对高斯新分布的再认识

孔建新
（云南省会泽县老年科技工作者协会，广东 中山 528400）

摘要：高斯新分布是建立在高斯分布原理基础上的一种新的分布形态。打破了高斯分布不能描述不对称分布的限定条件。达到精确地描述近似正态分布和偏斜分布的目的。从而补充和完善了高斯分布的理论。重新认识高斯分布的参数并对其进行分解与剖析。以正态分布的数学表达式为基础导出高斯新分布。从应用的角度对高斯新分布进行再认识。

关键词：高斯分布；高斯新分布；正态分布；偏斜分布；期望值；期望差
中图分类号：O211.3

Gaussian new-distribution of the again understanding

KONG Jianxin
(Yunnan province Huize county elderly associatiion of science and technology, GuangDong ZhongShan 528400)

Abstract: Gaussian new-distribution is based at the Gaussian distribution of the on a principle based of a new form. Breaking the Gaussian distribution can not describe of the asymmetric distribution. The accurately describe the approximate normal distribution and skewed distribution of the purpose. It supplement and perfect the theory of Gaussian distribution. Re-understand the parameters of Gaussian distribution and its decomposition and analysis. Mathematical expression of the normal distribution is derived based on the Gaussian new-distribution. From the application of angle view of re-understanding of the Gaussian new-distribution

Keywords: Gaussian distribution; Gaussian new-distribution; normal distribution; skewed distribution; expected value; expected deviation

0 引言

统计对象的广泛性使其具有不同的分布状态。随机变量的频数分布不仅存在对称的正态分布，还客观存在不对称的偏斜分布[1]。偏斜分布是指：随机变量的频数分布在单峰的条件下呈现出的不对称的偏斜形态。正态分布与偏斜分布是相互对应在对称与不对称关系之中的两种分布形式，归纳称为高斯新分布[2]。偏斜分布严格有别于泊松分布、对数分布、指数分布等一类的偏态分布。它与偏态分布是具有不同内涵且毫不相关的两个概念。上述界定的偏斜分布和正态分布是本课题探讨的要点不涉及以外的任何分布形态。

高斯分布精确地描述了对称的正态分布，则不能精确地描述近似的正态分布，而只能以正态分布来拟合近似的正态分布。高斯分布应用的局限性说明其不能描述不对称的偏斜分布。基于这样的认识，从高斯分布的位置特征值μ和离散特征值σ这两个参数入手对其具有对称性的限制条件进行分析，不难得出结论：若对称则期望值必然处在分布的中心位置，且两边的离散特征值相等；若不对称则期望值不在分布中心的位置，显然两边的离散特征值就不相等。若要反映期望值两边各异的离散变异指标需要用左右离散特征值来分别表达。由此引出对高斯分布离散特征值参数σ的分解，左边为σ_- 右边为σ_+，由它们分别来反映期望值两边的离散程度就能够精确计算随机变量分别从左右两边与期望值的变异指标。使（$\mu-\sigma_-$，$\mu+\sigma_+$）能精确地计算出分布曲线两边的拐点。从而达到精确描述近似正态分布乃至精确描述偏斜分布的目的，进而解决了计算偏斜分布任意区间概率的问题[2]。

本文基于中国科技论文在线众多同行专家提出的宝贵意见和建议。从传统正态分布性质

作者简介：孔建新, (1950-),男,高级统计师，主要研究方向：高斯分布. E-mail: kongfanjx@163.com

的特征入手进行剖析，并重新审视高斯分布的参数并对其进行重新认识，用规范严谨的方法以正态分布为依据通过对离散特征参数的分解来导出高斯新分布，对其再认识。

1 偏斜分布存在的普遍性是离散特征参数分解的统计背景

偏斜分布客观存在的普遍性主要表现在随机变量频数分布也同时具有随机分布的特征，这是不以人的意志为转移的自然属性。因为不同的统计对象各有不同的规范要求，就有不同的分布形态。随机变量频数分布的统计对象按不同的规范可分为“单侧规范、双侧规范、多侧规范、自然规范四大类。[2]”也称为单侧控制、双侧控制、多侧控制、自然控制。

属于单侧规范的统计对象如食品工业的产品，有效成分不能低，需要控制下限；有害元素不能高，需要控制上限。一般服从偏斜分布。

属于双侧规范的统计对象如加工一个零件，随机变量通过人为的控制容易达到标准值的要求，标准值处于规范限内的中心位置，所以容易围绕标准规定的尺寸波动，服从正态分布或近似正态分布。但是平均值不一定就等于标准值。还有可能服从不对称的偏斜分布。

属于多侧规范的统计对象如钻井、炮弹攻击目标的弹着点、卫星回收的着陆点、体育项目中的射击弹着点、射箭中靶点、跳伞着陆点等等，在平面坐标规定了标准值的点，需要四周的多侧控制，随机变量很难达到标准值的点上，若按随机变量距离标准值点的误差统计，一般服从偏斜分布。

属于自然规范的统计对象如气温的变化分布、降雨量分布，以及像人口年龄分布、人体身高分布等等。一般是指的是自然形成的随机变量的统计对象，难以进行人为的控制。其频数分布有可能服从正态分布或偏斜分布或双峰分布或多峰分布或其它类型的分布。

以上所述的四种分类的统计对象说明对称的正态分布在实际中并非普遍。“在过去的数十年里，国际性的、地区性的、各个国家的标准化组织以及行业协会就过程能力及性能颁布了许多标准，但是这些标准都假设过程是处于统计受控状态，即稳定的、正态分布的过程行为，然而人们在对生产过程进行全面分析后，发现只有非常少的过程表现为正态分布的稳定状态。[3]”它传达了一个重要的信息：正态分布不具有普遍性。由此对应偏斜分布数学模型的建立则是数学发展应用于统计的必然结果。它是以分解高斯分布的离散特征参数为统计背景。

2 重新认识高斯分布的参数

正态分布是“德国数学家高斯（C.F.Gauss，1777～1855）在研究误差理论时最早使用这一分布，所以正态分布又称高斯分布。[4]”高斯分布所描述的是随机变量的频数分布在单峰的条件下呈现对称的形态。它有两个参数：位置特征值 μ 和离散特征值 σ。

在正态分布的条件下“均值给出对分布中心的度量。[5]”对称性使平均值处于分布中心的位置。在此需要引出一个位置特征值的新概念——分布中心值（distribution center value）它是指：随机变量在分布区间里处于值域范围内中心位置的数。即：分布中最大值加最小值除以 2 。记为：dcv 。其作用是：量度其它位置特征值处在分布区间是否距中的标准。它与中位数是完全不同的概念。若呈现的是近似正态分布，说明平均值必然偏移分布中心值的位置，而期望值或众数则不一定在分布中心值的位置，在不对称的条件下尤其是这样。从另一方面来说不对称则众数不等于平均值 $M_O \neq \bar{x}$，传统偏度的计算公式就是应用众数与平均值离差来反映分布的偏斜程度。由于期望值 μ 与众数 M_O 是等价的同一概念，由 $\mu = M_O$ 可以从众数 M_O 与平均值 $\bar{x}$ 不是等价的概念容易推出期望值 μ 与平均值 $\bar{x}$ 也不是等价的概念。

