高斯随机变量

zcl.space

目录

八卦一下，高斯随机变量不是高斯发现的，是一个叫亚伯拉罕$cdot$棣莫弗的数学家发现的。无论哪个发现的高斯随机变量，都不影响高斯随机变量在概率论中的重要地位。

1 定义

如果随机变量X的概率密度函数是：

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty \tag{1.1}$$

称这个随机变量服从参数为μ和σ^2的高斯分布。稍后我们会发现μ和σ^2在完全控制一个高斯变量，他们分别是该高斯变量的均值和方差。无论μ和σ^2的值如何变化，高斯随机变量的图形始终是钟形。如图所示。

$f(x)$是一个概率密度函数，我们可以通过$\int_{-\infty}^{\infty} f(x)dx = 1$来验证。

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \tag{1.2}$$

变量替换，令$y = \frac{x-\mu}{\sigma}$，式 (1.2)可以变为：

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1 \tag{1.3}$$

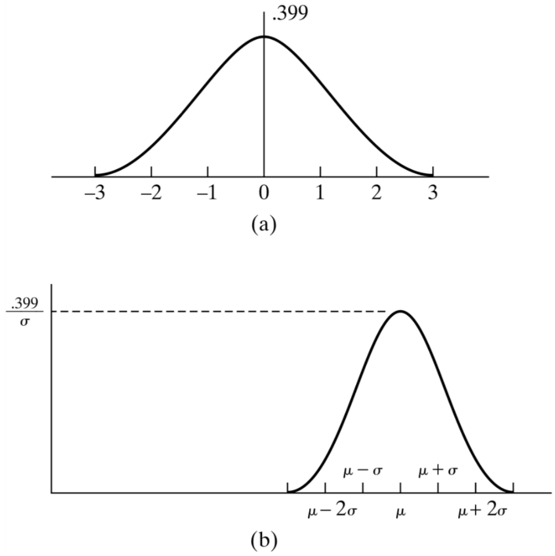

图 1: 高斯随机变量的形状

即，我们需要证明

$$\int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi} \tag{1.4}$$

这个积分非常有意思，令$I = \int_{-\infty}^{\infty} e^{-y^2/2} dy$，则：

$$I^2 = \int_{-\infty}^{\infty} e^{-y^2/2} dy \int_{-\infty}^{\infty} e^{-x^2/2} dx \tag{1.5}$$

我们利用左边变换来求解上面的二重积分，令：

$$x = r\cos\theta \tag{1.6}$$

$$y = r\sin\theta \tag{1.7}$$

则$dxdy = rd\theta dr$，所以有：

$$I^2 = \int_0^{\infty} \int_0^{2\pi} e^{-r^2/2} r d\theta dr = 2\pi \int_0^{\infty} r e^{-r^2/2} dr = 2\pi \tag{1.8}$$

因此$I = \sqrt{2\pi}$

2　正态分布的性质

2.1　期望和方差

先从标准正态分布的期望和方差开始，由于：

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx = 0 \tag{2.1}$$

$$\text{Var}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \tag{2.2}$$

分部积分，我们得到$\text{Var}(X) = 1$

2.2 正态分布的线性函数

如果X是一个服从参数为μ和σ^2的正态分布的随机变量，那么$aX + b$也服从正态分布，且参数为$a\mu + b$和$a^2\sigma^2$。设F_Y为Y的分布函数，则：

$$F^Y(x) = P\{Y \le x\} = P\{aX + b \le x\} = P\{X \le \frac{x-b}{a}\} = F_x(\frac{x-b}{a}) \tag{2.3}$$

其中$F_X(x)$为X的分布函数，求导可得Y的密度函数：

$$f_Y(x) = \frac{1}{a} f_x(\frac{x-b}{a}) \tag{2.4}$$

$$= \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(\frac{x-b}{a}-\mu)^2}{2\sigma^2}} \tag{2.5}$$

$$= \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(x-b-a\mu)^2}{2(a\sigma)^2}} \tag{2.6}$$

即，假设X是均值为μ方差为σ^2的高斯变量，则$aX+b$则是一个均值为$a\mu+b$方差为$a^2\sigma^2$的高斯变量。这个结论的一个重要应用是，如果X是一个参数为μ, σ^2的正态随机变量，那么$Z = \frac{X-a}{\sigma}$是一个参数为$(0,1)$的正态随机变量。参数为$(0,1)$的正态随机变量成为标准正态随机变量。

3 棣莫弗拉普拉斯极限定理

概率论中一个重要的结论就是棣莫弗拉普拉斯极限定理，它表明当n充分大时，参数为n, p的二项随机变量可以由正态随机变量来近似，其中正态随机变量的期望和方差与二项随机变量的期望和方差相同。更一般的表述：我们可以将二项随机变量标准化，先减去均值np，然后再除以标准差$\sqrt{np(1-p)}$，那么经过标准化后的随机变量的分布函数当$n \to \infty$时收敛于标准正态分布。

定理 3.1 棣莫弗拉普拉斯定理

在n次独立重复试验中，设每次成功的概率为p，记成功的总次数为S_n，那么对于任意$a < b$有：当(n→ ∞)时，

$$P\left\{a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right\} \to \Phi(b) - \Phi(a) \tag{3.1}$$

现在，二项分布有两种可能的近似：当n较大p较小时，泊松分布 是一个很好的近似；另外，可以证明，当$np(1-p)$较大时，正态分布近似的效果很好。一般情况下，当$np(1-p) \geq 10$时，正态近似就非常好。