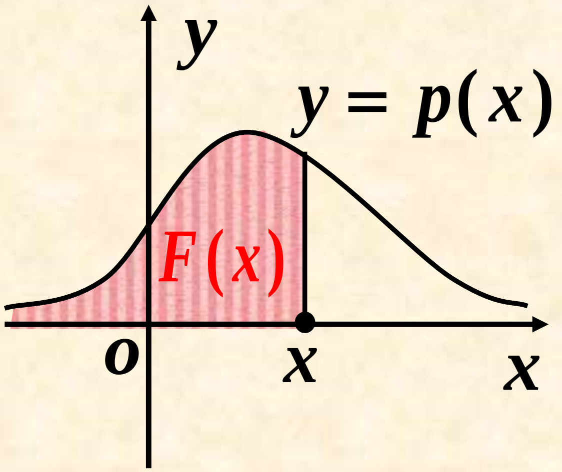
第一节 连续型随机变量及其分布密度 (3)

- 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、内容小结

一、概率密度的概念与性质

1.定义

对于随机变量X，若存在非负可积函数$p(x)$ ($x\in R$), 使得X 的分布函数



$$F(x)=\int_{-\infty}^{x} p(t)\,dt$$

则称X为连续型随机变量, 且称

p(x)为X的密度函数，或概率密度.

2. 密度函数的性质

设X为连续型随机变量 $p(x)$为X的密度函数，$F(x)$为X的分布函数，则

(1) (非负性) $p(x) \geq 0 \quad x \in R$;

(2) (规范性) $\int_{-\infty}^{+\infty} p(x)dx = 1$;

(3) $P\{a < X \leq b\} = F(b) - F(a) = \int_a^b p(x)dx$;

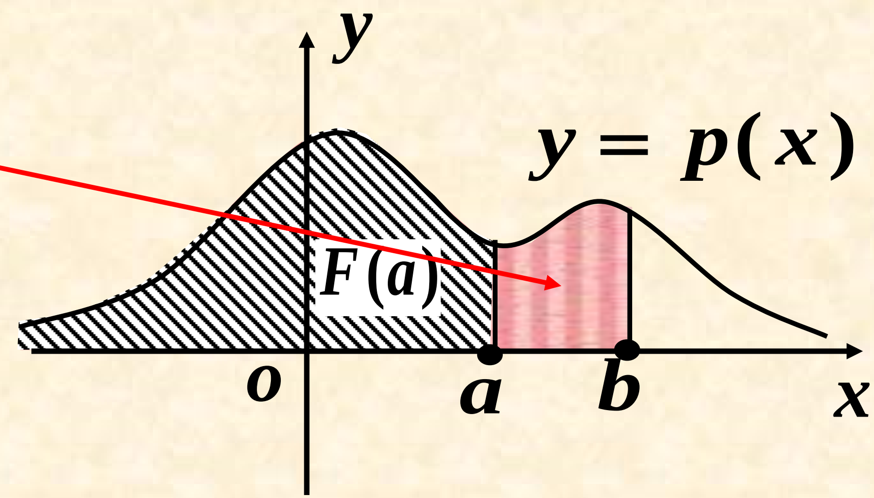
(4) $P\{X = c\} = 0$;

(5) 分布函数F(x)在$(-\infty, +\infty)$上连续;

(6) 若p(x)在点x处连续，则 $F'(x) = p(x)$.

