概率论与数理统计

正态分布

主讲人：曾华琳

信息科学与技术学院

三种重要的连续型随机变量

❸ 正态分布

若连续型 r .v X 的概率密度为

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

其中 μ 和 σ ($\sigma>0$)都是常数，则称 X 服从参数为 μ 和 σ 的正态分布或高斯分布。记作 $X\sim N(\mu, \sigma^2)$

三种重要的连续型随机变量

❸ 正态分布

1° $f(x)$具有下述性质：(1) $f(x) \geq 0$；

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

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

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad = \frac{2}{\sqrt{2\pi}\sigma} \int_{0}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

三种重要的连续型随机变量

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

令 $t = \dfrac{x-\mu}{\sqrt{2}\sigma}$ ， 则有 $\int_{0}^{+\infty} e^{-t^2}\,dt = \dfrac{\sqrt{\pi}}{2}$

$$\int_{-\infty}^{+\infty} f(x)\,dx = \frac{2}{\sqrt{\pi}}\int_{0}^{+\infty} e^{-t^2}\,dt = \frac{2}{\sqrt{\pi}}\cdot\frac{\sqrt{\pi}}{2} = 1$$

三种重要的连续型随机变量

(3) 曲线 $f(x)$ 关于 μ 轴对称;

$$\Downarrow$$

$$P(\mu - h < X \le \mu) = P(\mu < X \le \mu + h) \quad (h > 0)$$

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

三种重要的连续型随机变量

(4) 函数 $f(x)$ 在 $(-\infty, \mu]$ 上单调增加，在$[\mu, +\infty)$ 上单调减少，

在 $x=\mu$ 取得最大值；

$$f'(x)=\frac{\mu-x}{\sqrt{2\pi}\sigma^3}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

三种重要的连续型随机变量

(5) $x=\mu\pm\sigma$为$f(x)$的两个拐点的横坐标；

$$f''(x)=\frac{(x-\mu)^2-\sigma^2}{\sqrt{2\pi}\sigma^3}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$$

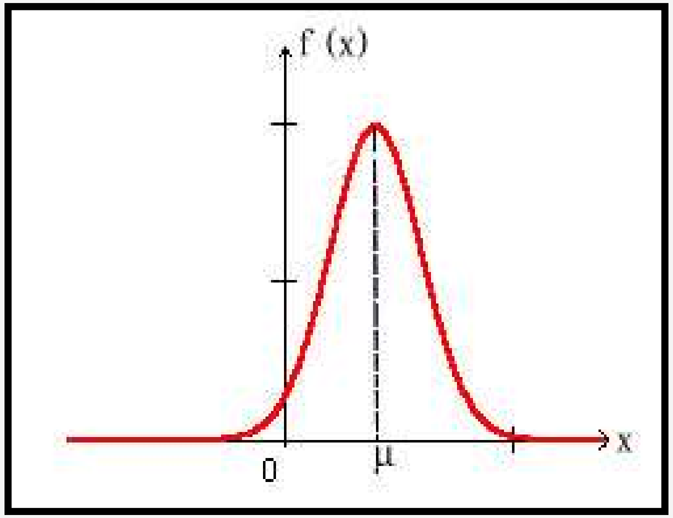
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<\infty$$

