中图分类号： 01-0

UDC： 510　　　　学校代码： 10094

河北师范大学

硕士学位论文

（学历硕士）

正态分布进入统计学的历史演化

The Historical Evolution of the Normal Distribution Entering into Statistics

作 者 姓 名： 吴江霞

指 导 教 师： 邓明立

学科专业名称： 基础数学

研 究 方 向： 近现代数学史

论文开题日期： 2006 年 12 月 10 日

二〇〇八年三月二十日

学位论文原创性声明

本人所提交的学位论文《正态分布进入统计学的历史演化》，是在导师的指导下，独立进行研究工作所取得的原创性成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中标明。

本声明的法律后果由本人承担。

论文作者（签名）：吴江霞　　　　指导教师确认（签名）：邓明立

2008年 6月 日　　　　年 月 日

学位论文版权使用授权书

本学位论文作者完全了解河北师范大学有权保留并向国家有关部门或机构送交学位论文的复印件和磁盘，允许论文被查阅和借阅。本人授权河北师范大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或其它复制手段保存、汇编学位论文。

论文作者（签名）：吴江霞　　　　指导教师（签名）：邓明立

年 月 日　　　　年 月 日

摘 要

正态分布由于其强大的普适性，是概率论中最重要的一种连续型分布。从形式上看，它属于概率论的范围，但同时又是统计学的基石，因此它的提出和应用具有其独特的双重理论背景和重要价值。

从历史上看，正态分布从问世到作为分析统计数据的概率模型经历了三个阶段：18 世纪 30 年代，狄莫弗最初在研究对一个概率作近似计算时发现了正态曲线，但由于多种原因它并没有作为刻画随机现象的概率分布；1809 年，高斯在研究测量误差时，第一次以概率分布的形式重新提出此分布，并赢得了人们的普遍关注和研究。然而人们对统计数据与观测数据不相容性的认识，使得它的应用范围却仅限于天文学、测地学等误差论领域；19 世纪中叶至末期，凯特莱在社会领域、高尔顿等人在生物学领域的工作，使正态分布迅速扩大到许多自然和社会科学领域，并最终进入统计学，成为一系列核心理论的基础和导火索。

本文以时间为顺序，以人物为依托，划分为四个部分：正态曲线在概率论中的提出（17 世纪中叶至 18 世纪中叶）、正态分布从误差论中的重生（18 世纪中叶至 19 世纪中叶）、正态分布向近代统计学的引入（19 世纪中叶至 19 世纪末期）以及正态分布对现代统计学的影响（19 世纪末期至 20 世纪 30 年代），并结合不同阶段的发展背景，对该理论的思想演化过程及其代表人物的重要工作做了系统的分析与总结，阐明了正态分布在发展的不同阶段同相关理论的相互影响，同时强调了该理论从产生到成熟发展的曲折性，诠释了概率统计理论由观念转变和技术创新来推动发展的微妙性及二者互为渗透、交融的发展规律，以期为现代历史研究和教育提供借鉴意义。

关键词：概率论 统计学 误差论 正态分布

Abstract

Due to its universality, Normal Distribution is the most important continuous distribution in Probability Theory. At the same time it is the cornerstone of the Statistics. So its proposal and application has special dual theoretical backgrounds and great value.

From a historical perspective, there are three stages from its coming out to being a probability model of analyzing statistical data: in 1730s, De Moivre initially found the normal curve when he calculated an approximation of a probability, but for certain reasons it didn't become a probability distribution describing random phenomena; in1809, while studying the theory of error, Gauss firstly proposed it in form of the distribution and caused people's attention and extensive research, but people then thought that the observed data is compatible with the statistical data, so its applied scope was limited to the field of error, such as astronomy, geodesy, etc; entering the middle of 19th century, the work of Quetelet in the social sphere and Galton's research in the field of biology made the distribution expand to many natural and social science fields rapidly, and eventually it entered the Statistics and became the basis and the blasting fuses of a series of core theories.

According to time sequence and based on the characters, this thesis consists of four parts: the proposal of the Normal Curve in Probability Theory(from the mid-17th century to the mid-18th century); the renascence of the Normal Distribution from the Theory of Error (from the mid-18th century to the mid-19th century); the entrance into the Recently Modern Statistics (from the mid-19th century to the late); and the impact on the Modern Statistics (from the end of the 19th century to the 1930s). According to the development backgrounds of different stages, this thesis gives an elaborate analysis and a summary on the evolution processes and the important work of the representative characters. Synchronously, it clarifies the interaction between the distribution and its related theories in different stages and underlines its sinuosity from generation to mature, the delicacy of the progress pushed by technique innovation and concept change and the general regular from penetration to blend, expecting to have some help on the modern historical research and education.

Key words: Probability Theory Statistics Theory of Error Normal Distribution

目　录

引 言

概率论是一门研究随机现象规律性的数学分支，统计学研究怎样有效地收集、整理和分析随机性数据，并对所考察的问题进行推断或预测，直至为采取一定的决策和行动提供依据和建议。一般来讲，统计学的发展可划分为三个阶段：古典统计时期、近代统计时期和现代统计时期；自然地，概率论也经历了古典概率论、分析概率论和公理化概率论等发展阶段，呈现出应用上由片面到全面、方法上由粗糙到严格、理论上由青涩到成熟的发展规律。从历史上看，概率论和统计学是一对姊妹学科：在时间上，二者的发展具有并行性和交织性；在理论上，二者互为渗透、影响并逐渐走向交融。概率论直接来源于赌博，其理论的深化推动了统计学的发展；而统计事业的萌芽又为概率论提供了新的现实模型，拓展了其研究和应用的范围。

形如“中间高，两头低”的钟形曲线所代表的正态分布，是概率论中最重要的一种连续型分布。许多实际问题中的变量，如测量误差、射击时弹着点与靶心间的距离、热力学中理想气体的分子速度、某地区成年男子的身高等都可以认为服从正态分布。研究表明，一个变量如果受到大量微小的、独立的随机因素的影响，那么这个变量一般是一个正态变量。从形式上看，正态分布属于概率论的范围，但同时又是统计学的基石，因此它的提出和应用具有其独特的双重理论背景和重要价值。笔者认为，正态分布从提出时的不受重视到之后在统计学中大行其道的历史过程，精辟地诠释了在发展的不同阶段概率论与统计学相关理论间的相互影响，成为相关研究的一个理论典范！按照这种基本思路，本文划分为四个部分：正态曲线在概率论中的提出（17 世纪中叶至 18 世纪中叶）、正态分布从误差论中的重生（18 世纪中叶至 19 世纪中叶）、正态分布向近代统计学的引入（19 世纪中叶至 19 世纪末期）、正态分布对现代统计学的影响（19 世纪末期至 20 世纪 30 年代）。

具体说来，17 世纪中叶到 18 世纪初概率论从帕、费通信到雅各布·伯努利全面拓展并开创数学概率论的起源过程，为狄莫弗成为历史上发现正态曲线的第一人奠定了理论基础，他做出了与雅各布·伯努利遥相辉映，并成奇葩的工作，但其成果并未引起人们足够的重视和应用，究其原因对古典统计时期概率论发展的相对孤立性可窥见一斑。18 世纪中叶以后，随着自然科学的空前发展，社会实践的需要以及纯粹数学形式化研究的深入，概率论思想的微妙性和挑战性吸引了许多数学家的注意；天文学、测地学、物理学等误差论领域的发展，使概率论思想和方法受到进一步发展的全面刺激。继伽利略、辛普森、拉格

朗日等人之后，拉普拉斯作为古典概率论的集大成者，首先围绕误差论的基本问题展开了深入研究，随后高斯力挽狂澜、另辟蹊径，以极其简单的手法，给了这个问题一个完满的解决，使正态分布首次具有了“概率分布”的身份。尤其令人鼓舞的是，正态分布理论的完善与中心极限定理、最小二乘法等理论的工作相辅相成，互为促进，激发了众多学者的兴奋点，他们纷纷把目光聚集过来，审视并完善这一理论，基本误差假设就是一个昭然若揭的范例！然而，在相当长的时期内，误差论和统计学被看成是两个不相干的领域，其研究对象和研究方法具有本质的区别，这样正态分布就陆续进入了发展的双轨道。19 世纪中叶，凯特莱大手笔地将包括大数法则、正态分布和误差法则在内的概率论广泛地用于自然现象和社会现象的研究之中，为国势学、政治算术和概率论三要素的相互交融扫平了道路。随后，统计学逐渐“走进了理论的领域，成为关于实验和观测的科学”，为正态分布的尽情释放提供了更加广阔的历史舞台。在凯特莱的启发下，高尔顿最早把统计方法应用于生物学，他提出了中位数、四分位数、百分位数及四分位偏差等概念，并创立了回归分析，对英国生物统计学派的兴起起到了奠基性作用。自此，正态分布逐步完成了从丑小鸭向白天鹅的蜕变，它亭亭玉立于世人的面前，等待着来自各个理论研究的人们的召唤。进入现代统计时期，以威尔顿、卡尔·皮尔逊和埃其沃斯等人为先导，引发了正态分布及其相关理论的一系列创新和深化，相关回归分析、多元分析、方差分析、因子分析等统计方法，陆续登上了历史舞台，成为推动现代统计学飞速发展的一个强大动力。

现有文献（参见[1],[2],[3]）大多只是谈论某一个人物或某一阶段对正态分布理论的工作，并且以详实地记录其理论上的推导和证明为主，对于正态分布从开始不受重视到之后大行其道的发展背景和历史根源并没有进行详尽地挖掘和明确地表述，而后者对于数学研究则更具有理论价值和指导意义。本文的重点是以正态分布的历史发展为题材，对其不同阶段的发展背景——同期概率论、误差论和统计学的发展状况及其代表人物的重要工作做了系统的分析与总结，并以此为线索来考证概率统计理论由互相孤立到彼此渗透，再到相互交融的发展特点。由于正态分布强大的普适性和发展的曲折性，使得如何选取最能体现其发展主脉络的史料，如何挖掘其在历史进程中同相关理论的相互影响，如何阐明该理论由观念转变和技术创新来推动发展的微妙性和一般规律成为本文的难点。笔者试图从这些角度入手，做出突破性尝试，对正态分布进入统计学的历史演化作一个更全面的梳理和考察，以期为相关历史研究和教育提供借鉴意义。

1 正态曲线在概率论中的提出（17 世纪中叶至 18 世纪中叶）

正态曲线的起源比大多数人认为的时间要早得多。美国统计学家沃克（Helen M.Walker）教授曾在一堂统计学教育课上，对 100 个学生做了一次试验，要求他们说出正态曲线是在哪 25 年发现的。结果 6—8 个学生把时间定在了 20 世纪；大部分学生可能想到了高尔顿，把时间定在了 19 世纪后半叶；只有 3—4 名学生冒然估计是在 19 世纪的前 25 年[1]。而正确时间应该是 1733 年 11 月 12 日。由此统计学专业学生历史知识的匮乏可窥见一斑，本文的写作同样希望能为学生提供一种背景知识，从而帮助他们正确理解和评价正态曲线所引起的讨论①。

1733 年，狄莫弗的二项分布正态逼近的研究，首次引入了两个事物：正态曲线和中心极限定理。在介绍这项工作之前，我们有必要先来回顾一下其前史知识，介绍概率论从帕、费通信到惠更斯第一部概率论著作的诞生，再到雅各布·伯努利全面拓展并开创数学概率论的起源过程。这是因为，就正态分布的提出过程而言，它主要受到了来自概率论的影响[4]：首先，狄莫弗的工作是解决某类赌博问题的直接结果，这与概率论的产生和发展是一脉相承的；其次，狄莫弗在做此工作之前只详细阅读过惠更斯的著作，这对他之后概率论方面的成就奠定了坚实的基础；最后，虽然狄莫弗做出了与雅各布·伯努利遥相辉映，并成奇葩的工作，但他只能被称为概率学家，而不是统计学家。

1.1 古典统计时期的概率论

概率论和统计学是一对姊妹学科，二者并行发展，互相交织并逐渐渗透。概率论直接来源于赌博，其理论的深化推动了统计学的发展，而统计事业的萌芽又为概率论提供了新的现实模型，拓展了其研究和应用的范围[5]。一般来讲，统计学的发展可划分为三个历史阶段：古典统计时期、近代统计时期和现代统计时期。所谓“古典统计时期”，是针对十七世纪中叶到十八世纪中叶这一百年间而言的[6]。随着资产阶级革命的蔓延和海外殖民地的争夺，欧洲各国在经济、政治、文化等领域都发生了巨大的变革，文艺复兴以及随之而来的社会进步为近代数学搭好了历史舞台，并注定要开创数学发展的新时代。概率论和古典统计学就是在这样的历史背景下产生的。

为了征兵或征税，欧洲的各国政府开始收集出生、死亡及结婚等人口统计学的资料，

① The articles on “Probability” in the 7th, 9 th, and 11th editions of the Encyclopaedia Britannica; the artical on “Probability” in the Library of Useful Knowledge, presumablely written by Lubbock and Bethune, but published anonymousely; the articles by De Morgan on “Theory of Probabilities” in the Encyclopaedia Metropolitana,1838,in Mathematical Papers of De Morgan.

并逐渐显示出 “统计”式处理这些资料的好处；随着殖民地的开拓，航海运输的频繁，为了减少诸如船难等意外造成的损失，具有赌博性质的保险业与精算业得到进一步发展；自然科学迅猛发展，观测和实验的重要性也日益增加，处理观测结果特别是估计观测中出现的误差的方法，变得也日益重要起来。在这些人口统计、人寿保险、误差等范畴中，需要整理和研究大量的随机数据资料，这就孕育出一种研究大量随机现象规律性的数学分支——概率论[5]。但当时刺激数学家们首先思考概率论问题的，却是赌博问题。

人们通常认为概率论始于两位法国人——帕斯卡（Blaise. Pascal,1623—1662）和费马（Pierre de Fermat,1601—1665）的工作。早期概率论中最重要的一个问题是分赌本问题，它通常被描述成下列形式：

两个人开始赌博。他们对结果下相同的赌注。所有的钱归获胜的一方。但是没到出现最后结果的时候，赌博被打断。这时，其中一个人领先。问该如何分配赌金呢？

这个问题涉及到包括期望在内的几个重要的概率论概念，一些学者认为它可能还受到了当时社会广泛关注的经济问题如保险业的催生[7]。因为任何问题的提出都不是凭空而来的，在历史上，赌博游戏早已有之，与之相随的问题也不是现在才有，只是在这时，由于社会发展和科学进步所引起的需要，才使它引起了人们的关注，尤其是引起了当时最顶尖数学家的注意和努力，这对这一问题的发展和理论的深化是至关重要的。

很快，素有“数学神童”之称的帕斯卡通过与费马通信的方式，首次使用组合工具、递推公式等方法，解决了这类赌博问题的多种形式，并引进了“值”的概念。3 年后，惠更斯（Christian Huygens,1629—1695）改“值”为“期望”，成为概率论中最重要的概念之一，为古典概率的形成提供了思想线索[8]。他发表的《论赌博中的计算》（On Reckoning at Games of Chance）一文，成为第一部概率论著作，为概率论的进一步发展奠定了坚实的基础。

在概率史界，存在着概率论诞生标志的争论，一般认为是“帕、费通信”，但也有作者认为，由于“帕、费通信”直到 1679 年才完全公布于世，故惠更斯的《论赌博中的计算》标志着概率论的诞生[9]。我个人认为，在概率论的起源这一点上，二者都做出了巨大贡献，起到了不可替代的作用。如果没有帕、费通信，恐怕兴趣广泛的惠更斯不会立刻想到去着手研究赌博问题，期间他也不可能急于寻求帕斯卡和费马的回信，从而至少会延迟他的工作；反过来，如果没有惠更斯的这一著作，概率理论的进一步系统化和一般化，也不知要等上多少年。所幸，二者没有让这种可能成为现实，概率论也正如我们所看到的这样发展了下来，从而出现了历史上第一部概率论著作，并且首次把该学科建立在公理、命

题和问题上而构成了一个较完整的理论体系。惠更斯所提出的“期望”概念以及一些具体的赌博中的推理技巧，后来都成为雅各布·伯努利（Jacob Bernoulli,1654—1705）、蒙特莫特（P.R.Montmort,1678—1719）和狄莫弗（De Moivre,1667—1754）等人建设概率论的基础。

此外，值得一提地是，1692 年在伦敦出版了一本编著者迄无定论的著作①，书名叫《论机会法则》（On the laws of Chance），其中包括一篇惠更斯著作的译文，若干篇有关机会游戏的论文，还有试图“根据经验计算概率数值”的文章，比如待产婴儿是男孩的概率或某人一年内死亡的概率等，这一定程度上也验证了我们之前所说的赌博问题来源于现实模型，反过来其一般化方法对现实问题也具有解决意义。

至此，我们可以看出，赌博游戏在概率概念的产生及其运算规则的建立中，起到了主导作用。而把概率论由局限于赌博问题的讨论全面拓展出去的转折点，应是 1713 年雅各布·伯努利的划时代著作《猜度术》②的出版，它标志着概率概念漫长行程过程的终结与数学概率论的开端。

微积分对当时数学的巨大冲击，怎么形容也不过分。许多曾被认为难以解决的问题，后来都变成更广泛数学背景下容易解决的特例，数学的前沿和边界被大大推进。由此，概率论从微积分的新思想和新方法中受益，也就成为历史的必然。雅各布·伯努利是历史上最著名数学家族的一员，对当时新兴的微积分、微分方程和变分法等领域的进展非常熟悉，这使他成为首批认识到微积分对概率论重要性的数学家之一[10]。同时，一位在纯数学的许多分支中都很有名望的数学家投入到一部全部是关于概率论的著作，恰恰是这个学科越来越重要的标志。

《推测术》写于雅各布·伯努利生命的最后两年，共 239 页，分四个部分。前三个部分再现了惠更斯的论文，并附有大量的评论和新解；给出了“排列”的定义、Bernoulli 数、前 r 项正数次幂的和公式以及多种组合方法；介绍了组合理论在各种机会游戏中的应用[2]。其中第四部分是本书最富有创造性的部分，它强调了可以把前述理论应用到政治、经济和道德事务中的思想，包括著名的结果——弱大数定律，被认为是统计学史上迈出的第一个巨大进步。

① Montucla, Lubbock and Bethune, and Galloway ascribe it to one Benjamin Motte. Todhunter is of the opinion that the author was Arbuthnot.