证(3)

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



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

$$= \int_{a}^{b} p(x)\,\mathrm{d}x.$$

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

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

(4)　对于任意可能值c ,连续型随机变量取 c 的概率等于零.即

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

证(4)　$\because \quad \{X=c\}\subseteq\{c-\varepsilon<X\le c\},\ \varepsilon>0$

而　$0\le P\{X=c\}\le P\{c-\varepsilon<X\le c\}$

$$\lim_{\varepsilon\to 0^+} P\{c-\varepsilon<X\le c\}$$

$$=\lim_{\varepsilon\to 0^+}\int_{c-\varepsilon}^{c} p(x)\,\mathrm{d}x=0.$$

$\therefore \quad P\{X=c\}=0.$

注. 1 若X为连续型随机变量，则

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

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

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

连续型随机变量的概率与区间的开闭无关

2° $P(A) = 0 \; \substack{\not\longrightarrow \\ \longleftarrow} \; A = \varnothing$

$P(A) = 1 \; \substack{\not\longrightarrow \\ \longleftarrow} \; A = \Omega$

例如：投针实验

例1 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,

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

故有 $F(-a-0)=F(-a+0)$

$=F(-a)$,

$F(a-0)=F(a+0)$

$=F(a)$,

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

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

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

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

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

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

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

(3) 随机变量 X 的概率密度为

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

二、常见连续型随机变量的分布

分布名称	记号	分布密度
均匀分布	$X \sim U[a,b]$	$p(x)=\begin{cases}\dfrac{1}{b-a}, & x\in[a,b]\\ 0, & x\notin[a,b]\end{cases}$
正态分布	$X \sim N(\mu,\sigma^2)$	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $(x\in R,\sigma>0)$

分布名称	记号	分布密度
指数分布	$X \sim E(\lambda)$	$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x<0\end{cases}$ (常数 $\lambda>0$)

1. 均匀分布

(1) 定义 设连续型随机变量 X 具有概率密度

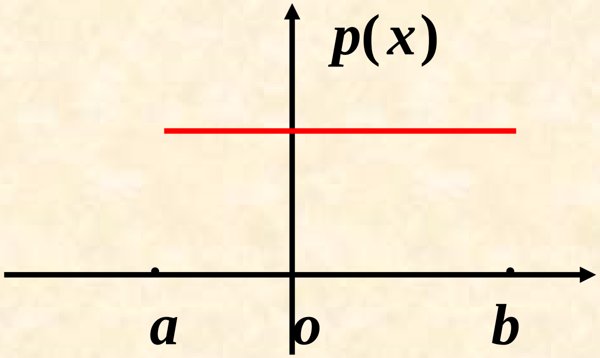
$$p(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\ 0, & \text{其它},\end{cases}$$

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



(2) 均匀分布的性质

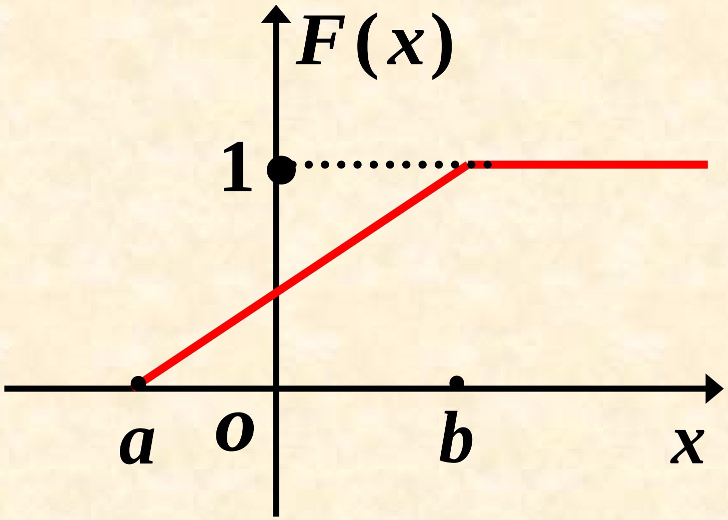
若 $X \sim U[a,b]$，则

① $P\{X < a\} = P\{X > b\} = 0$

② 当 $a \le c < d \le b$ 时, 有 $P\{c \le X < d\} = \dfrac{d-c}{b-a}$.

