正态分布的前世今生(下)

靳志辉

六、开疆拓土，正态分布的进一步发展

19 世纪初，随着拉普拉斯中心极限定理的建立与高斯正态误差理论的问世，正态分布开始崭露头角，逐步在近代概率论和数理统计学中大放异彩。在概率论中，由于拉普拉斯的推动，中心极限定理发展成为现代概率论的一块基石；而在数理统计学中，在高斯的大力提倡之下，正态分布开始逐步畅行于天下。

1 论剑中心极限定理

先来说说正态分布在概率论中的地位，这个主要是由于中心极限定理的影响。1776 年，拉普拉斯开始考虑一个天文学中彗星轨道的倾角的计算问题，最终的问题涉及独立随机变量求和的概率计算，也就是计算如下的概率值

$$S_n = X_1 + X_2 + \cdots + X_n$$
$$P(a < S_n < b) = ?$$

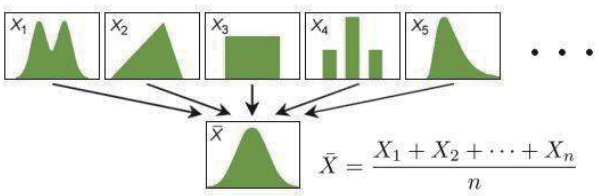
在这个问题的处理上，拉普拉斯充分展示了其深厚的数学分析功底和高超的概率计算技巧，他首次引入了用特征函数（也就是对概率密度函数做傅立叶变换）来处理概率分布的神妙方法，而这一方法经过几代概率学家的发展，在现代概率论里面占有极其重要的位置。基于这一分析方法，拉普拉斯通过近似计算，在他 1812 年发表的名著《分析概率论》中给出了中心极限定理的一般描述：

定理 0.6.1 (拉普拉斯，1812) $e_i\,(i = 1, \cdots n)$ 为独立同分布的测量误差，具有均值 μ 和方差 σ^2，如果 $\lambda_1, \cdots, \lambda_2$ 为常数，$\alpha > 0$, 则有

$$P\left(\left|\sum_{i=1}^{n}\lambda_i(e_i-\mu)\right| \le a\sqrt{\sum_{i=1}^{n}\lambda_i^2}\right) \approx \frac{2}{\sqrt{2\pi}\sigma}\int_0^a e^{-\frac{x^2}{2\sigma^2}}dx.$$

这已经是比棣莫弗 - 拉普拉斯中心极限定理更加深刻的一个结论了，理科专业的本科生学习《概率论与数理统计》这门课程的时候，通常学习的中心极限定理的一般形式如下：

定理 0.6.2 (林德伯格 - 列维中心极限定理) 设 $X_1, \cdots, X_n$ 独立同分布，且具有有限的均值 μ 和方差 σ^2，则在 $n \to \infty$ 时，有



中心极限定理

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \to N(0,1).$$

多么奇妙的性质，随意的一个概率分布中生成的随机变量，在序列和 (或者等价的求算术平均) 的操作之下，表现出如此一致的行为，统一地规约到正态分布。

概率学家们进一步的研究结果更加令人惊讶，序列求和最终要导出正态分布的条件并不需要这么苛刻，即便 $X_1, \cdots, X_n$ 并不独立，也不具有相同的概率分布形式，很多时候他们求和的最终归宿仍然是正态分布。一切的纷繁芜杂都在神秘的正态曲线下被消解，这不禁令人浮想联翩。中心极限定理恐怕是概率论中最具有宗教神秘色彩的定理，如果有一位牧师拿着一本圣经向我证明上帝的存在，我是丝毫不会买账；可是如果他向我展示中心极限定理并且声称那是神迹，我可能会有点犹豫，从而乐意倾听他的布道。如果我能坐着时光机穿越到一个原始部落中，我也一定带上中心极限定理，并劝说部落的酋长把正态分布作为他们的图腾。