② Ars Conjectandi, Opus Posthumum Accedit Tractatus de Seriebus Infinitis et Epistola Gallice Scripta de Ludo Pilae Reticularis, Basel.

在大数定律里，雅各布·伯努利考虑了一组独立随机事件中，某个特定事件出现的次数与实验总次数之间的比率，并以掷骰子为例说明：当投掷次数足够多时，这个比率接近于此事件出现的概率。为了证明这个结论，他明确给出了某一事件的比率接近其概率的方法：在二项式展开中，取出所有项的一个给定数目——比如为 n 项——最大项的每一边都有 n 项，并且使得这 $2n$ 项的和与展开式的两个尾部保留项 $(n\text{项})$ 的和的比值比某个任意确定的值 c 要大。（其中他忽略了最大项，但是它的包含项也只是增加了比值的大小，结果不会受到影响。）很清楚，如果 n 被取得足够大，无论 c 取何值，这个比值超过 c 的概率就可以被用来研究（道德）确定性，而且伯努利解法的目的就是去找到一个使得比值超出 c 的 n 必须超出的边界。最后他采用直接估计概率的方法，并通过一个巧妙的不等式获得了这个 n 应取的最小值，显然这样的 n 值并不能令人满意[11]。1713 年，他的侄儿尼古拉·伯努利（Nicolas Bernoulli）也得出一个改进的结果，但在方法上并没有大的突破。

虽然所求得的 n 值并不令人满意，所用到的推理较现在的观点看也稍显笨拙，但在伯努利身上有三个值得赞赏之点需要我们指出：

1、他在概率论发展的起步阶段，给出了问题的一个恰当提法。关于频率接近概率的问题有另外两个提法，一个是当试验次数趋于无穷时，频率的极限等于概率；另一个是我们现在教科书上的说法，即频率依概率收敛于概率。前者貌似显然，实则不真，后者虽属精确，但在当时的情况下，要想证明则难上加难。雅各布·伯努利选择了一个既符合当时的严格性要求，在可操作性上又可行的提法，这对问题的初步解决是大有裨益的。

2、他首次认识到了问题的两个方面——已知事件的概率，预测事件出现的频率；给定事件出现的频率，推导事件的概率。后者相对难度更大，以至于伯努利也没能取得太大的进展，但这些思想激发了以后一个多世纪里几代数学家和哲学家们的争论和探讨。

3、大数定律在单一的概率值与众多现象的统计度量之间建立起了演绎关系，为概率论推向更广泛的应用领域奠定了理论基础。美国统计史学者斯蒂格勒认为，伯努利证明了数学家不仅可以后验地认识世界，还可以用数学去估量他们知识的限度[5]。

由于家族问题，雅各布·伯努利逝世 8 年以后，这部巨著才由尼古拉·伯努利编辑出版。它是古典概率的系统化和深刻化，更加注重指导概率计算的一般规律及其数学证明，并强调指出概率论在国家事务，道德事务和经济事务中是有广阔应用领域的。可惜地是，在他去世时并没有完成这一部分，这个世界也许不得不等待 19 世纪凯特莱的到来。但是，笔者认为，由于雅各布·伯努利不是一位应用统计学家，被凯特莱及高尔顿所着魔的搜集资

料及数据分析在他身上也没有得到充分地体现，所以在这个理论的基础科学化，效用及对人类事物的各种应用被揭开之前，它只能作为天文学家和数学家的财产，即使雅各布·伯努利完成了第四部分，他也绝不会像凯特莱那样在数学和社会、经济之间的联合上建立深刻的理论，引发各个领域的革命。历史总是这样，偶然性与必然性共存，任何理论的发展都需要经历从具体到抽象，再到抽象的具体这样一个漫长曲折的过程，它往往需要几代人的共同努力，并不能一蹴而就。

尼古拉·伯努利也试图说服其他数学家沿着他叔叔的建议而把概率论应用到新的领域中去。这一点被狄莫弗很好地领会了。在《机会学说》（The Doctrine of Chances,1718）第一版的序言结尾处，他指出："《推测术》的作者所展示出的非凡技巧和洞察力，无愧于他所获得的声望。我希望我能够继续这项事业，把《机会学说》应用到经济和政治领域，这些是我和蒙特莫特一起，被尼古拉·伯努利先生所启示的方面.....”[1]。这样，我们就引出了正态曲线发展史上的第一位主人公——狄莫弗。毫无疑问，狄莫弗受到了惠更斯及伯努利家族的影响，但同时他又结合自己的工作，对概率计算的方法做出了技巧性推进，作为正态分布发展史的第一曲——正态分布的提出，值得我们花重墨来描述。

1.2　二项式正态逼近——狄莫弗

在十八世纪数学发展的历史上，狄莫弗虽称不上是最伟大的数学家，但也不失其夺目的光华。他于 1667 年 5 月 26 日生于法国的维特里，1754 年 11 月 27 日死于英国伦敦，是卡尔文派新教徒。19 岁那年，他因保护卡尔文教徒的南特兹赦令不被废除而遭监禁，度过了两年的铁窗生涯。南特法令废除之后，为了免遭迫害，他流亡伦敦[12]。随后，狄莫弗一直生活在英国，他对数学的所有贡献全是在英国做出的。

早在法国时，他就对惠更斯关于赌博的著作很感兴趣，尤其是《论赌博中的计算》一书，启发了他以后研究的灵感。他自学成才，学习了包括欧几里得（Enclid）的《几何原本》在内的一些重要数学著作，并精通牛顿的《自然哲学的数学原理》（Mathematical principles of natural philosophy）[10]，所有这些都为他后来的成就奠定了知识基础。

天文学家哈雷（Edmoncl Halley,1656—1742）是当时英国皇家学会的助理秘书，正是他为狄莫弗走向数学殿堂铺设了一条道路。1695 年，哈雷把狄莫弗的第一篇论文《关于牛顿的流数说》推荐给英国皇家学会，并帮助他于 1697 年初当选为英国皇家学会会员。狄莫弗一生主要在概率论、年金保险和复数理论方面具有突出表现。1710 年，他曾接受委托，去解决莱布尼兹、牛顿关于微积分发明权的争论，这是英国皇家学会一直想要弄清楚的问

题。1735 年，他当选为柏林科学院院士，并于 1754 年成为巴黎研究院会员。狄莫弗一生未娶，并且贫困潦倒，以一个收入不高的教师为职业，尽管在科学界他有许多有权威的朋友，可他得到的资助并不太多。在他八十七岁时，他得了一种嗜眠症，以至于在睡了整整一天后，再也没有醒来[12]。

前面我们已经谈到雅各布·伯努利在其《猜度术》中表明，经过多次的反复试验，在一定范围内存在着预期的概率。随着试验次数的增加，某特定事件出现的频率就越接近这个概率。因此，他指出："无限次地进行此试验，我们终能正确地计算任何事物的概率，并从偶然现象之中看到事物的秩序。"但是，他并没有表述出这种偶然现象中的秩序，而且给定 c 求 n 值的问题，也需要一种实质性的方法上的改进[2]。这些工作就需要依靠狄莫弗的二项概率逼近的研究来完成。那么，狄莫弗是在怎样的背景下，以怎样的方式得到正态分布的？他得到的结果具有怎样的统计意义，正态分布产生后又引起了怎样的反应呢？下面我们将主要围绕这几个问题来进行阐述。

早在《猜度术》出版之前，狄莫弗就对概率论进行了广泛而深入的研究。1711 年他在英国皇家学会的《哲学学报》（Philosophical Transactions）上发表了论文《抽签的数理》（De Mensure Sortis），后于 1718 年成功地进行扩展，完成了杰作《机会学说》，此书大大推进了帕斯卡、费马和惠更斯等人开创的古典概率的研究，但这时他并没有讨论上述雅各布·伯努利的问题，直到 1733 年，他才对此给出了重要解法[13]。有趣地是，吸引狄莫弗解决这一问题的并不是雅各布·伯努利，而是 1721 年，一个叫亚历山大·喀明的人向他提出的一个问题：

甲、乙二人在丙家赌博，甲获胜的概率为 a，乙获胜的概率为 $b=1-a$，进行 n 局后，设甲胜 m 局。约定：若 $m \ge na$，则甲付给丙 $m-na$ 元。若 $m<na$，这时 $n-m>nb$，则乙付给丙 $(n-m)-nb=na-m$ 元。问丙所得的期望值是多少？

按照定义，此期望值为

$$D_n = E(|m-na|) = \sum_{i=1}^{n} |i-na| C_n^i a^i b^{n-i}.$$

同年，狄莫弗首先得到当 na 为整数时，

$$D_n = 2nabC_n^{na} a^{na} b^{nb}.$$

但在 n 较大时，$C_n^{na} a^{na} b^{nb}$ 的计算很不易，于是狄莫弗就想找到一个便于计算的近似公式。他首先假定 n 为偶数，且 $m=\frac{n}{2}$，$a=\frac{1}{2}$，则 $C_n^{na} a^{na} b^{nb}$ 的中项就变为二项式 $(1+1)^n$，他计算

出当$n\to\infty$时，这个值约为：

$$\frac{2A\times(n-1)^{\frac{1}{n}}}{n^n\times\sqrt{n-1}} \tag{1.1}$$

其中A为双曲对数是级数

$$\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}+\cdots$$

的值。此外，他还发现，$(1+1)^n$的中项与任何不同于它的项（与中心项距离为l的项）的比率的对数可由式子

$$\left(m+l-\frac{1}{2}\right)\times\log(m+l-1)+\left(m-l+\frac{1}{2}\right)\times\log(m-l+1)-2m\times\log m+\log\frac{m+l}{m} \tag{1.2}$$

来逼近[2]。其中他用到了求导、对数、无穷级数、积分、变量代换等先进的分析技术，从而初步回答了喀明的问题。

1725 年，喀明把狄莫弗的结果告知了斯特灵（James Stirling,1692—1770），这激起了后者的兴趣，他算出$\frac{2^n}{C_n^m}$的级数表达式为

$$\pi m\left(1+\frac{1}{4(m+1)}+\frac{9}{32(m+1)(m+2)}+\cdots\right).$$

这是π这个符号首次被引入进来，我们可以看出，当$n\to\infty$时，此式只取主项1，就得到

$$\frac{C_n^m}{2^n}\sim\sqrt{\frac{2}{\pi n}} \tag{1.3}$$

把(1.3)与(1.1)进行对比：(1.1)取的是近似，(1.3)则为取极限所得，更符合数学严密性；另外，从后来标准正态分布的正确得出来看，(1.3)更是起到了关键性作用。所幸地是，狄莫弗与斯特灵也有联系。当他从斯特灵那里获知了这一结果后，感到非常的高兴，立即展开了进一步研究。很快，他发现应用 1655 年瓦利斯得到的无穷乘积结果就可以得到(1.3)。所以这一史实表明：在正态曲线的提出这一工作中，也有斯特灵的一份功劳。

狄莫弗首先得到

$$n!\sim Cn^{n+\frac{1}{2}}e^{-n}.$$

其中常数C是一个无穷级数和的极限，但他没有给出具体的极限值。后来斯特灵利用他的发现作了进一步探讨，求出

$$C = \sqrt{2\pi}\ .$$

这样就得到了现在教科书上常见的阶乘公式：

$$n! \sim (2\pi)^{\frac{1}{2}} n^{n+\frac{1}{2}} e^{-n}\ .$$

它被称作斯特灵公式[①]。从狄莫弗的文章及他自己的叙述中，我们可以知道，狄莫弗在斯特灵之前已做了大量研究，所以二人发现这个公式的贡献应该并举。此公式在概率论中具有很强的应用性，通过它得出了许多数值逼近，正态分布的得出就是其直接受益者。

利用此逼近，狄莫弗开始考虑二项式$(a+b)^n$从任意一项至中心项的总和。他认为在n次试验中，某一特定事件出现m次的概率，是通过$(a+b)^n$的表达式中含有m次的那一项（即第$m+1$项）表示出来的，即

$$C_n^m a^m b^{n-m} = \frac{n!}{m!(n-m)!} a^m b^{n-m}\ .$$

其中，a是某一事件出现的概率，而$b = 1 - a$。这样，他就得到了我们现在概率教材上常见的二项分布。狄莫弗注意到，可用一个字母代表一个事件发生的几种情况，如果用x表示一个事件发生的概率，那么$1-x$则意味着这一事件不发生的概率，同样，$x(1-y)(1-z)$则表示一事件发生，同时其它两个事件没有发生的概率。这种用字母来描述事件发生情况的方法，省去了麻烦的想象和繁琐的叙述，为更好地进行演算提供了方便。在此基础上，1733年狄莫弗走出了具有决定意义的一步，证明(1.2)可变形为

$$\frac{C_n^{m+l}}{C_n^m} \sim \exp\left(-\frac{2ll}{n}\right) \tag{1.4}$$

把(1.3)与(1.4)结合，就得到

$$\frac{C_n^{m+l}}{2^n} \sim \frac{2}{\sqrt{2\pi n}} e^{-\frac{2ll}{n}}\ .$$

最后，他以定积分代替和，得到了当$p = \frac{1}{2}$时极限定理的单边形式：

$$\sum\nolimits_{i:c_1\sqrt{n} \le m-i \le c_2\sqrt{n}} C_{2m}^i \sim \frac{1}{\sqrt{2\pi}} \int_{2c_1}^{2c_2} e^{-\frac{x^2}{2}} dx\ .$$

---

① Stirling announced the formula in his Methodus Differentialis, p. 137(London, 1730).

这里$-\infty < c_1 < c_2 < \infty$，$c_1, c_2$有界但与$n$有关，其中等式的右边就是当$\mu = 0$，$\sigma = 1$时的标准正态分布函数。

这个公式最初发表于 1733 年 11 月 12 日。1730 年，狄莫弗出版了他的《分析引论》（Miscellanea Analytica），三年后他私下给他的一些朋友一篇 7 页的题目为《二项式$(a+b)^n$展开的逼近定理》（Approximatio ad Summam Terminorum Binomii $(a+b)^n$ in Seriem expansi）的简要论文，并以拉丁文小册子的形式把它作为《分析引论》的附刊予以发表。这个文件的发现应归功于卡尔 •皮尔逊，1924 年他在《正态曲线史》（A History of the normal curve）一文中提到了它，现存的仅有两个复写本有该附录，所以极其难得[1]。1738 年，狄莫弗在《机会学说》的第二版中附加了它的翻译版。

同雅各布・伯努利关于大数定律的工作进行比较，可以得出狄莫弗研究的几点创新：

首先，思想上的先行者。正如他自己所言："我知道无人尝试这个问题，虽然他们①有很高的技巧，他们的勤奋受到人们的赞扬，但是这些是远远不够的，因为他们所做的无非是由于确定了一个很宽的范围，使他们所证明的各项的和都包含在其中，从而达到非常好的逼近……我专心致力这一问题的研究，并不是我比别人强，而是为了实现一位非常值得尊敬的先生、一位著名数学家②的愿望，他鼓励我去做这个问题，我现在给以前的东西增加一些新的思想，是为了使它们更加连贯，清楚……" [2]。

其次，技巧上的突破性。我们可以看出，从狄莫弗的结果很容易证明大数定律，并对n值的界限给出最为精确的确定；反过来，由大数定律到极限定理，则不是一朝一夕之力所能及的。雅各布・伯努利更早接触到二项式，但随着指数的增长，关于曲线的纵坐标，他没有试着去寻找二项式展开的一个近似函数，因此就没有得到方程的正态曲线；他也没有试着去寻找那个展开式的给定次数和的近似表达，因此他也没有得到任何与概率积分相关的结果。而这两者正是狄莫弗所做出的不可替代的贡献。卡尔・皮尔逊针对雅各布・伯努利方法的缺点曾说："如果采用它，将或者使保险团破产或者使它的顾客受损惨重"③，可知其方法不够精细，在统计应用上的受限性还有待改进。

① 雅各布・伯努利及其侄儿

② Presumably the Honorable Francis Roberts，who wrote "An Arithmetical Paradox，concerning the Chances of Lotteries" Philosophical Transactions, XⅦ(1693),677—681.

