

正态分布的历史八卦

zcl.space

正态分布最早由法国数学家亚伯拉罕·棣莫弗于1733年引入。他利用正态分布求出了有关抛掷硬币试验中随机事件的概率的近似值，当时称正态分布为钟形曲线。直到1809年，德国数学家高斯以正态分布作为主要工具预测天文学中星体的位置，正态分布才展现了它的应用价值。此后，正态分布就成为高斯分布。这说明高斯分布不是高斯发明的，但是高斯用这个分布做了一件与模式识别极其相关的事情。

在19世纪后半叶，大多数统计学家都开始相信大部分数据的直方图都具有高斯钟形的形状。事实上，大家认为正常的数据集应该具有这种形状。从英国统计学家卡尔·皮尔森开始，简称高斯曲线为正态曲线。大家认为正常的数据集应该具有这种形状 这个事实是通过 中心极限定理 来证明的。

现在统计学已经普及，统计学家具有很好的工作环境。然而，统计学的诞生却是在18世纪初伦敦一所黑暗，肮脏的赌窟中诞生的。这个赌窟叫做屠夫咖啡屋。亚伯拉罕·棣莫弗是一个来自天主教法国的耶稣教难民。为了生计，他要为各种赌博计算赔钱的概率。

虽然亚伯拉罕·棣莫弗在咖啡屋内谋生存，但他是一位著名的数学家，他发现了正态曲线。他还是皇家学会的会员，据说还是牛顿的朋友。

统计学家皮尔森想象棣莫弗在屠夫咖啡馆内工作的情景：“我想象棣莫弗坐在咖啡屋内肮脏的小桌边，旁边坐着一位破产的赌徒，而牛顿从嘈杂的人群走向棣莫弗的小桌边，拉出他的朋友。在艺术家的想象中，这是一幅多么伟大的艺术杰作！”

高斯是正态曲线的最早使用者。著名的数学史家 E.T.Bell在1954年的著作《数学人物》中，在“数学王子”那一章中写道：“阿基米德，牛顿和高斯”这三位位于最伟大数学家之列，我们不可以用通常的眼光来评价他们的贡献的大小。他们在纯数学和应用数学邻域内做出了重要的贡献。阿基米德特别推崇纯数学，牛顿把他的数学发现成功应用于科学研究；而高斯则宣称无论是纯数学还是应用数学对他而言都是一样的。



从这段八卦中我们可以看到，高斯最开始应用正态分布的过程就是看中了其 **正常的数据集合应该具有这种形状** 的特性，可以说高斯是第一个使用高斯分布进行模式识别研究的人（他从观测数据中预测星体的位置）。在模式识别和机器学习过程中，可以看到高斯分布的这一特性使其在这些领域有着非常广泛的应用。