离散特征值σ被定义为：随机变量与平均值离差平方和平均的平方根。传统的定义被锁定在规范术语“标准差”的概念中，已经满足不了分布呈现不对称时的要求，再者期望值与平均值不是等价的概念。所以重新认识的离散特征值σ是为了与原来的概念相区别，所以需要重新命名和重新定义。将其命名为：期望差[6]，重新定义为：随机变量与期望值离差平方和平均的平方根。容易理解，在对称的条件下“期望差”与“标准差”是等同的，在不对称的条件下“期望差”与“标准差”涵义则完全不同。可追溯如下。

在质量管理中“双侧公差且T的中心（标准中心）和分布中心μ相重合的情况，这仅是实际生产运作中的一种理想状态。在实际生产过程中往往很难保证标准中心与分布中心重合。[7]”以上所述的标准中心指的就是标准值。为了便于对问题的讨论有必要先对标准值(standard value) [8]进行定义：统计对象对应相关的质量指标达到的最好水平值作为对其为满足质量要求的准则和依据。简言之是指：产品质量指标达到最好水平目标值的标准称为标准值。

正态分布是高斯在研究误差理论时最早证明和使用的这一分布已经是一个无可争议的定论。追根溯源其研究的统计对象必然是误差分布。依据误差的定义：“某量值的给出与其客观真值之差为误差。[9]”容易理解客观真值就是指的标准值。所以误差是随机变量与标准值的离差，而不一定是随机变量与平均值的离差。因为平均值不一定等于标准值。当平均值等于标准值时，随机变量与平均值的离差是误差。当平均值不等于标准值时，随机变量与平均值的离差就不是误差。所以，随机变量与标准值离差平方和平均的平方根才是名副其实的标准差。随机变量与平均值离差平方和平均的平方根就不是“标准差”而其实是平均差[10]。

上述涉及到两个不同的位置特征值：标准值和平均值。它们具有不同的内涵，其作用也不尽相同。标准值是标准规定的值，是相对不变的常数。而平均值则是通过随机变量的计算结果来确定的变量，所以有可能等于标准值也有可能不等于标准值。

随着社会的进步与科学的发展促进了检测计量手段不断提高，人们越来越注重随机变量与标准值的离差，具体表现在质量管理中。实际上随机变量与平均值的离差在工作中是不常用的。比如加工一个零件，一定是用标准值来检验是否合格，不可能用平均值来检验。高斯分布从原来的误差理论推广到更普遍的反映离散程度的指标为什么会惯用随机变量与平均值的离差？原因之一“期望是随机变量最重要的位置特征，它表示随机变量的平均大小和位置，期望愈大，随机变量平均愈大，位置愈靠数轴的右边。它与算术平均在概念上有严格的区别，期望是对随机变量或分布而言，而算术平均是对样本而言。样本算术平均对期望来说是一个无偏、一致、有效的估计值。[11]”以上文献给出的论断是建立在正态分布的假设条件下，若在偏斜分布的条件下，样本算术平均对期望来说则是一个有偏、不一致、无效的估计值。所依据的是下一节给出期望值的新定义。

有文献指出：“位置数字特征，一般是指平均数而言，它是用以表达数据集中位置的，是实验室出具品质证书的主要指标。我们在实验中获得的数据是分散的，必须把它们集中起来，反映其共同趋向的平均水平。也就是说它代表了数据的集中位置。反映了数据的总的情况或者说它具有代表性。平均数一般包括算术平均、加权算术平均、几何平均值、中位数和众数等。[11]”文献把这些位置数字特征统称为：平均数，并把中位数和众数归属于平均值，这是锁定在正态分布条件下的真正原因。在对称的正态分布的情况下它们的确都代表平均的水平，况且它们还不一定都等于标准值。在不对称的情况下平均值、中位数和众数已经具有不同的涵义，表现在它们与分布中心值的离差有所不同，众数反映的是集众位置，是平均值集中还是中位数集中则需要与分布中心值的离差来决定。所以它们反映统计对象的变异指标

各有不同的作用。统计实践说明：平均值应用的广泛性使之在统计的实际应用中客观地存在一定的误区。“平均数是统计分析中使用频率最高的用来描述数据分布集中程度的统计指标，平均数的特点是容易受到数据取值中极端值的影响，这就使得平均数看似精确的背后往往隐藏着未知的陷阱[12]。”由此说明平均值的使用一定要考虑统计研究对象的适用条件，滥用带来的结果必然会出现矛盾。所以“统计方法的正确应用有助于我们认清事物的真像、发现事物变化的数量界限、揭示事物发展的内在规律。相反，统计方法的错误使用，将造成事实的扭曲、读者的误解，甚至决策的失败。统计在许多应用领域都存在着不同程度的误用[12]。”它给出一个重要的警示：准确把握概念内涵的意义是统计分析的重要原则之一，因为数据的来源所针对的是概念准确涵义的统计对象，而不能指鹿为马。同时也是为分解期望差找准位置特征值的理论依据。有关期望差与其它不同位置特征值的关系还将在下一节进一步说明。

以上对标准值、平均值，标准差、平均差概念的重新认识其目的是便于与原来的习惯概念相区别。

3 传统正态分布性质引出对问题的剖析

正态分布的对称性决定了期望值两边的变异指标相等。而近似正态分布则出现不对称的因素，显然期望值两边的变异指标就有差异而不相等。在对称的情况下仅需一个期望差足以描述期望值两边的离散程度。在不对称的情况下这个期望差就显然不能分别描述期望值两边的离散程度。所以需要将期望差 σ 分解为左期望差 σ_{-} 和右期望差 σ_{+}。统计学家范剑青教授早已指出：“统计的最大问题在于模型误差，局部建模的优点在于可以大大降低误差。[13]”利用左右期望差能准确计算分布曲线的拐点。

以下从传统正态分布的性质为理论依据，以平均值和期望值的离差关系探讨高斯分布离散特征参数期望差进行分解的必要性。

“正态分布密度函数 $f(x)$ 有以下性质：

（1） 曲线 $f(x)$ 关于直线 $x=\mu$ 对称；

（2） 当 $x=\mu$ 时，$f(x)=(\sqrt{2\pi}\,\sigma)^{-1}$ 是 $f(x)$ 的最大值；

（3） 在 $x=\mu\pm\sigma$ 处曲线有拐点，当 $x\rightarrow\pm\infty$ 时，$f(x)$ 以 x 轴为渐近线；

（4） μ、σ 分别是描述 ξ 的集中位置和离散度的参数。[14]”

