第十四章 连续型随机变量：正态分布

正态分布：如果随机变量 X 的概率密度函数是

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-\mu)^2/2\sigma^2},$$

那么 X 就服从均值为 μ 且方差为 σ^2 的正态分布, 并记作 $X \sim N(\mu, \sigma^2)$. 这个概率密度函数非常重要, 也说 X 服从**高斯分布**(均值为 μ, 方差为 σ^2). 如果 X 服从**标准正态分布**, 那么 $X \sim N(0,1)$. 如果 X 服从正态分布, 有时也说 X 遵循**钟形曲线**.

14.1 确定标准化常数.

标准化常数理论：设 g 是一个非负实值函数, 并且满足

$$c = \int_{-\infty}^{\infty} g(x)\mathrm{d}x > 0.$$

如果 $c < \infty$, 那么 $f(x) = g(x)/c$ 是一个概率密度函数.

可以通过放缩确定 $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx$

有上界 $< +\infty$

现只需证明标准化常数 $c = 1$.

① 记 $I(\mu, \sigma^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx$.

先证明 $I(\mu, \sigma^2) = I(0, 1)$

令 $x-\mu=u$. 有

$$原式 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\frac{u}{\sigma})^2/2} \cdot \sigma \cdot \frac{dx}{\sigma}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$$= I(0,1)$$

② 待证 $I(0,1)=1$.

$\pi \rightarrow$ 想到圆.

一个变量 $\rightarrow$ 极坐标.

$$x = r\cdot\cos\theta,\quad y = r\cdot\sin\theta$$

$$r \in [0,+\infty) \quad \theta \in [0,2\pi)$$

$$dx\,dy = r\,dr\,d\theta$$

$$I(0,1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{y^2}{2}} dy$$

$$\Rightarrow I^2(0,1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\frac{x^2}{2}} dx \cdot \int_{-\infty}^{\infty} e^{\frac{y^2}{2}} dy$$

$$= \frac{1}{2\pi} \int_{r=0}^{+\infty} \int_{\theta=0}^{2\pi} e^{-\frac{r^2}{2}} r\,dr\,d\theta$$

$$= \frac{1}{2\pi} \int_{r=0}^{+\infty} e^{-\frac{r^2}{2}} r\,dr \int_{\theta=0}^{2\pi} d\theta$$

$$= -\frac{1}{2\pi} e^{-\frac{r^2}{2}} \Big|_0^{+\infty} \cdot \theta \Big|_0^{2\pi}$$

$$= -\frac{1}{2\pi}(-1) \cdot 2\pi$$

$$= 1$$

故 $I(0,1) = 1$

得证.

14.2 均值和方差.

均值:

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

记 $x-\mu = u$

$$\mu_X = \int_{-\infty}^{+\infty} (u+\mu) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du$$

$$= \int_{-\infty}^{+\infty} u \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du + \int_{-\infty}^{+\infty} \mu \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du$$

$$= \int_{-\infty}^{+\infty} u \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du + \mu$$

$$= \mu.$$

（奇函数的对称区间上积分为0）

方差：

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

$$\overset{w=\frac{x-\mu}{\sigma}}{=\!=\!=} \sigma^2 \int_{-\infty}^{+\infty} w^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{w^2}{2}} dw$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} w^2 e^{-\frac{w^2}{2}} dw$$

$$= \;〃$$

$$= \sigma^2$$

14.3 服从正态分布的随机变量之和.

服从正态分布的随机变量之和：如果 $X \sim N(\mu_X, \sigma_X^2)$ 和 $Y \sim N(\mu_Y, \sigma_Y^2)$ 是两个相互独立的随机变量, 并且都服从正态分布 (X 和 Y 的均值分别是 μ_X 和 μ_Y, 方差分别是 σ_X^2 和 σ_Y^2), 那么 $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. 也就是说, 它们的和服从均值为 $\mu_X + \mu_Y$ 且方差为 $\sigma_X^2 + \sigma_Y^2$ 的正态分布.

更一般地说, 如果 $X_i \sim N(\mu_i, \sigma_i^2)$ 是相互独立且都服从正态分布的随机变量, 那么

$$X_1 + \cdots + X_n \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2).$$

满足上述规律的分布被称为"稳定分布".

(非常繁琐的证明, 未抄)

14.4 从正态分布中生成随机数

思路: 得到 cdf, 利用

$$X = F_X^{-1}(Y)$$

逆变换抽样生成随机数.

困难: 累积分布函数不易求得.

需要逼近 ⇒ 级数展开.

先考察标准正态分布

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

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

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{t^2}{2}\right)^n dt$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} \int_{-\infty}^{x} t^{2n}\, dt$$

记为 $\Phi(x)$.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} \left(\int_{-\infty}^{0} t^{2n}\, dt + \int_{0}^{x} t^{2n}\, dt \right)$$

$$= \Phi(0) + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \cdot n!} \int_{0}^{x} t^{2n}\, dt$$

$$= \frac{1}{2} + (\text{有限和}).$$

$$= \frac{1}{2} + \int_{0}^{x} \phi(t)\, dt \quad (x \geq 0).$$

$$\Phi(-x) = \frac{1}{2} - \int_{0}^{x} \phi(t)\, dt$$

$$\mathrm{Prob}(-2 \leq x \leq 2) = \Phi(2) - \Phi(-2)$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \cdot n!} \int_{0}^{2} t^{2n}\, dt$$

$$= \frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \cdot 2^{n+1}}{n! \cdot (2n+1)}$$

