

实验室比对数据处理中卡方统计量研究

杭晨哲, 徐定华, 原遵东

中国计量科学研究院, 北京 100029



【摘要】 卡方统计量是实验室比对数据处理的关键统计分析工具, 可用于比对数据的一致性核验和参考值不确定度的估计。本研究在比对量具有相同期望及比对数据服从正态分布的条件下, 提出了一种包含一般线性估计的卡方统计量, 研究了该统计量的性质及分布。该统计量可以实现一般线性参考值估计的一致性检验和不确定度估计, 为更广泛的线性参考值估计提供统计工具, 可用于比对数据分析或多实验室定值测量。作为示例, 针对算术平均值使用传统卡方检验需基于各实验室宣称相同不确定度的局限, 对任意不确定度组合下的算术平均值给出卡方统计量, 为这种常用线性参考值估计的扩展应用提供了新统计分析方法。

【关键词】 实验室比对; 参考值估计量; 卡方统计量; 一般模型; 随机影响模型; 蒙特卡洛模拟

DOI: 10.12338/j.issn.2096-9015.2020.9042

【引用本文】 杭晨哲, 徐定华, 原遵东. 实验室比对数据处理中卡方统计量研究 [J]. 计量科学与技术, 2021, 65(5): 108-114.

Research on Chi-Square Statistics in Data Analysis of Inter-laboratory Comparison

HANG Chenzhe, XU Dinghua, YUAN Zundong

National Institute of Metrology, Beijing 100029, China

【Abstract】 The chi-square statistics, which can be used in the consistency test of comparison results and the estimation of reference value uncertainty, is a key statistical tool in data analysis of inter-laboratory comparison. In this study, under the condition that the comparison results obey Gaussian distribution with a common mean, a chi-square statistic containing generalized linear estimation is proposed, and the properties and distribution of the statistic are investigated. The statistic enables consistency testing and uncertainty estimation of general linear reference value estimates, and provides a statistical tool for a wider range of linear reference value estimates, which can be used for comparison data analysis or multi-laboratory fixed value measurements. As an example, against the limitation of using the traditional chi-square test for the arithmetic means based on the same uncertainty claimed by each laboratory, this method gives the chi-square statistic for the arithmetic mean under any combination of uncertainties, providing a new statistical analysis method for the extended application of this common linear reference value estimation.

【Key words】 inter-laboratory comparison, reference value estimation, chi-square statistics, generalized model, random effects model, Monte Carlo simulation

0 引言

实验室比对可以验证参加者的校准或测量能力, 为测量结果的一致性提供证据^[1]。基于国际计量互认协议, 作为实验室比对的一种具体应用形式, 关键比对及其数据统计方法近年来得到了极大的关注和研究。关键比对的最终结果为每个参加实验室

的等效度, 是各实验室获得校准与测量能力 CMC 准入的数据基础。而等效度的获得需要基于关键比对参考值及其不确定度, 从而使参考值及其不确定度的分析方法成为研究热点。

参考值及其不确定度的分析方法种类繁多, 常用的统计分析方法大致可分为线性估计方法、鲁棒

性估计方法和贝叶斯方法^[2],在线性估计方法中常用的参考值估计为不确定度权重平均值和算术平均值。其中,不确定度权重平均值在分析模型中的应用较为系统,大致分为一般模型(CMM)^[3-4]、随机影响模型(REM)^[5]和确定影响模型(FEM)^[5-6]。由于 CMM 和 REM 均假设比对结果服从正态分布并具有相同期望,可从理论上得出不确定度权重平均值的卡方统计量在两种模型条件下均成立。此卡方统计量在 CMM 中用以检验比对数据的一致性:当检验成功时,即可使用 CMM 进行统计分析;当检验失败时,CMM 失效,可采用 REM 进行分析。REM 分析的核心是应用卡方统计量期望为确定值的性质估计随机影响方差。由此可见,卡方统计量是数据分析的核心统计工具。