根据以上正态分布的性质，以此为基础结合客观存在不对称的分布对其进行剖析。

性质（1）确定了正态分布的对称性，毫无疑问期望值等于平均值 $\mu=\bar{x}$。对称性使传统平均值作为高斯分布位置特征值的参数无疑是对的。但是，近似的正态分布并不完全对称，平均值偏移期望值呈现偏斜，所以期望值就不等于平均值 $\mu\neq\bar{x}$。平均值 $\bar{x}$ 就与高斯分布位置特征值的参数脱离了相等的关系。说明高斯分布位置特征参数与平均值无关，这是问题之一。

性质（2）确定了期望值是分布曲线最大值点上的取值。以此为依据对期望值（expected value）重新定义：随机变量的频数分布在单峰的条件下是分布曲线最大值点上的取值。满足随机变量 $x=\mu$ 时，$f(\mu)=(\sqrt{2\pi}\,\sigma_{-})^{-1}$ 或 $f(\mu)=(\sqrt{2\pi}\,\sigma_{+})^{-1}$ 为 $f(x)$ 最大值的条件。期望值的定义与众数的定义“使概率（或密度）达到极大值的随机变量的取值。[14]”是完全等价的。期望值和众数都是分布的峰值，所以期望值、众数、峰值都是等价的不同称谓。由此说明：期望值反映的不一定是集中的位置特征而一定是集众的位置特征。平均值是集中的位置特征但不一定是集众的位置特征。这是期望值与平均值不同的显著区别之一。上一节所述在正态分布的条件下平均值、中位数、众数三数相等。此时，这三个位置特征值在单峰

对称的条件下都满足期望值的定义。若在单峰不对称的条件下就仅有众数满足期望值定义的要求。定义确定了分布位置特征值的参数是分布曲线最大值点上的取。根据性质（1）（2）和期望值的定义，容易推出：当 $x=\bar{x}$ 时，

$$f(\bar{x}) = (\sqrt{2\pi}\,\sigma_-)^{-1}exp[-(\bar{x}-\mu)^2(2\sigma_-^2)^{-1}] \text{ 或} \quad (1)$$

$$f(\bar{x}) = (\sqrt{2\pi}\,\sigma_+)^{-1}exp[-(\bar{x}-\mu)^2(2\sigma_+^2)^{-1}] \quad (2)$$

不一定是 $f(x)$ 的最大值。

证明：根据性质（3）当 $x=\mu$ 时， $f(x)=(\sqrt{2\pi}\,\sigma)^{-1}$ 是 $f(x)$ 的最大值。

若 $\bar{x}=\mu$ 时，

（1）式右边自然数 e 的指数 $-(\bar{x}-\mu)^2(2\sigma_-^2)^{-1}=0$ ，有 $e^0=1$

则（1）式 $f(\bar{x})=(\sqrt{2\pi}\,\sigma_-)^{-1}$ 是 $f(x)$ 的最大值。

若 $\bar{x}\neq\mu$ 时，（1）式右边自然数 e 的指数 $-(\bar{x}-\mu)^2(2\sigma_-^2)^{-1}\neq 0$ ，

则（1）式 $f(\bar{x})=(\sqrt{2\pi}\,\sigma_-)^{-1}exp[-(\bar{x}-\mu)^2(2\sigma_-^2)^{-1}]$ 不是 $f(x)$ 的最大值。

以上证明平均值不一定是分布曲线最大值点上的取值是由平均值与期望值是否相等决定的。在不对称的情况下平均值与分布位置特征值的参数产生偏移而不相等。期望值是分布曲线最大值点上的取值与平均值不一定是分布曲线最大值点上的取值，其内涵意义随条件的改变而不同。从以上剖析性质（1）的论述和近似的正态分布可以说明它们的对比关系，在正态分布条件下必然有 $\mu=\bar{x}$ ，在近似正态分布条件下期望值接近平均值，有 $\mu\approx\bar{x}$ 而导致 $\mu\neq\bar{x}$ 。因为，近似的正态分布必然是不完全对称的分布形态，表明期望值两边已经有差异而出现不对称的元素，从问题之一得出：只要呈现不对称，期望值就不等于平均值，就可以导出平均值必然不是分布曲线最大值点上的取值的结论。这是问题之二。又由于期望值的定义是依据正态分布性质（2），它反映了期望值 μ 的本质属性，而与平均值 $\bar{x}$ 只有在特定的正态分布的条件下有相等的数量关系 $\mu=\bar{x}$；而在偏斜分布下则有不相等的数量关系 $\mu\neq\bar{x}$ 。以上从不同的角度反复讨论平均值与期望值的目的是要证实传统离散特征值 σ 的定义中所确定的平均值显然是有误的不准确的，只有众数（期望值）取而代之才能精确地表达其内涵。还可以从以下左右期望差的定义的论述中得到进一步的验证。

性质（3）确定了在 $x=\mu\pm\sigma$ 处分布曲线的拐点。当期望值等于平均值时，且对称无疑 $x=\mu\pm\sigma$ 是分布曲线的拐点。由问题之二引出的当期望值不等于平均值时，则不对称 $x=\mu\pm\sigma$ 就不是分布曲线的拐点。因为当期望值不等于平均值时，分布不对称，由此得出以期望值为中心左右两边不对称。而传统的 σ 是随机变量与平均值离差平方和平均的平方根，而不是随机变量与期望值离差平方和平均的平方根。这是问题之三。上节所引出的新概念期望差（expected deviation）及定义源于性质（3）。由于存在不对称的因素期望值两边离散程度的变异指标各异，期望差就需要分解为左右两个。

将左期望差[6]（expected left-deviation）定义为：小于等于期望值随机变量与期望值离差平方和平均的平方根。符号记为 σ_-。

将右期望差[6]（expected right-deviation）定义为：大于等于期望值随机变量与期望值离差平方和平均的平方根。符号记为 σ_+。

不对称所引出的左期望差 σ_- 和右期望差 σ_+ 。满足 $\mu-\sigma_-$，$\mu+\sigma_+$，分别为不对称分布曲线两边的拐点。以上问题的提出还说明：当分布对称时，左期望差与右期望差相等，则有 $\sigma_-=\sigma_+=\sigma$ 。由此导出一个重要的结论：随机变量的频数分布若要精确反映不对称的

分布形态，就有必要将高斯分布离散特征的参数 σ 分解为左 σ_-和右 σ_+ 两个，以判定 $\sigma_- = \sigma_+$对称，$\sigma_-\neq\sigma_+$不对称。从而解除受限于对称的条件。

性质（4）仅仅限于描述对称的正态分布，而不能描述不对称的分布，这是问题之四。由问题之三引入的不对称元素。性质（4）可以表述为：μ 是描述 ξ 的集众位置，σ_-和 σ_+ 分别是描述期望值左右两边离散度的参数。这个表述不但适合于正态分布还适合于偏斜分布。这是补充完善高斯分布理论的一个重要的基本内容。