分布函数

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

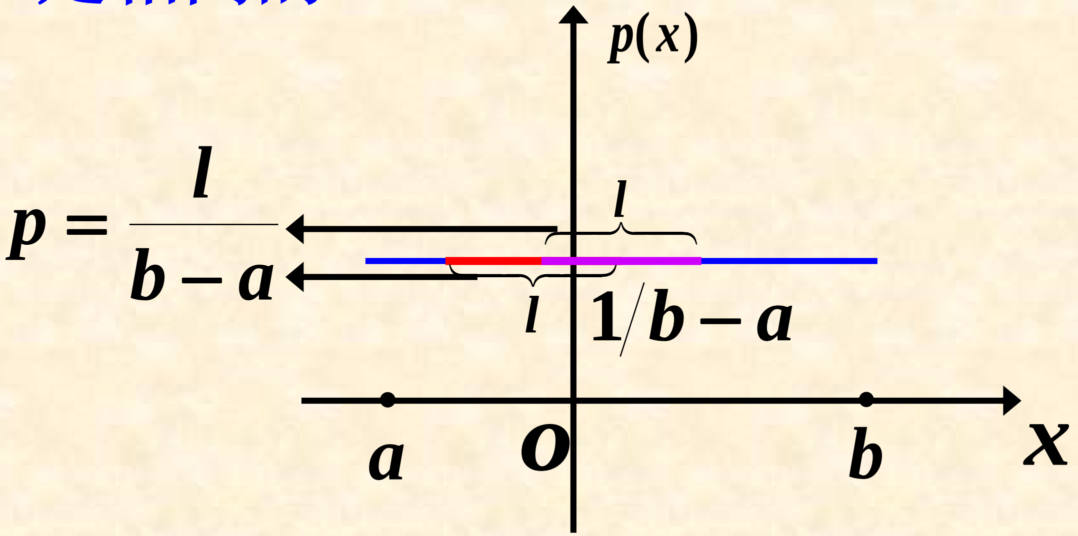


(3) 均匀分布的意思

背景：几何概型

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.



例2 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

$$p(x)=\begin{cases}\frac{1}{3}, & 2\le x\le 5,\\ 0, & \text{其它}.\end{cases}$$

设 A 表示 "对 X 的观测值大于 3",

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表示对 **X进行3次独立观测中，观测值大于3的次数,**

则 **$Y \sim B(3, \frac{2}{3})$.**

因而有

$$P\{Y\geq 2\}=C_3^2\left(\frac{2}{3}\right)^2\left(1-\frac{2}{3}\right)+C_3^3\left(\frac{2}{3}\right)^3\left(1-\frac{2}{3}\right)^0=\frac{20}{27}.$$

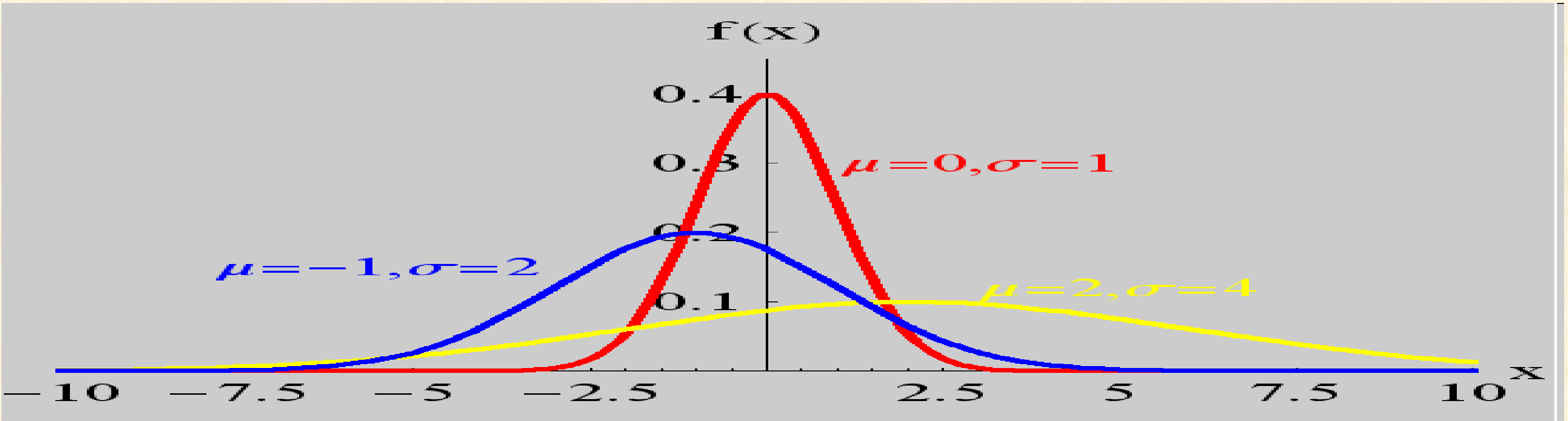
2. 正态分布(或高斯分布)