③ Pearson, "James Bernoulli's Theorem," Biometrika XⅦ(1925),201-210(this is a critical analysis of one particular aspect of the work of Bernoulli，namely his work on the use of the binomial expansion in calculating probabilities).

最后，工作上的再思考。狄莫弗本人意识到了其结果的重要性，并花费大量时间研究其蕴含的思想及一些简单结论。他虽然没有用钟形曲线描述连续型分布，但对其形状、拐点、最高点等基本性质都有重要认识。他还考虑了积分$\int_0^a e^{-\frac{x^2}{2}}dx$的数值计算问题，发现正态曲线下正负标准差$\sigma$间的面积与总面积的比值为0.682688，并且一个误差在数值上小于3σ的概率是0.99874[2]，这在没有计算机的当时，实在是令人望而生畏的事情，狄莫弗致力于数学研究的勇气昭然若揭。

狄莫弗的二项式正态逼近的工作，是概率论应用及统计学中最富有成效及最具有指导意义的发现，之后，经拉普拉斯（Pierre Simon Laplace,1749—1827）、俄罗斯学派等学者的努力，到20世纪30年代独立变量和的中心极限定理最一般的形式最终完成，嗣后统计学家发现，一系列的重要统计量，在样本量时，其极限分布都具有正态形式，这构成了大样本方法的基础。如今，大样本方法在统计方法中占据了很重要的地位，饮水思源，狄莫弗的工作可以说是这一重要发展的源头。我们喜欢想象着在伦敦的某个咖啡屋里狄莫弗和一帮好友坐在桌旁，向他们讲述他所掌握的有限级数的性质以及他对赌博运气的兴趣；我们同样喜欢带着一份疑问和惊讶去想象，他们将会做出一种预言：这样一个时代将要到来，这个定理将会影响这个世界关于所有社会问题的思考，它会进入学校，形成教材，科学中成千上万的研究都将诉诸它的帮助，虽然这些科学研究的确切名字当时还不知道。而事实上，这个愿望的实现并不是一帆风顺的，伴随着概率统计发展的错综复杂性，它同样经历了一段漫长曲折的道路，而最直接的原因就是狄莫弗的工作当时并没有引起更多的关注。

1.3 正态曲线影响微小的原因浅析

二项概率逼近的研究固然重要，可在当时狄莫弗的工作并未引起人们更多的重视，正态曲线也仅仅停留在一个数学表达式的层面，在实际应用中还没有找到其适合存活的土壤，挖掘其内在的理论和现实原因具有重要的历史意义。陈希孺先生认为：狄莫弗本人并不是一个统计学家，他从未从统计学的观点去考虑其工作的意义。狄莫弗的出发点始终是：把p作为一个已知值，如何在数值上去逼近二项式概率及其指定次数的和，而不是把p看作未知，如何通过观察结果去对p进行推断的问题[13]。这当然是很重要的一点，但除此之外，我个人认为还有一些附加原因可归纳如下：

首先，特定舆论氛围的影响。在古典统计时期对概率论存在两种不正确的看法：一是认为概率论来源于赌博，因而轻视并反对将其作为科学研究的手段；二受经院哲学的影响，

把大数法则看成是神定秩序[4]。狄莫弗本人也受到了这种思潮的影响，在后期的著作中，他表现出一种既想把概率学科从与赌博有关的诽谤中解脱出来，又想确立一种以稳定的统计比例来表现的神学秩序去在人类事务中发挥作用的热切。由此看，这段时期的概率论始终不能挣脱赌博和神学的羁绊，获得普遍应用。

其次，缺乏现实问题的刺激。在很长一段时期，统计方法在社会问题中的应用主要限于人口统计，特别是典型的与二项分布有关的男、女婴出生的比例问题[14]。人们的注意力总是被引向这个领域，而最终放弃其它方面的兴趣。这样，应用上的重要性及其较简单的形式，使得二项分布得到了众多学者的关心，它作为一个概率方法可有用武之地的模型，成为 19 世纪以前推断统计最主要的研究对象。而正态分布作为一种连续性分布，也必须要等到它所适合的现实问题（如误差论）发展到一定阶段才能引起人们的重视和研究，从这一点来看，这也是历史发展的必然规律。

再次，收效甚微的相关研究。与狄莫弗同期的有三位学者：英国牧师贝叶斯（Thomas Bayes,1702—1761）、数学家辛普森（Thomas Simpson,1710—1761）以及瑞士数学家欧拉（Leonhard Euler,1707—1783），他们都或多或少地研究过狄莫弗的工作，并试图给予发展，但都收效不大。其中，贝叶斯有一篇遗著，题目为《论机会游戏中的一个问题》（An Essay Towards Solving a Problem in the Doctrine of Chances），据说他写此文的动机就是为解决狄莫弗没能解决的二项分布概率 p 的估计问题[13]。这篇文章提出了一种逆概率的思想，后经追随者发展成为一种有效的统计推断方法，并自成一派。然而在贝叶斯定理诞生初期，也没能得到学术界的注意。辛普森（Thomas Simpson,1710—1761）是一位天文学家，具有接触误差论问题的客观条件，并重视概率论的应用，熟悉狄莫弗的著作，应该说他是在此问题上做出突破最具备良好条件的人。1755 年，他曾试图从概率论的角度去论证在应用天文学中取若干个观测值的平均值的好处，但令人遗憾地是，他并没有从狄莫弗的著作中得到丝毫有关概率分布的启发，而是假设了误差的三角形分布[2]。欧拉是 18 世纪脱颖而出的最伟大的数学家，他也曾考虑过本质上类似狄莫弗考虑的问题，但兴趣的广泛性使他并没有在这一领域做过多的纯粹性研究，因此正态曲线作用的发现和进一步发展也没能得到欧拉的垂青。

最后，史学家的疏忽和误导。丹麦统计学家 Todhunter 在其《概率论史》（History of the Theory of Probability）中完全遗漏了狄莫弗的“逼近”在《机会学说》中的扩展及其划时代性。他没有强调这是斯特灵公式的起源，是正态曲线的第一次出现，是先于拉普拉斯和

高斯的工作；他也没有提出狄莫弗扩展了牛顿的术语以及指导了统计学几乎一个世纪的发展路线。卡尔·皮尔逊认为 Todhunter 在抓住科学革命的趋势，以及挖掘当时与科学革命相关的思想方面几乎全部失败，这对科学思想的影响与科学对公众思想的影响是同等重要的[3]。正是在“History”的影响下，狄莫弗的这篇论文才没有被作为对统计学的数学理论和统计实践有广泛影响的原则被使用，我们现在所知的正态分布也没有被称作“狄莫弗分布”。在这种意义下，我们可以看出史学家在科学的传播和发展中所起到的向导性作用。

由此看，在狄莫弗时代，使正态分布作为一种有效的概率模型的时机还远不成熟。然而，18 世纪中叶以后概率论思想的微妙性和挑战性吸引了许多数学家的注意，他们思考如何依据崭新的思想去求解基础算术问题，如何给出一个好的概率定义，如何应用概率论去得到新的发现，新的解释，为概率论走向更广阔的应用做出了自己的贡献。

那么究竟正态分布是如何被人们发现其重要性的，以之命名的高斯在其中发挥了怎样的作用，这个时期的概率统计理论又进入了怎样的发展阶段呢？对于这些问题的回答将作为本文第二部分的内容。由于正态分布是沿着另一条截然不同的道路被重新发现并引起人们重视的，因此我们称之为——正态分布的重生。

2 正态分布从误差论中的重生（18 世纪中叶至 19 世纪中叶）

无论是对自然现象还是对社会现象进行观测，总会产生误差，这一点在很早以前人们就注意到了，但是对于其观测值所呈现出的随机性人们总是认识模糊，众说纷纭，不能达成统一。进入 18 世纪，数学呈现出一个很重要的特征：“工作的目标不是数学，而是求解社会问题；数学是实现实践目的的一种方法”。这段时期，数学、物理学和天文学是自然科学的主体[15]。数学的主流是由微积分发展起来的数学分析，由于分析方法广泛而卓有成效的应用，致使天文学和物理学变得数学化起来；随着科学数学化进程的深入，概率论的地位也就日益增强，它不仅得到传统哲学和神学的认可，而且为众多的科学成果所证实。这样，概率论就进入到广泛应用的时期，而把概率论应用到描述误差的问题中来就推动了误差论的发展。

误差论的基本问题是：随机测量误差应服从怎样的概率规律，即有怎样的概率分布。这个问题的提出和讨论，是天文学者的功劳[13]，由于正态分布的重新提出始终是围绕误差论基本问题的解决来做出重大突破的，因此我们将首先介绍误差论在天文学领域的早期发展情况，侧重阐述伽利略的早期思想萌芽及辛普森、拉格朗日（Lagrange,1736—1813）对问题的进一步明确化和初步研究，这为高斯（Gauss,1777—1855）提出“正态误差定律”，以及拉普拉斯、勒让德（Legendre,1752—1833）等人发展误差论，推广中心极限定理，并开创最小二乘法理论，提供了一定的现实氛围和知识储备。

2.1 误差论在天文学中的酝酿

丹麦统计史学家哈尔德在其著作《1750 年以前概率统计及其应用史》中，指出：“天文学自古代至 18 世纪是应用数学中最发达的领域，观测和数学天文学，给出了建模及数据拟合的最初例子。”[16]正态分布的重生过程为此评论提供了一个典型的例子。天文学的问题使得人们最早关心使用算术平均的合理性问题，并从误差分布论的角度来进行考察。

早在 17 世纪初，开普勒在著作《和谐的世界》中提出了一些建模（model building）的原则，其中有一条是“模型选择的最终标准是其与观测数据的符合程度”，这实质上蕴含了误差概率理论的问题，但他并没有使之明确化。伽利略（G. Galileo,1564—1642）在 1632 年出版的著作《关于两个主要世界系统的对话——托雷密和哥白尼》中给出了有关天文学家测量天体距离的结论。他推断：观测误差在仪器观测中是无法避免的；小误差比大误差更有可能发生；测量值在真值两测有相同的出错倾向；大多数观测值都聚集在真值周

围[17]。由此看，虽然他没有提出“随机”和“分布”这样的概念，而是使用“观测误差”的名称，但他揭示了正态概率定律的许多特征。他宣称在观测中引起的（随机）误差和由计算引起的最后误差是不同的，并提到误差的传递性质：所算出的（天体间的）距离是一些观测值的函数，观测值小的变化，可以引起距离值的大变动。伽利略似乎是第一位在著作中提出观测误差这个概念，并对之有所讨论的学者。

但是由于当时概率论发展的水平，没有能够提供进行有意义的定量式研究所必须的工具，因此伽利略的工作并没有超出定性式讨论的范围。这又使我想到，如果狄莫弗能够看到伽利略的这部著作，或者他本人也是一位天文学家，是否会加速正态分布时代的到来呢？以我们前面的分析看，恐怕也不好成定论，而且从后面对此诉诸努力的几位学者来看，同样经历了尝试—失败—再尝试的漫长过程。18 世纪中期，辛普森和拉格朗日开始了最初的尝试。

辛普森的工作上面我们已经有所介绍，他在一封题为《在应用天文学中取若干个观测值的平均值的好处》的信中指出：通常天文学家们对取平均值的做法并不大能接受，他们认为，应选择其中那个“谨慎地观测”所得的值，这比平均值要可靠。辛普森意识到这是一件有意义的事情，于是决定使用数学方法来验证一下，是否取平均这种做法有更大的可信度[18]。现在，我们认为，当时天文学界对此持怀疑态度，有其现实性原因：不同天文台的设备、观测条件以及人员素质都难免会有差异，故其观测结果的可靠性也有差异，取平均将会使结果受到“坏”的观测值的干扰，而不如其中的优良值。但是他们忽略了一个问题，那就是并不是总能够有足够的根据去鉴定其观测值的优劣，当这种情况发生时，应如何利用和选择观测结果呢？这就是辛普森工作的出发点[19]。

辛普森重新提出了与伽利略相同的观点，并把注意力放在观测误差上。他假定在没有系统误差的情况下，随机误差有一个由具体条件所限定的界限，在这界限内它依其与 0 距离的增大而递减。这样，辛普森就排除了真值这个未知参量而使问题提法得以简化。

设观测量的真值为 θ，n 次（独立同分布）观测值为 $x_1,\cdots,x_n$，则各次测量的误差为 $e_i = x_i - \theta$，$1 \le i \le n$。若用 x_i 的平均值 $\bar{X}$ 去估计 θ，则其误差为 e_i 的平均值 $\bar{e}$。辛普森想要证明的是：$|\bar{e}|$ 取小值的机会比 e_i 大，即：

$$P\left(|\bar{e}| \le c\right) \ge P\left(|e_i| \le c\right), c > 0 \tag{2.1}$$

如图 1 所示，辛普森首先假定误差只能取 $0, \pm1, \cdots, \pm5$ 这 11 个值，且取这些值的概率，在 0

处最大，然后在两边按比例下降，直到±6处为 0，即

$$P\left(e_i = j\right) = \left(6 - \left|j\right|\right) r, j = 0, \pm 1, \cdots, \pm 5 .$$

其中 $r = \frac{1}{36}$ 。

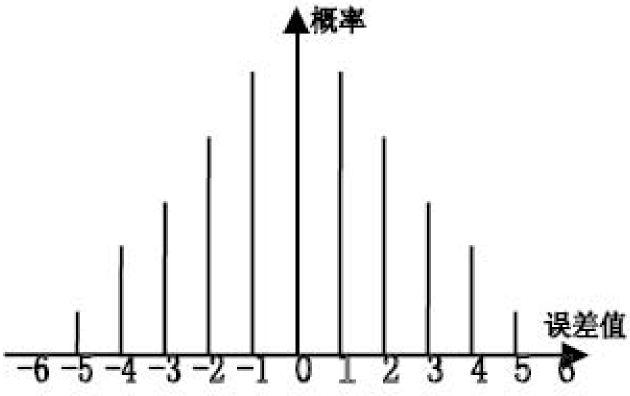

图 1 概率分布

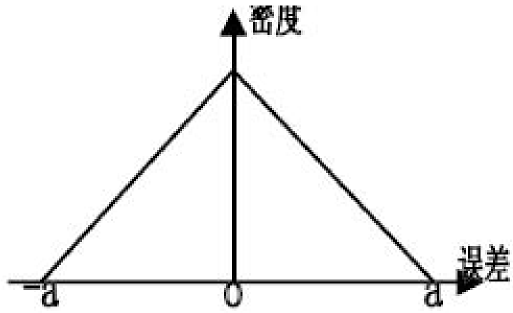

图 2 三角形分布

辛普森取 n=6，通过计算验证了(2.1)。接着他利用母函数的思想，把误差和$\sum_{i=1}^{6} e_i$的分布看作是12个形如

$$P\left(\xi = \frac{i}{2}\right) = \frac{1}{6}, i = \pm 1, \pm 3, \pm 5$$

的均匀分布的叠加，而$\sum_{i=1}^{6} e_i = j$的概率就可以通过二项式展开来求得，这被认为是第一次在一个特定情况下从概率论的角度严格证明了算术平均的优良性。

之后，辛普森把图 1 横轴上的分点加密，并指出其极限形式是一个连续的三角形分布（如图 2 所示）。同上述方法一样，辛普森算出了$\sum_{i=1}^{n} e_i = j$的分布形式（是$2n$个均匀分布$R\left(-\frac{a}{2}, \frac{a}{2}\right)$的叠加），即现在我们所熟知的独立均匀分布和的密度公式。

但是，辛普森选择这样一个特例，显然是出于计算上可能性的考虑。当他经过计算在这一特例上证实了(2.1)时，他推测，这个结果对任何符合上述性质的误差分布（对称，随$|x|$增大而下降）都会成立，但是实践证明这种推测是错误的，后来的柯西分布就是一个反例[20]。由此可以看出辛普森所做的工作，实际上并未触及建立一般的误差概率理论的问题。他只是在假设误差（假定为独立同分布）满足某种特定分布的前提下，去计算平均误差的

分布，从而证明在某种概率的意义上，平均误差小于个别误差。

随后拉格朗日沿袭辛普森的研究方法，在1776年发表的一篇论文①中，考察了其他一些离散情况及个别的连续情况，如误差有矩形分布的情形，但他没有做更深入的研究，不久就转向了别的研究领域。至此，我们更有理由相信，狄莫弗关于正态曲线的声明没有引起任何的注意。在数学史上不乏这样的例子，有时候一个相当成熟的工具就摆在人们的面前，但往往由于缺乏发现的眼睛和创新性观点的提出，使得它与人们失之交臂，这不能不使人感到惋惜。

而对此问题做长时间深入研究的，是大数学家拉普拉斯。

2.2 误差论的创立

1809年，德国大数学家高斯，在研究测量误差的概率分布时，提出正态误差理论，使狄莫弗在约 70 年前引进的那个函数首次取得了“概率分布”的身份，为赢得人们的重视和普遍应用提供了可能。与此同时，与正态分布相关的一系列概率理论的完善，正态分布从天文学向社会学、生物学领域渗透的理论基础，概率论与统计学的交融发展...... 这一切都与一个伟大的名字——拉普拉斯密不可分，他对统计学的早期发展做出了他人无法比拟的贡献。

2.2.1 拉普拉斯的成就与初步尝试

拉普拉斯是法国天文学家、物理学家、数学家和统计学家。他才华横溢，著作如林，十分喜欢用归纳和类比的研究方法，是一位分析学大师。在概率论史上，拉普拉斯被认为是古典概率论的集大成者，他的代表性著作《概率分析论》(1812年）汇集了40年以来概率论的进展以及拉普拉斯自己在这方面的发现，他运用微积分、微分方程、差分方程、特殊函数等分析方法，对概率论的基本理论作了系统性整理，为概率论走向新的历史阶段起到了承上启下的作用。有学者称，它在19世纪概率论发展中的作用可以与欧几里得的《几何原本》在几何学中的作用相媲美[21]。他秉承了自 17 世纪雅各布·伯努利开始的对于概率论的传统理解，即概率不仅可以描述人类理性而且也可以将理性推理系统化，从而把概率论应用于当时如天文地理、人口统计、赌博输赢、人寿保险、法庭判决等各个领域中去。

在《概率分析论》的第二卷第四章中，拉普拉斯讨论了误差论问题，这是他 30 年来有关误差方面工作的总结。他提出了观测值“真值”估计的多种观点；比较了多种估计值的（方差）精度；并给出了中心极限定理的推导提纲。拉普拉斯叙述了一种极其多产的思

① Mémoire sur l’utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations.

