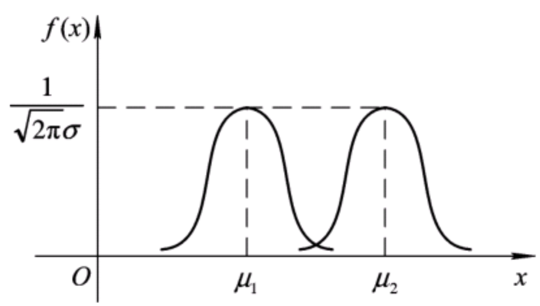


图 2-6

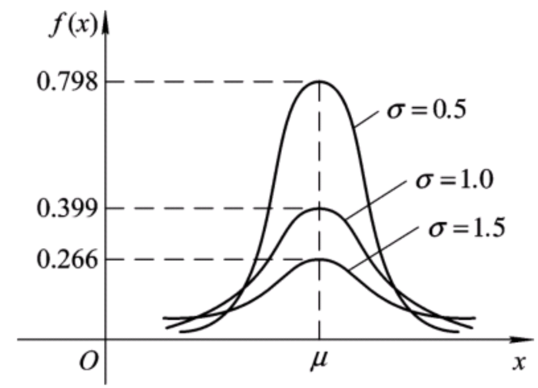
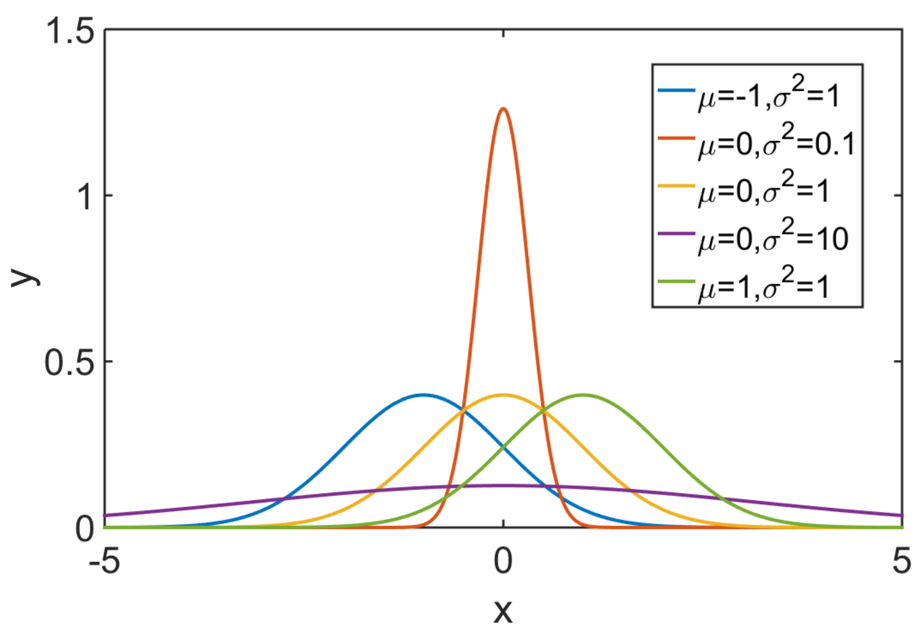
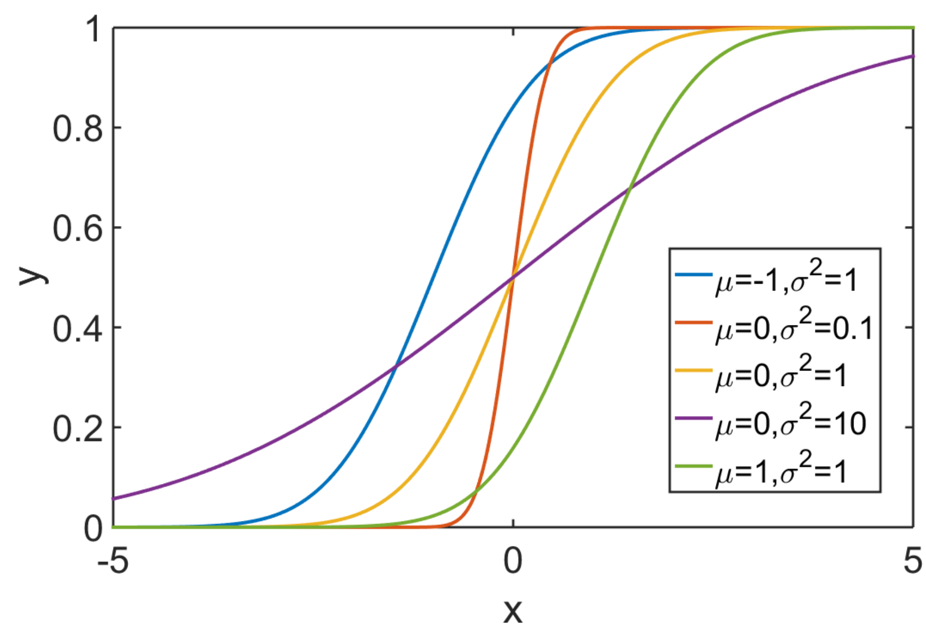