然而,不确定度权重平均值的卡方统计量对于其他线性估计并不适用。关键统计工具的缺失,造成了算术平均值等其他线性估计难以应用 CMM 和 REM 展开系统地分析,限制了线性估计作为参考值的应用。

本研究基于 CMM 提出了包含一般线性估计的卡方统计量,并通过理论分析和蒙特卡洛方法模拟研究了该统计量的统计性质和分布,得到该统计量可作为一般线性估计的统计分析工具,在 CMM 中检验比对数据的一致性,在 REM 中采用矩估计方法获得随机影响方差。作为示例,基于 CMM 和 REM 的分析思路,给出了新卡方统计量在算术平均值作为参考值估计时的比对数据分析步骤,为算术平均值扩展应用提供了新的统计分析方法。

1 传统卡方统计量简介

CMM 是一种基础模型。如果比对有 n 个实验室参加,所有实验室均提供了比对结果 x_i 及其标准不确定度 u_i ,且不同实验室的测量结果相互独立。CMM 可表示为

$$x_i = \mu + \varepsilon_i \quad (1)$$

式中: μ 为被测量;随机量 ε_i 为第 i 个实验室的测量误差且 $\varepsilon_i \sim N(0, u_i^2)$ 。在该模型条件下,不确定度权重平均值 $x_w = \sum_{i=1}^n w_i \cdot x_i$ 是参考值的一种比对数据线性估计,其中权重值 $w_i = \frac{1}{u_i^2} \cdot \left(\sum_{i=1}^n \frac{1}{u_i^2} \right)^{-1}$ 。同时,对于 x_w 的 $n-1$ 阶卡方统计为

$$\chi_{\text{obs}}^2 = \sum_{i=1}^n \frac{(x_i - x_w)^2}{u_i^2} \quad (2)$$

可用于比对结果一致性的检验^[3,7]。

REM 是在一般模型的基础上,为每个实验室加入了被忽视的、相等的随机性系统影响。REM 可表示为

$$x_i = \mu + b_i + \varepsilon_i \quad (3)$$

式中: b_i 为随机性系统影响且 $b_i \sim N(0, \tau^2)$; τ 为随机影响的标准差。因此,第 i 个实验室的不确定度 $u^2(x_i) = \tau^2 + u_i^2$, 并以此得到不确定度权重平均值 x_w^{REM} 。对于 x_w^{REM} 的 $n-1$ 阶卡方统计量为

$$\chi_{\text{REM}}^2 = \sum_{i=1}^n \frac{(x_i - x_w^{\text{REM}})^2}{u^2(x_i)} \quad (4)$$

在已知比对结果及不确定度的条件下,基于 χ_{REM}^2 并采用矩估计方法可对 τ^2 进行估计^[8-9],具体方法有 Mandel-Paule 法^[8,10]和 DerSimonian-Laird 法^[4,8];Mandel-Paule 法使用迭代方式计算随机影响方差的无偏估计;DerSimonian-Laird 法给出了随机影响方差的有偏估计,但不涉及数值计算。

实验室间比对的线性参考值估计^[11]表示为

$$x_y = \sum_{i=1}^n \gamma_i \cdot x_i \quad (5)$$

式中: γ_i 为权重值,且满足 $0 < \gamma_i < 1$ 和 $\sum_{i=1}^n \gamma_i = 1$ ($i = 1, \dots, n$)。由于 $\sum_{i=1}^n \gamma_i = 1$, 可得 $E(x_y) = \mu$, 即 x_y 为 μ 的无偏估计。然而,在 x_y 为非不确定度权重平均值时,采用 x_y 替换 x_w 和 x_w^{REM} , 则 χ_{obs}^2 和 χ_{REM}^2 都不服从卡方分布且分布没有明显规律^[5]。此现象导致 CMM 和 REM 分析只适用于 x_w , 而不适用于非等不确定度数据的算术平均值和其他线性参考值估计。