想：观测误差是大量独立基本误差共同作用的结果，如果这些误差一致小，则一般来说，观测误差将具有正态分布的形式。利用这种思想，他得到了均值绝对误差，均值平方误差的大样本分布，解决了线性模型中的许多参数估计问题[22]。他也提出了最小二乘法的原理，即不管误差的频率定律（一种依赖于相对频率k的定律）是什么，必须说明这种方法（最小二乘法）是应当优先选择的。这部分内容受到了众多概率学者和数学史家的推崇和赞扬：德摩根（De Morgan）说它是“我们已经遇到的最困难的数学工作”，Ellis 说①，“必须承认很少有数学研究比《概率分析论》的第四章更引人注目”[23]。

然而事实上，拉普拉斯在误差论的这些成就也是逐步完成的，他同样经历了众多的尝试与失败，进行了理论上的反复思考。我们首先介绍他在高斯推导正态误差分布之前的工作。

与辛普森、拉格朗日等人不同，拉普拉斯一针见血，直接涉及误差论的基本问题，即应取怎样的分布为误差分布，以及在决定了误差分布以后，如何根据未知量的多次测量结果去估计它。

1772 年，对于第一个问题，拉普拉斯给出了一种符合当时科学界潮流的想当然的推理：因为随着x的递增，$f(x)$和它的下降率即$-f'(x)$都呈下降趋势（因为曲线$f(x)$越来越平缓），于是他假定：$-f'(x)=mf(x)$，其中$x\geq 0$，$m>0$为常数，解此微分方程就得到$f(x)=ce^{-mx}$，其中$c>0$为常数。再根据$\int_{-\infty}^{\infty}f(x)dx=1$，就有$c=\frac{m}{2}$，于是得到

$$f(x)=\frac{m}{2}e^{-mx}, x>0 \tag{2.2}$$

这样拉普拉斯就给出了误差分布密度②。

接下来，拉普拉斯就着手解决通过未知量θ的观测值$x_1,\ldots,x_n$去估计它的问题。拉普拉斯基于“不充分推理”的原则：即如果一个问题中存在若干个不同的原因$A_1,\ldots,A_n$，则在没有理由认为其中哪一个更有优势时，其先验概率应各取为$\frac{1}{n}$，若以E记在这种原因下可能产生的事件（如在某参数值下观察到的样本），则有$\frac{P(A_i|E)}{P(E|A_i)}$与$i$无关。由此得到了后验分布

① “On the Method of Least Squares,”Cambridge Philosophical Transactions Ⅷ,PartⅠ（1844）,204—240.

② Recherches sur le calcul intégral aux différences infiniment petites,& aux différences finies.Misc.Taurinensia 4,273—345.

$$f\left(\theta \middle| x_1,\ldots,x_n\right)=\frac{f\left(x_1-\theta\right)f\left(x_2-\theta\right)\cdots f\left(x_n-\theta\right)}{\int_{-\infty}^{\infty}\left(\prod_{i=1}^{n}f\left(x_i-\psi\right)\right)d\psi}.$$

为估计θ，拉普拉斯提出了“均概原则”：

$$\int_{-\infty}^{\hat{\theta}}f\left(\psi \middle| X_1,\ldots,X_n\right)d\psi=\int_{\hat{\theta}}^{\infty}f\left(\psi \middle| X_1,\ldots,X_n\right)d\psi=\frac{1}{2}$$

和“绝对平均误差最小原则”：

$$M\left(\hat{\theta}\right)=\min_{\theta}M\left(\theta\right)=\min_{\theta}\int_{-\infty}^{\infty}\left|\psi-\theta\right|f\left(\psi \middle| X_1,\ldots,X_n\right)d\psi.$$

并发现这两种原则殊途同归，可得到相同的估计值。但是当他把(2.2)式代入计算时，却遇到了计算上的麻烦，即使对$n=3$的简单情况，也得到一个极其复杂的方程。

这项工作，拉普拉斯一直持续到 1774 年[24]。他给出的误差分布密度在误差论中也没有起到什么作用，只是以拉普拉斯分布的名称流传下来，有时也被称为重指数分布。到了 1777 年，他从某种奇特的考虑出发，又提出以$f\left(x\right)=\frac{1}{2a}\ln\frac{x}{a}$为误差密度，终究也是竹篮打水一场空①。不过，在这篇论文中他计算出$\int_0^{\infty}e^{-t^2}dt$，这似乎承认了误差定律方程的存在性，并且被认为是继狄莫弗之后对它的第一次声明[25]。1783 年，拉普拉斯建议由于$\int_0^{\infty}e^{-t^2}dt$在演算中经常会被碰到，于是制作一个关于此积分值的表格将会极其有用，他还给出了做这种计算的一个近似公式②。至此可以说，寻找误差分布的问题，依旧进展甚微。现在轮到高斯出场，出人意料的是，他以极其简单的手法，给了这个问题一个完满的解决。

2.2.2 高斯的创新性思想与相关理论

素有“数学王子”美称的高斯，是德国数学家、天文学家和物理学家，他的研究，遍及纯粹数学和应用数学的各个领域，并开辟了许多新的领域。高斯区别于 18 世纪数学家的主要特征，是对工作精益求精。他曾说：“宁可发表少，但发表的东西是成熟的成果。”美国著名数学家贝尔(E. T. Bell)曾这样批评高斯：在他死后，人们才知道他早就预见到一些十九世纪的数学，如果他能把所知道的一些东西泄漏，很可能现在数学早比目前还要先进半个世纪或更多的时间；阿贝尔(Abel)和雅可比(Jacobi)可以从高斯所停留的地方开始工

① Recherches sur le calcul intégral et sur le systéme du monde.Mém.Acad.R.Sci.Paris,pp.267—376(1772),OC8,369—477.

②“Suite du Mémoire sur les approximations des Formules qui sont fonctions de très-grands Nombres,”Histoire de l'Académie Royale des Sciences, Année MDCCLXXXIII(1786),p.434.

作，而不是把他们最好的努力花在发现高斯早在他们出生时就知道的东西；而那些非欧几何学的创造者，也可以把他们的天才用到其他方面去[26]。这就是高斯，其实在概率论的研究方面，又何尝不是如此呢？

高斯在天文学和误差论上都受到拉普拉斯的影响。他有关概率的工作分散在大量拉丁文和德文论文中，它们并没有像拉普拉斯的《概率分析论》那样被整理成册，也没有像拉普拉斯的工作那样停留在非常宽泛的原则上，为概率的数学研究提供详尽的理论基础。由于从 1807 年到 1855 年逝世，他一直担任格丁根天文台台长，因此他非常关心天文学的实践和理论问题，尤其对那些处理各种观测结果的问题比较感兴趣，由此高斯独创了观测值的误差理论，并发展了一套相关的知识体系[27]。

1809 年，高斯发表了数学和天体力学专著《绕日天体运动的理论》（Theoria motus corporum coelestium）。它共有两册，第一册包含了微分方程、圆椎截痕和椭圆轨道，第二册展示了如何估计行星的轨道问题。在参考拉普拉斯的基础上，高斯在第二册的第III部分中利用概率定律，给出了首先被勒让德在 1805 年发表的最小二乘法的第一次说明，这部分内容与从任意数目的观测值中来确定行星的轨迹有关[22]。在此书末尾，他写了一节有关“数据结合”（date combination）的问题，实际涉及的就是这个误差分布的确定问题。

高斯最初采取和拉普拉斯同样的做法，设真值为θ，n个独立测量值为$x_1,\dots x_n$，把后者的概率取为

$$L(\theta)=L(\theta;x_1\dots x_n)=f(x_1-\theta)\dots f(x_n-\theta).$$

其中f为待定的误差密度函数[28]。

但再往下进行时，高斯提出了两个创新性想法：

1、他没有采取“不充分推理”的原则，而是直接把使得$L\left(\hat{\theta}\right)=\max\limits_{\theta}L(\theta)$成立的$\hat{\theta}$，作为$\theta$的估计。

2、他把问题倒过来，先承认算术平均值$\bar{x}$是应取的估计，然后去寻找误差密度函数以迎合这一点。

最终高斯证明：只有在

$$f(x)=\frac{1}{\sqrt{2\pi}h}e^{-\frac{x^2}{2h^2}} \tag{2.3}$$

时才能成立，其中 $h>0$ 是常数，这就是正态分布 $N\left(0,h^2\right)$。

尤其令人鼓舞的是，高斯连同拉普拉斯、勒让德等人关于中心极限定理、最小二乘法等理论的工作与正态分布理论的完善相辅相成，互为促进，激发了测地学家、天文学家、物理学家等众多学者的兴奋点，他们纷纷把目光聚集过来，审视并完善这一理论，对正态分布的广泛应用奠定了良好的理论基础。

最小二乘法实际上是考虑误差在整体上的平衡，让它均匀地分布于各方程内，从而取使得

$$\sum_{i=1}^{n}\left(x_{0i}+x_{1i}\theta_1+\cdots+x_{ki}\theta_k\right)^2 \tag{2.4}$$

最小的 $\theta_1,\ldots,\theta_k$ 来估计待定参数，这对天文学、测地学中经常面临的数据分析问题具有重要的解决意义[29]。勒让德是法国大数学家，在数学的许多领域都有重大贡献，他曾授命参加量测经过巴黎的子午线长的工作，这为他发现最小二乘法提供了现实模型。

1805 年，勒让德发表了一本题为《计算彗星轨道的新方法》的著作，在其附录中他描述了最小二乘法的思想、具体作法及三方面优点：

第一，最小二乘法使误差平方和达到最小，并在各方程的误差之间建立了一种平衡，从而防止了某一极端误差（对决定参数的估计值）取得支配性地位。

第二，在具体计算上，为实现(2.4)式最小只需对各 θ_i 求偏导，从而求解新的只涉及加、乘运算的线性方程组即可，并且如果事后打算抛弃某些观测值，或者增加新的观测值，对于新方程组的修改也易完成。

第三，当 $k=1$ 时，利用最小二乘法会得到测量值的算术平均值作为 θ 的估计值，并且如果观测值全部严格符合某一线性方程，则它必是最小二乘法的解[30]。

这些观念一经点破，我们会感到理所当然，可见科研中观念上的革新是如何的不易，在没有发现之前，梅耶、波斯科维奇、欧拉等顶级学者①努力了几十年也无功而返。那么高斯对最小二乘法又做了哪些贡献，它究竟引发了怎样的争论，以及他比勒让德又有怎样的突破呢？这是我们接下来要阐述的内容。

高斯对最小二乘法的最大贡献在于：他利用已有原则，结合误差密度（2.3），得到 $(e_1\cdots e_n)$ 的概率为

① Theory of Errors of Observations, 2d ed.（1875）,15.

$$\left(\sqrt{2\pi}h\right)^{-n}\exp\left\{-\frac{1}{2h^2}\sum_{i=1}^{n}\left(x_{0i}+x_{1i}\theta_1+\cdots+x_{ki}\theta_k\right)^2\right\}.$$

其中$e_i=x_{0i}+x_{1i}\theta_1+\cdots+x_{ki}\theta_k, 1\le i\le n$可视为误差，$\left(x_{0i},\cdots,x_{ki}\right), i=1,\cdots,n$是观测数据。为使此式最大，必须取使得

$$\sum_{i=1}^{n}\left(x_{0i}+x_{1i}\theta_1+\cdots+x_{ki}\theta_k\right)^2$$

达到最小的$\hat{\theta}_1,\cdots,\hat{\theta}_k$，这样就得到了与待定系数$h$无关的$\theta_1,\cdots,\theta_k$的最小二乘估计[28]。

此外，他还考虑了最小二乘法的误差分析问题。他把通过对各θ_i求偏导得到的方程组称为正则方程组，并设计了一种消元法来求解，被称为高斯消元法，得到的解就是$\theta_1,\cdots,\theta_k$的线性无偏估计，高斯断言：在这个解的线性函数的一切线性无偏估计类中，唯有其最小二乘估计的方差达到最小[27]。这是最小二乘法理论中最重要的理论结果，它从统计学的角度回答了最小二乘法在缩小误差这个根本点上的优势所在，使得在勒让德那里只是被作为一个处理测量数据的代数方法逐渐渗透到统计数据分析的领域，与统计学的发展产生了千丝万缕的联系。一些学者甚至认为，最小二乘法之于统计学，犹如微积分之于数学，其影响是深远的，这一点在 20 世纪哥塞特、费歇尔等人发展了正态小样本理论后，尤其看得明显[13]。

拉普拉斯虽然在得到误差正态密度这一环节上技不如高斯，但这绝不能作为衡量二人贡献大小的唯一标准。事实上，拉普拉斯在最小二乘法、中心极限定理，包括正态分布方面一直在不遗余力地进行着全方位的考虑。

早在 1780 年左右，拉普拉斯就推广了狄莫弗的结果，得到了中心极限定理较一般的形式[31]。我们知道，测量误差是由诸多因素形成，每种因素影响都不大，按照中心极限定理，其分布近似于正态是势所必然。可惜的是，拉普拉斯未能把这一成果用到确定误差分布的问题上来。但是当他得知高斯的工作后，马上就将其与他发现的中心极限定理联系起来。1810 年，他在即将发表的一篇文章①中为此加了一点补充：如果误差可看成是许多量的叠加，则根据他的中心极限定理，误差理应有正态分布。这是历史上第一次提到所谓的“元误差学说”——误差是由大量的、由种种原因产生的元误差叠加而成的[32]。

① Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilities. Mém.Acad.R.Sci, Paris, pp.353—415, OC12, 301—345.

这一点对于给正态误差论一个更自然合理、更令人信服的解释有重大意义。因为，高斯以算术平均是优良的作为出发点，推出误差必须服从正态分布；反过来，由后一结论又推出算术平均及最小二乘估计的优良性，故必须认定这二者之一（算术平均的优良性，误差的正态性）为出发点。但算术平均到底并没有自行成立的理由，以它作为理论中一个预设的出发点，终觉有其不足之处。拉普拉斯的理论把这断裂的一环连接起来，使之成为一个和谐的整体，为正态分布之后在大样本理论中的基础性地位奠定了基础。

此外，高斯对最小二乘法的误差分析，最初也源自拉普拉斯的考虑。1811 年，拉普拉斯考虑了可测量值为一维时参数的无偏估计，证明了当误差为正态时，利用方差尽量小原则和施瓦茨不等式，可知或然误差最小的估计是最小二乘估计，以及若误差不服从正态分布但 n 很大时，可利用中心极限定理近似地得出相同结论①。由此拉普拉斯表明当样本规模很大时，线性回归系数的一类线性无偏估计值是渐近正态分布的。当他得知了高斯的误差正态分布后，又从期望的角度，利用中心极限定理给出了论证[32]。随后，高斯注意到，只有在正态或至少是对称的情况下，或然误差才有意义，若只要求估计的方差最小，则像误差为正态或 n 很大这些限制就都可以去掉[28]。于是他给出了最一般性的结论，并得出了最小二乘估计的方差表达式，即我们现在称的高斯——马尔科夫定理。

正是由于拉普拉斯、勒让德以及高斯之间这种错综复杂的理论联系和工作时间的交叉性，导致三者发生了与牛顿、莱布尼兹的微积分发明优先权之争声名相若的争论，引发了很多学者的研究，然而直到现在也没有做出判然的结论。其实正如我们之后所看到的那样，19 世纪的统计研究经常都是独立地进行，研究者之间彼此不很关注其他人的成就，再加上高斯精益求精的研究风格，自 1795 年就开始使用的最小二乘法没有被他发表出来，也不足为奇；而勒让德 1805 年首次发表他的成果时，对大数学家高斯的研究工作一无所知，也就理所当然；至于拉普拉斯，他发现最小二乘法确实在二人之后，只是他最先得到了狄莫弗中心极限定理的一般形式，并在此基础上考虑了最小二乘法的误差估计问题，这对最小二乘法向统计学的渗透起到了关键性作用，同时也体现了拉普拉斯作为一位统计学家的过人之处。

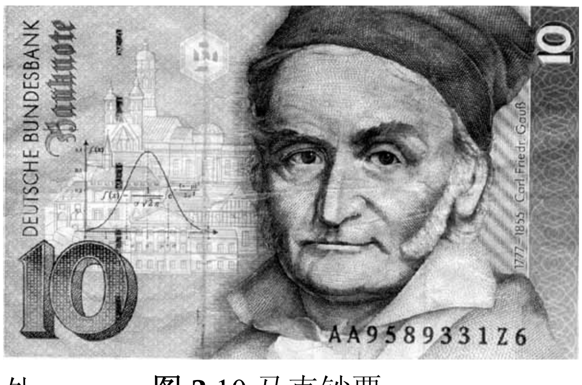

图 3 10 马克钞票

① Du milieu qu'il faut choisir entre les résultats d'un grand nombre d'observations, Connaiss, Temps, pp.213—223, OC13, 78.