图2-7

不同参数设置正态分布的概率密度



不同参数设置正态分布的分布函数



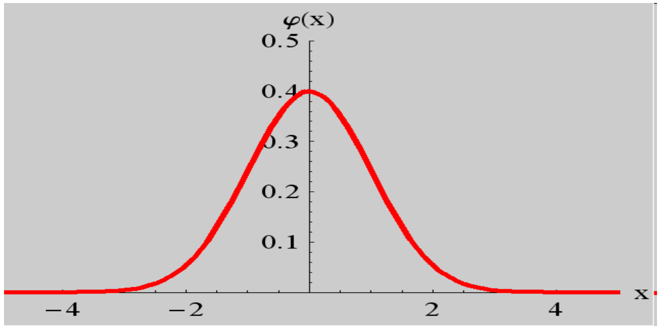
定义 设$X \sim N(\mu, \sigma^2)$，若$\mu=0$，$\sigma^2=1$，则称X服从**标准正态分布**，记为$X \sim N(0, 1)$。

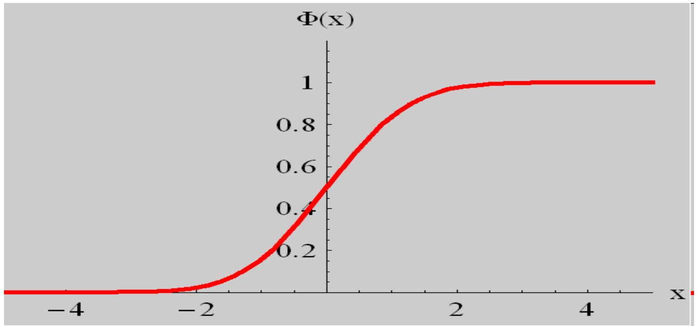
标准正态分布的概率密度$\varphi(x)$和分布函数$\Phi(x)$分别为：

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

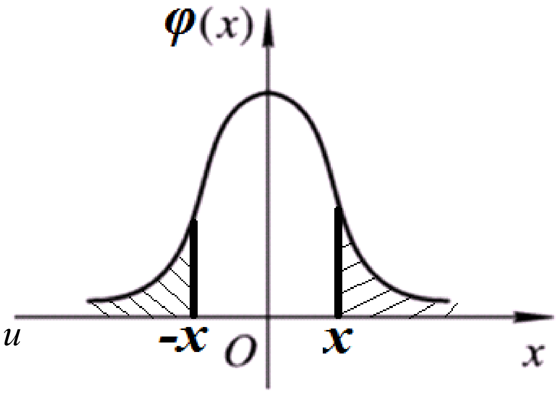
$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,\mathrm{d}t,\quad -\infty<x<+\infty.$$

$\varphi(x)$和$\Phi(x)$的图像





定理 设$X\sim N(0,1)$，则$\Phi(-x)=1-\Phi(x)$.



证

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

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

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

由于标准正态分布在工程技术中有着重要的应用，因此人们为了使用方便，通过计算$\Phi(x)$的值，编制了$\Phi(x)$的函数表(见附表 1)，供实践中查用。这样，就解决了标准正态分布的问题，但在实际工作中，人们也会经常遇到一般的正态分布$X\sim N(\mu,\sigma^2)$，要解决一般正态分布的问题，自然需要知道其分布函数，那么如何得到$F(x)$的值呢？这时可以运用如下定理。

XIDIAN UNIVERSITY

标准正态分布$N(0, 1)$的 x与$\Phi(x)$ 可以查表可知，那么其他正态分布$N(\mu, \sigma^2)$的$F(x)$呢？

定理 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

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

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mu+\sigma x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt \overset{令 v=\frac{t-\mu}{\sigma}}{=\!=} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-v^2/2}dv =\Phi(x)$$

$$P\left(\frac{X-\mu}{\sigma}\le x\right)=\Phi(x) \qquad Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

定理 若$X\sim N(\mu,\sigma^2)$，则 $F(x)=P(X\leq x)=P\left(\frac{X-\mu}{\sigma}\leq\frac{x-\mu}{\sigma}\right)=\Phi\left(\frac{x-\mu}{\sigma}\right)$

对任意区间$(x_1,x_2]$有 $P(x_1<X\leq x_2)=P\left(\frac{x_1-\mu}{\sigma}<\frac{X-\mu}{\sigma}\leq\frac{x_2-\mu}{\sigma}\right)=\Phi\left(\frac{x_2-\mu}{\sigma}\right)-\Phi\left(\frac{x_1-\mu}{\sigma}\right)$