2 线性估计的卡方统计量

由于权重值 γ_i 确定了线性估计,因此线性估计的分析就可变为权重值组合的分析。故此,将权重值组成向量,并记权重值向量为 $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)^T$, 则 $\boldsymbol{\gamma}$ 的基本取值空间为 $0 < \gamma_i < 1$ 且 $\sum_{i=1}^n \gamma_i = 1$ 。另外,记 x_w 对应的权重向量 $\boldsymbol{w} = (w_1, w_2, \dots, w_n)^T$, 记平均值 \bar{x} 对应的权重向量 $\boldsymbol{a} = (1/n, \dots, 1/n)^T$ 。

基于 w_i 的定义公式,式(2)可变形为

$$\chi_{\text{obs}}^2 = (n-1) \cdot \frac{\sum_{i=1}^n w_i \cdot (x_i - x_w)^2}{\sum_{i=1}^n w_i \cdot u_i^2 - \sum_{i=1}^n w_i^2 \cdot u_i^2} \quad (6)$$

将式(6)中的 w_i 替换为 γ_i , 并记该统计量 χ_y^2 为

$$\chi_y^2 = (n-1) \cdot \frac{\sum_{i=1}^n \gamma_i \cdot (x_i - x_y)^2}{\sum_{i=1}^n \gamma_i \cdot u_i^2 - \sum_{i=1}^n \gamma_i^2 \cdot u_i^2} \quad (7)$$

式中: $i = 1, \dots, n$ 。

由于 u_i^2 的取值大小并不会对此随机变量的性质产生实质性影响, 因此采用更简单的 CMM 模型来分析基本性质及分布。在 CMM 模型下, χ_γ^2 有以下性质:

1) 由于 $x_i = \mu + \varepsilon_i$ 和 $x_\gamma = \mu + \sum_{i=1}^n \gamma_i \cdot \varepsilon_i$, 可得 $x_i - x_\gamma = \varepsilon_i - \sum_{i=1}^n \gamma_i \cdot \varepsilon_i$, 即 $x_i - x_\gamma$ 已将 μ 的影响去掉, 也就是 μ 的取值不影响 χ_γ^2 的分布。

2) 在 γ 取基本空间的任意确定点条件下都满足 $E(\chi_\gamma^2 | \gamma) = n - 1$ 。证明如下:

只将式(7)中的 $x_i (i = 1, \dots, n)$ 视为随机变量, 对于第 i 个实验室 γ_i 和 u_i 均视为常量。已有 $E(x_\gamma) = \mu$, 则

$$\sum_{i=1}^n \gamma_i (x_i - x_\gamma)^2 = \sum_{i=1}^n \gamma_i (x_i - \mu)^2 + (x_\gamma - \mu)^2 - 2(x_\gamma - \mu) \sum_{i=1}^n \gamma_i (x_i - \mu)$$

又有 $\sum_{i=1}^n \gamma_i = 1$ 和 $\sum_{i=1}^n \gamma_i (x_i - \mu) = x_\gamma - \mu$, 得到

$$\sum_{i=1}^n \gamma_i (x_i - x_\gamma)^2 = \sum_{i=1}^n \gamma_i (x_i - \mu)^2 - (x_\gamma - \mu)^2$$

两边对实验室的测量结果取期望可得:

$$E\left(\sum_{i=1}^n \gamma_i (x_i - x_\gamma)^2\right) = \sum_{i=1}^n \gamma_i E(x_i - \mu)^2 - E(x_\gamma - \mu)^2$$

其中,

$$E(x_i - \mu)^2 = \text{Var}(x_i) + (E(x_i) - \mu)^2$$

$$E(x_\gamma - \mu)^2 = \text{Var}(x_\gamma)$$

整理以上三式, 并将 $\mu = E(x_\gamma)$ 带入后, 可得下式:

$$E\left(\sum_{i=1}^n \gamma_i (x_i - x_\gamma)^2\right) = \sum_{i=1}^n \gamma_i (E(x_i) - E(x_\gamma))^2 + \sum_{i=1}^n \gamma_i \cdot \text{Var}(x_i) - \text{Var}(x_\gamma)$$