三种重要的连续型随机变量

(6) $f(x)$ 以 x 轴为渐近线

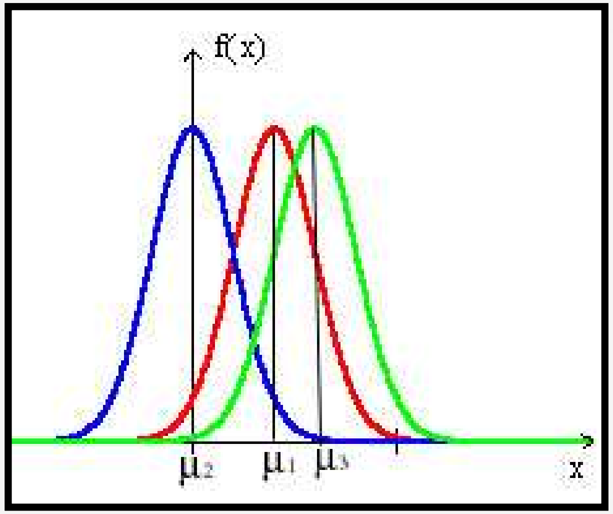
当$x \to \infty$时，$f(x) \to 0.$

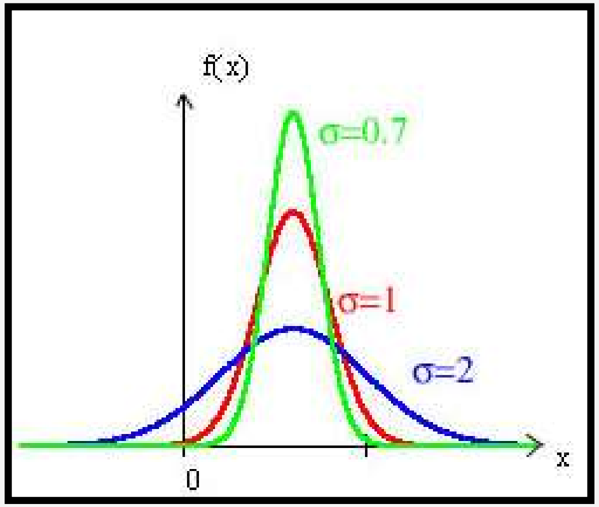
根据对密度函数的分析，也可初步画出正态分布的概率密度曲线图。



三种重要的连续型随机变量

正态分布$N(\mu, \sigma^2)$的图形特点





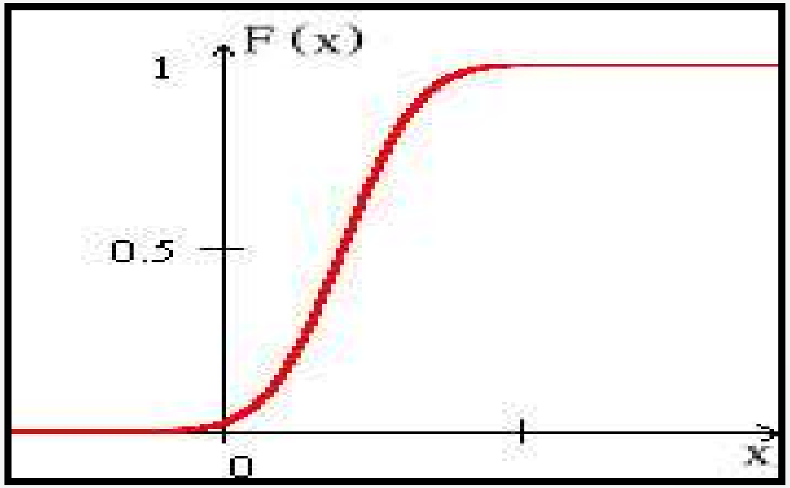
- μ 决定了图形的中心位置；
- σ 决定了图形中峰的陡峭程度。

三种重要的连续型随机变量

2° **正态分布$N(\mu, \sigma^2)$的分布函数**

设 $X \sim N(\mu, \sigma^2)$，X 的分布函数是

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



三种重要的连续型随机变量

正态分布由它的两个参数 μ 和 σ 唯一确定，当 μ 和 σ 不同时，是不同的正态分布。

下面我们介绍一种最重要的正态分布

标准正态分布

三种重要的连续型随机变量

3° 标准正态分布

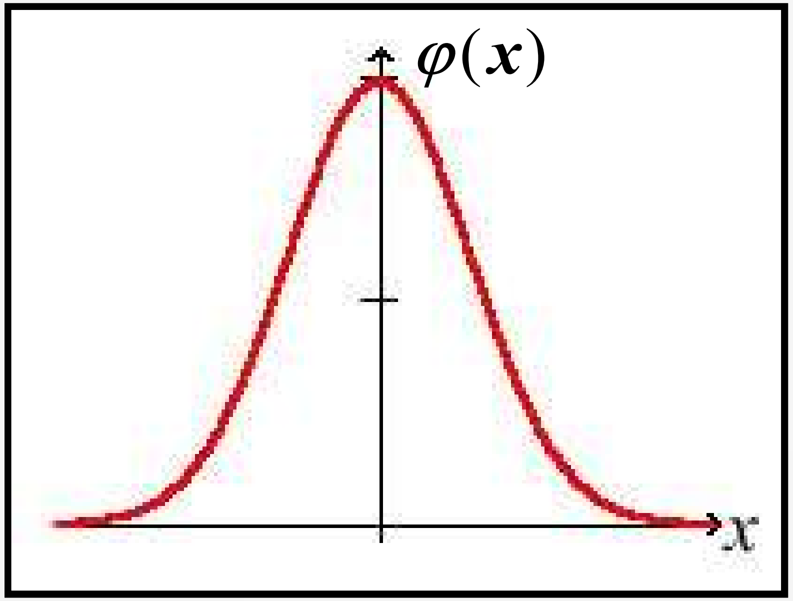
$\mu=0$，$\sigma=1$ 的正态分布称为**标准正态分布**。其密度函数和分布函数常用 $\phi(x)$ 和 $\Phi(x)$ 表示：