以上对高斯分布参数的剖析满足正态分布的条件可归纳为以下三条：

（1）期望值等于平均值或期望值等于分布中心值；

（2）期望值两边的左右频数相等；

（3）期望值两边的左期望差与右期望差相等。

呈现正态分布必须同时满足以上三个条件说明：对称是相对的，不对称是绝对的。由此验证一个事实，服从正态分布受到诸多因素的限制，不具有普遍性。偏斜分布则是客观存在的普遍现象。所以重新审视正态分布的目的是要证明对应不满足以上三个条件时所呈现不对称偏斜分布的普遍性从而说明高斯分布参数分解扩展的必要性。

偏斜分布存在的客观性需要建立一个在正态分布理论基础上能描述不对称的新分布形态。正如德国数学家汉克尔（H. Hankel, 1839～1873）所言：“在大多数数学科里，一代人的建筑往往被另一代人所摧毁，一个人的创造被另一个人所破坏；唯独数学，每一代人都在古老的大厦上添加一层楼。[15]”新分布是建立在高斯分布离散特征值参数分解扩展的基础上，由于源于高斯分布原理所以新分布命名为：高斯新分布。

高斯分布“由于具有分布密度形式的误差分布对后来的数理统计及其它相关学科的发展影响极大, 故此命名为高斯分布或正态分布。德国 10 马克纸币上有一高斯头像且在头像后面印有高斯分布钟型曲线，这表明在高斯的一切科学贡献中, 对人类文明影响最大的莫过于这一项。到了 19 世纪可以说成了高斯分布统治的年代，而到了 20 世纪小样本理论建立后，高斯正态分布又显示了它的强大的优越性。[16]”“正态分布是所有概率分布中最重要的一种分布，这有实践与理论两方面的原因。实践方面的原因在于其常见性。从理论方面说，正态分布可以导出一些其他的分布，而某些分布在一定条件下又可用正态分布来近似，因此正态分布在理论研究中具有重要地位。[17]”以上文献中的论述一方面确立了高斯分布的重要地位，另一方面说明了高斯新分布正是由高斯分布的原理所导出。以下给出推论。

4 高斯新分布数学表达式的导出

以上对高斯分布参数分解扩展的推论，其目的是建立一个既能描述对称分布又能描述不对称分布的数学表达式。“自然界的事物基本上都很简单，所有的基础原理及主要问题都可以用数学方式表达，这是应用数学家的一个信仰。”这句话出自应用数学大师林家翘的一次演讲。他归国八年以来绝少面对媒体，但几乎每一次出现在镜头前他都会强调这一点——应用数学的意义在于揭示自然界和社会实际问题的规律。[18]由于新分布的数学表达式源于高斯分布参数分解后的扩展，所以称为：高斯新分布。其理由和根据可以从以下对正态分布数学表达式的展开推论中得到验证。

4.1 正态分布可导出的新定义及性质

若$-\infty<\mu<\infty$，$\sigma>0$ 为两个实数，则由下列密度函数

$$f(x)=(\frac{1}{\sqrt{2\pi}\sigma})exp[-\frac{(x-\mu)^2}{2\sigma^2}],\ -\infty< x <\infty \quad (3)$$

确定的随机变量 X 的分布称为正态分布，记为 $N(\mu,\sigma^2)$。[4]

以上是对传统正态分布的定义，根据上述对问题的剖析将离散程度的参数整体期望差（以下简称为：期望差）σ 分解为左期望差 σ_-，右期望差 σ_+，将整体期望方差（以下简称为：期望方差）σ^2 分解为左期望方差 σ_-^2、右期望方差 σ_+^2，将分解的不对称参数的元素溶入（3）式，由此导出以下（4）式，为此就可以对正态分布重新定义如下：

若 $-\infty<\mu<\infty$，$\sigma_->0$，$\sigma_+>0$ 为三个实数，点在 μ 处连续，且 $\sigma_-=\sigma_+$ 则由下列密度函数：

$$f(x)=\begin{cases}(\sqrt{2\pi}\sigma_-)^{-1}exp[-(x-\mu)^2(2\sigma_-^2)^{-1}] & x\leqslant\mu\\ \\ (\sqrt{2\pi}\sigma_+)^{-1}exp[-(x-\mu)^2(2\sigma_+^2)^{-1}] & x\geqslant\mu\end{cases} \quad (4)$$

确定的随机变量 X 的分布称为正态分布。记为 $N(\mu,\sigma_-,\sigma_+)$。

以上由（3）式导出的（4）式表明：在正态分布的条件下（3）式与（4）式完全等价。反过来说，当（4）式中的 $\sigma_-=\sigma_+$ 则（4）式便还原为（3）式。所以，（4）是由（3）推导而出。

由（4）式可以重新对正态分布的性质补充描述如下：

根据正态分布的新定义，其密度函数 $f(x)$有以下性质：

（1） 当 $\sigma_-=\sigma_+$时，曲线 $f(x)$关于直线 $x=\mu$ 对称，呈正态分布；

（2） 当 $x=\mu$ 时，且点在 μ 处连续，$f(\mu)=(\sqrt{2\pi}\sigma_-)^{-1}$ 或 $f(\mu)=(\sqrt{2\pi}\sigma_+)^{-1}$ 为 $f(x)$的最大值；

（3） 在 $x=\mu-\sigma_-$，$x=\mu+\sigma_+$ 处曲线有拐点，当 $x\to\pm\infty$时，$f(x)$以 x 轴为渐近线；

（4）μ 是描述 ξ 的集众位置，σ_-和 σ_+分别是描述期望值左右两边离散度的参数。

将以上（3）式分解为（4）式还可以推论，在单峰的条件下随机变量的频数分布不论对称与否，容易从（4）式分解为以下左右函数表达式：

左函数：$f_-(x)=(\sqrt{2\pi}\sigma_-)^{-1}exp[-(x-\mu)^2(2\sigma_-^2)^{-1}]\quad x\leqslant\mu$ 单增函数；

右函数：$f_+(x)=(\sqrt{2\pi}\sigma_+)^{-1}exp[-(x-\mu)^2(2\sigma_+^2)^{-1}]\quad x\geqslant\mu$ 单减函数。

由于以上左右两个函数来自于正态分布的（4）式，所以可以分别称为正态分布左函数 $f_-(x)$和正态分布右函数 $f_+(x)$。又由于（4）式源自于（3）式，所以以上左右函数的表达式中点在 μ 处是连续的。从 $\sigma_-=\sigma_+$ 得出以 μ 为对称轴对称。如果当 $\sigma_-\neq\sigma_+$时，则以 μ 为中轴不对称。就此推广为任意偏斜分布都可以用以上左右两个函数来表达。由此得出这样的结论：应用正态分布的数学表达式通过分解离散特征值参数使之化为左正态分布函数和右正态分布函数就能推导出以下偏斜分布。

4.2 偏斜分布的定义及性质

通过以上对正态分布的分解由此导出偏斜分布的定义如下：

若 $-\infty<\mu<\infty$，$\sigma_->0$，$\sigma_+>0$ 为三个实数，点在 μ 处连续，且 $\sigma_-\neq\sigma_+$ 则由下列密度函数：

$$f(x)=\begin{cases}(\sqrt{2\pi}\ \sigma_-)^{-1}\exp[-(x-\mu)^2(2\sigma_-^2)^{-1}] & x\leqslant\mu\\ (\sqrt{2\pi}\ \sigma_+)^{-1}\exp[-(x-\mu)^2(2\sigma_+^2)^{-1}] & x\geqslant\mu\end{cases}\qquad(5)$$