在一般模型条件下, 有 $E(x_i) = E(x_\gamma) = \mu$ 和 $\text{Var}(x_\gamma) = \sum_{i=1}^n \gamma_i^2 \cdot u_i^2$, 带入上式得:

$$E\left(\sum_{i=1}^n \gamma_i (x_i - x_\gamma)^2\right) = \sum_{i=1}^n \gamma_i \cdot u_i^2 - \sum_{i=1}^n \gamma_i^2 \cdot u_i^2 \quad (8)$$

由式(7)和式(8)易见 $E(\chi_\gamma^2 | \gamma) = n - 1$ 。

3) 在 $\gamma = \mathbf{w}$ 条件下, 可得 $\chi_\gamma^2 |_{\gamma=\mathbf{w}} = \chi_{\text{obs}}^2$, 即 χ_{obs}^2 是 χ_γ^2 的一种特殊情况。

4) 在 $\gamma \neq \mathbf{w}$ 条件下, $\chi_\gamma^2 |_\gamma$ 也有可能服从或近似服从 $n - 1$ 阶卡方分布。

例如, 在 $\gamma = \mathbf{a}$ 的条件下, 式(7)可变为

$$\chi_\gamma^2 |_{\gamma=\mathbf{a}} = n \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n u_i^2} \quad (9)$$

假设有 4 个实验室参加比对, 测量结果都服从正态分布, 提供的标准不确定度 u_i 分别为 1.4、2.0、1.6 和 1.8。采用蒙特卡洛方法分析 $\chi_\gamma^2 |_\gamma$ 的分布情况。由于 μ 不影响 χ_γ^2 的分布, 因此 μ 在模拟过程中取值为 0^[7]。

具体模拟步骤如下:

a) 按照正态分布 $N(0, u_i^2)$ 生成 4 个测量结果 x_i 的模拟样本序列, 每个序列样本量为 10^6 ;

b) 使用这些样本序列以及不同的权重值, 分别按照式(6)和式(9)分别计算得到 $\chi_\gamma^2 |_{\gamma=\mathbf{w}}$ 和 $\chi_\gamma^2 |_{\gamma=\mathbf{a}}$ 的样本序列;

c) 使用直方图法分别得到概率密度曲线。

由图 1 可见, 两条模拟曲线与 3 阶卡方分布的概率密度曲线非常吻合, 即线性估计 \bar{x} 或者权重向量 \mathbf{a} 也使 $\chi_\gamma^2 |_\gamma$ 服从或者近似服从 3 阶卡方分布。由此可见, 除不确定度权重值以外, 对其他线性估计也可能存在卡方统计量。

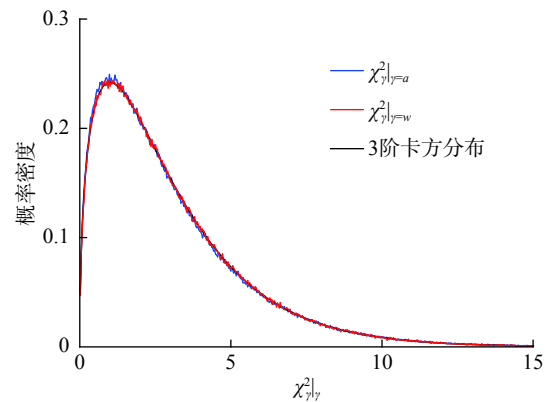


图 1 $\chi_\gamma^2 |_{\gamma=\mathbf{w}}$ 和 $\chi_\gamma^2 |_{\gamma=\mathbf{a}}$ 的模拟概率密度
Fig.1 Simulated density of $\chi_\gamma^2 |_{\gamma=\mathbf{w}}$ and $\chi_\gamma^2 |_{\gamma=\mathbf{a}}$