高斯资料

(1) 定义 设连续型随机变量X的概率密度为

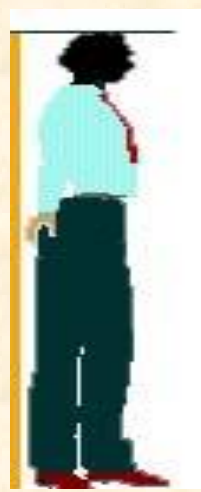
$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ -\infty<x<\infty,$$

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.



目录 上页 下页 返回 结束

(3) 正态分布下的概率计算

原函数不是初等函数

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

$$= ?$$

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

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 $N(0,1)$.

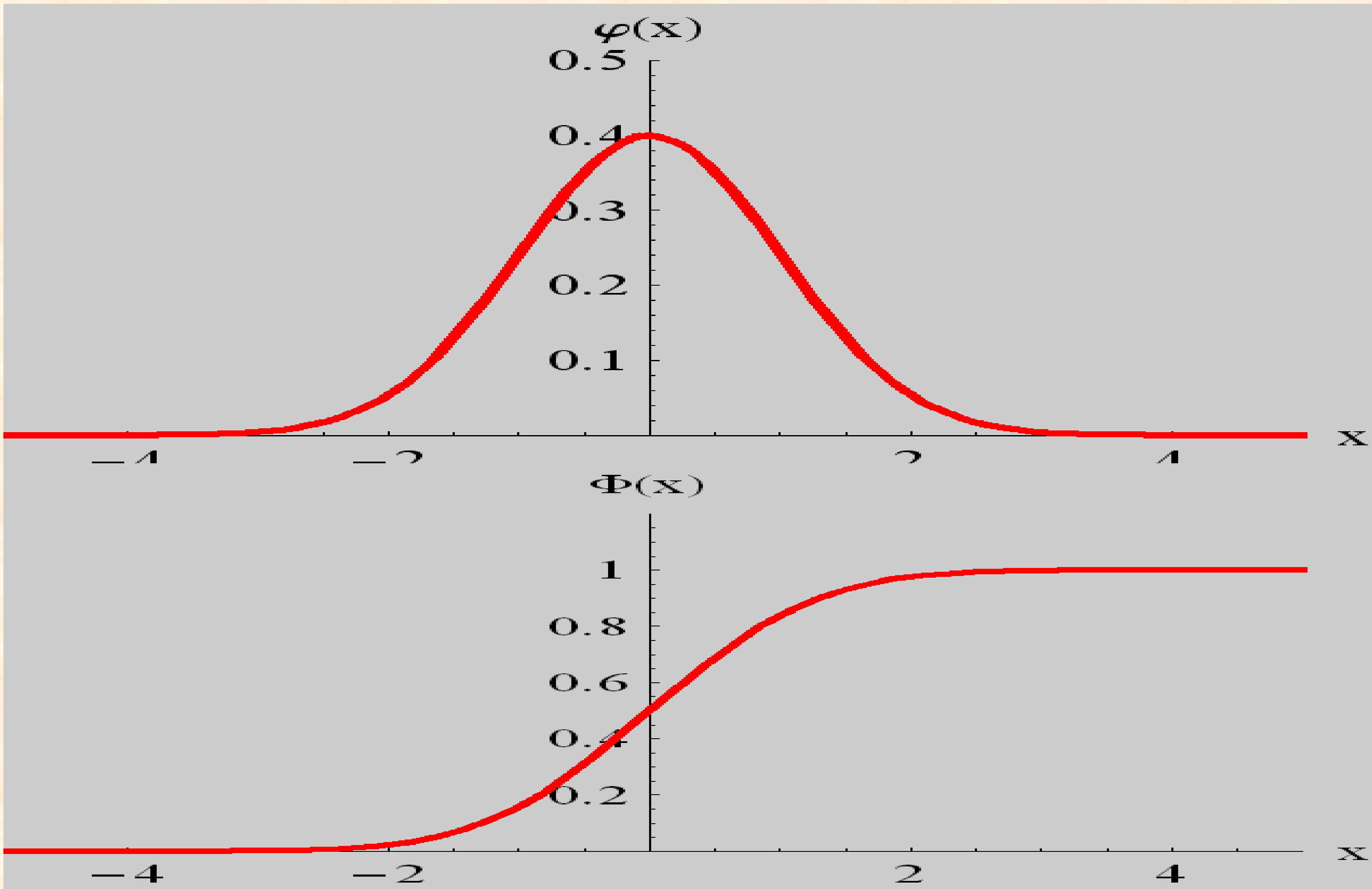