确定的随机变量 X 的分布称为偏斜分布。记为 $N(\mu，\sigma_-，\sigma_+)$。

以上对偏斜分布定义的（5）式与重新定义正态分布的（4）式完全相同，不同的仅是以限定的条件 $\sigma_-=\sigma_+$ 与 $\sigma_-\neq\sigma_+$ 来确定对称与否以示区别。

根据偏斜分布的定义，其密度函数 $f(x)$有以下性质：

（1） 当 $\sigma_-\neq\sigma_+$时，曲线 $f(x)$关于直线 $x=\mu$ 不对称，呈偏斜分布；

（2） 当 $x=\mu$ 时，且点在 μ 处连续，$f(\mu)=(\sqrt{2\pi}\,\sigma_-)^{-1}$ 或 $f(\mu)=(\sqrt{2\pi}\,\sigma_+)^{-1}$ 为 $f(x)$ 的最大值；

（3） 在 $x=\mu-\sigma_-$，$x=\mu+\sigma_+$ 处曲线有拐点，当 $x\to\pm\infty$时，$f(x)$以 x 轴为渐近线；

（4）μ 是描述 ξ 的集众位置，σ_-和 σ_+分别是描述期望值左右两边离散度的参数。

以上对偏斜分布密度函数 $f(x)$性质的描述中，性质（1）与不同于正态分布的限制条件确定描述了偏斜分布。性质（2）（3）（4）则描述了正态分布和偏斜分布共同具有的性质。它反映了正态分布和偏斜分布的区别及相互间的关系与联系，说明了它们的数学表达式都来自于高斯分布的原理，是同一类中以对称与不对称区分的不同的两种分布形态。

4.3 高斯新分布的定义及性质

综合以上正态分布和偏斜分布的定义及其性质表明：重新定义的正态分布和偏斜分布数学表达式的导出仅仅是对高斯分布离散特征参数期望差 σ 分解所进行的扩展。使其不但能描述对称分布还能描述不对称分布，将其归纳为高斯新分布正是通过对正态分布数学表达式的推导来确定的。据此对高斯新分布定义如下：

若$-\infty<\mu<\infty$，$\sigma_->0$，$\sigma_+>0$ 为三个实数，点在 μ 处连续，则由下列密度函数：

$$f(x)=\begin{cases}(\sqrt{2\pi}\ \sigma_-)^{-1}\exp[-(x-\mu)^2(2\sigma_-^2)^{-1}] & x\leqslant\mu\\ (\sqrt{2\pi}\ \sigma_+)^{-1}\exp[-(x-\mu)^2(2\sigma_+^2)^{-1}] & x\geqslant\mu\end{cases}\qquad(6)$$

由以上密度函数确定的随机变量 X 的分布称为高斯新分布。当 $\sigma_-=\sigma_+$时，呈正态分布；当 $\sigma_-\neq\sigma_+$ 时，呈偏斜分布。记为 $N(\mu，\sigma_-，\sigma_+)$。

高斯新分布定义中的（6）式与重新定义正态分布与偏斜分布的数学表达（4）式和（5）式完全相同，不同的是在（6）式前取消了限定的条件，在（6）式后分别确定：当 $\sigma_-=\sigma_+$ 时呈正态分布，当 $\sigma_-\neq\sigma_+$ 时呈偏斜分布。定义明确了高斯新分布概念的外延包括正态分布和偏斜分布两种形态。其内涵 $\sigma_-=\sigma_+$及 $\sigma_-\neq\sigma_+$确定了数学表达（6）式所反映高斯新分布的本质属性。

通过以上推论可以得出结论：高斯新分布的数学表达（6）式包括了正态分布和偏斜分布两种形态。

根据高斯新分布的定义，其密度函数 $f(x)$ 的性质归纳如下：

（1） 当 $\sigma_-=\sigma_+$时，曲线 $f(x)$ 关于直线 $x=\mu$ 对称，呈正态分布；

（2） 当 $\sigma_-\neq\sigma_+$时，曲线 $f(x)$ 关于直线 $x=\mu$ 不对称，呈偏斜分布；

（3） 当 $x=\mu$ 时，且点在 μ 处连续，$f(\mu)=(\sqrt{2\pi}\,\sigma_-)^{-1}$ 或 $f(\mu)=(\sqrt{2\pi}\,\sigma_+)^{-1}$ 为 $f(x)$的

最大值；

（4） 在 $x = \mu - \sigma_-$，$x = \mu + \sigma_+$ 处曲线有拐点，当 $x \to \pm\infty$ 时，$f(x)$ 以 x 轴为渐近线；

（5） μ 是描述 ξ 的集众位置，σ_- 和 σ_+ 分别是描述期望值左右两边离散度的参数。

以上对高斯新分布密度函数 $f(x)$ 性质的描述中，归纳了正态分布和偏斜分布的性质。性质（1）以限制条件 $\sigma_- = \sigma_+$ 确定描述了正态分布。性质（2）以限制条件 $\sigma_- \neq \sigma_+$ 确定描述了偏斜分布。性质（3）（4）（5）则描述了正态分布和偏斜分布具有的共同性质。

高斯新分布归纳了正态分布和偏斜分布。从数学的严谨性审视呈现正态分布必须同时满足的三个条件看，可将正态分布、偏斜分布、高斯新分布的关系概括如下：

正态分布是高斯新分布的特殊形态，偏斜分布是高斯新分布的普遍形态，高斯新分布是正态分布与偏斜分布的统一体。

5 期望差及分解的频数分配、计算公式、三者间的关系

期望差的计算公式是将传统的“标准差”计算公式 $\sigma = \{[\sum_{i=1}^{n}(x_i - \bar{x})^2] \div n\}^{1/2}$ 其中的平均值 $\bar{x}$ 改为期望值 μ 即可。根据的是期望差的定义。其计算公式如下：

$$\sigma = \{[\sum_{i=1}^{n}(x_i - \mu)^2] \div n\}^{1/2} \tag{7}$$

在对称的情况下，（7）式可化简转换为以下（8）式：

$$\sigma = \{[\sum_{i=1}^{n/2}(x_i - \mu)^2] \div n/2\}^{1/2} \tag{8}$$

需要对（8）式说明的是：当 n 是偶数时，求和符上标 $n/2$ 无疑是整数；当 n 是奇数时，求和符上标 $n/2$ 取整数，舍 0.5 小数。理由是：随机变量等于期望值的频数有多个，当 $x_i = \mu$ 时，有 $x_i - \mu = 0$，所以求和符上标频数取整，舍 0.5 小数对于计算的结果没有丝毫影响。求和符的上标一定是整数。而作为分母的 n/2 则不能舍去小数。