3 模拟研究方法

为了进一步分析 γ 在基本空间中的哪些区域使 $\chi_\gamma^2 |_\gamma$ 服从或近似服从 $n - 1$ 阶卡方分布, 需要将第 2 节中的模拟步骤进行扩展, 形成以下模拟步骤:

1) 按照正态分布 $N(0, u_i^2) (i = 1, \dots, n)$ 生成每个测量结果 x_i 的模拟样本序列, 每个序列样本量为 10^5 ;

2) 在基本空间中生成一定量的 γ 样本点。为了避免 γ 样本点在基本空间中的某些区域过少导致模拟分析结果不具有代表性, 因此在基本空间中均匀地生成 γ 样本点。然而, 这并不意味着 γ 在基本空间内服从均匀分布。具体的生成样本方法为:

a) 使用 0 到 1 的均匀分布生成 $n - 1$ 个不相等的

数值且每个数值都不等于 0 或 1, 记为 $y_i (i = 1, \dots, n-1)$;

b) 按照由小到大的顺序排列这组数值, 记为 $0 < y_{(1)} < \dots < y_{(n-1)} < 1$;

c) 排列后的数值就将 0 到 1 分为了 n 份, 每一份对应一个权重值, 可得 $\gamma_1 = y_{(1)} - 0, \dots, \gamma_i = y_{(i)} - y_{(i-1)}, \dots, \gamma_n = 1 - y_{(n-1)}$, 并将这组权重值记为 γ 的一个样本。

重复上述步骤 a) 到 c) 就可得到 γ 的样本序列。例如, 当 $n = 3$ 时, 10^4 个 γ 样本如图 2 中蓝色样本点所示, 图中蓝色线所围成区域为 γ 的基本取值空间。由图可见, 按照此方法生成的 γ 样本均匀的分布于基本空间。

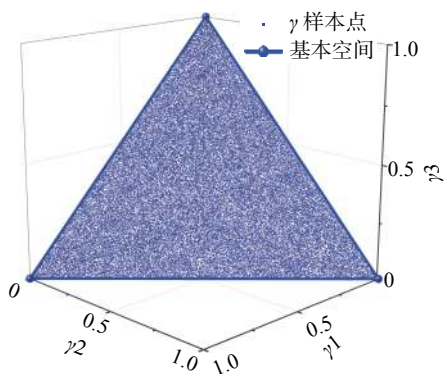


图 2 基本空间中的 γ 样本 ($n = 3$)
Fig.2 Samples of γ in basic space ($n = 3$)

3) 选取 γ 样本序列中的第一个样本 S_1 , 使用此样本和步骤 1) 中生成的所有 x_i 样本序列, 由式(7)计算得到 $\chi^2_{\gamma|_{\gamma=S_1}}$ 的样本序列, 并采用直方图法得到其模拟概率密度曲线。将模拟曲线与 $n-1$ 阶卡方分布理论曲线进行比较, 得到两条曲线偏差最大值和最小值的差值, 记为 D 。当 D 小于某限值时, 则认为 $\chi^2_{\gamma|_{\gamma=S_1}}$ 服从或近似服从 $n-1$ 阶卡方分布。

4) 重复步骤 3), 逐个模拟分析 γ 样本序列中所有样本, 并统计 D 小于给定限值的 γ 样本, 则可认为 γ 在这些样本覆盖空间内取值条件下能使 $\chi^2_{\gamma|_{\gamma}}$ 服从或近似服从 $n-1$ 阶卡方分布, 并将此子空间简记为卡方子空间。

4 模拟结果与讨论

4.1 $n = 2$ 的模拟结果

此条件下, 有 $\gamma_1 + \gamma_2 = 1$, 并式(7)可变为

$$\chi^2_{\gamma|_{n=2}} = \frac{(x_1 - x_2)^2}{u_1^2 + u_2^2} \quad (10)$$

在 CMM 或 REM 中, 由于比对结果 x_i 服从正态分布, 易见 $\chi^2_{\gamma|_{n=2}}$ 服从 1 阶卡方分布, 而且不受 γ 取值

