

## Ch 4.4 常用连续型随机变量



# 回顾前一次课

---

期望:  $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$  及其性质

- 线性性、凸函数、对非负随机变量 $X$ 的期望、 $E(g(X))$

方差:  $\text{Var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$

均匀分布  $X \sim U(a, b)$ :  $E(X) = \frac{a+b}{2}$        $\text{Var}(X) = \frac{(b-a)^2}{12}$

指数分布  $X \sim e(\lambda)$ :  $E(X) = \frac{1}{\lambda}$        $\text{Var}(X) = \frac{1}{\lambda^2}$

指数分布的无记忆性:  $P(X > s + t | X > t) = P(X > s)$

# 正态分布

---

正太分布是概率统计中最重要的一种分布, 最早由法国数学家棣莫弗在1730s提出, 用于近似抛硬币试验中随机事件的概率

高斯在1800s首次将正太分布应用于预测天文学中星体的位置, 由此才展示出正太分布的应用价值, 正太分布因此被称为**高斯分布**

正太分布的重要性主要体现在以下三个方面:

- 现实生活中很多随机现象可用正太分布进行描述, 如人的身高等
- 很多分布可以通过正太分布来进行近似计算
- 数理统计中常用的统计分布都是由正太分布导出的, 如 $\chi^2$ 分布、 $t$ -分布和 $F$ -分布

## 正态分布 (Normal distribution/Gaussian distribution)

给定  $u \in (-\infty, +\infty)$  和  $\sigma > 0$ , 随机变量  $X$  的概率密度

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

称  $X$  服从参数为  $\mu, \sigma^2$  的**正态分布**, 记  $X \sim N(\mu, \sigma^2)$

特别地, 若  $\mu = 0$  和  $\sigma = 1$ , 称  $N(0,1)$  为**标准正态分布**, 密度函数

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

验证  $f(x)$  构成一个分布

## 标准正态分布与一般分布的相互转换

---

若  $X \sim N(\mu, \sigma^2)$ , 则

$$Y = \frac{X - \mu}{\sigma} \sim N(0,1)$$

若  $X \sim N(0,1)$ , 则

$$Y = \sigma X + \mu \sim N(\mu, \sigma^2)$$

## 正态分布的期望和方差

---

若  $X \sim N(\mu, \sigma^2)$ , 则

$$E(X) = \mu \quad \text{Var}(X) = \sigma^2$$

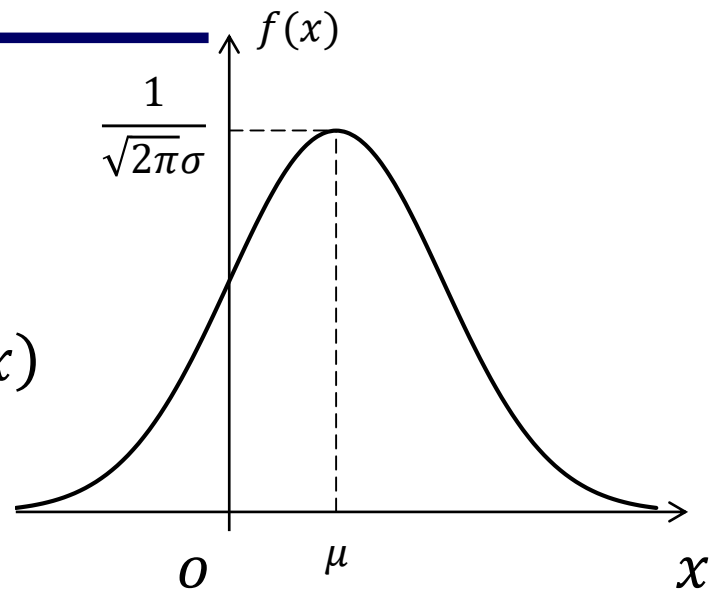
若  $X \sim N(0,1)$ , 则

$$E(X) = 0 \quad \text{Var}(X) = 1$$

## 正态分布的图像

当 $x = \mu$ 时取最大值 $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$

关于直线 $x = \mu$ 对称, 即 $f(\mu - x) = f(\mu + x)$

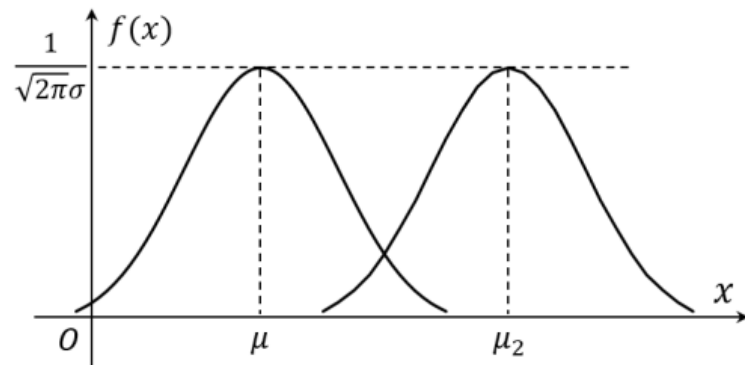


根据 $\lim_{x \rightarrow \infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 0$ 可得渐近线为 $y = 0$