中心极限定理虽然表述形式简洁，但是严格证明它却非常困难。中心极限定理就像一张大蜘蛛网，棣莫弗和拉普拉斯编织了它的雏形，可是这张网上漏洞太多，一个多世纪来，数学家们就像蜘蛛一样前赴后继，努力想把所有的漏洞都补上。在 19 世纪，泊松（Siméon Denis Poisson, 1781-1840）、狄利克雷（Gustav Lejeune Dirichlet, 1805-1859）、柯西（Augustin-Louis Cauchy, 1789-1857）、贝塞尔（Friedrich Bessel, 1784-1846）这些大蜘蛛都曾经试图把这张网的漏洞补上。从现代概率论的角度来看，整个 19 世纪的经典概率理论并没有能输出一个一般意义下的严格证明。而最终把漏洞补上的是来自俄罗斯的几位蜘蛛侠：切比雪夫（Pafnuty Chebyshev, 1821-1894）、马尔可夫（Andrey Andreyevich Markov, 1856-1922）和李雅普诺夫（Aleksandr Mikhailovich Lyapunov, 1857-1918）。俄罗斯是一个具有优秀数学传统的民族，产生过几位顶尖的数学家，在现代概率论的发展中，俄罗斯的圣彼得堡学派可以算是顶了大半边天，而切比雪夫正是圣彼得堡数学学派的奠基人和领袖。给中心极限定理补漏的方案雏形是从切比雪夫 1887 年的工作开始的，切比雪夫提出了一个基于矩法的证明，矩法是概率分析中比较传统的方法，使用的数学工具比较基础，不过切比雪夫这个证明也还存在一些漏洞。马尔可夫和李雅普诺夫都是切比雪夫的学生，两人在中心极限定理的严格证明上展开了竞赛。马尔可夫在概率论里面可算是大名鼎鼎，马尔可夫链是应用最为广泛的概率模型之一。马尔可夫和他的老师切比雪夫一样，他们在数学中的研究风格都偏向于使用初等、简单易懂的数学工具来证明复杂艰深的定理。马尔可夫沿着老师的基于矩法的思路在蜘蛛网上辛勤编织，在证明上做了很多的修补，但洞还是补得不够严实。李雅普诺夫不像马尔可夫那样深受老师的影响，他沿着拉普拉斯当年提出的基于特征函数的思路，于 1901 年给出了一个补洞的方法，切比雪夫对这个方法大加赞赏，李雅普诺夫的证明被认为是第一个在一般条件下的严格证明；而马尔可夫也不甘示弱，在 1913 年基于矩法也把洞给补严实了。

20 世纪初期到中期，中心极限定理的研究几乎吸引了所有的概率学家，这个定理俨然成为了概率论的明珠，成为了各大概率论武林高手华山论剑的场所。不知道大家对中心极限定理中的“中心”一词如何理解，许多人都认为“中心”这个词描述的是这个定理的行为：以正态分布为中心。这个解释看起来确实合情合理，不过并不符合该定理被冠名的历史。事实上，20 世纪初概率学家大都称呼该定理为极限定理（Limit Theorem），由于该定理在概率论中处于如此重要的中心位置，如此之多的概率学武林高手为它魂牵梦绕，于是数学家波利亚于 1920 年在该定理前面冠以“中心”一词，由此后续人们都称之为中心极限定理。

数学家们总是极其严谨苛刻的，给定的一个条件下严格证明了中心极限定理，数学家们就开始探寻中心极限定理成


切比雪夫（1821-1894）


马尔可夫（1856-1922）


李雅普诺夫（1857-1918）


费勒（1906-1970）


列维（1886-1971）

华山论剑

立的各种条件，询问这个条件是否为充分必要条件，并且进一步追问序列和在该条件下以什么样的速度收敛到正态分布。1922 年林德伯格（Jarl Waldemar Lindeberg, 1876-1932）基于一个比较宽泛且容易满足的条件，为中心极限定理提出了一个很容易理解的初等证明，这个条件我们现在称之为林德伯格条件。然后概率学家费勒（William Feller, 1906-1970）和列维（Paul Pierre Lévy, 1886-1971）就开始追问：林德伯格条件是充分必要的吗？基于林德伯格的工作，费勒和列维都于 1935 年独立地得到了中心极限定理成立的充分必要条件，这个条件可以用直观的非数学语言描述如下：

定理 0.6.3（中心极限定理充要条件）假设独立随机变量序列 X_i 的中值为 0，要使序列和 $S=\sum_{i=1}^{n} X_i$ 的分布密度函数逼近正态分布，以下条件是充分必要的：

* 如果 X_i 相对于序列和 S 的散布（可以理解为标准差）是不可忽略的，则 X_i 的分布必须接近正态分布；

* 对于所有可忽略的 X_i，取绝对值最大的那一项，这个绝对值相对于序列和也是可忽略的。

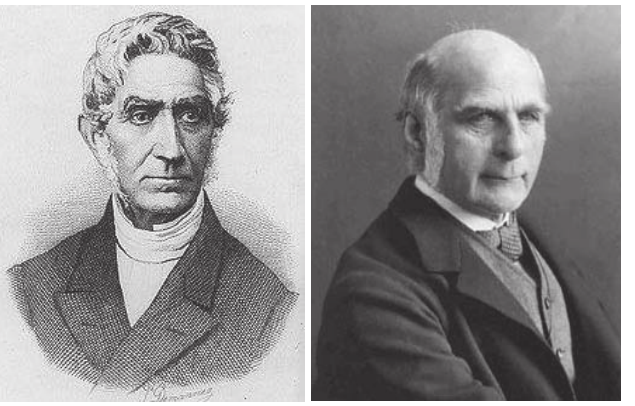
事实上这个充分必要条件发现的优先权，费勒和列维之间还着实出现了一些争论，当然他们俩都是独立的几乎在同一时间解决了这个问题。在列维证明这个充分必要条件的过程中，列维发现了正态分布的一个有趣的性质。我们在数理统计中都学过，如果两个独立随机变量 X, Y 具有正态分布，则 $S=X+Y$ 也具有正态分布；奇妙的是这个定理的逆定理也成立：