变化的影响。因此, γ 在全部基本空间内取值都能使 $\chi^2_{\gamma|_{n=2}}$ 服从 1 阶卡方分布。

4.2 $n \geq 3$ 且各实验室不确定度相等的模拟结果

此条件下, 式(7)可变为

$$\chi^2_{\gamma|_{u,n \geq 3}} = (n-1) \cdot \frac{\sum_{i=1}^n \gamma_i \cdot (x_i/u - x_{\gamma}/u)^2}{\sum_{i=1}^n \gamma_i - \sum_{i=1}^n \gamma_i^2} \quad (11)$$

式中: $x_i/u \sim N(0, 1), i = 1, \dots, n$ 。

此时, 有 $w = a$, 并且通过模拟可得: γ 在 w 或 a 周围的一个子空间内能够使 $\chi^2_{\gamma|_{u,n \geq 3}}$ 服从或近似服从 $n-1$ 阶卡方分布。

以 $n = 3$ 时为例, 将第 3 节中步骤 1) 改为按照 $N(0, 1)$ 生成 x_i/u 的样本, 并在步骤 3) 中按照式(11)计算得到 $\chi^2_{\gamma|_{u,n \geq 3}}$ 的样本序列。通过模拟发现, 当 $D = 0.088$ 时模拟曲线与理论曲线偏离较小(图 3a), 可认为 $\chi^2_{\gamma|_{u,n \geq 3}}$ 近似服从 $n-1$ 阶卡方分布, 故此限值取为 0.09。按照此限值划分 γ 样本构成的子空间(图 3b), 其中红色和绿色样本分别为 $D \leq 0.09$ 的样本和 $D > 0.09$ 的样本, 则 γ 在红色样本占据的区域内使 $\chi^2_{\gamma|_{u,n \geq 3}}$ 服从或近似服从 χ^2_2 分布, 即红色样本占据区域为卡方子空间, 且此卡方子空间包含 w 和 a 。

4.3 $n \geq 3$ 的模拟结果

此条件下, $u_i > 0 (i = 1, \dots, n)$, 各个 u_i 可能相等, 也可能不相等。通过模拟也可得到类似的结果: γ 在 w 周围的卡方子空间内能够使 $\chi^2_{\gamma|_{\gamma}}$ 服从或近似服从 $n-1$ 阶卡方分布, 而且卡方子空间的范围受到 u_i 取值的影响。

为了更加具体的分析, 同样以 $n = 3$ 时为例, 针对 4 种实际比对中可能出现的 u_i 取值情况(表 1), 并采用第 3 节的模拟方法对卡方子空间进行了分析。通过 4.2 节已知可通过 D 划分基本空间, 而且模拟的目的是获得空间区域而非具体样本。因此, 将第 3 节的步骤 2) 中生成样本的数量由 10^4 降低为 3×10^3 , 以增加模拟效率。另外, 在模拟过程中, D 的限值同样取为 0.09。

按照第 3 节中的方法, 4 种情况下的模拟结果见图 4。在情况 1 时, 3 个参加实验室的不确定度相当, 模拟得到卡方子空间区域较大且包含 w 和 a , 如图 4(a); 在情况 2 至 4 时, 三个参加实验室的不确定度有明显差别, 模拟得到的卡方子空间较小且不包含 a , 如图 4(b) 到图 4(d) 所示。由此例可知, 在参加比对实验室测量能力相当时, 卡方子空间范围更大且其中更有可能包含 a 。