标准正态分布的概率密度表示为

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

标准正态分布的分布函数表示为

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

标准正态分布的图形



性质:

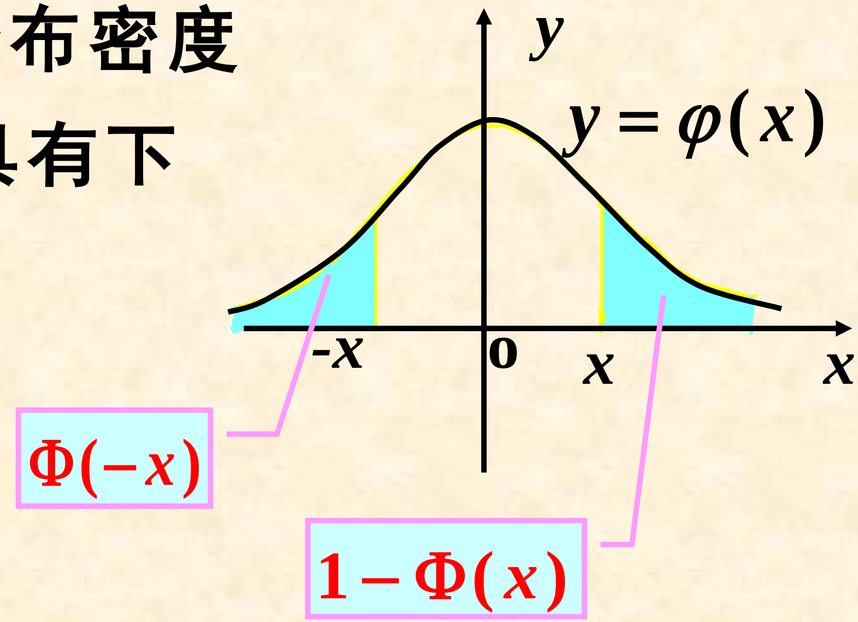
若 $X \sim N(0,1)$,则其分布密度 $\varphi(x)$及分布函数$\Phi(x)$具有下述性质:

① $\phi(x)$为**偶函数**;

② $\Phi(-x) = 1 - \Phi(x)$,

$\Phi(0) = 0.5$;

③ $\phi(x)$的原函数不能用初等函数表示;



❹ 由 $\int_{-\infty}^{+\infty}\varphi(x)dx=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=1$

