

# 1. 正态分布

正态分布（normal distribution），或者说叫高斯（Gaussian）分布、钟形曲线（bell curve）统计学甚至任何科学领域中最重要概念，推论统计几乎完全是以正态分布为基础的。根据数据点进行统计，很大程度上都是基于正态分布。

正态分布在生活中无处不在，是高斯发现的。它是一种概率密度函数，公式：

$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

$\frac{X-\mu}{\sigma}$  表示的是X离均值的距离然后除以标准差，所以这项的意思是离均值有多少个标准差。这又称作标准z分数（z score）

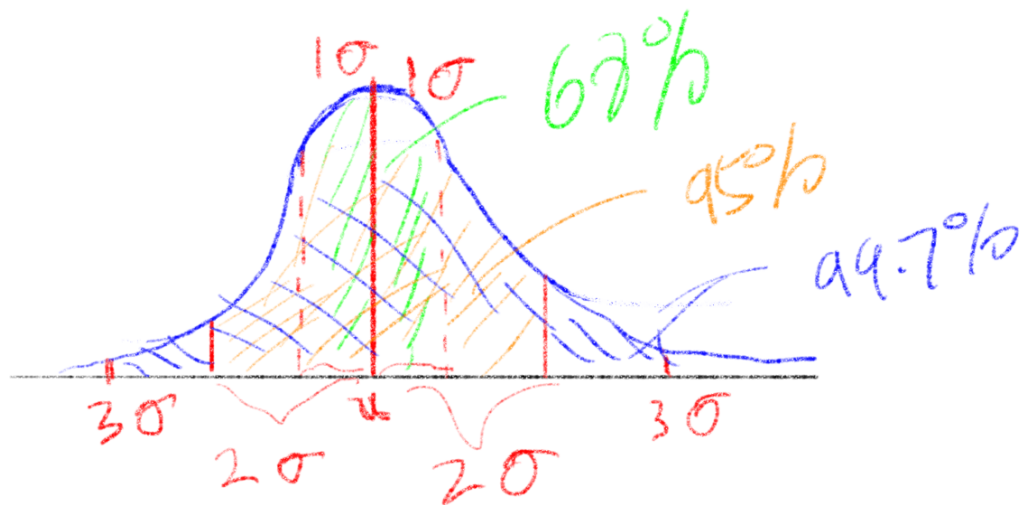
二项分布是正态分布的很好的近似。二项分布中试验次数足够时会很接近正态分布。

## 2. 正态分布的经验法则

1个标准差内的概率为68%

2个标准差的概率为95%

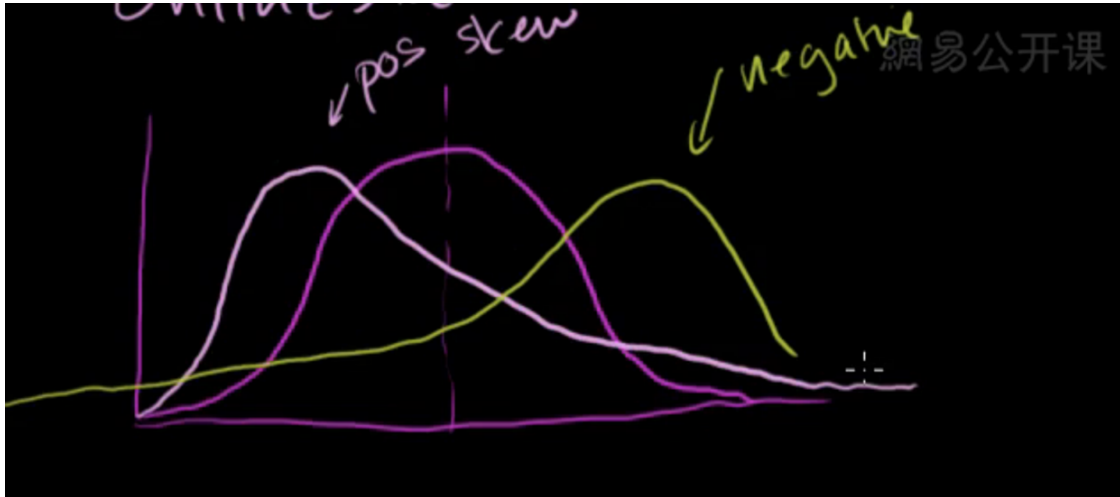
3个标准差的概率为99.7%



## 3. 衡量正态分布的指标

### 3.1 偏度(skew)

如果是完美的正态分布，偏度是0，如果偏度为正，意味着右侧尾部较长，叫正偏态分布。如果偏度为负，则左侧尾部较长



### 3.2 峰度(kurtosis)

如果峰度为正，尾部较平（fatter tails），同时峰值会较尖（more pointy peak）。这叫做正峰态；负峰态则尾部较小，中间更平滑