总之，高斯对误差正态分布的提出，以及与之相伴的最小二乘法和中心极限定理一般形式的诞生，对后世的影响极大，这也使正态分布同时有了“高斯分布”的名称。如图 3 所示，现今德国 10 马克的印有高斯头像的钞票，其上还印有正态分布 $N\left(\mu,\sigma^2\right)$ 的密度曲线，这无疑传达了一种观念：在高斯的一切科学贡献中，对人类文明影响最大者就是正态分布；这也充分预示了正态分布之后在概率论和统计学中所取得的重要性地位。

值得注意的是，1808 年美国人艾德里安（Robert Adrain, 1775—1843）在所著《观测误差的概率研究》一文中，也为误差法则提出两项力证[33]。但是他的工作似乎没有影响这个问题的发展。

2.2.3 艾德里安的工作及其评论

艾德里安第一个推导的背景①是勘测中一个设奖的问题，即通过测量边和角去得到一块多角形土地的面积。因为测量误差，多角形不封闭，因此在面积求出之前不得不先调整观测值。他首先把一条已知长度为 γ 的直线分成 n 条未知长度的直线 $\mu_1,\ldots,\mu_n,\mu_i>0$，就有 $\sum\mu_i=\gamma$。设 μ_i 的测量值是 x_i，则有误差 $\varepsilon_i=x_i-\mu_i$，则观测之后总体误差为 $\sum\varepsilon_i=\sum x_i-\gamma=c$。这样问题就转化为寻找估计值 $\hat{\mu}_i$，使得 $\sum\hat{\mu}_i=\gamma$。

艾德里安首先考虑了 $n=2$ 的情况。他利用一个“显然”的准则：

$$\frac{\varepsilon_1}{\mu_1}=\frac{\varepsilon_2}{\mu_2} \tag{2.5}$$

用估计值代替真值，就得到 $\hat{\mu}_1=\left(\hat{\mu}_1+\hat{\mu}_2\right)\frac{x_1}{x_1+x_2}$。令误差分布在这两种情况下有相同的函数形式使得

$$p\left(\varepsilon_1,\varepsilon_2\right)=f\left(\varepsilon_1,\mu_1\right)f\left(\varepsilon_2,\mu_2\right) \tag{2.6}$$

艾德里安指出在条件 $\varepsilon_1+\varepsilon_2=c$ 下，f 可由使此密度最大化而得到。于是有方程

$$\frac{d\ln f\left(\varepsilon_1,\mu_1\right)}{d\varepsilon_1}d\varepsilon_1+\frac{d\ln f\left(\varepsilon_2,\mu_2\right)}{d\varepsilon_2}d\varepsilon_2=0$$

和 $d\varepsilon_1+d\varepsilon_2=0$。令

① R.Adrain, Research concerning the probabilities of the errors which happen in making observations, Analyst1, 93—109, Reprinted in Stigler, 1980.

$$\frac{d\ln f(\varepsilon,\mu)}{d\varepsilon}=m\frac{\varepsilon}{\mu} \tag{2.7}$$

其中m表示一个任意常量，就得到

$$f(e,\mu)=ae^{\frac{m\varepsilon^2}{2\mu}} \tag{2.8}$$

其中a是一个取决于μ的积分常量，所以得到ε服从$N\left(0,\frac{\mu}{m}\right)$。这样，$n$个观测值的联合密度就变成

$$a_1\ldots a_n\exp\sum_{i=1}^{n}\frac{m\varepsilon_i^2}{2\mu_i}.$$

它应该是在条件$\sum\varepsilon_i=c$下的一个最大值。艾德里安指出m必须是负的，令$m=-2$，就得到$f(\varepsilon,\mu)=a(\mu)e^{\frac{\varepsilon^2}{\mu}}$。他没有指出$a(\mu)=(\pi\mu)^{-1/2}$，但提到："表现概率曲线本质的最简单的方程形式"是$\exp(-\varepsilon)^2$[34]。

为了找到在条件$\sum\varepsilon_i=c$下的最大联合误差密度，须把$\sum\varepsilon_i^2/\mu_i$最小化，这导致方程

$$\varepsilon_i=\frac{\mu_i}{\sum\mu_i}c,i=1,\ldots n \tag{2.9}$$

令$\varepsilon_i=x_i-\mu_i$并且解μ_i，自然就得到

$$\hat{\mu}_i=\frac{\gamma}{\gamma+c}x_i=\gamma\frac{x_i}{\sum x_i}.$$

这就是艾德里安的最小二乘解。

对此，1872年Glaisher给出了三点评论[35]：

1、假设(2.5)是不合理的，因为在许多测量过程中，误差都是独立于观测值或者和μ的函数是不成比例的。

2、与(2.7)相符合的完全解是

$$\frac{d\ln f(\varepsilon,\mu)}{d\varepsilon}=g\left(\frac{\varepsilon}{\mu}\right).$$

这里g是一个任意（可积）函数。

3、艾德里安对于密度 $f(\varepsilon,\mu)$ 使用术语“概率”是不恰当的。

对此笔者认为这里需补充说明三点：

1、没有假设 (2.5)，艾德里安的证明也可以导出正态分布。

2、艾德里安把分母上的 μ_i 看作已知，并通过把 $\sum \varepsilon_i^2/\mu_i$ 关于 ε_i 最小化来导出方程 (2.7) 是不合理的，因为 $\varepsilon_i = x_i - \mu_i$ 且 μ_i 是未知的。

3、他的推导过程是合理的，并且可以被看作是这个问题的一个迭代解法的第一步。

最后 Glaisher 总结说：“艾德里安不得要领的推导使人们相信，他首次注意到了使用最小二乘法来处理此方程的便利性，并且随后他努力去利用实例来论证这一点。但是，无论他的论据会使人想到什么，我们必须‘把独立地发现误差定律归功于艾德里安’”①，这一点已普遍被其它评论家所接受。

对艾德里安和高斯的推导进行比较，可以看出其中存在着如下异同。相同点在于，他们都没有研究估计值的分布，甚至没有提到算术平均值的不确定性，并且都采用了极大似然原则。不同之处有两点：

1、艾德里安是寻找一个导致具体化估计规则的密度，而高斯是寻找一个使得观测值的算术平均值作为估计的密度。

2、高斯假设误差分布有形式 $f(\varepsilon)$，这比艾德里安的模型 $f(\varepsilon,\mu)$ 更简单。

在完成了第一个推导之后，艾德里安指出“可以通过一种不同的研究方法来证实我所说的②”。他再一次以勘测中一段长度和一个方向的测量为例，把这两种误差表示成以真值为原点的直角坐标系中的一点。他假设这两个误差分布在恰当地定标之后有相同的函数形式并且相互独立。此外，在误差分布是单峰的、对称的一般性假设基础上，他通过 (2.6) 给出了误差的联合分布。他证明中的新思想就是在 $(\varepsilon_1,\varepsilon_2)$ 平面上去寻找一条或然率相同的曲线，并证明这条曲线应满足三个性质：

1、它必须位于一个中心在原点，四条边分别与坐标轴平行的正方形的内部和一个内接于第一个正方形并且四角分别位于其四边中点的正方形的外部。

① In Glaisher's paper “On the Law of Facility of Errors of Observations and on the Method of Least Squares,” Memoirs of the Royal Astronomical Society, XXXIX(1872),75—124.

② The second proof is reproduced in Merriman's “List of Writings relating to the Method of Least Squares”.

2、这条曲线必须是连续的、对称的以及关于坐标轴是凸的。

3、它必须是最可能简单的[35]。

由此，他推测出这条曲线应该是一个圆周，并指出如果误差是不同精度的，这条曲线就变成一个椭圆。这样艾德里安就要求，联合密度（2.6）当 $\varepsilon_1^2+\varepsilon_2^2=r^2,(r\geq 0)$ 时应该是常数。接着有

$$\frac{d\ln f\left(\varepsilon_1,\mu_1\right)}{d\varepsilon_1}d\varepsilon_1+\frac{d\ln f\left(\varepsilon_2,\mu_2\right)}{d\varepsilon_2}d\varepsilon_2=0$$

和 $\varepsilon_1 d\varepsilon_1+\varepsilon_2 d\varepsilon_2=0$，因此有

$$\frac{d\ln f\left(\varepsilon,\mu\right)}{d\varepsilon}=k\varepsilon$$

和 $f\left(\varepsilon,\mu\right)=a\left(\mu\right)e^{k\varepsilon^2/2}$。令 $k=m/\mu$，他就得到了先前的结果(2.8)。他指出在只有一个未知参数的情况下，误差分布才和 $\exp\left(-\varepsilon^2\right)$ 是成比例的，使得未知的 μ 可以通过最小化 $\sum\left(x_i-\mu\right)^2$ 来估计，从而推导出算术平均值。

艾德里安的第二个推导建立在密度形式简单化要求的基础之上，同样受到了 Coolidge 的批评："在我看来，这个证明无疑比另一个更弱，'这个曲线必须是最简单'的要求是不恰当的，如果存在一个误差定律，我们没有理由必须给出最简单的曲线。"[23]而笔者认为，艾德里安利用简单的数学性质寻找到一个误差密度，它导致了被实干者认为是明智的并且实用的估计值，并且在一些可能性中，他选择了明显满足这些要求的那个，这表现了他选择 $\exp\left(-\varepsilon^2\right)$ 而不像拉普拉斯那样选择 $\exp\left(-|\varepsilon|\right)$ 的直觉和一般判断，他的工作可认为是二元正态分布研究的雏形，因此在历史上对他的工作应给予更多的肯定。

虽然如今看来艾德里安用的方法颇有不正确之处，但他终究推导出了最小二乘法以及单变量和双变量的正态分布，产生出与高斯异曲同工的思想，应该说是难能可贵的。但是他的工作并没有引起人们过多的关注，也没有影响到误差论和相关问题的发展，究其原因应该也是多方面的。笔者认为，由于艾德里安研究的孤立性，使得他并没有把其结果同误差论紧密地结合起来，且其证明同高斯相比也没有特别的优势所在，因此它只能成为历史长河中溅起的一朵浪花，并没有像高斯的误差理论那样形成波澜壮阔的气候。随着高斯对正态分布的发明，统计学开始了一个新的时代。自然科学界首次有了一个可导致算术平均值作为真值估计的双参数分布，在限制测量精度的条件下，对尺度参数有了一个更易理解

的解释，并得到了关于观测值线性组合的数学表达式的一个简单标准，对观测值的分布给出了一个很好的拟合[36]。这样，在狄莫弗时代被忽视的正态分布，就获得重生，这一概率模型及其相关理论逐渐成熟，在误差论领域引起了广泛关注，并带动了一系列新理论的产生，其中基本误差假设就是一个典型的范例。

2.3 基本误差假设

高斯在一个“可靠性众说不一”的假设下通过解一个函数方程推导出了正态概率密度，他的目的是想找到一个观测值分布，使得算术平均值可以作为真值的自然估计。1810 年，拉普拉斯通过对一组有任意分布的观测值的算术平均的分布进行大样本逼近，得到了相同的密度形式，这是理论上得到的结论。任何理论成果，终究要经过实践的检验，并提炼出新的结论，在这方面做出努力的是几位德国学者。

1818 年，德国贝塞尔（Friedrich Wilhelm Bessel,1784—1846）通过对一个分组的经验分布和一个理论分布进行比较，指出经验误差分布和正态分布之间有很好的一致性，从而表明天文学观测值的误差几乎都是正态分布的①。1834 年，德国天文学家恩克（Johann Franz Encke,1791—1865）对贝塞尔的例子重新分析，也得到了一致的结论。但这些都是经验性的，从数学的严密性要求来看，并不能得到令人信服的说法，为了完成正态分布作为误差理论基础的描述，有一个问题必须被回答，那就是为什么观测误差应该是正态分布的？前面我们已经谈到，拉普拉斯首次给出了所谓的“元误差学说”，但当时这只是一篇用于其它目的的文章的补充，提法过于粗糙，沿着这个思想努力的还有其它一些作者。

1837 年，德国工程师海根（G. Hagen,1797—1884）在关于误差论和勘测中应用最小二乘法的初等教科书中正式提出了元误差学说②。他指出，实际的观测误差是由测量仪器、环境、天气条件以及观测者等许多因素造成的大量相互独立的微小误差构成的，假定观测误差是这些微小误差的代数和，且每个误差只取 $\pm a$ 两值，概率都是 $\frac{1}{2}$，由此出发，按狄莫弗的中心极限定理，立刻就得出误差（近似地）服从正态分布。这个假设被叫做海根的基本误差假设。作为一个物理解释很显然它是错的，没有测量或观测的过程可以以这种简单的形式被描述。但是海根选择这样简单的陈述也许是由于教学法上的原因，因为观测误差的正态分布遵循对称二项式的极限形式。在该著作的不同地方，海根提到了勒让德、高斯、拉普拉斯和恩克，但是他没有指出观测误差的正态性是基于基本误差假设和中心极限定理

① Fundamenta astronomiae pro anno MDCCLV, Regiomonti.
② Grundzüge der Wahrscheinlichkeits—Rechnung.Dümmler, Berlin.

的结合。这种观点由贝塞尔在 1838 年给出，他系统地阐明了基本误差的一般性假设，并给出了中心极限定理一个新的证明①。

贝塞尔指出他的目的是去解释观测到的误差由哪种方式从它们的成因中产生出来。作为一个实验观察者，他详细地描述了一个天文观测值误差的 13 种不同成因，并补充说还有更多。他写道："因此我得到了这个令人惊奇的结果，即许多独立成因的相同量阶的误差通过它们的联合，产生了总体误差，它们的概率和按照算术平均假设以及最小二乘法得到的概率渐近相同。"贝塞尔假设基本误差是关于有限矩的对称分布的随机变量，并推导出它们的和的分布的渐近展开。他指出，只要基本误差互相独立，并且没有一个基本误差的方差对误差和的方差起支配作用，正态分布就可以被看作是实际误差的分布。

基本误差假设在 19 世纪被自然科学家狂热地接受。1872 年，Glaisher 在他的著名论文中写到：如果我们把每一个实际误差都看成是由大量来自不同独立成因的误差的线性组合，我们将得到同样的定律 $\exp\left(-h^2x^2\right)$，这种关于误差的自然观的看法似乎是最自然和真实的[23]。19 世纪末，Glaisher 的乐观主义被泊松（Poisson,1781—1840）的怀疑主义所取代，泊松在完善拉普拉斯证明的基础上，通过反例指出：当矩条件不满足时，存在其它的极限分布②。这样，就导致正态性不能被假定为一个先决条件，而是对于每一个数据集合必须被检验。

此外，随着误差论的发展，早在狄莫弗和丹尼尔•伯努利时就被自然使用的概念——概差（probable error），在 1815 年被贝塞尔首次正式提出[25]。高斯在此基础上，给出了一个表示误差偏离均值不超过定值的概率表格，使概差逐渐成为误差理论中一个标准术语，并建立起一般性的理论③。1832 年，恩克写了一本关于最小二乘法的书，包括了各种统计量的概差公式和标准误差公式，为误差论的相关计算提供了实用方面的准则。

1838 年英国数学家德摩根（De Morgan,1806—1871）在理论之外做出简化和普及概率论的首次尝试。他在《概率在偶然事件和保险事业中的应用》一文中指出：到目前为止，概率论的研究只是专业数学家的事情，但它也应该允许那些已经承认概率论急待研究，并对数学不能付出太多精力的人们，也能从该理论结果中受益[37]。此书对于人们把概率论应

① Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler.Astron.Nachr.15,369—404,Reprinted in Abhandlungen,2.