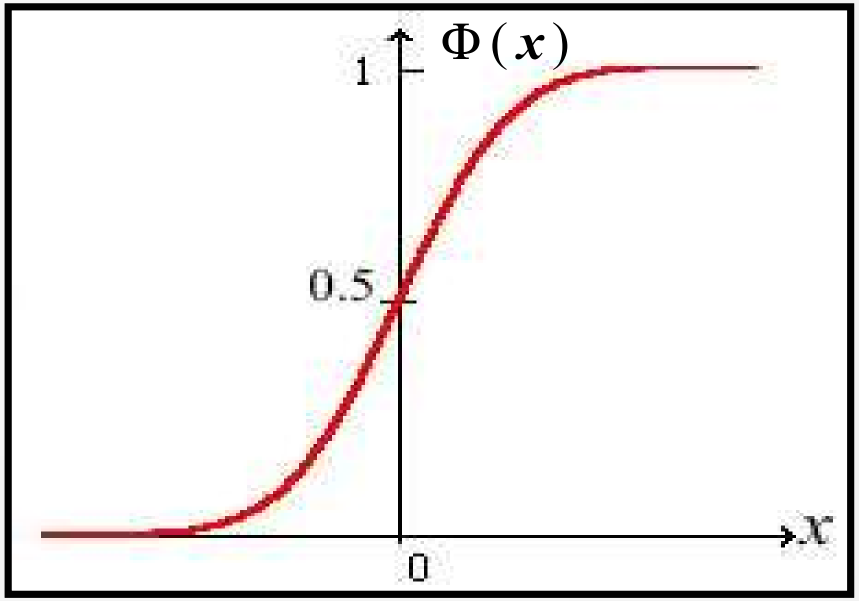
$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\quad,-\infty<x<\infty$$

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

三种重要的连续型随机变量

3° 标准正态分布





三种重要的连续型随机变量

3° 标准正态分布

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt\ \ (-\infty<x<\infty)$$ **的性质：**

(1) $\Phi(0)=\dfrac{1}{2}$ ；

$$\left(\ \Phi(0)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}e^{-\frac{t^2}{2}}dt=\frac{1}{2}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\frac{1}{2}\ \right)$$

三种重要的连续型随机变量

3° 标准正态分布

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

(2) $\forall x \in R,\ \Phi(-x)=1-\Phi(x)$；

事实上， $\Phi(-x)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{-x} e^{-\frac{t^2}{2}}\,dt$

$$\underline{\underline{u=-t}}\ \frac{1}{\sqrt{2\pi}}\int_{x}^{\infty} e^{-\frac{u^2}{2}}\,du \quad =1-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du=1-\Phi(x)$$

三种重要的连续型随机变量

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

标准正态分布的重要性在于，任何一个一般的正态分布都可以通过线性变换转化为标准正态分布。

证：Z 的分布函数为

$$P\{Z \le x\} = P\left\{\frac{X-\mu}{\sigma} \le x\right\} = P\{X \le \mu+\sigma x\} = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu+\sigma x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

令 $u = \dfrac{t-\mu}{\sigma}$ ，**则有** $P\{Z \le x\} = \dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du = \Phi(x)$

三种重要的连续型随机变量

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

于是 $X\sim N(\mu,\sigma^2)$

$$\Rightarrow F_X(x)=P\{X\le x\}=P\left\{\frac{X-\mu}{\sigma}\le\frac{x-\mu}{\sigma}\right\}=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