定理 0.6.4（正态分布的血统）如果 X, Y 是独立的随机变量，且 $S=X+Y$ 是正态分布，那么 X, Y 也是正态分布。

正态分布真是很奇妙，就像蚯蚓一样具有再生的性质，你把它一刀切两断，它生成两个正态分布；或者说正态分布具有极其纯正的优良血统，正态分布的组成成分中只能包含正态分布，而不可能含有其它杂质。

一流的数学家都是接近上帝的人，善于猜测上帝的意图。1928 年列维就猜到了这个定理，并在 1935 年使用这个定理对中心极限定理的充分必要条件作了证明。有意思的是列维却无法证明正态分布的这个看上去极其简单的再生性质，所以他的证明多少让人觉得有些瑕疵。不过列维的救星很快就降临了，1936 年概率学家克拉美（Harald Cramér, 1893-1985）证明列维的猜想完全正确。

中心极限定理成为了现代概率论中首屈一指的定理，事实上中心极限定理在现代概率论里面已经不是指一个定理，而是指一系列相关的定理。统计学家们也基于该定理不断地完善拉普拉斯提出的元误差理论，并据此解释为何世界上正态分布如此常见。而中心极限定理同时成为了现代统计学中大样本理论的基础。



凯特勒（1796-1874）和高尔顿（1822-1911）

2 进军近代统计学

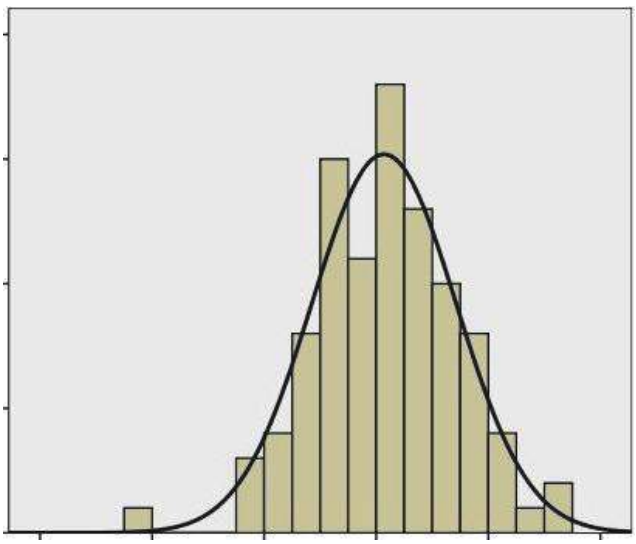
花开两朵，各表一枝。上面说了正态分布在概率论中的发展，现在来看看正态分布在数理统计学中发展的故事。这个故事的领衔主演是凯特勒（Adolphe Quetelet, 1796-1874）和高尔顿（Francis Galton, 1822-1911）。

由于高斯的工作，正态分布在误差分析中迅速确定了自己的地位。有了这么好的工具，我们可能拍脑袋就认为，正态分布很快就被人们用来分析其它的数据，然而事实却出乎我们的意料，正态分布进入社会领域和自然科学领域，可是经过一番周折的。

首先我要告诉大家一个事实：误差分析和统计学是风马牛不相及的两个学科，当然，这个事实存在的时限截止到 19 世纪初。统计学的产生最初是与“编制国情报告”有关，主要服务于政府部门。统计学面对的是统计数据，是对多个不同对象的测量；而误差分析研究的是观测数据，是对同一个对象的多次测量。观测数据和统计数据在当时被认为是两种不同行为获取得到的数据，适用于观测数据的规律未必适用于统计数据。19 世纪的数据统计分析处于一个很落后的状态，和概率论没有多少结合。概率论的产生主要和赌博相关，发展过程中与误差分析紧密联系，而与当时的统计学交集非常小。将统计学与概率论真正结合起来推动数理统计学发展的便是我们的统计学巨星凯特勒。

凯特勒这名字或许不如其它数学家那么响亮，估计很多人不熟悉，所以有必要介绍一下。凯特勒是比利时人，数学博士毕业，年轻的时候曾追随拉普拉斯学习过概率论。此人学识渊博，涉猎广泛，脑袋上的桂冠包括统计学家、数学家、天文学家、社会学家、国际统计会议之父、近代统计学之父、数理统计学派创始人。凯特勒的最大贡献就是将法国的古典概率理论引入统计学，用纯数学的方法对社会现象进行研究。

1831 年，凯特勒参与主持新建比利时统计总局的工作。



用正态分布拟合数据

他开始从事有关人口问题的统计学研究。在这种研究中，凯特勒发现，以往被人们认为杂乱无章的、偶然性占统治地位的社会现象，如同自然现象一样也具有一定的规律性。凯特勒搜集了大量关于人体生理测量的数据，如体重、身高与胸围等，并使用概率统计方法来对数据进行数据分析。但是当时的统计分析方法遭到了社会学家的质疑，社会学家们的反对意见主要在于：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。凯特勒大胆地提出：**把一批数据是否能很好地拟合正态分布，作为判断该批数据是否同质的标准**。