（8）式还说明：随机变量等于期望值的频数一半属于左边，一半属于右边。有文献对正态分布性质的对称性进行了这样的描述：“正态曲线对于纵轴是对称的。曲线在任一 z 值上的高度，正好与曲线在该值的负数上的高度相同。[19]”由此可以推论，任意偏斜分布以期望值左边的分布曲线向右旋转 180 度，就成为以期望值左边为对称的正态分布。同样期望值右边的分布曲线向左旋转 180 度，也成为以期望值右边为对称的正态分布。所形成两个不同曲线对称的分布图形呈现的正态分布，由于是以期望值为对称轴旋转所形成的正态分布，所以期望值是相等的一个。其依据的是 3.1 中所分解的正态分布左函数 $f_-(x)$ 和正态分布右函数 $f_+(x)$，以及函数在期望值 μ 点处连续的条件。

从期望差的计算公式为（7）式化简转换为（8）式可以得出：任意偏斜分布随机变量等于期望值的频数一半属于左边，一半属于右边，在前期的课题中已经得出验证不再赘述。

已知期望值为：μ。

由于高斯新分布的数学表达式的参数还包括有期望方差、左右期望方差。所以根据以上条件应先给出计算期望方差 σ^2、左期望方差 σ_-^2、右期望方差 σ_+^2 的计算公式如下：

$$\text{期望方差：}\sigma^2 = [\sum_{i=1}^{n}(x_i - \mu)^2] \div n \tag{9}$$

$$\text{左期望方差：}\sigma_-^2 = [\sum_{i=1}^{n_-}(x_i - \mu)^2] \div n_-,\ x_i \leqslant \mu \tag{10}$$

$$\text{右期望方差：}\sigma_+^2 = [\sum_{i=n_-+1}^{n}(x_i - \mu)^2] \div n_+,\ x_i \geqslant \mu \tag{11}$$

根据以上对期望方差的计算容易得出对期望差 σ 、左期望差 σ_-、右期望差 σ_+ 的计算公式。

期望差：$\sigma=\{[\sum_{i=1}^{n}(x_i-\mu)^2]\div n\}^{1/2}$ （12）

左期望差：$\sigma_-=\{[\sum_{i=1}^{n-}(x_i-\mu)^2]\div n_-\}^{1/2}\quad x_i\leqslant\mu$ （13）

右期望差：$\sigma_+=\{[\sum_{i=n-+1}^{n}(x_i-\mu)^2]\div n_+\}^{1/2}\quad x_i\geqslant\mu$ （14）

从以上计算公式来看，期望方差、左期望方差、右期望方差和整体期望差、左期望差、右期望差它们两组各自的三者之间有着密切的数量关系。需要从对称与不对称两方面来表述。

在对称的条件下：

期望方差 σ^2、左期望方差 σ_-^2、右期望方差 σ_+^2 三者之间的关系：$\sigma_-^2=\sigma_+^2=\sigma^2$。

期望差 σ 、左期望差 σ_-、右期望差 σ_+ 三者之间的关系：$\sigma_-=\sigma_+=\sigma$。

在不对称的条件下容易用文字来描述：期望方差等于左期望方差的离差平方和加右期望方差的离差平方和之和除以左期望方差的频数加右期望方差的频数之和。换言之：期望方差是左期望方差和右期望方差的加权平均值。期望差等于左期望方差的离差平方和加右期望方差的离差平方和之和除以左期望方差的频数加右期望方差的频数之和的平方根。换言之：期望差是左期望差和右期望差的加权平均值。

以上三者的关系在不对称的条件下从文字描述中看出，简单且容易理解，但是应用现有的数学符号则无法表达它们之间的数学关系，所以需要提出一个新的数学符号来连接表达它们之间的关系。可以应用前期的研究成果“加权符：$\widehat{+}$”来表达三者之间的关系。加权符$\widehat{+}$实际上是加权平均值计算过程数学表达式与各个单项平均值的连接符，其运算法则是：在两个及以上的分数中，分子相加除以分母相加等于加权平均值。[20]

表达式：$\sigma^2=\sigma_-^2\ \widehat{+}\ \sigma_+^2$ 就能简单、清晰地表达期望方差 σ^2、左期望方差 σ_-^2、右期望方差 σ_+^2 三者之间的关系。验证如下：

已知：$\sigma_-^2=[\sum_{i=1}^{n-}(x_i-\mu)^2]\div n_-$；

$\sigma_+^2=[\sum_{i=n-+1}^{n}(x_i-\mu)^2]\div n_+$ 。

由：$\sigma^2=\sigma_-^2\ \widehat{+}\ \sigma_+^2$

将已知条件带入上式右边，根据加权符的运算法则：分子相加除以分母相加等于加权平均值。

$$
\begin{aligned}
\text{上式右边}&=\{[\textstyle\sum_{i=1}^{n-}(x_i-\mu)^2]\div n_-\}\ \widehat{+}\ \{[\textstyle\sum_{i=n-+1}^{n}(x_i-\mu)^2]\div n_+\}\\
&=[\textstyle\sum_{i=1}^{n-}(x_i-\mu)^2+\textstyle\sum_{i=n-+1}^{n}(x_i-\mu)^2]\div(n_-+n_+)\\
&=[\textstyle\sum_{i=1}^{n}(x_i-\mu)^2]\div n\\
&=\sigma^2
\end{aligned}
$$

同理，期望差 σ 、左期望差 σ_-、右期望差 σ_+ 三者之间的关系：$\sigma=\sigma_-\ \widehat{+}\ \sigma_+$ 。

以上可以用频数或用频率（比率）作权数计算出具体的值。或者直接用期望方差、左期望方差、右期望方差分别的平方根来求得。具体的计算方法和推论参见《关于对左期望方差

和右期望方差与期望方差数学关系表达符的探讨》[20]一文，在此不再赘述。

6 高斯分布离散特征值分解的参数与半方差和半标准差的区别

有关左右方差的概念专家在《左右方差计算公式的推导与应用》[21]一文中给出的评议是“关于左右方差的概念，从大的方面来说，在风险管理或风险度量中或许能有些应用，但须实证检验。左右方差这种度量的应用需要进一步阐述。另外，关于不对称分布，在概率论中已有斜度（skewness）的概念，也有分位数、甚至完整分布的概念，暂时看不出引入左右方差的意义。”为此需要与学界在风险管理中提出的方差、下半方差、上半方差、标准差、下半标准差、上半标准差的概念进行比较，以严格区别避免混淆和误解。

以下是风险管理中半方差的定义及描述。

“设 X 是一随机变量，EX 是它的数学期望，令 $(X-EX)^-=\min(X-EX,0)$，$(X-EX)=\max(X-EX,0)$，称 $E[(X-EX)^-]^2$、$E[(X-EX)^+]^2$ 分别为随机变量的下半方差和上半方差，分别记为 $D^-(X)$、$D^+(X)$。把 $\sigma^-(X)=\sqrt{D^-(X)}$ 和 $\sigma^+(X)=\sqrt{D^+(X)}$ 称为随机变量 X 的下半标准差和上半标准差。显然，$D(X)=D^-(X)+D^+(X)$。下半方差 $D^-(X)$ 表示 X 对于 EX 的左偏离程度；上半方差 $D^+(X)$ 表示 X 对于 EX 的右偏离程度。[22]”