密度函数的二阶导数 $f''(x) = \frac{1}{\sqrt{2\pi}\sigma^5} e^{-\frac{(x-\mu)^2}{2\sigma^2}} ((x-\mu)^2 - \sigma^2)$ ,  
可得其拐点为 $x = \mu \pm \sigma$

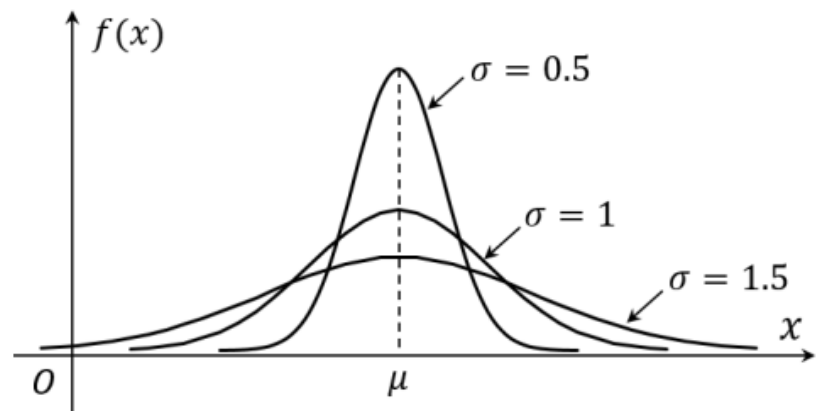
## 正态分布的图像

固定标准差 $\sigma$ 而改变期望 $\mu$ 的值, 曲线 $f(x)$ 形状不变, 仅沿 $x$ 轴左右平行移动



当 $\mu$ 固定改变 $\sigma$ 的值, 根据 $f(x)$ 最大值 $f(\mu) = 1/\sqrt{2\pi}\sigma$ 知:

- 当 $\sigma$ 越小, 图形越陡,  $X \sim N(\mu, \sigma)$  落入 $\mu$ 附近概率越大
- 反之 $\sigma$ 越大, 图形越平坦,  $X$ 落入 $\mu$ 附近的概率越小





## 正态分布的估计

---

若 $X \sim N(0,1)$ , 对任意 $\epsilon > 0$ 有

$$P(X \geq \epsilon) \leq \frac{1}{2} e^{-\epsilon^2/2}$$

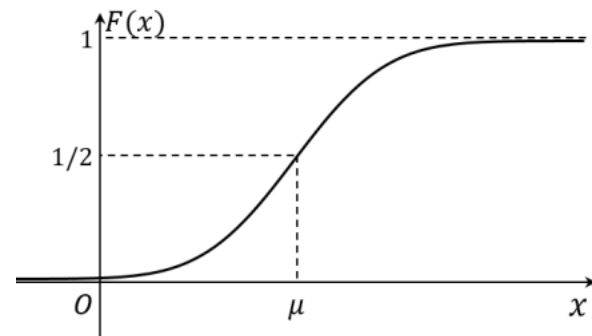
**[Mill不等式]** 若 $X \sim N(0,1)$ , 对任意 $\epsilon > 0$ 有

$$P(|X| \geq \epsilon) \leq \min \left( 1, \sqrt{\frac{2}{\pi}} \frac{1}{\epsilon} e^{-\frac{\epsilon^2}{2}} \right)$$

## 分布函数

若随机变量  $X \sim N(\mu, \sigma^2)$ , 则有分布函数

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



该分布函数没有显示的表达式, 只能求数值解

将其它正太分布转化为标准正太分布  $N(0,1)$ , 其分布函数为

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

通过查表给出了标准正太分布  $\Phi(x)$  的分布函数

# 标准分布函数表

表 4.1 标准正态分布表  $\Phi(x) = \int_{-\infty}^x e^{-t^2/2}/\sqrt{2\pi}dt$ .