凯特勒提出了一个使用正态曲线拟合数据的方法，并广泛地使用正态分布去拟合各种类型的数据。由此，凯特勒为正态分布的应用拓展了广阔的舞台。正态分布如同一把屠龙刀，在凯特勒的带领下，学者们挥舞着这把宝刀在各个领域披荆斩棘，攻陷了人口、领土、政治、农业、工业、商业、道德等社会领域，并进一步攻占天文学、数学、物理学、生物学、社会统计学及气象学等自然科学领域。

正态分布的下一个推动力来自生物学家高尔顿，当正态分布与生物学联姻时，近代统计学迎来了一次大发展。高尔顿是生物统计学派的奠基人，他的表哥达尔文的巨著《物种起源》问世以后，触动他用统计方法研究遗传进化问题。受凯特勒的启发，他对正态分布怀有浓厚的兴趣，开始使用正态分布去拟合人的身高、胸围以至考试成绩等各类数据，发现正态分布拟合得非常好。他因此相信正态曲线是适用于无数情况的一般法则。

然而，对高尔顿而言，这个无处不在的正态性给他带来一些困惑。他考察了亲子两代的身高数据，发现遵从同一的正态分布，遗传作为一个显著因素是如何发挥作用的？1877年，高尔顿设计了一个叫高尔顿钉板（quincunx, 或者 Galton board）的装置，模拟正态分布的性质，用于解释遗传现象。

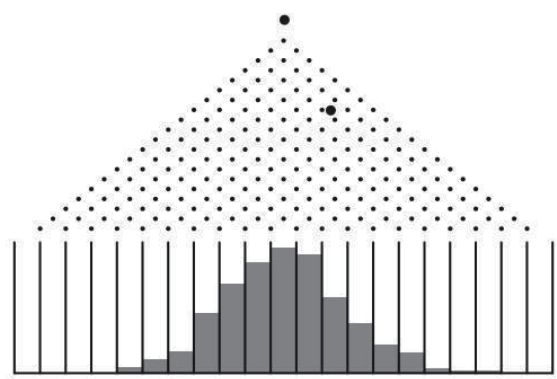
如下图中每一点表示钉在板上的一颗钉子，它们彼此的距离均相等。当小圆球向下降落过程中，碰到钉子后皆以1/2 的概率向左或向右滚下。如果有 n 排钉子，则各槽内最终球的个数服从二项分布 $B(n, 1/2)$，当 n 较大的时候，接近正态分布。

设想在此装置的中间某个地方 AB 设一个挡板把小球截住，小球将在 AB 处聚成正态曲线形状，如果挡板上有许多阀门，打开一些阀门，则在底部形成多个大小不一的正态分布，而最终的大正态分布正是这些小正态分布的混合。

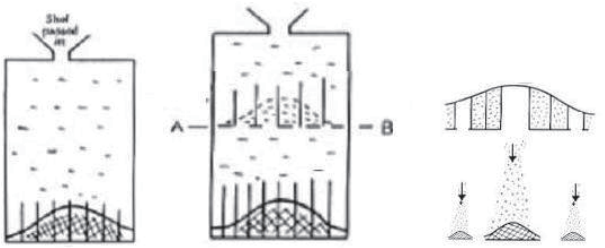
高尔顿利用这个装置创造性地把正态分布的性质用于解释遗传现象。他解释说身高受到显著因素和其它较小因素的影响，每个因素的影响可以表达为一个正态分布。遗传作为一个显著因素，类似图中底部大小不一的正态分布中的比较大的正态分布，而多个大小不一正态分布累加之后其结果仍然得到一个正态分布。

网络上有人开发了一些好玩的程序用于模拟高尔顿钉板，我们可以动态地观察这些小球自上而下滚动的时候，是如何形成正态分布的。一个使用 Java 开发的很漂亮的动态模拟可以在如下网页中观察到：http://www.math.psu.edu/dlittle/java/probability/plinko/index.html。

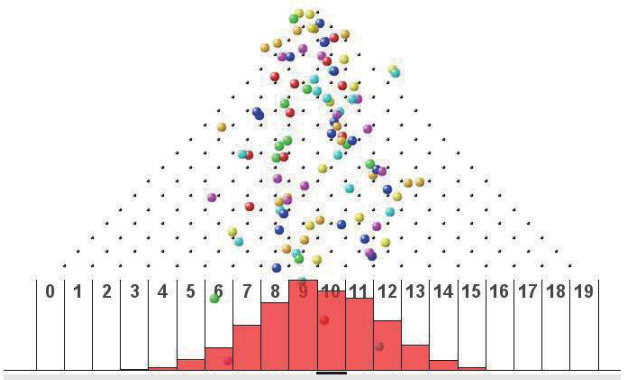
高尔顿在研究身高的遗传效应的时候，同时发现一个奇特的现象：高个子父母的子女，其身高有低于其父母身高的趋势，而矮个子父母的子女，其身高有高于其父母的趋势，即有“回归”到普通人平均身高去的趋势，这也是“回归”一词最早的含义。高尔顿用二维正态分布去拟合父代和子代身高的数据，同时引进了回归直线、相关系数的概念，从而开创了回归分析这门技术。