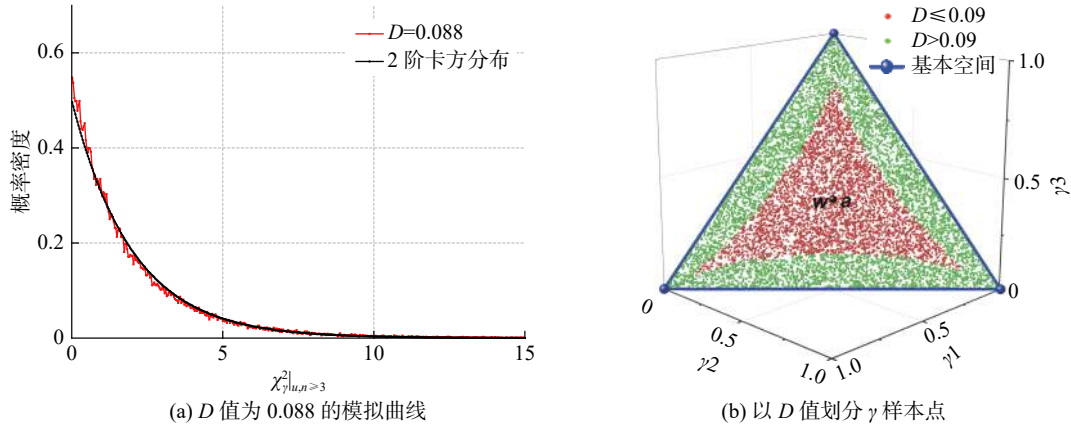


图3 $\chi^2_{\gamma|u_i, n \geq 3}$ 服从或近似服从2阶卡方分布的 γ 样本

Fig.3 Samples of γ contribute or approximately contribute $\chi^2_{\gamma|u_i, n \geq 3}$ to the 2nd order chi-square distribution

表1 u_i 的4种情况 ($n=3$)

Tab.1 Four conditions of u_i ($n=3$)

序号	模拟条件	u_1	u_2	u_3	序号	模拟条件	u_1	u_2	u_3
1	u_i 数值大致相当	1.4	1.8	1.6	3	一个 u_i 数值很大	4.0	1.8	1.6
2	一个 u_i 数值很小	0.2	1.8	1.6	4	u_i 数值差异较大	0.5	1.8	3.5