“设 $f(x)$ 是 X 的密度函数，$\mu=EX$，则 σ_-^2 定义为 $[(X-\mu)^-]^2$ 的期望值，

也就是 $\sigma_-^2=E[(X-\mu)^-]^2=\int_{-\infty}^{\mu}(x-\mu)^2 f(x)\,dx$ （下偏差） （15）

类似地定义 $\sigma_+^2=E[(X-\mu)^+]^2=\int_{\mu}^{\infty}(x-\mu)^2 f(x)\,dx$ （上偏差） （16）

若 $f(\mu-x)=f(\mu+x)$，则称 X 具有对称分布。

设 $X_i(i=1,2,\cdots,n)$ 是取自总体 X 的 $iid.$ 随机样本，则 σ_-^2 的估计 $\hat{\sigma}_-^2$ 如下造：

$$\hat{\sigma}_-^2=\frac{1}{n}\sum_{Xi\le Xn}(X_i-\bar{x}_n)^2，其中\ \bar{x}_n=\frac{1}{n}\sum_{i=1}X_i \tag{17}$$

同样方式构造 σ_+^2 的估计为：

$$\hat{\sigma}_+^2=\frac{1}{n}\sum_{Xi>Xn}(X_i-\bar{x}_n)^2 \tag{18}$$

对于半方差，容易验证下面的性质成立：① $\sigma_-^2+\sigma_+^2=\sigma^2$；② 若 $Y=aX+b$，则 $(\sigma_-^2)_Y=a^2(\sigma_-^2)_X$；③ 若 X 具有对称分布，则 $\sigma_-^2=\sigma_+^2=\sigma^2/2$。性质③隐含了半方差在风险管理中的一个重要应用，即如果能设法检验 $\sigma_-^2=\sigma_+^2$ 成立，则可大致判定该起风险因素的变动总体上是平衡的，而判定这一点特别对于高风险的投资媒介的宏观调控来说是十分重要的，这正是半方差研究的重要背景。[23]”

根据文献给出的计算公式和方差、上、下半方差和标准差、上、下半标准差的数学关系可以清楚表明与本课题的不同之处如下：

公式采用的期望值不同：本课题的期望值采用的是峰值或众数；文献的期望值采用的是平均值。

公式采用的频数不同：本课题的期望方差和期望差所采用的是样本的频数之和，左期望方差和左期望差所采用的是属于期望值左边的频数，右期望方差和右期望差所采用的是属于期望值右边的频数。文献采用的均是样本的频数之和。就是说下半方差、上半方差的计算不需要进行频数的左右分配。

反映它们的数学关系不同。

本课题反映它们的数学关系：

$\sigma^2=\sigma_-^2 \overset{\frown}{+} \sigma_+^2$、$\sigma_-^2=\sigma_+^2=\sigma^2$、$\sigma=\sigma_- \overset{\frown}{+} \sigma_+$、$\sigma_-=\sigma_+=\sigma$。

文献反映它们的数学关系：

$\sigma^2=\sigma_-^2+\sigma_+^2$、$\sigma_-^2=\sigma_+^2=\sigma^2/2$、$\sigma=\sigma_-+\sigma_+$、$\sigma_-=\sigma_+=\sigma/2$。

它们的作用不同。

左右期望方差和左右期望差的作用是：反映偏斜分布期望值两边随机变量与期望值的离散程度；能精准确定分布曲线两边的拐点；能精确计算偏斜分布任意区间的概率。

上、下半方差和上、下半标准差则是用于风险管理中的决策判定与调控分析。

通过对比分析说明，本课题分解的离散特征参数与风险管理中的下半方差、上半方差和下半标准差、上半标准差虽然表达的符号有相似之处，但是它们完全是具有不同的涵义和作用且是不相关的两个概念，不可同日而论。

7 高斯新分布的偏度与偏斜指数

“偏度反映频数分布中的非对称程度。平均数和标准差相同的数列，频数分布形态不一定全都一样，如频数分布的对称程度不同。所以，偏度也是反映频数分配性质的一个重要指标[4]。”根据满足正态分布的三个条件可以对高斯新分布呈现的偏斜程度进行偏度的量度，在前期《关于对高斯新分布偏度与偏斜指数的研究》[24]中已经提出具体的计算方法，以此为基础归纳补充如下：

满足正态分布的三个条件的第一条是期望值等于平均值，由于平均值不一定等于分布中心值，所以期望值 μ 与分布中心值 dcv 的关系更能说明期望值在分布中的具体位置所在。当期望值小于分布中心值时呈现正偏斜，期望值偏向左边。当期望值大于分布中心值时呈现负偏斜，期望值偏向右边。当期望值等于分布中心值时是满足正态分布的必要条件之一。

“偏度的计算是以标准差为度量单位计量的众数与算术平均数的离差，计算公式为：

$$SK=(\bar{x}-M_O)\div\sigma \qquad (19)$$

式中 SK 代表偏系数；M_O 为众数；σ 为标准差[4]。”是传统偏度的定义和计算公式。需要重新定义为“偏度的计算是以期望差为度量单位计量的期望值与分布中心值的离差。”计算公式改为：

$$SK=(dcv-\mu)\div\sigma \qquad (20)$$

式中 SK 代表偏系数；dcv 为分布中心值；μ 为期望值；σ 为期望差。若期望值小，分布中心值大，称正偏，即 $SK>0$。反之，当期望值大，分布中心值小，就称为负偏，即 $SK<0$。当期望值等于分布中心值时，说明期望值处在分布中心的位置，即 $SK=0$。偏度所反映期望值处于分布的位置说明，满足正态分布的三个条件的第一条的表达应该是期望值等于分布中心值更能说明问题。由此又证实了一个结论：有关随机变量居中的位置特征分布中心值较平均值和中位数更为有效，这对于统计研究和分析的对象选择有效的指标完全可以降低误导，从而避免对习惯的某一指标的滥用尤为重要。由于偏度还不足于说明分布是否对称，还需要考察满足正态分布的其它两个条件，从而需要引出偏斜指数[24]来进行判定。

以偏度为基础编制偏斜指数的计算公式如下：

当 $SK<0$ 时，呈现负偏斜，偏斜指数公式为：

$$SK\,I_{-}=\frac{dcv}{\mu}\times\frac{n_{+}}{n_{-}}\times\frac{\sigma_{+}}{\sigma_{-}} \tag{21}$$

当 $SK>0$ 时，呈现正偏斜，偏斜指数公式为：

$$SK\,I_{+}=\frac{\mu}{dcv}\times\frac{n_{-}}{n_{+}}\times\frac{\sigma_{-}}{\sigma_{+}} \tag{22}$$