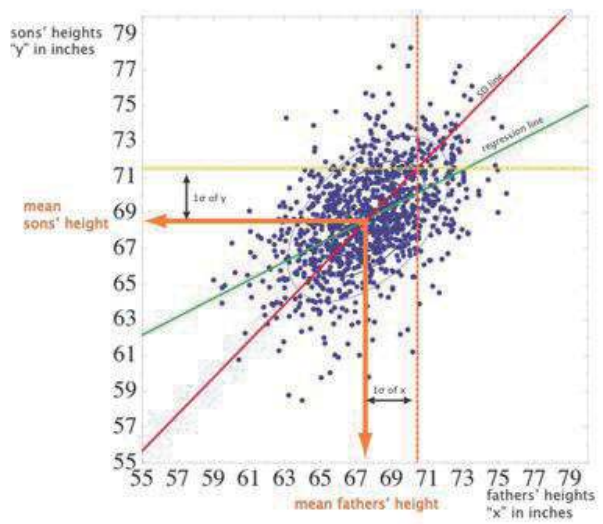
高尔顿钉板



高尔顿钉板解释遗传现象



高尔顿钉板动态模拟程序



儿子与父亲的身高回归线

可以说，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勒的思想，为数理统计学的产生奠定了基础。无论是凯特勒还是高尔顿，他们的统计分析工作都是以正态分布为中心的，在他们的影响下，正态分布获得了普遍认可和广泛应用，甚至是被滥用，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。

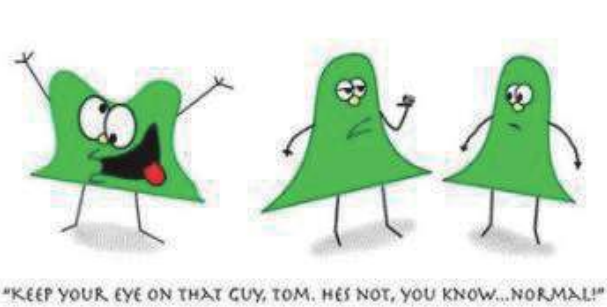
3 数理统计三剑客

最后，我们来到了20世纪，正态分布的命运如何呢？如果说19世纪是正态分布在统计学中独领风骚的话，20世纪则是数理统计学蓬勃发展、百花齐放的时代。1901年，高尔顿和他的学生卡尔•皮尔逊（Karl Pearson,1857-1936）、韦尔登（Walter Frank Raphael Weldon, 1860-1906）创办《生物计量》（*Biometrika*）杂志，成为生物统计学派的一面旗帜，引导了现代数理统计学的大发展。统计学的重心逐渐由欧洲大陆向英国转移，使英国在以后几十年数理统计学发展的黄金时代充当了领头羊。

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。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；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，逐渐被人们所重视。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受质疑，这促使人们研究这种情况下如何能找到更加准确的统计方法。

在这个背景之下，统计学三大分布 χ^2 分布、t 分布、F 分布逐步登上历史舞台。这三大分布现在的理科本科生都很熟悉。在历史上，这三个分布和来自英国的现代数理统计学的三大剑客有着密切的关系。

第一位剑客就是卡尔•皮尔逊，手中的宝剑就是 χ^2 分布。χ^2 分布这把宝剑最早的锻造者其实是物理学家麦克斯韦，他在推导空气分子的运动速度的分布的时候，发现分子速度在三个坐标轴上的分量是正态分布，而分子运动速度的平方 v^2 符合自由度为3的 χ^2 分布。麦克斯韦虽然造出了这把宝剑，但是真正把它挥舞得得心应手、游刃有余的是皮尔逊。在分布曲线和数据的拟合优度检验中，χ^2 分布可是一个利器，而皮尔逊的这个工作被认为是假设检验的开山之作。皮尔逊继承了高尔顿的衣钵，统计功力深厚，



非正态分布

卡尔·皮尔逊（1857-1936）

戈塞特（1876-1937）

费希尔（1890-1962）

数理统计三剑客

在19世纪末20世纪初很长的一段时间里，一直被数理统计武林人士尊为德高望重的第一大剑客。

第二位剑客是戈塞特（William Sealy Gosset, 1876-1937），笔名是大家都熟悉的学生氏（Student），而他手中的宝剑是 t 分布。戈塞特是化学、数学双学位，依靠自己的化学知识进酿酒厂工作，工作期间考虑酿酒配方实验中的统计学问题，追随卡尔·皮尔逊学习了一年的统计学，最终依靠自己的数学知识打造出了 t 分布这把利剑而青史留名。1908年，戈塞特提出了正态样本中样本均值和标准差的比值的分布，并给出了应用上极其重要的第一个分布表。戈塞特在 t 分布的工作开创了小样本统计学的先河。

第三位剑客是费希尔（Ronald Aylmer Fisher, 1890-1962），手持 F 分布这把宝剑，在一片荒芜中开拓出方差分析的肥沃土地。F 分布就是为了纪念费希尔而用他的名字首字母命名的。费希尔剑法飘逸，在三位剑客中当属费希尔的天赋最高，各种兵器的使用都得心应手。费希尔统计造诣极高，受高斯的启发，系统地创立了极大似然估计剑法，这套剑法现在被尊为统计学参数估计中的第一剑法。