$x$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706

## 标准正太分布函数的性质

---

1) 根据对称性有  $\Phi(x) + \Phi(-x) = 1$

2) 若随机变量  $X \sim N(\mu, \sigma^2)$ , 则对任意实数  $a < b$  有

$$P(X < a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right),$$

$$P(X > b) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right),$$

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

## 标准正态分布函数的性质

3) 若随机变量  $X \sim N(\mu, \sigma^2)$ , 则对任意实数  $k > 0$  有

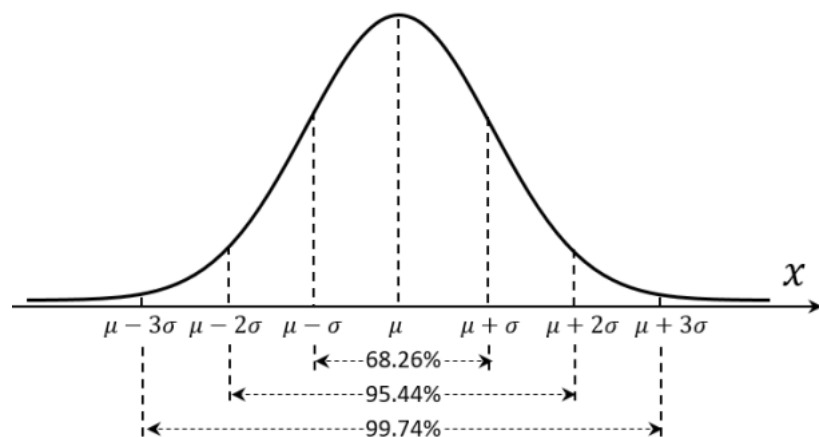
$$P(|x - \mu| < k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

特别的, 当  $k = 1, 2, 3$  时通过表有

$$P(|x - \mu| < \sigma) = 0.6826$$

$$P(|x - \mu| < 2\sigma) = 0.9544$$

$$P(|x - \mu| < 3\sigma) = 0.9974$$



随机变量  $X \sim N(\mu, \sigma^2)$  的取值范围为整个实数域, 但其取值落在  $[\mu - 3\sigma, \mu + 3\sigma]$  之外的概率不超过千分之三, 即  $X$  的取值几乎总在  $[\mu - 3\sigma, \mu + 3\sigma]$  之内, 这就是人们所说的  $3\sigma$  原则

在实际的统计推断, 特别是产品质量检测中具有重要的应用

## 标准正太分布函数的性质

---

4) 若随机变量 $X \sim N(\mu, \sigma^2)$ , 且已知 $P(X < c) = p$ , 则有

$$p = P(X < c) = P\left(\frac{X - \mu}{\sigma} < \frac{c - \mu}{\sigma}\right) = \Phi\left(\frac{c - \mu}{\sigma}\right)$$

由此可反解出 $c = \mu + \sigma\Phi^{-1}(p)$ . 这里 $\Phi^{-1}(p)$ 表示标准正太分布函数 $\Phi(x)$ 的反函数, 可根据表由里向外查得, 例如

$$\Phi^{-1}(0.5871) = 0.22$$

## 例题

---

已知某公司员工每个月的工资服从正太分布 $N(6000, \sigma^2)$ , 问题:

- i) 若已知标准差 $\sigma = 500$ , 求工资在5000与7000之间的员工在公司中占比多少?
- ii) 当标准差 $\sigma$ 为何值时, 工资在5000与7000之间的员工在公司中占比为0.803?

## 随机变量的函数

当知道一个随机变量的概率分布后, 经常会考虑它的一些函数的分布, 例如当知道一个圆的直径 $X$ 服从均匀分布 $U(a, b)$ , 需要考虑圆的面积 $Y = \pi(X/2)^2$ 的分布

一般地, 若已知随机变量 $X$ 的分布函数, 以及 $g(x)$  是定义在随机变量 $X$ 所有可能取值的集合上的函数, 则称 $Y = g(X)$ 为随机变量 $X$ 的函数, 很显然 $Y$ 也是随机变量

研究的问题: 若已知随机变量 $X$ 的概率分布和函数 $g(x)$ , 如何求解随机变量 $Y = g(X)$ 的概率分布



## 离散型随机变量的函数

---

若 $X$ 为离散型随机变量, 分布列为 $p_k = P(X = x_k)$  ( $k = 1, 2, \dots$ ), 则随机变量 $Y = g(X)$  的分布列较为简单, 将相等的项 $g(x_i) = g(x_j)$  进行合并, 相应的概率相加即可

例题: 随机变量 $X$ 的概率分布列为 $P(X = k) = 1/2^k$  ( $k = 1, 2, \dots$ ), 求随机变量 $Y = \cos(\pi X/2)$  的分布列