可得

$$\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx=\sqrt{2\pi}.$$

若 $X\sim N(\mu,\sigma^2)$, 则其分布密度

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

$(x\in R,\sigma>0)$

计算方法:

情形1. 若 $X \sim N(0,1)$, 则

① 当 $x>0$时, 可查表2（P307 行+列= λ）, 得

$$\Phi(x)=P\{X\le x\}$$

如: $P\{X<0.25\}=\Phi(0.25)=0.5987$

② 当 $x<0$时, 可利用 $\Phi(x)=1-\Phi(-x)$;

③ $P\{a<X\le b\}=\Phi(b)-\Phi(a)$.

例3 已知 $X \sim N(0,1)$,求 $P\{1.25 \le X < 2\}$.

解 $P\{1.25 \le X < 2\}$

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

$$= 0.9772 - 0.8944$$

$$= 0.0828.$$

情形2. 若 $X \sim N(\mu, \sigma^2)$, 则

① $\dfrac{X-\mu}{\sigma} \sim N(0,1)$;

② 当 a, b 为常数，且 $a \neq 0$ 时，有

$$aX + b \sim N(a\mu + b, a^2\sigma^2);$$

③ $P\{a \le X \le b\} = \Phi(\dfrac{b-\mu}{\sigma}) - \Phi(\dfrac{a-\mu}{\sigma})$,

$$P\{X \le b\} = \Phi(\frac{b-\mu}{\sigma}).$$

证① 令 $Y=\dfrac{X-\mu}{\sigma}$, 则 $\forall y\in R$,

$\because \quad F_Y(y)=P\{Y\le y\}=P\{\dfrac{X-\mu}{\sigma}\le y\}$

$=P\{X\le \sigma y+\mu\}$

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

令 $u=\dfrac{t-\mu}{\sigma}$

$$\underline{\underline{\qquad}}\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{u^2}{2}}\cdot\sigma\,\mathrm{d}u=\Phi(y)$$

$\therefore \quad Y \sim N(0,1) \quad$ 即 $\dfrac{X-\mu}{\sigma} \sim N(0,1).$

③ $\because \quad Y = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

$$\therefore \quad P\{a \le X \le b\} = P\{\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\}$$

$$= P\{\frac{a-\mu}{\sigma} \le Y \le \frac{b-\mu}{\sigma}\}$$

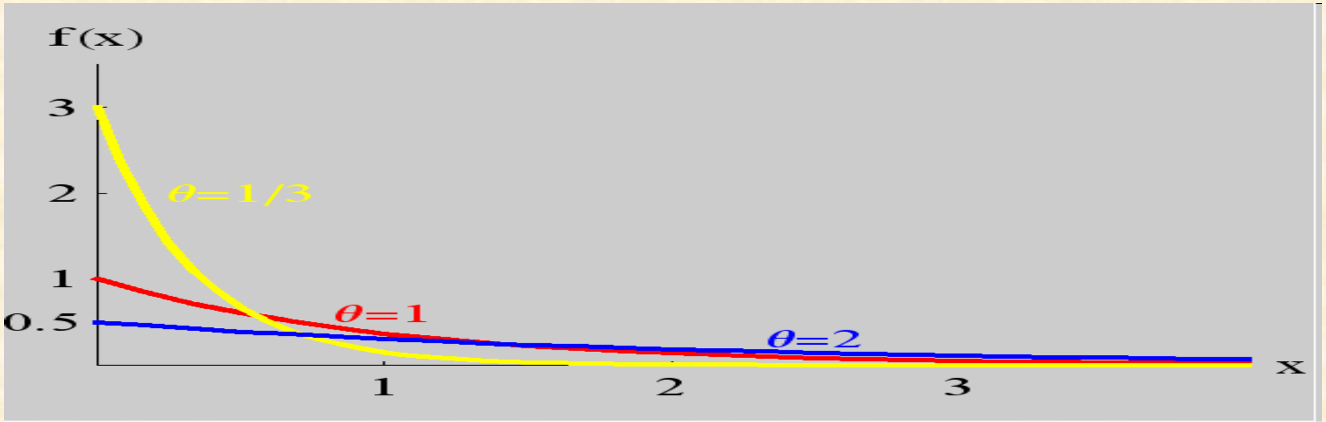
$$= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}).$$