② Sur la probabilité des résultats moyens des observations, pp.273—302.

③ In a paper on the determination of the accuracy of observations "Bestimmung der Genauigkeit der Beobachtungen,"Zeitschrift für Astronomie und verwandte Wissenschaften, Ⅰ,185—196.

用到生活保险等领域并获得良好建议具有丰富的实践意义。此外，关于正态曲线他说，“将不断会有场合涉及到它：事实上，利用它几乎对整个高等数学的应用都将被简化”[1]。同年，他又讨论了最小二乘法、定积分的应用以及一些有关证据、奇迹、道德、审判等复杂问题的概率论处理，表明了当时科学研究者思想转变的方向。值得注意的是，在德摩根夫人所写的《德摩根回忆录》中提到一封凯特莱写给德摩根的信，信中感谢他寄了一个表格和一本著作的复写本，并说这给了他许多乐趣[2]。这个表格是否是关于概率积分的表格，现在已无从知晓，但很显然这时二人已经认识，并且很自然让我们引发一个猜测：凯特莱很可能是受到德摩根先生思想的启发，把正态分布从误差论领域中提炼出来，大张旗鼓地引入到描述社会现象中去。这无疑为我们提供了一条思想线索。

应当指出的是，误差论领域在几十年内向前发展，对统计理论的发展并没有产生更多进一步的影响，高斯似乎没有打算去确立一种适用于其它研究领域的原则，正态分布在统计学的理论和应用分支中也还没有找到它的位置。这其中很重要的一个原因在于，在很长一段时期，误差论和统计学被看成是两个不相干的领域，其根本区别在于，前者研究的是对同一对象作重复研究的观测数据，用到的方法是高等数学；而后者处理的则是对同类个体的同一指标（如身高）所做测量得到的相对频率的统计数据。这样，正态分布就陆续进入了发展的双轨道，以上我们介绍了误差论领域的前期情况，而正态分布正式进入统计领域则要依靠凯特莱的威望和高尔顿的工作，这就进入本文第三部分的主旨——正态分布的引入。

3 正态分布向近代统计学的引入（19 世纪中叶至 19 世纪末期）

近代统计学，是指十八世纪中末叶至十九世纪中末叶这一百年期间的统计学，是古典统计学向现代统计学的过渡[6]。古典统计时期的概率论基本上是独立发展的，与统计学的联系多属间接，并缺乏现实应用的有力刺激。虽然当时的学者如雅各布·伯努利、狄莫弗等人曾提出用概率论研究道德、法律和经济现象，但由于此时概率论理论上的零散性和方法的局限性，并没有发挥出其应有的影响力，因此概率论与统计学的互相渗透终究只是海市蜃楼，缺乏理论的深化作为其坚实的基础。

进入近代统计时期，拉普拉斯带着对人类的馈赠适宜的到来了。他首次明确给出了概率的古典定义，系统阐述了概率论的基本定理，证明了中心极限定理，完善了包括最小二乘法在内的观测误差理论，并把概率论有效地应用于人口统计等社会生活领域，他的工作是继承性与开创性并具的。18 世纪的概率论，最终以拉普拉斯的《分析概率论》宣告其研究的最终结果，它不仅实现了概率论方法上的革命，而且系统整理了 18 世纪之前概率论所处理过的所有重要问题，构造了整个概率论的理论体系。包括正态误差定律在内的概率数学理论的进展，在欧洲激发了广泛的兴奋点，有关论文和相关著作如雨后春笋般涌来。德国的贝塞尔、高斯、恩克，法国的勒让德、傅立叶、泊松，英国的德摩根，美国的艾德里安和俄国的契比雪夫等人都做出了自己的贡献[37]。但是正如我们上文所说的，由于误差论和统计领域的不相干性，使得把概率论同社会现象联系起来的尝试仍然没能取得有效地进展；试图用概率论来研究法律、政治、道德、经济等社会问题的人和思想虽一直在跃跃欲试，但却形不成燎原之势，概率论与统计学的有力结合还有待进一步发展。

概率的数学理论，计数、计算和测量机构的产生，现代国家日益增长的关于国内外资源统计信息的需要，被认为是现代统计理论和实践的先驱。十九世纪中叶，在政府统计方面，沿着格兰特开拓的道路前进的人口统计，出现了保险统计、卫生统计或医疗统计等分支；沿着配第指引的方向发展的经济统计，出现了农业统计、工商业统计和物价指数计算方法的研究等[6]。许多统计协会犹如雨后春笋般地建立起来，并与政府统计机关保持密切合作，推动了政府统计的进一步发展；政府统计机关由间接地，单纯地搜集、整理统计资料向连续地，分析地整理统计资料、概括统计规律的方向发展。热衷于探索和揭示统计规律，成为近代统计学区别于古典统计学的显著特点之一[6]。在这种社会氛围的推动下，概率论用于社会现象的研究有了实质性进展，最终这三者在一个伟大的比利时人凯特莱

（Adolphe Quetelet,1796—1874）身上汇聚一点，他不仅是一位数学家、天文学家、人体测量学家，而且是统计学组织委员会的主要发动者、国家官方统计的负责人以及首次国家范围内人口调查的倡导者。他大手笔地将包括大数法则、正态分布和误差法则在内的概率论广泛地用于自然现象和社会现象的研究之中，为国势学、政治算术和概率论三要素的互相交融扫平了道路。随后，统计学逐渐“走进了理论的领域，成为关于实验和观测的科学”[38]，为正态分布的尽情释放提供了更加广阔的历史舞台。

3.1 近代统计学巨星凯特莱

凯特莱出生于比利时的甘德市。应该说，他在世的年代是与近代统计学史相始终的。1834—1848 年是比利时的黄金时代，经济上突飞猛进的发展为他在统计领域做出贡献提供了有利的现实环境。

17 岁时他开始在一所私立学校教学，后由于一篇数学论文，获得甘德大学博士学位，并被聘为布鲁赛尔学院的数学教授，1820 年被选为布鲁赛尔皇家学会会员。1823 年，他建议比利时政府在布鲁塞尔建立天文台，并被派往法国购置天文仪器和学习观测方法。在巴黎，他与拉普拉斯、泊松、傅立叶等人建立了友谊，在此期间，他分别跟随傅立叶、拉普拉斯学习了数学和概率论，并越发的强调把概率论应用到科学研究中去①。1829 年，凯特莱设计了一个普查方案，并举办了第一次全国人口普查，从此以后，他与统计结下了不解之缘。比利时独立后，于 1831 年成立了中央统计局，由斯米茨（E. Smits）任局长，他和凯特莱共同负责公布 1829 年普查结果的工作。1832 年，布鲁塞尔皇家天文台正式建成，凯特莱被任命为台长。1834 年，他又被选为比利时科学院秘书。1841 比利时成立了由凯特莱任终身主席的中央统计委员会，成为其它国家效仿的典范。1853 年在他的倡议下，由比利时政府邀请，在布鲁塞尔召开了第一次国际统计会议。他在统计事务中为国际合作不知疲倦地工作着，成为当时的中心人物[39]。

凯特莱不仅学识渊博、勇于创新，而且善于融会贯通各种科学方法、兼收并蓄各派所长。他二十三岁开始发表文章，且题材广泛，数目繁多，只与统计学有关的著作，就达 60 多种； 他连续几年主编《数学和物理学学报》，邀请欧美各地著名学者撰稿；他是世界上一百多个统计学会的会员，成为当时国际上最活跃的人物之一；他旅行了欧洲的大部分国家，和当时健在的大多数科学家都有私交。凯特莱统计方法广泛的涉猎面与他的研究风格和思维模式有直接联系，他曾在科学史课程上重复声明这样一种思想：“科学变得越进步，

① J.F.W.Herschel,Quetelet on probabilities,Edinburgh Rev.92,1—57,1850.

就越倾向于进入数学领域，这是它们汇聚的中心。我们可以通过计算来判定一门科学所达到的完美程度，并利用这种完美性来展开研究[38]。”他以一种惊人的勤奋坚持不懈地工作着，在理论和实践两方面对统计学做出了重要贡献。

在理论方面，他首次广泛、深入地把自然科学的研究精神和研究方法（实验法、归纳法）带到社会现象的研究中来，用数理方法来研究社会现象，为统计学开辟了广阔的新天地。他指出，社会统计学从性质上说，是一门实质的社会科学；从对象上说，既研究统计规律也研究统计方法；从方法上说，它就是对同性质的社会现象进行大量观察。他尝试用力学原理研究社会现象，用生物学原理研究年龄与身高的关系。尽管其研究存在着某些缺点和错误，但启示后人：统计学的方法论是由先进的科学方法所构成的；只有不断地、有抉择地汲取先进科学的方法来充实方法论，统计学才能得到发展。如果说，数理统计学派是沿着凯特莱“论数学与大量观察的关系”向前发展的，那么，社会统计学派则是按照其“统计基本理论”向前迈进的[19]。

在实践方面，他扩大了国情普查的统计范围、改进了统计调查方法。他主持了比利时几次包括人口、农业和工业的国情普查，是行政目的和研究目的相结合的，成为统计体制上的创举和当时各国效法的样板。在促进国际统计学术交流方面，他主编了《数学和物理学学报》，倡议并积极参加国际统计会议的召开，并成为国际上一百多个学会的成员。他对于促进政府统计工作和国际统计学术交流起到了非常积极的作用。

凯特莱大胆设想、大胆创新的精神在受到称赞的同时也遭到了不可避免的非议。他在《社会物理学》中提出一些新颖的见解，比如强调客观规律、否定自由意志的作用，特别是提出犯罪现象中也存在规律性之后，出现了凯特莱派和非凯特莱派。凯特莱派以丹克瓦特、瓦格纳为代表，他们不仅支持凯特莱的观点，还进一步阐发凯特莱的学说。非凯特莱派则以纳普和鲁米林为代表。纳普反对凯特莱等同社会现象和自然规律，他说：“凯特莱喜欢看到死的自然和活的社会规律之间的类似，其根本错误在于脱离了科学的范围，这与熟睡了的天文学家的梦寐毫无区别”[38]。鲁米林则在其大学就职演说——《社会法则的概念问题》中说：“对于死亡之类的自然现象，固然可以从数学上揭示出一个公式来预测其概率，但是无论如何做出犯罪的预算，它总不能约束个人的力量[39]。”

此外，史学家沃克认为，是凯特莱把统计学这一概念发展成对任何观察科学都适用的一种研究方法。韦斯特加德说：“才能出众的学者凯特莱，把得到的统计资料深入浅出地写成书，把显然枯燥无味的观察结果用引人入胜的笔法予以描述。他的著作显示出这一时代的力量，也显示出这一时代的缺点。因为他急切的把统计考察扩展到可能扩展的广阔范

围中去，就必然会不加批判地接受一些基础很坏的计算结果。”[6]笔者认为，韦斯特加德的评论是中肯的：他一方面指出凯特莱的优点——把统计学通俗化和通用化；同时也点破他的缺点——无批判地根据一些统计资料，概括出一些轻率的结论，由此，我们应该一分为二地来看待凯特莱的工作。不过就正态分布的引入，以及融合误差论和社会统计这一环节来说，凯特莱的思想是开拓性的，举措是前所未及的，体现出了他横扫千军马的一统气概。

3.2 凯特莱对正态曲线的拓展

18 世纪所涉及的统计数据分析问题，主要与二项分布有关，狄莫弗引入的正态分布被人们生生地抛在脑后；到了 19 世纪初，拉普拉斯中心极限定理与高斯正态误差理论的问世，使这种情况逐渐出现转机，正态分布开始崭露头角，发挥其作为统计模型的优势，只是由于观测数据与统计数据不相容性的认识，使得它的影响仍受限于其发源地；而真正把正态分布的舞台拓展出去的当属凯特莱，他首次强调了正态分布的用途，并将以它为基础的统计方法应用到天文学、数学、物理学、生物学、社会统计学及气象学等研究范围，在他的影响下，正态分布获得了普遍认可和广泛应用，以至有些学者认为 19 世纪是正态分布在统计学中占统治地位的时代。那么凯特莱究竟是如何做到这些的呢？

凯特莱与正态曲线有关的工作分为两个方面：一是把误差理论应用到新的领域，二是“平均人”的概念。前者的灵感来源于人口普查，后者则是使统计方法获得广泛应用的理论基础[2]。二者相辅相成，共同构成了凯特莱思想的根基。

1826 年，凯特莱成为比利时国家统计局的地区通信员，并做了很多与人口调查有关的统计工作，其中有一项工作就是利用拉普拉斯的“比例法”来估计比利时的人口总数。拉普拉斯 “比例法”在概念上很简单，是一种根据局部地区的人口调查结果来估计全国人口总数的方法：把全国人口总数与过去一段时间内人口出生总数的比值记为 r ，一段时间内人口出生数 a 可以从有关的登记资料中查出，若知道 r ，则人口总数为 ar [13]。拉普拉斯提出在国内选定若干被认为有代表性的地区，通过实地调查将其人口总数与过去一段时间中人口出生数的比值定出，作为 r 的估计，这种做法现在我们称之为“代表性抽样”。

凯特莱用出生人数及死亡人数来估计 r ，但得到了不理想的结果，并且这种抽样法引起了很多社会学家的反对。他们指出，影响人口出生率和死亡率的因素很多，如居住环境、人口密度、文化水平、职业性质、饮食情况以及生活习惯等，都对此有影响，把这些受到极大数目因素影响的数据放在一起去处理，理论上不合理，也不可能得出有用的结果。于是凯特莱放弃了用抽样法来估计人口数的计划而回到普查。不过，这项工作促使他注意到

了数据的同质性问题。

社会学家们的反对意见主要在于：对不同质的数据进行统计分析没有意义。社会问题与科学实验不同，其数据一般由观察得到，无法控制且经常不了解其异质因素，这样数据的同质性连带其分析结果的解释往往就有了问题，于是社会统计工作者就面临一个如何判断数据同质性的问题。凯特莱大胆地提出：把一批数据是否能很好地拟合一个正态分布，作为判断该批数据同质的一个标准。凯特莱是拉普拉斯的学生，他很熟悉中心极限定理和高斯的正态误差理论[39]。从形式上看，他不过是在已有的基础上向前迈进一步——把高斯发现的测量误差分布的规律推广到其它数据；但在当时这需要突破观念上的很大障碍，因为人们普遍认为，适用于误差的规律未必一定适用于其他数据。

为验证这一想法，凯特莱以 5738 个苏格兰士兵胸围的数据为例，发明了一种基于中心极限定理，用正态曲线拟合数据的方法，这种间接的手段避免了使用微积分或正态概率表格。他首先制造了一个二项分布的表格，作为某正态分布的近似，然后利用这个表格拟合数据，通过分析和计算他指出：如果数据严格符合某正态分布，则各胸围相应的概率，应接近二项分布所得，这可以作为数据与正态符合的一个初步检察。这种方法在原理上不复杂，但实行起来很繁琐，需要相当的计算和一定的技巧，凯特莱在此例中得到了满意的拟合。之后，他注意到由于许多性质都是可以量化的，若用柱形图来表示诸如身高、胸围此类的数据，则很容易看出这些数据与平均值偏差的分布与打靶的弹洞在靶心周围散布的方式类似，其图像很像正态曲线[38]。这样凯特莱就引入了正态曲线，把正态分布从观测误差推广到各种来源的数据，为在社会科学与人文学中使用统计方法迈出了决定性的一步。

之后，凯特莱发展了大量观察法。在他看来，对任何现象的观察都有误差，而任何现象通过大量观察都可以发现规律；前者符合误差法则，后者符合大数法则。他通过将质量、数量、形状、大小等完全相同的黑白两种球置于盒中，然后不断地取出来放进去，发现试验次数越多，黑白二球取出的次数就越接近，从而验证了这两项法则。为了使大量观察法更精密化、科学化，他把概率论、大数法则、误差法则、正态分布等概念和计算方法引进社会统计学，丰富了它的方法论，并且认为这是适用于“任何观察科学的通用方法”[1]。

1835 年，凯特莱发表了巨著《人及其天赋的发展》，出于建立“社会物理学”这样一个大胆的设想，他引进了“平均人”的概念，成为书中论述及人们关注的焦点。

凯特莱说：“这里所研究的人，他在社会中的存在就象物体的重心，是那些社会因素围绕着波动的平均数”。他指出，研究人事的学问，应当对性质相同的事物作大量的观察和数量的比较，从而求出平均人。对于那些与平均人相违者，则应进一步推究其发生改变

的原因[6]。因此，他把误差法则、正态分布的原理引入到人事研究中去，试图通过广泛地搜集资料，来求出人的生理特征（身高、体重、寿命等）和精神特征（智力、道德、犯罪倾向等）的平均数，从而构成所谓的“平均人”。由此可以看出这虽然是一个社会学概念，但其含义是统计性的，尽管凯特莱在实践上从未得出过平均人，但是这对以大数法则、正态分布和误差法则为主体的统计方法获得广泛应用，起到一定的影响。