根据定理1，只要将标准正态分布的分布函数制成表，就可以解决一般正态分布的概率计算问题。

三种重要的连续型随机变量

4° 正态分布表

书末附有标准正态分布函数数值表，有了它，可以解决一般正态分布的概率计算查表。

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

表中给的是 $x>0$ 时，$\Phi(x)$的值。

当 $x<0$ 时，$\Phi(-x)=1-\Phi(x)$

三种重要的连续型随机变量

4° 正态分布表

若 $X \sim N(0,1)$, $P(a < X < b) = \Phi(b) - \Phi(a)$

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

$$P(a < X < b) = P(\frac{a-\mu}{\sigma} \le Y \le \frac{b-\mu}{\sigma})$$

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

三种重要的连续型随机变量

5° 3σ 准则

由标准正态分布的查表计算可以求得，

当 $X\sim N(0,1)$ 时，$P(|X|\leq 1)=2\Phi(1)-1=0.6826$

$$P(|X|\leq 2)=2\Phi(2)-1=0.9544$$

$$P(|X|\leq 3)=2\Phi(3)-1=0.9974$$

这说明，**X 的取值几乎全部集中在 $[-3,3]$ 区间内，超出这个范围的可能性仅占不到0.3%。**

三种重要的连续型随机变量

将上述结论推广到一般的正态分布，

$X \sim N(\mu, \sigma^2)$ 时，$Y = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

$$P(|X-\mu| \le \sigma) = 0.6826$$

$$P(|X-\mu| \le 2\sigma) = 0.9544$$

$$P(|X-\mu| \le 3\sigma) = 0.9974$$

可以认为，X 的取值几乎全部集中在 $[\mu - 3\sigma, \mu + 3\sigma]$ 区间内。

这在统计学上称作“3 σ 准则”。

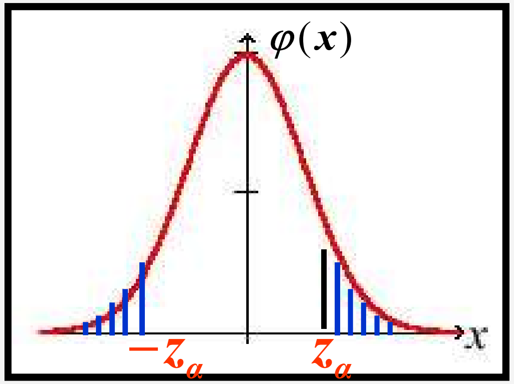
三种重要的连续型随机变量

6° 标准正态分布的上 α 分位点

设 $X \sim N(0,1)$, 若数 Z_α 满足条件

$$P\{X > z_\alpha\} = \alpha, 0 < \alpha < 1 \Rightarrow P\{X < z_{-\alpha}\} = \alpha$$

则称点 Z_α 为标准正态分布的**上 α 分位点。**



$$P\{X > z_{1-\alpha}\} = 1 - \alpha$$

$$\Downarrow$$

$$P\{X < z_{1-\alpha}\} = \alpha$$

$$\boxed{z_{1-\alpha} = z_{-\alpha}}$$

三种重要的连续型随机变量

看一个应用正态分布的例子:

公共汽车车门的高度是按男子与车门顶头碰头机会在 0.01 以下来设计的。设男子身高$X\sim N(170,6^2)$, 问车门高度应如何确定?

解: 设车门高度为h cm,按设计要求

$$P(X\geq h)\leq 0.01$$

或 $$P(X< h)\geq 0.99,$$

下面我们来求满足上式的最小的 h 。

三、三种重要的连续型随机变量

求满足$P(X< h) \geq 0.99$的最小的 h 。

因为 $X \sim N(170,6^2)$, 所以 $\dfrac{X-170}{6} \sim N(0,1)$.

故 $P(X< h)=P\left(\dfrac{X-170}{6}<\dfrac{h-170}{6}\right)=\Phi\left(\dfrac{h-170}{6}\right)$

查表得 $\Phi(2.33)=0.9901>0.99$

因而 $\dfrac{h-170}{6}=2.33,$

即 $h=170+13.98 \approx 184$

设计车门高度为184 cm时，可使男子与车门碰头机会不超过0.01。

归纳总结

这一节，我们介绍了连续型随机变量及三种重要分布。即均匀分布、指数分布、正态分布。其中正态分布的应用极为广泛，在本课程中我们一直要和它打交道。

谢 谢 大 家