3. 指数分布

定义　设连续型随机变量 X 的概率密度为

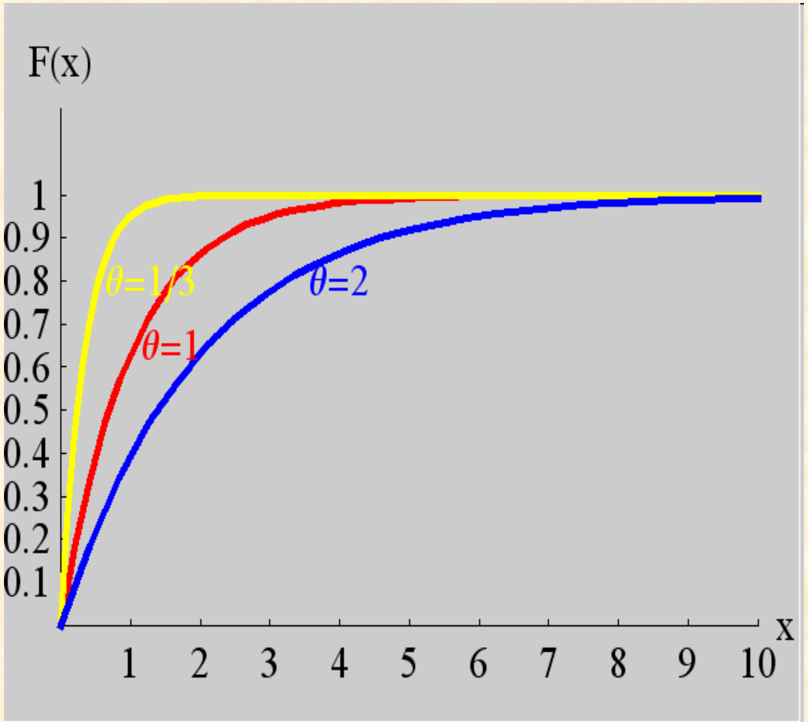
$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$$

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



分布函数

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



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 $\lambda=1/2000$ 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