费希尔还未出道，皮尔逊已经是统计学的武林盟主了，两人岁数相差了33岁，而戈塞特介于他们中间。三人在统计学擂台上难免切磋剑术。费希尔天赋极高，年少气盛；而皮尔逊为人强势，占着自己武林盟主的地位，难免固执己见，以大欺小，费希尔着实受了皮尔逊不少气。而戈塞特性格温和，经常在两位大侠之间调和。毕竟是长江后浪推前浪，一代新人换旧人，在众多擂台比试中，费希尔都技高一筹，而最终取代了皮尔逊成为数理统计学第一大剑客。

由于这三大剑客和统计三大分布的出现，正态分布在数理统计学中不再是一枝独秀，数理统计的领地基本上是被这三大分布抢走了半壁江山。不过这对正态分布而言并非坏事，我们细看这三大分布的数学细节：假设独立随机变量 $X_i \sim N(0, 1)$, $Y_j \sim N(0, 1)(i = 1 \cdots n,\ j = 1 \cdots m)$，则满足三大分布的随机变量可以如下构造出来

$$1.\ \chi_n^2 = X_1^2 + \cdots + X_n^2$$

$$2.\ t = \frac{Y_1}{\sqrt{\frac{X_1^2+\cdots+X_n^2}{n}}}$$

$$3.\ F = \frac{\frac{X_1^2+\cdots+X_n^2}{n}}{\frac{Y_1^2+\cdots+Y_m^2}{m}}$$

你看这三大分布哪一个不是正态分布的嫡系血脉，χ^2、t、F 这三大分布最初都是从正态分布切入进行研究的。所以正态分布在19世纪是武则天，进入20世纪就学了慈禧太后，垂帘听政了。或者，换个角度说，一个好汉三个帮，正态分布如果是孤家寡人恐怕也难以雄霸天下，有了统计学三大分布作为开国先锋为它开疆拓土，正态分布真正成为了傲视群雄的君王。

20世纪初，统计学这三大剑客成为了现代数理统计学的奠基人。以戈塞特为先驱，费希尔为主将，掀起了小样本理论的革命，事实上提升了正态分布在统计学中的地位。在数理统计学中，除了以正态分布为基础的小样本理论获得了空前的胜利，其它分布上都没有成功的案例，这不能不让人对正态分布刮目相看。在随后的发展中，相关回归分析、多元分析、方差分析、因子分析、布朗运动、高斯过程等等诸多概率统计分析方法陆续登上了历史舞台，而这些和正态分布密切相关的方法，成为推动现代统计学飞速发展的一个强大动力。

七、正态魅影

Everyone believes in it: experimentalists believing that it is a mathematical theorem, mathematicians believing that it is an empirical fact.

—— Henri Poincaré

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，使得偶然性现象在数量上被计算和预测成为可能。杰恩斯在《概率论沉思录》中提出了两个问题：

1. 为什么正态分布被如此广泛的使用？
2. 为什么正态分布在实践使用中非常的成功？

杰恩斯指出，正态分布在实践中成功地被广泛应用，主要是因为正态分布在数学方面具有多种稳定性质，这些性质包括：

1. 两个正态分布密度的乘积还是正态分布
2. 两个正态分布密度的卷积还是正态分布，也就是两个独立正态分布随机变量的和还是服从正态分布
3. 正态分布 $N(0, \sigma^2)$ 的傅立叶变换正规化为密度分布后还是正态分布
4. 中心极限定理保证了多个随机变量的求和效应将导致正态分布
5. 正态分布和其它具有相同均值、方差的概率分布相比，具有最大熵

前三个性质说明了正态分布一旦形成，就容易保持该形态的稳定，兰登对于正态分布的推导也表明了，正态分布可以吞噬较小的干扰而继续保持形态稳定。后两个性质则说明，其它的概率分布在各种的操作之下容易越来越靠近正态分布。正态分布具有最大熵的性质，所以任何一个对指定概率分布的操作，如果该操作保持方差的大小，却减少已知的知识，则该操作不可避免地增加概率分布的信息熵，这将导致概率分布向正态分布靠近。

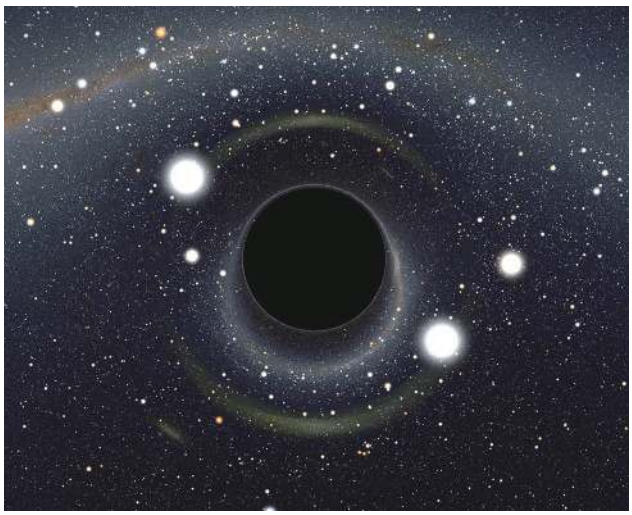
正由于正态分布的多种稳定性质，使得它像一个黑洞一样处于一个中心位置，其它的概率分布形式在各种操作之下都逐渐向正态分布靠拢，杰恩斯把它描述为概率分布中的重力现象（gravitating phenomenon）。