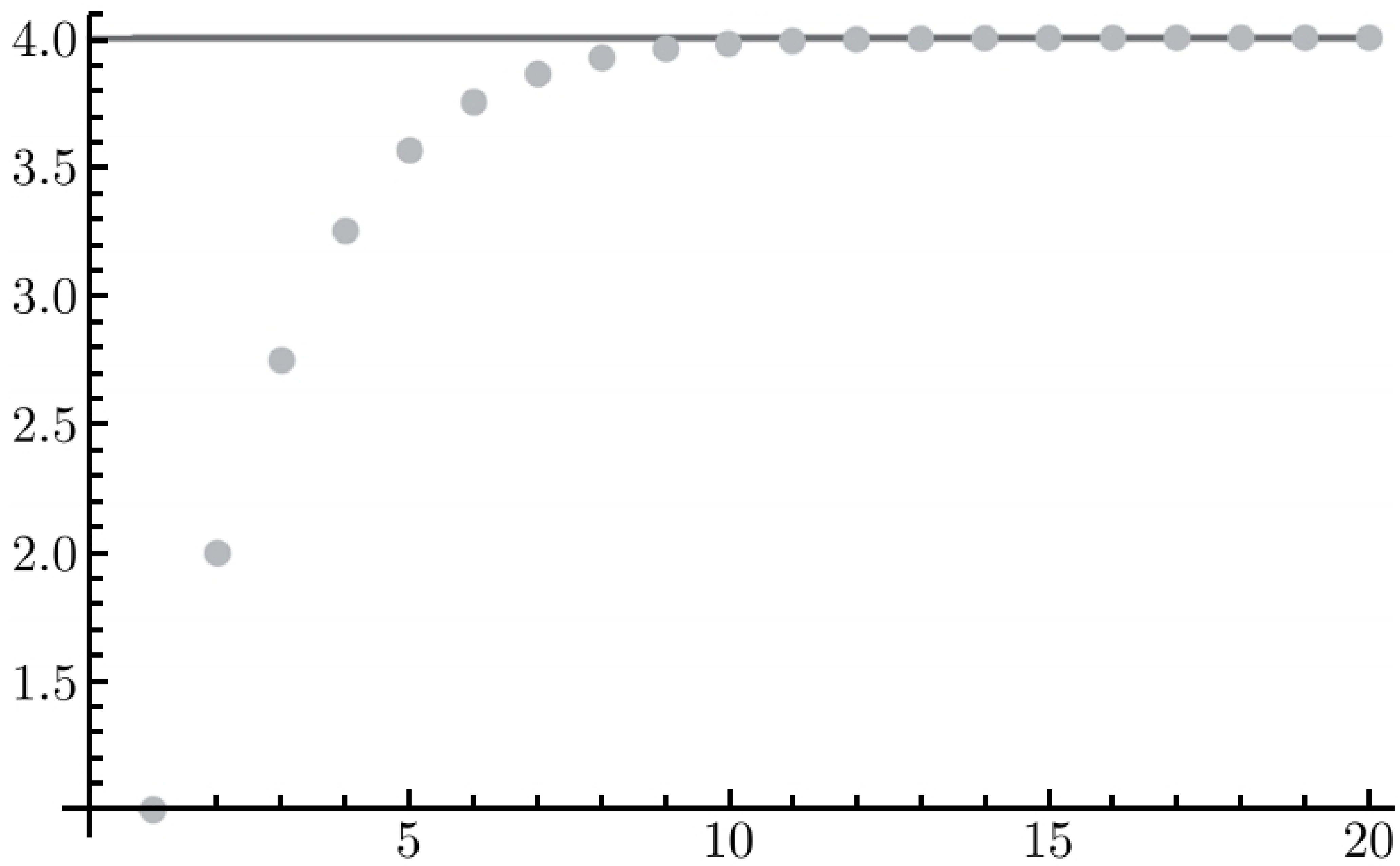

图 14-2　$\Phi(2) - \Phi(-2)$ 级数近似的收敛, 其中 $\Phi(2) - \Phi(-2)$ 表示服从标准正态分布的随机变量在 $[-2, 2]$ 上取值的概率. 不难看出, 只需要 6 项就能得到很好的近似

对于这个无法求出解析解的积分，我们命名为误差函数.

$$\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

$$\begin{aligned} \mathrm{Prob}(-2 \le x \le 2) &= F_X(2) - F_X(-2) \\ &= 2 \int_0^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{2}} e^{-u^2} du. \\ &= \mathrm{Erf}(\sqrt{2}) \end{aligned}$$

如果 Φ 是标准正态分布的累积分布函数, Erf(x) 是误差函数, 那么

$$\Phi(x) - \Phi(-x) = \mathrm{Erf}\left(\frac{x}{\sqrt{2}}\right) = \frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{n+1}}{n!(2n+1)}.$$

14.5 例子与中心极限定理.

考察随机变量 $X \sim N(5, 16)$

求 $\Pr(2 \le X \le 7)$

$$Pr(2 \le X \le 7) = Pr\left(\frac{2-5}{4} \le \frac{X-5}{4} \le \frac{7-5}{4}\right)$$

$$= Pr\left(-\frac{3}{4} \le \frac{X-5}{4} \le \frac{1}{2}\right)$$

再套用标准正态分布的数值表即可.

$$\text{原式} = \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{3}{4}\right)$$

$$\doteq 0.8944$$