证 由定理5，得

$$F(x)=P(X\leq x)=P\left(\frac{X-\mu}{\sigma}\leq\frac{x-\mu}{\sigma}\right)=P\left(Z\leq\frac{x-\mu}{\sigma}\right)=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

于是由分布函数的性质可得：

$$P\{x_1<X\leq x_2\}=F(x_2)-F(x_1)$$
$$=\Phi\left(\frac{x_2-\mu}{\sigma}\right)-\Phi\left(\frac{x_1-\mu}{\sigma}\right)$$

例 若$X \sim N(\mu, \sigma^2)$，则

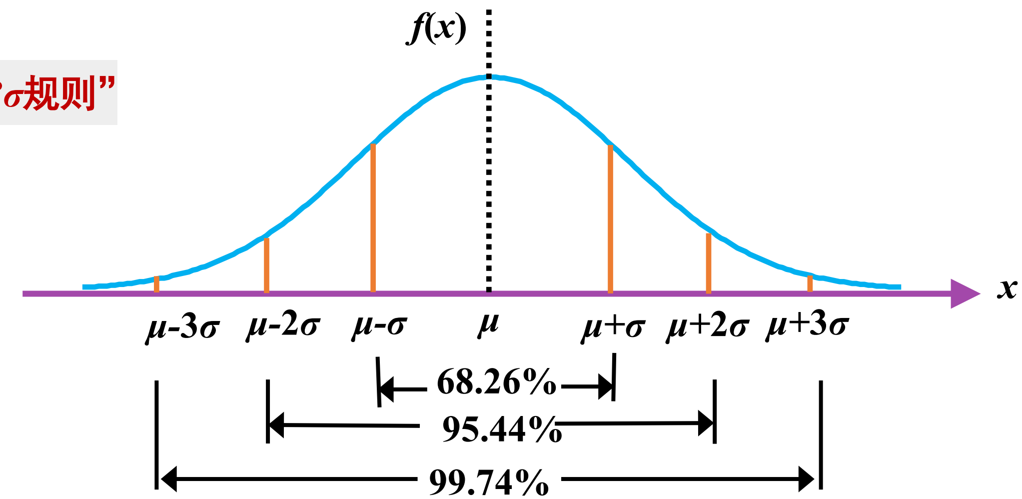
$$P(\mu-\sigma < X < \mu+\sigma) = \Phi(1)-\Phi(-1) = 2\Phi(1)-1 = 68.26\%$$

$$P(\mu-2\sigma < X < \mu+2\sigma) = \Phi(2)-\Phi(-2) = 95.44\%$$

$$P(\mu-3\sigma < X < \mu+3\sigma) = \Phi(3)-\Phi(-3) = 99.74\%$$

在$(-\infty, \infty)$ 落在$(\mu-3\sigma, \mu+3\sigma)$区间几乎是必然的

“3σ规则”



例1：工厂生产的电子管寿命X（小时）$\sim N(1600, \sigma^2)$，若要求寿命在1200小时以外的概率不小于0.96，求 σ。

$$P(X > 1200) \geq 0.96$$

$$1 - P(X \leq 1200) \geq 0.96$$

$$1 - \Phi(\frac{1200-1600}{\sigma}) = 1 - \left[1 - \Phi(\frac{400}{\sigma})\right] \geq 0.96$$

$$\Rightarrow \quad \Phi(\frac{400}{\sigma}) \geq 0.96 \approx \Phi(1.76)$$

由于 $\Phi(x)$**是单调递增的函数，故** $\frac{400}{\sigma} \geq 1.76$

$$\sigma \leq \frac{400}{1.76} = 227.27$$

例2：测量距离产生的随机误差X(cm)满足正态分布

$f(x)=\frac{1}{40\sqrt{2\pi}}e^{-\frac{(x-20)^2}{3200}}$， 求三次测量中至少有一次误差不超过30m的概率。

n=3重Bernoulli试验，每一次成功的概率 $p=P(|X|\le 30)$

设成功的次数为Y $(Y\le 3)$,求$P(Y\ge 1)$ $Y\sim B(n,p)$

$$P(|X|\le 30)=P(-30\le X\le 30)=F(30)-F(-30)$$

$$=\Phi(\frac{30-20}{40})-\Phi(\frac{-30-20}{40})$$

$$=\Phi(0.25)-\Phi(1.25)-1=0.4931$$

$$P(Y\ge 1)=1-P(Y=0)=1-C_3^0 0.4931^0(1-0.4931)^3=0.87$$

例3：某电子元件的寿命X~$f(x)=\frac{1000}{x^2}$, $x\in(1000,\infty)$， 任取5个元件，求恰好有两个元件寿命不超过1500h的概率。

$$P(X\le 1500)=\int_{-\infty}^{1500} f(x)dx=\int_{-\infty}^{1500}\frac{1000}{x^2}dx=\frac{1}{3}$$

$$P(Y=2)=C_5^2\left(\frac{1}{3}\right)^2\left(1-\frac{1}{3}\right)^3=\frac{80}{243}$$