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

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

$= 1 - F(1000)$

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

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

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

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

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

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

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

指数分布的重要性质 :“无记忆性”．

三、内容小结

1. **连续型随机变量** $F(x) = \int_{-\infty}^{x} p(t)\,\mathrm{d}t$

分布函数 概率密度

2. 常见连续型随机变量的分布

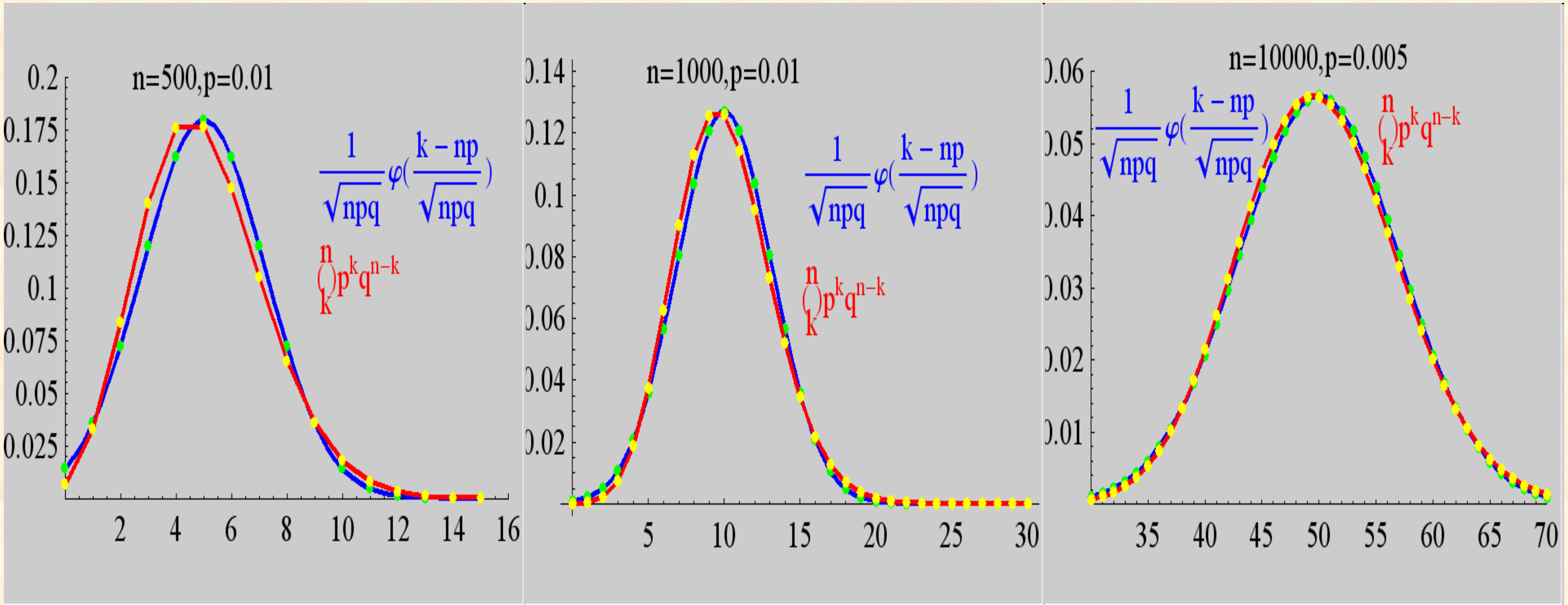
- **均匀分布**
- **正态分布(或高斯分布)**
- **指数分布**

3. 正态分布是概率论中最重要的分布

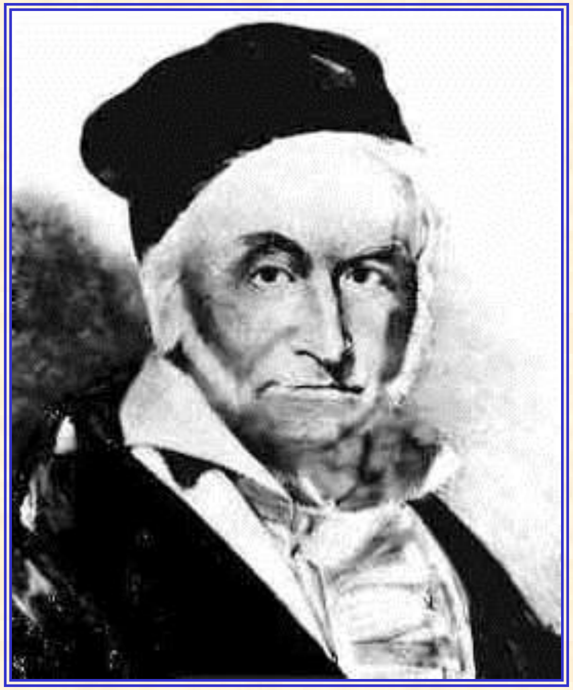
正态分布有极其广泛的实际背景, 例如测量误差; 人的生理特征尺寸如身高、体重等 ; 正常情况下生产的产品尺寸:直径、长度、重量高度; 炮弹的弹落点的分布等, 都服从或近似服从正态分布.可以说,正态分布是自然界和社会现象中最为常见的一种分布, 一个变量如果受到大量微小的、独立的随机因素的影响, 那么这个变量一般是一个正态随机变量.

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



Gauss



Carl Friedrich Gauss

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

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