观察周围的现象会发现，大部分随机变量具有“中间大、两头小、左右对称”的分布特点。如测量误差，产品的质量指标，农作物的产量，人的身体和心理指标（和身高、智力水平）等。这种随机变量所服从的分布称为“正态分布”，即“正常”状态下的分布。正态分布又称为高斯分布。但高斯并不是正态分布的发现者,高斯是正态曲线的最早应用者之一.第一个发现者应归功于法国数学家亚伯拉汗.棣莫弗(1667-1754).

1.正态概率曲线的发现过程

棣莫弗是法国的加尓文教徒.由宗教和政治的原因,他迁居到政治气氛较好的伦敦.为生计,他做家庭教师,替人计算各种赌博中赔钱的概率(现在统计学己经普及,统计学家具有很好的工作环境.然而,统计学的诞生地却是18世纪初的伦敦,一所黑喑的、肮脏的赌窟,称为屠夫咖啡屋的地方).由于他的学术成就,1697年当选为英国皇家学会会员,也是牛顿的朋友.1710年,被委派参与英国皇家学会委员会,调查牛顿-莱布尼茨微积分优先权问题.1735年,当选为柏林科学院院士.1754年,被法国巴黎科学院接纳为会员.

1733年，棣莫弗将一篇7页的论文送给了几位朋友。后来听取了朋友的意见作了些修改并增加一些内容，收录在《机会学说》。这篇文章中他第一次导出了正态概率曲线的表达式。

棣莫弗是解决如下问题的过程中导出了正态概率曲线。

考虑成功概率为的重伯努利试验中成功次数的概率问题，用今天的话说就是假设随机变量～,计算概率



在棣莫弗时代这个概率的精确计算是不可能的,因此只能近似计算.这就是棣莫弗考虑的问题.棣莫弗解决这个问题的大致过程是这样的.

假设(为正整数),先对二项分布中的中心项(也即最大项)作出近似,然后考察其他项与中心项的比.中间项为

.

棣莫弗得到了结果～.

该结果利用斯特林公式 ～容易得到：



棣莫弗最初的结果并不是这个,而是～.在有了斯特林公式后才得到这个结果(斯特林是棣莫弗的朋友) .

棣莫弗进一步证明了

～.

同样地利用斯特林公式可得上面结果:

又

，



因此。

。

进而棣莫弗得到近似式



即得

。

1774年，拉普拉斯证明了



并对棣莫佛的结果进行推广，建立了较一般的形式。

如果要给学生展示这个发现过程可以对以上推导作一些修改，以使学生更好地理解。

设随机变量～,记，，那么有



先证如下结论：对于满足的，一致地有

～。

证明：对满足的，有

，

从而有～，～，

于是

～

～

又

，

因此

～～。

再证结论

 .

证明：

记，则

～



注意到，便有

.

注意，以上证明省略了很多技术细节，比如一致性问题，近似误差的估计。另外，上面推导方法完全适用于一般情形，即成功概率为的情形，只须令，往后的推导类似。

棣莫弗认为自己解决了哲学问题：在以为纯粹偶然的事件中，人们可以寻找其规律和必然。正如他在《机会学说》第三版所言，尽管机会具有不规则性，由于机会无限多，随着时间的推移，不规则性与秩序性相比将显得微不足道。

2.最小二乘法与正态分布

天文学和测地学的发展促进了概率论与数理统计学的发展。郝德曾指出天文学在数理统计学发展中所起的作用：

天文学自古至18世纪是应用数学中最发达的领域。数学观测和天文学给出了建立数学模型及数据拟合的最初例子，在此意义下，天文学家就是最初的数理统计学家。天文学的问题逐渐引导到算术平均，以及参数模型中的种种估计方法，其中以最小二乘法为顶峰。

这说明了最小二乘法的重要地位。最小二乘法理论最早出现在勒让德1805年出版的论著《计算慧星轨道的新方法》。勒让德的成功之处在于他从新的角度看待问题，不像前辈那样致力于找出几个方程（方程个数等于未知数个数）再去求解，而是考虑误差在整体上的平衡。最初的最小二乘法是一个处理观测值的纯粹代数方法。要将其纳入统计推断框架中则需要考虑观测误差，并给误差分布建立统计模型。

1809年，高斯出版了论著《天体运动理论》.在该书末尾,他写了一节有关“数据结合”的问题，以极其简单的手法导出误差分布—正态分布，并用最小二乘法加以验证。关于最小二乘法，高斯宣称自1795年以来他一直使用这个原理。这立刻引起了勒让德的强烈反击，他提醒说科学发现的优先权只能以出版物来确定，并严斥高斯剽窃了他人成果。他们之间的争执持续了多年。

高斯较之于勒让德把最小二乘法推进得更远，他由误差密度函数推导出这个方法并详尽阐述了最小二乘法的理论依据。其推导过程是这样的：设误差密度函数为，真值为，那么测量结果的密度为。设次测量结果为，则测量结果的密度为

，

高斯认为既然是的公认的优良的估计，那么时出现测量结果的可能性最大，即时取得最大值，也即是的最大值点，从而满足

，

这等价于

，

令，则有

，

令，上式变为

，

上式对求导，并结合，得



可见诸相等,因此，从而有

，

又，

故，因此

，

解此方程得

.

由于可积，因此有，取。再由，得，所以



反过来，若，则



容易验证确是的最大值点，而最大值点正是



最小值点，这正是最小二乘估计。

高斯是正态曲线的最早应用者之一，也正是他的工作才展现了正态分布的应用价值，此后正态分布又称为高斯分布（之前称正态分布为指数钟形曲线）。高斯是第一个不用极限推导出正态分布的学者。从推导过程可以看出他已经有了最大似然估计的思想，并且提出了似然函数。高斯逆向地使用了“最大似然估计”，即先承认是最大似然估计，然后找似然函数。所以一般地认为高斯是第一个提出最大似然估计的学者。今天基于似然的推断在数理统计学中具有统治地位（两个学派分歧很大，但都承认似然原理）。这个功劳要归功于费希尔，他于1922年再次提出了似然的思想和方法，并深入地讨论了其性质，从此后这种思想和方法得到了广泛的应用。

在十九世纪后半叶，大部分统计学家认为大部分数据集的直方图都具有高斯钟形曲线的形状。大家认为正常的数据集应该具有这种形状。由英国统计学家卡尔.皮尔逊开始,将高斯分布称为正态分布.