我们在实践中为何总是选择使用正态分布呢，正态分布在自然界中的频繁出现只是原因之一，杰恩斯认为还有一个重要的原因是正态分布的最大熵性质。在很多时候我们其实没有任何的知识知道数据的真实分布是什么，但是一个分布的均值和方差往往是相对稳定的。因此我们能从数据中获取到的比较好的知识就是均值和方差，除此之外没有其它更加有用的信息量。因此按照最大熵的原理，我们应该在给定知识的限制下，选择熵最大的概率分布，而这恰好就是正态分布。即便数据的真实分布不是正态分布，由于我们对真实分布一无所知，如果数据不能有效提供除了均值和方差之外的更多的知识，按照最大熵的原理，正态分布就是这时候的最佳选择。

无处不在的正态分布

当然正态分布还有更多令人着迷的数学性质，我们可以欣赏一下：

* 二项分布 $B(n, p)$ 在 n 很大时逼近正态分布 $N(np, np(1-p))$
* 泊松分布 $Poisson(\lambda)$ 在 λ 较大时逼近正态分布 $N(\lambda, \lambda)$
* $\chi^2_{(n)}$ 在 n 很大的时候逼近正态分布 $N(n, 2n)$
* t 分布在 n 很大时逼近标准正态分布 $N(0, 1)$
* 正态分布的共轭分布还是正态分布
* 几乎所有的极大似然估计在样本量 n 增大的时候都趋近于正态分布
* 克拉美分解定理（之前介绍过）：如果 X, Y 是独立的随机变量，且 $S = X + Y$ 是正态分布，那么 X, Y 也是正态分布
* 如果 X, Y 独立且满足正态分布 $N(\mu, \sigma^2)$，那么 $X + Y$，$X - Y$ 独立且同分布，而正态分布是唯一满足这一性质的概率分布
* 对于两个正态分布 X, Y，如果 X, Y 不相关则意味着 X, Y 独立，而正态分布是唯一满足这一性质的概率分布



正态黑洞

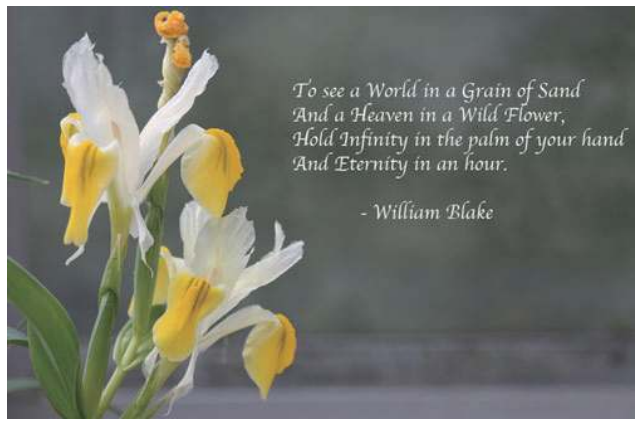
八、大道至简，大美天成

$$\overline{X}=\frac{X_1+X_2+\cdots+X_n}{n}$$

算术平均，极其简单而朴素的一个式子，被人们使用了千百年，在其身后隐藏着一个美丽的世界，而正态分布正是掌管这个美丽世界的女神。正态分布的发现与应用的最初历史，就是数学家们孜孜不倦地从概率论和统计学角度对算术平均不断深入研究的历史。中心极限定理在 1773 年棣莫弗的偶然邂逅的时候，它只是一粒普通的沙子，两百多年来吸引了众多的数学家，这个浑金璞玉的定理不断地被概率学家们精雕细琢，逐渐地发展成为现代概率论的璀璨明珠。而在统计学的误差分析之中，高斯窥视了造物主对算术平均的厚爱，也发现了正态分布的美丽身影。殊途同归，那是偶然中的必然。一沙一世界，一花一天国，算术平均或许只是一粒沙子，正态分布或许只是一朵花，它们却包含了一个广阔而美丽的世界，几百年来以无穷的魅力吸引着科学家和数学家们。

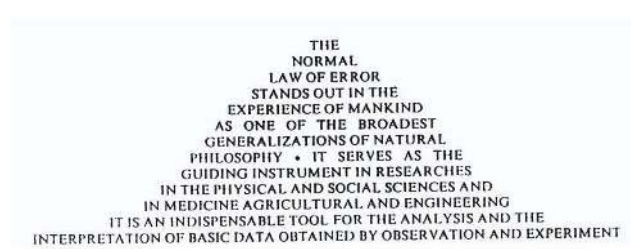
高尔顿对正态分布非常推崇与赞美，1886 年他在人类学研究所的就职演讲中说过一段著名的话：“我几乎不曾见过像误差呈正态分布这么美妙而激发人们无穷想象的宇宙秩序。如果古希腊人知道这条曲线，想必会给予人格化乃至神格化。它以一种宁静无形的方式在最野性的混乱中实施严厉的统治。暴民越多，无政府状态越显现，它就统治得越完美。他是无理性世界中的最高法律。当我们从混沌中抽取大量的样本，并按大小加以排列整理时，那么总是有一个始料不及的美妙规律潜伏在其中。”