当 $SK=0$ 时，期望值在分布的中心位置，偏斜指数公式为：

$$SK\,I=\frac{n_{+}}{n_{-}}\times\frac{\sigma_{+}}{\sigma_{-}} \quad 或 \quad SKI=\frac{n_{-}}{n_{+}}\times\frac{\sigma_{-}}{\sigma_{+}} \tag{23}$$

偏斜指数公式反映了满足正态分布三个条件所涉及的分布频数和两个参数。若满足三个条件则 $SK\,I=1$，呈现正态分布。当 $1>SK\,I_{-}$ 或 $SK\,I_{+}\geqslant 0.95$ 时则可判定为近似正态分布。当 $SK\,I_{-}$ 或 $SK\,I_{+}<0.95$ 时则可判定为偏斜分布。有关偏斜指数的应用可从以下的 8.2 中说明。

8 高斯新分布提出的目的

统计学界的教育大家张尧庭先生生前一直认为：“无论是一元统计或多元统计，统计分析的中心内容都是数据变异程度的度量和分解，从而解释变异的来源与影响它的因素是否重要、重要的程度如何[25]。”期望差作为变异指标的一种，左期望差和右期望差较传统意义的“标准差”更能把数据间存在的差异解释得更清楚一些。统计技术和统计方法在实践中的应用需要改进的相关问题均依据高斯新分布的理论。其实践作用及应用的成果已经在前期的相关课题研究中有详细的论述。为此不再赘述。

综上所述，高斯新分布再认识的目的是：理论应用于实践的过程必须满足达到精确描述、降低统计误差的结果。

2011 年 2 月 8 日 CCTV10 科教频道《地理·中国》主持人胜春在“澜沧江探源”科考节目播完后的结束语说道：“它们地理位置的精确程度达到了厘米级，而这将是未来澜沧江源头科学研究的基础。在澜沧江近五千千米的长度当中这两千米的影响似乎是微乎其微，但是在澜沧江源头的科学探索中人们又向精确接近了一步。还在考察队刚刚组成时，有一位科学家在评价考察澜沧江意义的时候就这样说：精确是一个民族进步的标志，而追求精确的过程正是我们向科学靠近的过程。”

2007 年 11 月 11 日“统计学与应用”全国博士生论坛中，东北师范大学校长史宁中教授的演讲谈到数学与统计的关系时，简洁而又深刻揭示了它们的区别：数学在于对与错，而统计没有对和错，只有好与不好，误差小为好误差大则不好。所说的“误差小为好”正是统计追求精确的过程。

综合科学大师们理论与实践总结的经典论述表达出了这样一个思想：自然界所有事物的基础原理及主要问题都可以用数学方式表达。而科学研究的终极目标是向精确靠近，需要应用数学的研究成果精确地测量、反映、描述、揭示自然界和社会实际问题的本质属性和客观规律。以便人们更深刻地认识、了解世界和自然，以促进科学的创新和社会的发展，进而使科学的创新成果尽快形成生产力而造福于人类。这是在高斯分布的研究探索中向精确描述近似正态分布和精确描述偏斜分布朝前靠近的一步。

9 结论

高斯新分布是建立在高斯分布离散特征参数分解的基础上，不论是正态分布或是近似正态分布或是偏斜分布都能以正态分布左右两个分部函数来精确表达。其理论意义在于：完善

高斯分布的理论使之能表达不对称分布的内涵。其学术价值在于：为改进建立在正态分布假设条件下的统计技术和方法提供理论依据；并为分部描述能降低统计误差的理念在不对称分布的实际应用中得出新的佐证；应用满足正态分布的三个条件检验分布的对称性，以偏斜指数替代传统 χ^2（卡方）拟合优度检验法。其实践作用在于：解决随机变量的频数分布在单峰的条件下能精确计算任意区间的概率问题；以高斯新分布的理论为依据以改进的统计技术和统计方法应用于生产实践。

[参考文献] (References)

[1] 孔建新,孔建勇.偏斜分布密度函数的提出与讨论[OL].[2008-3-5]. http://www.paper.edu.cn

[2] 孔建新. 高斯新分布扩展应用的研究[OL].[2010-9-7]. http://www.paper.edu.cn

[3] 斯特芬 康罗德，孙大炜.ISO21747 对过程性能及能力评价的新定义和新规定[J].中国质量,2010, 343(1):37-39.

[4] 茆诗松. 统计手册[M]. 北京：科学出版社，2003.

[5] 张公绪,孙静.质量工程师手册[M].北京:企业管理出版社,2002.

[6] 孔建新,孔建新,孔粤.高斯分布内涵扩展的研究[OL].[2008-12-16]. http://www.paper.edu.cn

[7] 苏秦.现代质量管理学[M].北京:清华大学出版社 2007.

[8] 孔建新,孔璐,孔粤. 标准差内涵扩展在统计实践中的意义[OL].[2009-2-11]. http://www.paper.edu.cn

[9] 肖明耀.误差理论-常见问题与解答[M].北京:科学出版社，1980.

[10] 孔建新,孔璐,何光伟. 关于对标志变异指标概念的重新认识[OL].[2009-5-18]. http://www.paper.edu.cn 中国科技论文在线

[11] 周尊英.实用统计技术指南[M].北京:中国标准出版社,2003.

[12] 王琪延,白日荣.统计在法律中的应用与展望[J].统计研究 2008,25（5）:104-105.

[13] 百度快照.范剑青:把数学作为解决社会问题的工具[OL].[2011-2-3]. http://www.pinggu.org/bbs/b5i164172.html

[14] 张公绪.全面质量管理词典[M].北京:经济科学出版社,1991.

[15] 王青建.数学史简编[M].北京:科学出版社,2004.

[16] 崔恒建,陈秋华.高斯分布的启示[J].数学通报,2000,(4):40-42.

[17] 唐国兴.高等数学(二)第二分册:概率统计[M].武昌:武汉大学出版社, 1991.

[18] 刘文嘉.林家翘:大师之忧[N].光明日报,2010-5-7(12).

[19] G H 维恩堡,J A 休麦克,D 奥尔特曼.数理统计初级教程[M].常学将,胡文明,王明生,等. 太原:山西人民出版社,1986.

[20] 孔建新.关于对左期望方差和右期望方差与期望方差数学关系表达符的探讨[OL].[2010-4-21]. http://www.paper.edu.cn

[21] 孔建新.左右方差计算公式的推导与应用[OL].[2010-8-24]. http://www.paper.cdu.cn

[22] 胡小文,惠军.加权半方差风险度量模型[J].合肥工业大学学报(自然科学版),2007,30(4):518-520.

[23] 徐克军.半方差风险管理中的应用及估计的统计性质[J].同济大学学报,1998,26(1):67-70.

[24] 孔建新.关于对高斯新分布偏度与偏斜指数的研究[OL].[2010-3-31].http://www.paper.edu.cn

[25] 龚凤乾.化神奇为平易──张尧庭统计教育思想研究[J].统计研究,2008,25(9):93-94.