概率分布的研究

朱梦

初稿于 2025-06-07，修改于 2025-06-12

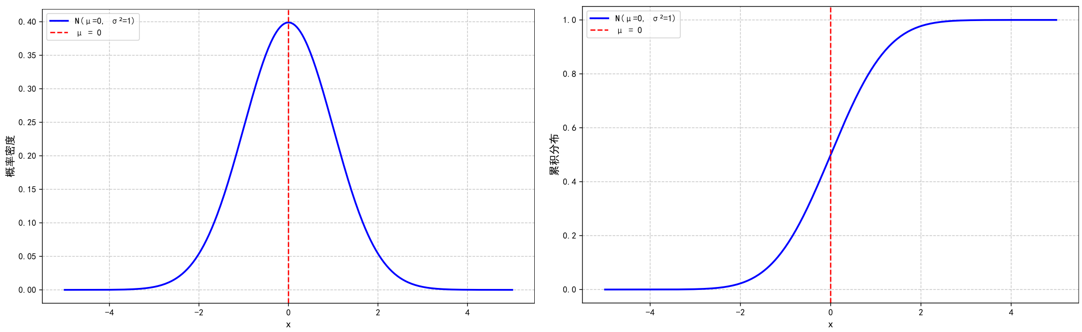
1.　标准正太分布

标准正太分布（Standard Normal Distribution）的概率密度函数和累积分布函数分别如式(1)，(2)所示：

$$f(x;0,1)=\phi(x)=\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}, \tag{1}$$

$$F(x;0,1)=P(X\leq x)=\Phi(x)=\int_{-\infty}^{x}\phi(t)=\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{2}$$

其中，$P(X\leq x)$ 表示随机变量 X 取值小于或等于某个特定值 x 的概率，erf(.) 表示误差函数。累积分布函数为概率密度函数的积分，概率密度函数为累积分布函数的导数。



(a) 概率密度曲线　　(b) 累积分布曲线

图 1　标准正太分布

2. 一般正太分布

一般正太分布（Normal Distribution）的概率密度函数和累积分布函数分别如式(3)，(4)所示：

$$f(x;\mu,\sigma)=\frac{\phi(\widehat{x})}{\sigma},\ \widehat{x}=\frac{x-\mu}{\sigma}, \tag{3}$$

$$F(x;\mu,\sigma)=\int_{-\infty}^{x}f(t;\mu,\sigma)=\int_{-\infty}^{\widehat{x}}\phi(\widehat{t})=\Phi(\widehat{x}),\ \widehat{x}=\frac{x-\mu}{\sigma} \tag{4}$$

可得：求解一般正太分布的概率密度和累积分布，可以先将输入标准化，然后查阅标准正太分布概率密度表和累积分布表，即可获得结果。一般正太分布的期望和方差分别为：

$$\mathbb{E}[X]=\mu, \tag{5}$$

$$\mathrm{Var}[X]=\sigma^2 \tag{6}$$

3. 均匀分布

均匀分布（Uniform Distribution）的概率密度函数和累积分布函数分别如式(7)，(8)所示：

$$f(x;a,b)=\begin{cases}\frac{1}{b-a}, & \text{如果}x\in[a,b],\\ 0, & \text{其它},\end{cases} \tag{7}$$

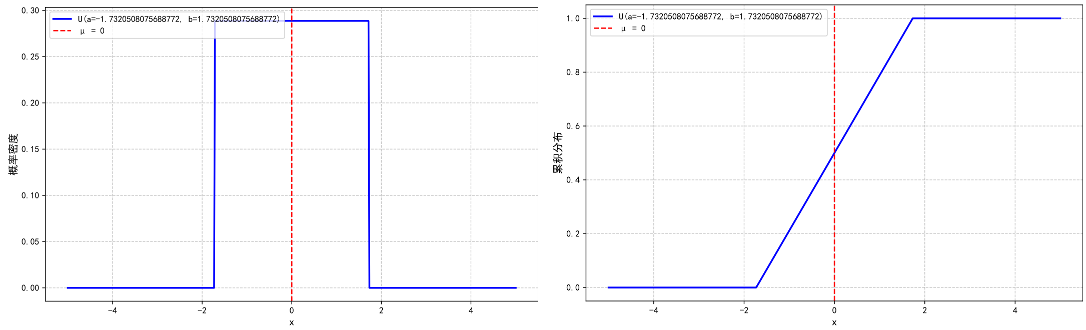
$$F(x;a,b)=\begin{cases}0, & \text{如果}x<a,\\ \frac{x-a}{b-a}, & \text{如果}x\in[a,b],\\ 1, & \text{如果}x>b\end{cases} \tag{8}$$