概率学家卡茨（Mark Kac, 1914-1984）在他的自述传



记《机遇之谜》（*Enigmas of chance: An autobiography*）中描述他与正态分布的渊源：“我接触到正态分布之后马上被他深深地吸引，我感到难以相信，这个来自经验直方图和赌博游戏的规律，居然会成为我们日常生活数学的一部分。”另一位概率学家米歇尔 • 洛伊（Michel Loève, 1907-1979）说：“如果我们要抽取列维的概率中心思想，那我们可以这样说，自从 1919 年以后，列维研究的主题曲就是正态分布，他一而再再而三地以她为出发点，并且坚决地又回到她……他是带着随机时钟沿着随机过程的样本路径作旅行的人。”美国国家标准局的顾问约登（W. J. Youden ）用如下一段排列为正态曲线形状的文字给予正态分布极高的评价，意思是说：误差的正态分布规律在人类的经验中具有“鹤立鸡群”的地位，它在物理、社会科学、医学、农业、工程等诸多领域都充当了研究的指南，在实验和观测数据的解读中是不可或缺的工具。

几乎所有的人都或多或少地接触数学，虽然各自的目的不同，对数学的感觉也不同。工程师、科学家们使用数学是因为它简洁而实用，数学家们研究数学是因为它的美丽动人。像正态分布这样，既吸引着无数的工程师、科学家，在实践中被如此广泛地应用，又令众多的数学家为之魂牵梦绕的数学存在，在数学的世界里也并不多见。我在读研究生的时候，经常逛北大未名 BBS 的数学板，有一个叫 ukim 的著名 ID 在精华区里面留下了一个介绍数学家八卦的系列



正态误差态分布律

“Heroes in My Heart”，写得非常的精彩，这些故事在喜欢数学的人群中也流传广泛。最后一个八卦是关于菲尔兹奖得主法国数学家托姆（René Thom）的，它曾经令无数人感动，我也借用来作为我对正态分布的八卦的结语：

> 在一次采访当中，作为数学家的托姆同两位古人类学家讨论问题。谈到远古的人们为什么要保存火种时，一个人类学家说，因为保存火种可以取暖御寒；另外一个人类学家说，因为保存火种可以烧出鲜美的肉食。而托姆说，因为夜幕来临之际，火光摇曳妩媚，灿烂多姿，是最美最美的……

九、推荐阅读

> *All knowledge is, in the final analysis, history.*
> *All sciences are, in the abstract, mathematics.*
> *All methods of acquiring knowledge are, essentially, through statistics.*
>
> 在终极的分析中，一切知识都是历史；
> 在抽象的意义下，一切科学都是数学；
> 在理性的基础上，所有的判断都是统计学。
>
> —— *C. R. Rao*

本人并非统计学专业人士，只是凭个人兴趣做一点知识的传播。对统计学历史知识的介绍，专业性和系统性都不是我的目的，我更在乎的是趣味性，因为没有趣味就不会有传播。如果读完这段历史会让你觉得正态分布更加亲切，不再那么遥不可及，那我的目的达到了。如果正态分布是一滴水，我愿大家都能看到它折射出的七彩虹。

本文所使用的大多是二手资料，有些历史细节并没有经过严格地考证，对于历史资料一定程度上按照个人喜好做了取舍，本文主要基于如下的资料写成，若对历史细节感兴趣，推荐阅读。

* 陈希孺，数理统计学简史，湖南教育出版社，2000
* 蔡聰明，誤差論與最小平方法，数学传播 21(3):3-13，1994
* 吴江霞，正态分布进入统计学的历史演化，2008
* E.T. Jaynes, Probability Theory: The Logic of Science，Cambridge University Press，2003
* Saul Stahl, The Evolution of the Normal Distribution, Mathematics Magazine, 1996
* Kiseon Kim, Georgy Shevlyakov, Why Gaussianity, IEEE Signal Processing Magazine, 2008
* Stephen M. Stigler, The History of Statistics: The Measurement of Uncertainty before, Belknap Press of Harvard University Press, 1990
* L. Le Cam, The Central Limit Theorem Around 1935, Statistical Science 1(1):78-91, 1986
* Hans Fischer, A History of the Central Limit Theorem: From Classical to Modern Probability Theory, Springer, 2010

作者简介：靳志辉，北京大学计算机系计算语言所硕士，日本东京大学情报理工学院统计自然语言处理方向博士，目前在腾讯科技（北京）有限公司担任研究员，主要参与计算广告学相关的业务，工作内容涉及统计自然语言处理和大规模机器学习方面的工程研究工作。