按凯特莱的设想，“社会物理学”是对支配社会的规律的量化研究。在《社会制度》一书中，他直言不讳地说统计学就是社会物理学，并主张用物理学的力学规律来研究和解释社会现象，通过类比指出，随着了解的深入社会现象的规律同样可以达到物理学那样精密的程度。他长时间研究社会现象之间潜在的关系，并首先把心理事实、社会事实和风俗习惯进行量化，由于个体的变异很大，在考虑单个人的基础上，这种关系就不易揭露出来，平均值的稳定性是他引进“平均人”概念的动力。他说：“这个社会体与造物主一手造成的万物没有区别，都是因为其中有不灭的理由才得以存续。当我们达到发现法则的最高点时，我们就可以发现一种与天体法则有同样价值的法则”[38]。由此，他认为社会规律与自然规律一样是固定不变的。

此外，凯特莱还把自然规律性推及到人的犯罪现象。他研究了年龄、性别、季节、气候、人种、职业、教育、贫穷以及意志与犯罪的关系，强调一切社会现象的规律，都“不是来自个人的自由意志，而是来自一种与个人意志相独立的民众的精神力量”，“在那种个人自由意志全然被忽视的地方，造物主的作用便畅行无阻、为所欲为地发挥出来”。[6]这种犯罪规律和自由意志的论述，后成为一场激烈争论的导火索。

这样，正态曲线的使用引发了凯特莱五个方面的工作：基本理论、大量观察法的数学基础、“社会物理”理论、“平均人”理论和“犯罪”理论[21]。他写出了大量的统计学著作，发展了以概率论为基础的统计学作为适用于任何观测科学的研究的一般方法的观念，阐明了统计学探索社会想象规律性的本质、研究社会静态事实和动态事实的对象以及“对相同性质的事物进行大量观察”的研究方法。1841 年，凯特莱在普利茅斯召开的英国统计委员会上，提交了一份希望通过统计方法去尽快研究的 40 多个课题目录，这包括气象学、物理学、化学、教育学、植物学、农业、动物学和人科等，迎来了统计学发展的又一个春天。

凯特莱的努力使正态分布在 19 世纪统计应用中大为流行，并造出了“凯特莱主义”、“凯特莱时代”这些名词[13]。他既是社会统计学的奠基者，又是数理统计学的先驱，在他之后出现了各种课题统计学领域的追随者，他们在不同程度上继承和发展了他的思想和方法。但是，需要指出的是，任何理论的提出都不是尽善尽美的，凯特莱的工作同样有其弊

端和武断之处，因此它引发了统计学界前所未有的争论。将统计方法用于研究社会问题存在着实践和道德考虑的双重困难：一方面由于统计方法的贫乏和社会问题的复杂性，使得相关研究在 19 世纪进步有限；另一方面，基于政治和伦理方面的反对意见接踵而来，这被认为是玩弄数字游戏，并不能揭示社会真正本质。对于凯特莱思想所引发的无数意见，我无力给出统一的看法，但我认为，无论怎样，凯特莱终究利用它渊博的知识、创新的精神、融会贯通的方法以及无上的威望为统计方法开辟了一席之地，赢得了各个领域专家的关注，不管是赞赏还是批判，他们都设法寻找令人信服的理由，这无疑会带动理论的深化和完善，引发新理论的产生。所以，从这个意义上来说，凯特莱推动了现代统计学的到来。

统计学不是一门孤立的学科，而是研究其它学科的工具。当科学家把注意力转向生物科学时，统计学有了一次大的发展。在凯特莱的启发下，高尔顿最早把统计方法应用于生物学，他继续研究和推广正态曲线，提出了中位数、四分位数、百分位数及四分位偏差等概念，并创立了回归分析，对英国生物统计学派的兴起起到了奠基性作用。

3.3 相关与回归思想的诞生

高尔顿（Francis Galton,1822—1911）早年在伯明翰医学院学医，后进入剑桥大学，有了从事试验和观察研究的机会，并主攻数学。他祖父、叔祖父都是英国皇家学会会员，父亲热衷于科学研究和统计调查工作，这对他的成长不无影响。毕业后，他放弃了不感兴趣的从医生涯，开始去非洲考察，并由于成就突出获得 1853 年英国皇家地理学会的金质奖章。他所研究的问题和所写的论文，涉及面十分广泛，多达二十几种。1865 年，受其表兄、《物种起源》的作者达尔文（Darwin）的影响，他的主要兴趣转向了遗传学，并首次提出一种改良人种的优生学说。他是一个不知疲倦的事实资料搜集者，在非洲考察期间，他搜索了大量资料，并投入很大精力钻研其中所隐藏的模型与关系，这对他一生的成就有直接影响[40]。高尔顿继承并发展了凯特莱的研究课题，如果说凯特莱是从物理学角度研究社会现象，那么高尔顿则是从生物学角度研究社会现象的。

高尔顿是英国利用欧洲大陆更早著作的第一人，受凯特莱的启发，他对正态分布怀有浓厚的兴趣。用他自己的语言可以最生动地描述这一点："我第一次对误差的高斯定律[正规曲线]产生兴趣应归功于一篇讨论某高山群海拔的地理学论文，我充分理解了他所展示的这个漂亮定律的广泛应用，之后我更是尝到了熟悉凯特莱工作的乐趣"，[1]"我几乎不知道还有什么可以像'误差的频数定律'所表示的和谐秩序这样使人印象深刻，如果希腊人知道它的话，一定会把它奉为神灵来祭拜。它在最激烈的杂乱中平静地盛行着，并且不求闻

达。[22]”

他与凯特莱一样相信正态曲线是“适用于无数情况”的一般法则[43]。他最早把统计方法应用于生物学，并作了两点突破：第一，他指出，若干个同质数据的混合体，可借助正态分布分离开，他极其广泛地搜集资料，目的就在于探索把大量已知数目归纳为能够用于描述和比较的几个简单公式的途径和方法；第二，引进了“统计尺度”的概念，把非数量性指标（如智力）数量化，从而能够进行计算和比较，并指出：同一物种若其某数量性指标（如身高）可用正态曲线拟合，则其他指标也可用正态分布拟合，这种思想在心理学和教育学中出现了很多追随者①。

然而，这种无所不在的正态性也给高尔顿带来一些困惑：“...... 身高曲线和正态曲线之间的一致性，形成了我对自然遗传定律研究的主要依靠。数学家基于描述误差的目的发现了误差定律，我们现在正带着另一种目的去利用他们的劳动成果......”，“但是总有机会去正当地疑惑，建立在误差定律严格性基础上的结论在逼近变量之间的相互作用时是否正确。当我们把理论上的结论放在频率试验中检验时，发现人类学者可能在使用误差频数定律的性质上比以前少了很多犹豫②”。高尔顿首先发现亲子两代身高数据服从同一正态分布，这引发了他两方面的考虑：

1、按照中心极限定理所述，正态分布是由大量微小因素的影响而形成的，但众所周知遗传是一个显著性因素，这应如何解释？

2、身高作为一种遗传性状，其优势传递给下一代，应出现两极分化的态势，但子代身高稳定的正态分布与此相悖[41]。

这成为他相关回归思想产生的萌芽，带着这些困惑，他开始对由实验和抽样得来的数据进行统计分析。

高尔顿首先借助两个类比实验，分别回答了这两个问题。第一个是“正态漏斗”实验。他利用许多小球从漏斗中落下，途经有规律安置的障碍物，最终形成的正态曲线为例，以及在途中安插适当的阀门，形成大小不同的球源，继续下落，最终仍形成正态分布的结果，指出，遗传作为一个显著性因素，仍可以分解为大量微小因素作用的叠加，这就与中心极限定理相一致了。高尔顿的发现表明，同质性表面的背后包含了许多“异质”的成分，这进一步回答了正态分布得以广泛应用的原因。第二个是种豌豆试验。1875 年，他挑选了大

① Natural Inheritance(1889a),Macmillan,London,pp.54,55.

② “Co-relations and their Measurement,”Proceedings of the Royal Society,ⅩLV（1880）,p.144.

小不同的豌豆种子，并分派不同的人去种。1877 年，他对亲子代的数据分别进行分析，得到重大发现：相同大小的种子的后代仍符合正态分布，且方差与种子大小无关；子代的平均与母代的大小有对应关系，且有向母代平均线性收缩的趋势，总朝着一般平均数发展[41]，这就初步回答了子代均值与母代一样的原因。

八年后，为进一步验证他的解释，高尔顿成立了一个“人体测量实验室”①。他向社会征求了 205 对夫妇及他们的 928 个成年子女的身高数据，并进行了统计分析。他把父母的平均身高作为母代变量 x，把子代变量记为 y（其中，女子身高乘以1.08），并把相应的数据绘制成二维的，就发现相等强度的数据点出现在一条椭圆曲线上。于是问题就转化为去寻找一个(x,y)的二维分布，来解释这一现象。在数学家狄克逊的帮助下，他很快得到了二维正态分布的答案。虽然早在 1846 年，法国天文学家布雷瓦斯已经对二元和三元正态分布进行过探讨，但他并没有提到任何“相关”的术语。1886 年，高尔顿发表了有关这种观察的论文，提出父子之间的身高，有显著的相关性：父代身材高，则子代的平均身材也高；但是从子代的组别观察中，发现有退步现象，即“回归”到父代平均数去。1888 年，高尔顿在《自然遗传》（Natural Inheritance）中，提出了中位数、四分位数、百分位数及四分位偏差等概念，引进了回归直线，并赋予相关概念前所未有的重要性，成为相关回归理论的奠基者[40]。这一理论后经埃其沃斯、卡尔·皮尔逊等人的发展，成为一种得力的统计方法。

这样，由于凯特莱和高尔顿的创新思想与实干精神的结合，正态分布逐步完成了从丑小鸭向白天鹅的蜕变，它亭亭玉立于世人的面前，是如此地谦恭，如此地多能，等待着来自各个理论研究的人们的召唤。如果说，充斥着偶然性的世界是一个纷乱的世界，那么，正态分布为这个纷乱的世界建立了一定的秩序，它如同一位少女在嘈杂的人群中翩翩起舞，直达理论的核心地位，使得偶然性现象在数量上被计算和预测成为可能。

---

① Gattell was associated with Galton and helped him set up it.

4 正态分布对现代统计学的影响（19 世纪末期至 20 世纪初）

从 19 世纪中叶起，以契比雪夫（Tchebyshev）、马尔可夫（Markov）、李雅普诺夫（Lyapunov）为代表的俄罗斯学派，对概率论进行了认真的研究。他们通过引入随机变量的概念，建立了使得独立和非独立的随机变量的标准和收敛到正态分布的足够条件，从而把数学的严格性应用到大数定律和中心极限定理上。此后这项工作随概率论一起，被伯恩斯坦（Bernstein）、肯钦（Khinchin）、柯尔莫戈洛夫（Kolmogorov）和利维（H.Levy）等人发展，被建立在一个牢固的公理化体系之上，至此，概率论才真正成为一门演绎的数学理论，并同数理经济学、生物计量学和应用数学的发展相互促进、密切相关，为数理统计学的形成奠定了坚实的理论基础[42]。

可以说，高尔顿对英国生物学派的兴起及主要进步的取得起到了催化剂的作用。首先，高尔顿是用统计方法研究生物学的第一人，他用实际行动开拓了凯特莱的思想领域，并得到学术界的广泛承认；其次，相关回归理论创新性思想的产生，是统计学上的突破性进展，它进一步沟通了误差论和统计学两个领域的联系，为数理统计学的产生奠定了基础；再次，由于高尔顿的工作尚有许多不成熟之处，人们也怀疑它是否能用于遗传以外的领域，以卡尔•皮尔逊为代表的一批学者为此付出了努力，由于其涉及的问题复杂而又多方面，从而促进了一个严整统计学框架的建立[43]；第四，高尔顿的工作，产生了一个明显的后果，此后，统计学的重心逐渐由欧洲大陆向英国转移，使后者在以后几十年统计学发展的黄金时代充当了火车头。

进入现代统计时期，以威尔顿、卡尔・皮尔逊和埃其沃斯等人为先导，引发了正态分布及其相关理论的一系列创新和深化；20 世纪初，以哥塞特为先驱，费歇尔为主将，掀起了小样本理论的革命，大大提升了正态分布在统计学中的地位，使得用正态分布拟合数据继续占据应用的主流；相关回归分析、多元分析、方差分析、因子分析等统计方法，陆续登上了历史舞台，成为推动现代统计学飞速发展的一个强大动力。

在 20 世纪以前，统计学所处理的数据一般都是大量的、自然采集的，所用的方法以拉普拉斯中心极限定理为依据，总是归结到正态[44]。这种大样本统计学的顶峰和押阵大将，当属卡尔・皮尔逊。到了 19 世纪末期，数据与正态拟合不好的情况也日渐为人们所注意：凯特莱曾建议用 $p \neq \frac{1}{2}$ 的二项分布去拟合“偏态”数据——这个思想后来成为卡尔・皮尔逊引进其著名的曲线族的出发点；高尔顿也认识到，一组观测值的几何均值可能更好地表

示估计的最可能值，如果是这样的话，观测值的对数可以被假设服从正态分布，这就导致了对数正态分布，这也激发了卡尔·皮尔逊的工作；而直接吸引卡尔·皮尔逊注意力的是 1892 年威尔顿的提问，他在用正态曲线去拟合一组那波里蟹体宽数据时，得到一个双峰分布，并告知了皮尔逊，这引发了卡尔·皮尔逊的一系列成就[43]。

皮尔逊首先认为这可能是两个正态分布的混合，于是给出复合分布函数去拟合，并提出用矩方法去估计其参数。随后，卡尔·皮尔逊认为，统计学需要的是一种能把观测数据转化为一个预测模型的方法，于是他希望找出一族曲线，去拟合从实际问题中得来的数据，以便在正态分布不适用时可供选择使用。1892 年—1895 年，他依赖清晰地表达能力、精湛的数学功底以及坚持自我的精神在这项工作中获得了成功，并以《数学用于进化论》为总题目发表了一系列论文[3]。由于正态分布在人们心中根深蒂固的地位，卡尔·皮尔逊的曲线族在当时同样引发了一场争论，并由于各种理由不被绝大多数学者承认，但它同对数正态分布、极值分布等分布一样，终究为人们提供了一种有用的工具，扩大了统计方法的武库。

在理论上，多元正态分布的重要意义在于：它把起初纯属于误差分析的线性模型理论与“统计数据”的分析沟通起来[43]。1892 年，埃其沃斯（Francis Edgeworth）从两元、三元以及四元变量开始工作，对于多元正态分布给出了第一个陈述。1896 年，卡尔•皮尔逊在此基础上给出了更加明确的推导，并发展了一套比高尔顿的相关理论更一般化和精确化的复相关和多重回归理论，他指出，这样一个理论对于回答威尔顿提出的那类问题是很有必要的。在试图把数据拟合到他的频数曲线上时，卡尔·皮尔逊面临对拟合优度检验的需要，这使他于 1900 年建立了作为现代统计学牢固基础的χ^2拟合优度检验。

进入 20 世纪之后，人工试验条件下所得数据的统计分析问题，日渐被人们所重视。由于试验数据量有限，那种依赖于近似正态分布的传统方法开始招致质疑，这促使人们研究这种情况下正确的统计方法问题[44]。卡尔·皮尔逊的理论指导和在酿酒公司工作的经历使得哥塞特具备了研究小样本问题的良好条件。1908 年，他发表著名论文《均值的或然误差》（The Probable Error of a Mean），提出了正态样本中样本均值和标准差的比值t的分布，并给出了应用上及其重要的第一个分布表格[45]。虽然其推导有一些漏洞，但这并不影响其在历史上的功绩。随后费歇尔为解决这些漏洞，开始了与哥塞特的通信及长达二十余年的友谊和研究。

费歇尔发展了“n维几何”的方法，这成为正态样本统计量的抽样分布中一个极为有

力的方法[46]。自 1915 年开始，沿用这个方法费歇尔获得了一些应用上极重要的统计量如相关系数r、正态样本中绝对偏差$\sum\left|X_i-\bar{x}\right|/n$、回归系数、相关比、多重回归和偏相关系数等的分布，以及来自两个正态总体的样本方差的比值F的分布。同时，他发展了估计，充分性，似然，推断，方差分析和实验设计的思想，使得正态性假设在统计分析中发挥了关键性作用[45]。此后，有关相关回归分析中一些重要统计量的精确分布如t分布、F分布以及χ^2分布的产生，与多维正态分布一起，始终雄踞于统计学的要津，发挥关键性作用。至此，这场革命再次从理论上鉴定了正态分布的基础性作用，在费歇尔（Fisher）的推动下，现代统计学开始出现各种分支并迅速发展起来。自 20 世纪 60 年代开始，越来越多的注意力又转向了对正态性假设的质疑，当此假设不满足时，估计值同样是健康的要求，有待人们进一步发明关于估计值有效性的检验[46]。这是后话，不再赘述。