均匀分布的期望和方差分别为：

$$\mathbb{E}[X]=\frac{b+a}{2}, \tag{9}$$

$$\mathrm{Var}[X]=\frac{(b-a)^2}{12} \tag{10}$$

如果指定期望 μ 和方差 σ^2，那么 $a=\mu-\sqrt{3}\sigma$ 且 $b=\mu+\sqrt{3}\sigma$。



(a) 概率密度曲线　　(b) 累积分布曲线

图 2　均匀分布（$a = -\sqrt{3}$，$b = \sqrt{3}$）

4.　截断正太分布

截断正态分布（Truncated Normal Distribution）是指将正态分布限制在某个有限区间 $[a, b]$ 内，并重新规范化概率密度函数后得到的分布。截断正太分布的概率密度函数和累积分布函数分别如式(11)，(12)所示：

$$f(x;\mu,\sigma,a,b) = \begin{cases} \frac{\phi(\overset{\frown}{x})/\sigma}{Z} = \frac{\phi(\overset{\frown}{x})/\sigma}{\Phi(\overset{\frown}{b})-\Phi(\overset{\frown}{a})}, & \text{如果} x \in [a,b], \\ 0, & \text{其它}, \end{cases} \tag{11}$$

$$F(x;\mu,\sigma,a,b) = \begin{cases} 0, & \text{如果} x < a, \\ \frac{\Phi(\overset{\frown}{x})-\Phi(\overset{\frown}{a})}{\Phi(\overset{\frown}{b})-\Phi(\overset{\frown}{a})}, & \text{如果} x \in [a,b], \\ 1, & \text{如果} x > b \end{cases} \tag{12}$$

其中，Z 表示归一化常数，以保证在截断区间上的积分为 1。截断正太分布的期望和方差分别为：

$$\mathbb{E}[X] = \mu + \sigma \frac{\phi(\overset{\frown}{a}) - \phi(\overset{\frown}{b})}{\Phi(\overset{\frown}{b}) - \Phi(\overset{\frown}{a})}, \tag{13}$$

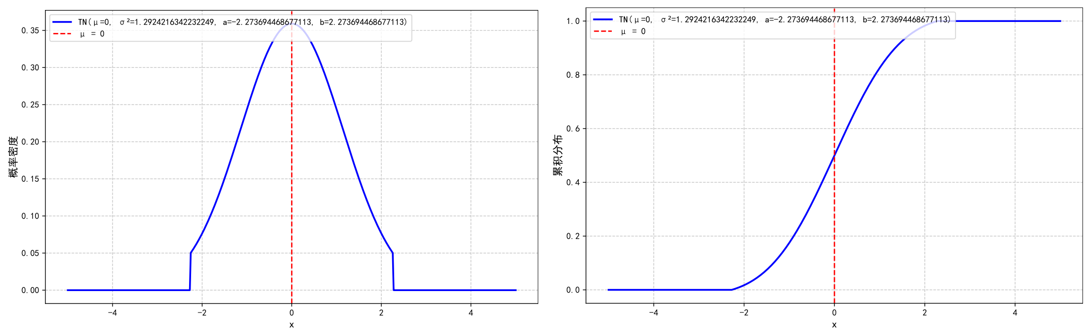
$$\mathrm{Var}[X] = \sigma^2 \left(1 + \left(\frac{\overset{\frown}{a}\phi(\overset{\frown}{a}) - \overset{\frown}{b}\phi(\overset{\frown}{b})}{\Phi(\overset{\frown}{b}) - \Phi(\overset{\frown}{a})} \right) - \left(\frac{\phi(\overset{\frown}{a}) - \phi(\overset{\frown}{b})}{\Phi(\overset{\frown}{b}) - \Phi(\overset{\frown}{a})} \right)^2 \right) \tag{14}$$

如果指定 $a=\mu-2\sigma$，$b=\mu+2\sigma$，那么截断正太分布的期望和方差分别为：

$$\mathbb{E}[X]=\mu+\sigma\frac{\phi(-2)-\phi(2)}{\Phi(2)-\Phi(-2)}=\mu, \tag{15}$$

$$\begin{aligned}\operatorname{Var}[X]&=\sigma^2\left(1+\left(\frac{-2\phi(-2)-2\phi(2)}{\Phi(2)-\Phi(-2)}\right)-\left(\frac{\phi(-2)-\phi(2)}{\Phi(2)-\Phi(-2)}\right)^2\right)\\&\approx 0.7737413035499232\sigma^2\end{aligned} \tag{16}$$

可得：截断正太分布的期望等于未截断正太分布的期望，截断正太分布的方差等于未截断正太分布的方差的 0.7737413035499232 倍。如果希望截断正太分布的方差等于 1，那么需要未截断正太分布的方差约等于 1.2924216342232249，即标准差约等于 1.1368472343385565。同理，如果指定 $a=\mu-2.5\sigma$，$b=\mu+2.5\sigma$，那么截断正太分布的方差等于未截断正太分布的方差的 0.9290050887483136 倍。如果希望截断正太分布的方差等于 1，那么需要未截断正太分布的方差约等于 1.0764203685335467，即标准差约等于 1.0375068040902415。



(a) 概率密度曲线　　(b) 累积分布曲线

图 3　截断正太分布（$\mu=0$，$\sigma^2=1.2924216342232249$，$a=\mu-2\sigma$，$b=\mu+2\sigma$）

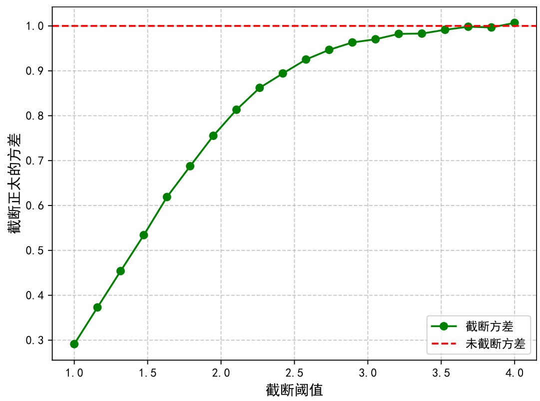


图 4　截断正太分布的方差 *vs.* 截断阈值