最后，笔者希望交代一下用于形容正态分布的专用名词的相关历史。由于正态分布曾引起众多学者的研究，以及统计学理论统一之前彼此研究的孤立性，使得它曾以许多科学家，包括拉普拉斯、高斯、凯特莱和麦克斯韦的名字命名，但距今所知没有一个现代作者以它的提出者狄莫弗的名字为它命名。高尔顿曾用“误差的频数定律”、“指数定律”、“与一个均值的偏差定律”、“观测误差定律”以及“正态定律”等名字来称呼它，说明在 1900 年之前还没有一个术语能得到普遍地承认。最终“正态定律”或者“正态分布”得到英国生物统计学派的承认，但卡尔·皮尔逊认为这个词从外表上看不出它更早的用法，并且容易使人们产生误会：一切其他的频率分布曲线，都是在这个或那个意义上“非正常”，这会导致许多作者把不管什么样的分布硬扭成正态的[3]。这一意见也得到了一些学者的响应，于是直到现在，还有很多人称之为“高斯分布”。

结 束 语

应该说正态分布的早期历史，主要是数理统计学开始作为一门科学的历史。从最初数学表达式的提出，到之后成为一种连续型概率分布的应用，再到构成各种统计方法的理论基础，每一次蜕变都蕴含着该阶段最著名数学家的心血和努力，体现了观念转变和技术创新的决定性影响，印证了人们对该理论由无知到有知，由有知到多知，再由多知到熟知的认识规律，同时展示了正态分布同概率论的起源、误差论的发展、近代统计学的形成以及现代统计学的飞跃等理论背景相互促进、相互制约的辩证关系。概率论和统计学是一对姊妹学科，是集天文学家、数学家、物理学家、生物学家、心理学家、人类学家、经济学家、政治学家、审计学家和职业赌博家的劳动成果为一体的学科。二者彼此交织，互相渗透，并行发展。正态分布进入统计学的历史过程可认为是精辟诠释这种关系的一个典范！几百年来，统计学中没有哪一个理论，象正态分布那样被广泛地发展。人们总是设法利用改进了的新观念和新方法，去研究该理论的应用基础及其数学性质，它的产生和发展，它的新理论和新应用，它的传递性和同内外理论的相互作用，推动了整个统计学的发展，在共同的研究中，将不断引起新的突破。

统计学中与正态分布相关的成果累累，使笔者难以单枪匹马地论述这一理论的每一个细节，并圆满说明所有的新发现，只能对该领域的文献资料进行以时间和人物为轴，以理论的相互影响为图像的梳理，对正态分布从提出、到重生、再到统计学中大行其道的发展过程做一概述，使读者对概率统计理论由观念转变和技术创新来推动发展的微妙性以及二者互为渗透、并走向交融的发展规律，有一个较为全面和生动的印象。至于某个阶段的内容细节，如第二部分中高斯和拉普拉斯的误差理论对其概率工作的影响，第四部分中以哥塞特为先导、费歇尔为主将掀起的小样本革命如何从理论上鉴定了正态分布的基础性作用，都值得进一步研究。

参 考 文 献

[1]Walker H M. Studies In The History Of Statistical Method. Baltimore The Williams& Wilkins Company，1929，Reprinted，1931.4、12-28、21-49、92-125.

[2]Hald A. A History of Mathematical Statistics from 1750 to 1930. John Wiley & Sons，Inc. New York：Wiley，1998.18-25、35-40、303-317、360-373.

[3]Pearson K. Historical Note on the Origin of the Normal Curve of Errors. Biometrika，Vo.l16，1924.402-404.

[4]Todhunter I. A History of the Mathematical Theory of Probability，from the Time of Pascal to that of Laplace. Cambridge，1865.1-4、4-6、7-21.

[5]Sheynin O. The Theory of Probability: Its Definition and Its Relation to Statistics， Archive for History of Exact Science，1998，152(2)：49-67.

[6]高庆丰.欧美统计学史.中国统计出版社，1987.57-61、79-82、86-93.

[7]郭贵春，宋尚玮.对概率论起源的思考.科学技术与辩证法，2006，23（2）：89-93.

[8]Daston L J. Probabilistic Expectation And Rationality In Classical Probability Theory. Historia Mathematical[J]（1980），vol.7：134-157.

[9]徐传胜，曲安京.惠更斯与概率论的奠基，自然辩证法通讯，2006（6）：76-80.

[10]Samuel K，Johnson N L. Leading Personalities In Statistical Sciences：From The 17th Century To The Present. John Wiley & Sons，Inc.1997.89-92.

[11][美]约翰·塔巴克著，杨静译.概率论和统计学.商务印书馆，2007.29—50.

[12]吕淑红.十八世纪的数学家狄莫弗.数学史研究文集，第一辑，九章出版社，1989.172-178.

[13]陈希孺.数理统计学简史.湖南教育出版社，2002.28-58，73-88，101-130.

[14]Johnson P O，Robert W B，Jackson，Modern Statistical Methods：Descriptive And Inductive.1-6.

[15]李文林著.数学史概论[M] .第二版，高等教育出版社，2002.286-292.

[16]Richard William Farebrother，Fitting Linear Relationship，A History of the Calculus of Observations，1750—1900. Springer，1998.49-67.

[17]Peters W S. Counting For Something: Statistical Principles And

Personalities. Springer-Verlag New York Inc，1987.100-107，127-136.
[18]Frances Marguerite Clarke. Thomas Simpson and His Times.1929.24-47.
[19]Kendall M，Plackett R L. Studies in the History of Statistics and Probability. Vol.2，Griffin，London，1977.64-77、92-105.
[20]Kendall M G，Stuart A. The Advanced Theory of Statistics. Griffin，London，1958.14-17.
[21]吴文俊主编.世界著名数学家传记[M].科学出版社，1995.
[22]Kolmogorov A N，Yushkevich A P. Mathematics of The 19th Century，Basel；Boston；Berlin：Birkhäuser，1992.226-230、268-282.
[23]Abbe C. A Historical Note On the Method of Least Squares. Am.J.Sci.Arts[3]1，1871.411-415.
[24]Aitken A C. On Least Squares and Linear Combination of Observations. Proc. R.Soc.Edinburgh，A55，1935.42-47.
[25]On the Probable Errors of Frequency Constans. Biometrika，Part Ⅱ，1903:273-281，Part Ⅱ，Vol.Ⅸ，1913:1-10，PartⅢ，Vol.XⅢ，1920-21:113-132.
[26]李文林.数学珍宝[M].科学出版社，1998.
[27]C.F.Gauss. Gauss's Work on the Theory of Least Squares. 1957.53-69、83-87.
[28]Linnik Y V. Method of Least Squares and Principles of the Theory of Observations. Pergamon Press，London，1958.101-112.
[29]Abbe. A Historical Note on the Method of Least Squares. American Journal of Science，3rd series，Ⅰ:411-415.
[30]胡作玄，邓明立著.大有可为的数学.河北教育出版社，2006.
[31]Adams W J. The Life and Times of the Central Limit Theorem. Kaedmon. New York，1974.151-172.
[32]Molina E C. The theory of probability：Some comments on Laplace's Théorie analytique. Bull.Am.Math.Soc.[2]36，1930.369—392.
[33]Dutka J. Robert Adrain and The Method of Least Squares. Arch.Hist.Ex.Sci.41，1990.171—184.
[34]Coolidge. Robert Adrain and the Beginnings of American Mathematics. American Mathematical Monthly，XXXIII，1926.61-76.

[35]Eisenhart C. Laws of error. In Encyclopedia of Statistical Sciences（S.Kotz，N.L.Johnson，and C.B.Read，eds）Vol.4，Wiley，New York，1983.530—566.
[36]中国大百科全书（数学卷）.中国大百科全书出版社，1988 年.
[37]林恩·阿瑟·斯蒂恩著，胡作玄等译.站在巨人的肩膀上.上海教育出版社，1997.109-156.
[38]Quetelet A as a Statistician. Archive for History of Exact Science，1986，36(4):282-325.
[39]Hankins F H. Adolphe Quetelet as Statistician. New York，1908.16、55、83-85、102-121.
[40]Eisenhart C. Galton in Dictionary of Scientific Biography. Vol.5.Scribners，New York，1974.158-166、182-211.
[41]Pearson. Francis Galton，1822-1922，a Centenary Appreciation，No.Ⅺ in the series Questions of the Day and of the Fray，1922.171-243.
[42]胡作玄.20 世纪的数学.见《20 世纪科学技术简史》[M]第十三章，科学出版社，1999，第二版.317-322.
[43]Magnello M E. Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician. The Rutherford Journal，Vol.1.2005.68-124.
[44]胡作玄，邓明立著.20 世纪数学思想[M].山东教育出版社，2001.9-44.
[45]Fisher R A. On an Absolute Criterion for Fitting Frequency Curves. Messenger of Mathematics，Vol.41，1912.155-160.
[46]MacKenzie D A. Statistics in Britain 1865—1930. Edinburgh University Press，Edinburgh，1981.66-98.

附 录：大 事 年 表

年代	作者	文章及刊名
1733	De Moivre	Approximatio ad Summam Terminorum Binomii $(a+b)^n$ in Seriem expansi，首次提出正态曲线。
1755	Simpson	A Letter on the Advantage of Taking the Mean of a number of Observations，in practical Astronomy，在应用天文学中首次提出取观测值均值的可靠性问题。
1774	Laplace	Determiner le milieu que l'on doit prendre entre trois observations données d’un même phénomène，假设误差分布为重指数分布。
1778	Daniel Bernoulli	Acta Academiae Scientiarum Imperialis Petropolitanae pro Anno MDCCLXXVII，Pars Prior，I，3—23，提出用一个圆来表示机会误差的分布。
1783	Laplace	Suite du Mémoire sur les approximations des Formules qui sont fonctions de très-grands Nombres，首次声明积分 e^{-t^2} 的重要性。
1799	Kramp，Strasburg，Leipzig	Analyse des Réfractions Astronomiques et Terrestres，该文给出了第一个正态曲线表。
1808	Adrain	Research concerning the probabilities of the errors which happen in making observations，也给出了正态曲线的两个力证。
1809	Gauss	Theoria Motus Corporum Coelestium in Sectionibus Coniciis Solem Ambientium，BookII，SectionIII，给出了正态误差定律。
1812	Laplace	《Théorie Analytique des Probabilités》，分析概率论的里程碑式著作。
1815	Bessel	“Ueber den Ort des Polarsterns”，首次提出了概差的概念。
1816	Gauss	“Bestimmung der Genauigkeit der Beobachtungen，”讨论了观测值的精确度问题，并进一步发展了概差理论。
1827	Quetelet	A.Quetelet，Recherches sur la population，les naissances，les décès，les prisons，les dépôts de mendicité，etc，dans le royaume des Pays-Bas.Nouv.Mém，Acad.R.Bruxelles4，117—192。
1831	Quetelet	“Recherches sur la penchant au crime aux différens ages”，研究了应用在犯罪上的统计学。
1832	Encke	“Über die Methode der kleinsten Quadrate”，是第一部有关统计学的数学理论的著作。
1836	Quetelet	A.Quetelet，Sur l'Homme et le Développement de ses Facultés，ou Essai de Physique Sociale，Bachelier，Paris.Pirated ed.，Hauman，Bruxelles。

1837	Poisson	Recherches sur la probabilité des jugements en matière criminelle et en matière civile，précédées des règles générales du calcul des probabilités，有关把概率论应用到证据和法庭判决中的讨论。
1838	De Morgan	“Encyclopaedia Metropolitana”，对概率论的数学理论及其发展做了总结。
1842	Quetelet	A Treatise on Man and the Development of His Faculties，Chambers，Edinburgh。
1845	Quetelet	Sur l'appréciation des documents statistiques，et particulier sur l'appréciation des moyennes，Bull.Comm.Cent.Stat.，Belg.2，205—286。
1846	Quetelet	Lettres à S.A.R.le Duc Régnant de Saxe-Cobourg et Gotha，sur la Théorie des Probabilités，appliquée aux Sciences Morales et Politiques，Hayez，Bruxelles。
1849	Quetelet	Letters Addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha，on the Theory。
1861	Airy	“Combination of Observations”，具体化了 Encke 的著作。
1871	Glaisher	“On a class of Definite Integrals，”Philosophical Magazine，4th Series，ⅩLⅡ（1871），421—436。
1872	De Morgan	“An Essay on Probabilities and on Their Application to Life Contingencies and Insurance Offices”，做出了首次化简概率论的尝试。
1888	Galton	Natural Inheritance，提出了相关回归的思想。
1898	K.Pearson，L.N.G.Filon	“Mathematical Contributions to the Theory of Evolution.—Ⅳ.On The Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation”，Philosophical Transactions，A,CXCI，229—311。
1903	K.Pearson	“On the Probable Errors of Frequency Constants”，PartⅠ，Biometrika，Ⅱ，273—281。
1906	K.Pearson，J.Blakeman	“On the Probable Error of the Coefficient of Mean—Square Contingency”，Biometrika，V，191—197。
1908	F.Y.Edgeworth	“On the Probable Errors of Frequency Constants”，Journal of the Royal Statistical Society，LXXI，381—397，499—512，651—678。
1908	“Student”	“Probable Error of a Correlation Coefficient”，Biometrika，VI，302—310；“The Probable Error of a Mean”，Biometrika,VI，1—25。
1910	“Student”	“The Distribution of the Means of Samples which are not drawn at Random”，Biometrika，VII，210—214。
1912	K.Pearson	“On the General Theory of the Influence of Selection on Correlation and Variation”，Biometrika，VIII，437—443。

1913	K.Pearson	“On the Probable Error of a Coefficient of Correlation as found from a Fourfold Table”，Biometrika，IX，22—27。
1915	R.A.Fisher	“Frequency Distribution of the Values of the Correlation Coefficient in Samples from an indefinitely large Population”，Biometrika，X，507—521。
1915	K.Pearson	“On the Distribution of the Standard Deviations of Small Samples：Appendix to papers by‘Student’and R.A.Fisher”，Biometrika，X，522—529；“On the Probable Error of a Coefficient of Mean—Square Contingency”，Biometrika，X，570—573。
1917	“Student”	“Tables for Estimating the Probability that the Mean of a Unique Sample of Observations lies between $-\infty$ and any given Distance from the Mean of the Population from which the Sample is Drawn”，Biometrika，XI，414—415。
1918	F.Y.Edgeworth	“On the Value of a Mean as Calculated from a Sample”，Journal of the Royal Statistical Society，LXXXI，628—632。
1920	K.Pearson	“On the Probable Errors of Frequency Constants，III”，Biometrika，XIII，113—130。
1921	R.A.Fisher	“On the‘Probable Error’of a Coefficient of Correlation deduced from a Small Sample”，Metron，I，3—32。
1921	“Student”	“An Experimental Determination of the Probable Error of Dr.Spearman's Correlation Coefficients”，Biometrika，XIII，263—282。
1923	E.S.Pearson	“The Probable Error of a Class-Index Correlation”，Biometrika，XIV，261—280。
1924	R.A.Fisher	“The distribution of the partial correlation coefficients”，Metrron，III，329—332。
1924	E.S.Pearson	“Note on the Approximations to the Probable Error of a Coefficient of Correlation”，Biometrika，XVI，196—197。
1927	K.Pearson	“Note on the Relation of the (P, χ^2)Goodness of Fit Test to the Distribution of Standard Deviations in samples from a Normal Population”，Biometrika，XIX，215。

注：此附录按照时间顺序，注明了与本论文相应时间（自 1773 年狄莫弗首次提出正态曲线——20 世纪 20 年代相关回归思想的兴起）内主要代表人物关于正态分布的相关性工作，以期使读者对此 150 年间正态分布的发展情况有一个脉络性的直观了解。随后以哥塞特为先导、费歇尔为主将掀起的小样本革命再次从理论上鉴定了正态分布的基础性作用，笔者希望在今后的学习中再做进一步探讨。

致　谢

本文是在我的导师邓明立教授的悉心指导下完成的。从论文选题到搜集资料，从论述篇幅到材料取舍，从写作大局到叙述细节，无不渗透着邓老师的心智。他高屋建瓴的启发对该研究的进展起到画龙点睛的作用，他茅塞顿开的指点使我增强了克服困难的信心，并较快走上了研究的正轨。

同样，在三年的硕士研究生学习期间，邓老师用严谨求实的治学态度，孜孜不倦的工作作风，以及谦逊朴实的为人之道向我诠释了专业与旁通的完美结合：他经验丰富，博学多才；他思想精妙，行为高远；他充满关爱、热情大度；他精益求精，勇于创新。从他那里我不仅学会了科学研究的方法，更领悟到了做人的真谛。邓老师的人格魅力乃至一言一行都将在我心中留下难以磨灭的痕迹，成为我成长道路上的躬行示范者。

与此同时，在学习过程中，中科院数学与系统科学研究所的胡作玄研究员给予我很大帮助，在论文的选题、文章的修改等方面都提出了宝贵的修改性意见。毕业在即，师恩难忘。在此，谨致以最衷心的感谢和最真诚的祝福。

此外，河北师范大学数信学院同门的师兄、师姐以及同学也给予我很多帮助，在此一并表示感谢！