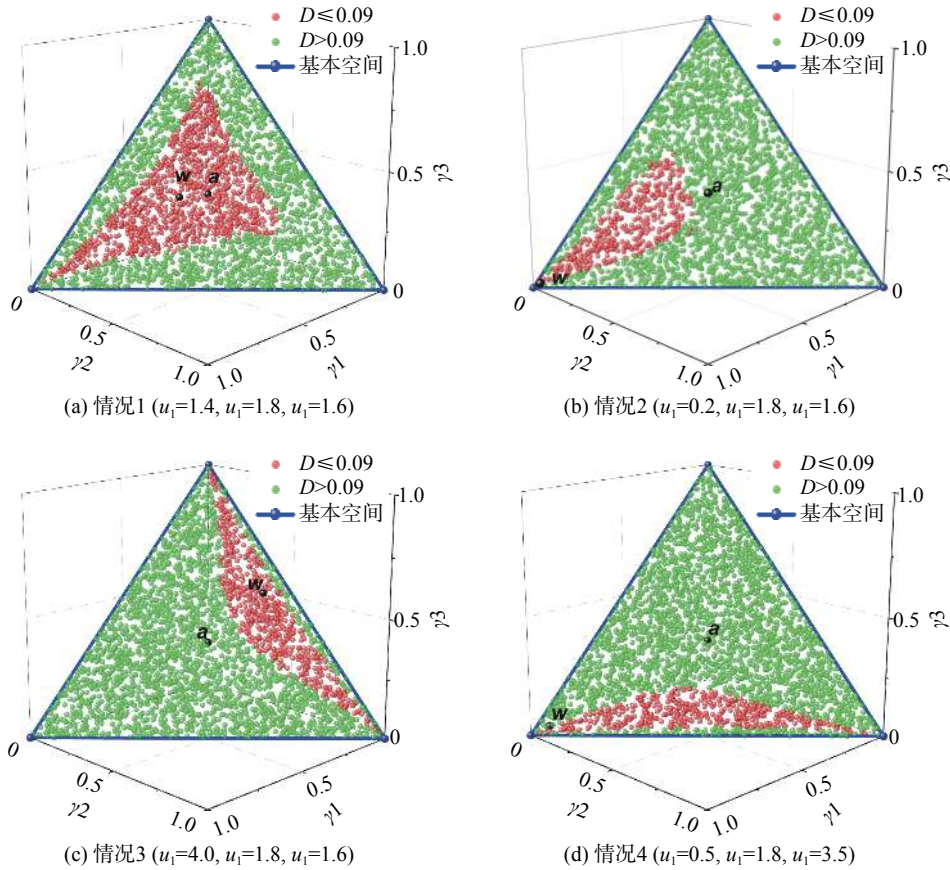


图4 χ^2_{γ} 服从或近似服从卡方分布的 γ 样本

Fig.4 Samples of γ contribute χ^2_{γ} to the chi-square distribution

由于第 5 节涉及到算术平均值的卡方统计 $\chi^2_{\gamma|y=a}$ 的应用, 采用第 2 节中的模拟步骤, 在表 1 的 4 种不确定度情况下, 对 $\chi^2_{\gamma|y=a}$ 的密度曲线分别进行模拟, 概率密度曲线模拟结果见图 5。在情况 1 时, 模拟密度曲线(绿色)与 2 阶卡方分布理论曲线(黑色)非常吻合, 即 a 处于卡方子空间内时, $\chi^2_{\gamma|y=a}$ 服从或近似服从 2 阶卡方分布。在情况 2 至 4 时, 3 条模拟密度曲线(蓝色、蓝绿色、品红色)与 2 阶卡方分布理论曲线(黑色)有不同程度的偏离, 即 a 不处于卡方子空间内时, $\chi^2_{\gamma|y=a}$ 不服从 2 阶卡方分布。然而, 这 3 条曲线也具有相似的规律, 它们都界于 2 阶卡方分布与 1 阶卡方分布的理论曲线之间。对更高维度情况进行模拟也可得到类似结论: 当 a 不处于卡方子空间内时, $\chi^2_{\gamma|y=a}$ 的概率密度曲线界于 $n-1$ 阶卡方分布与 1 阶卡方分布之间, 具体曲线需要通过模拟获得。

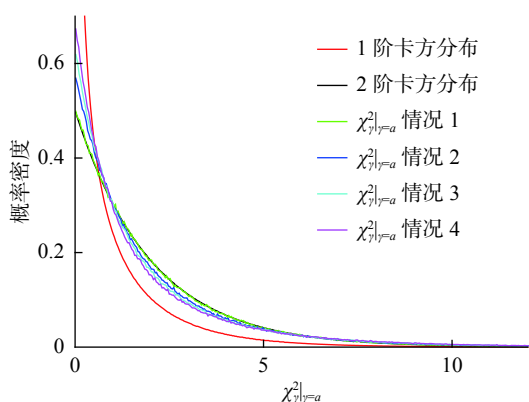


图 5 4 种情况下 $\chi^2_{\gamma|y=a}$ 的模拟概率密度曲线
Fig.5 Simulated density curves of $\chi^2_{\gamma|y=a}$ under the 4 uncertainties conditions

5 应用示例

实际应用中, 除不确定度权重平均值 x_w 外, 常用的线性估计还有算术平均值 \bar{x} ^[12]。本节将在 CMM 和 REM 条件下, 对 \bar{x} 的统计量 $\chi^2_{\gamma|y=a}$ 进行举例应用。

5.1 CMM 中检验数据一致性

假设有 4 个实验室参加比对, 测量结果及不确定度见表 2。

表 2 比对结果和不确定度 (正态分布且 $n = 4$)
Tab.2 Comparison results and uncertainties (Gaussian distribution and $n = 4$)

结果	实验室1	实验室2	实验室3	实验室4
x_i	9.5	13.9	7.2	11.6
u_i	1.4	2.0	1.6	1.8

应用统计量 $\chi^2_{\gamma|y=a}$ 之前, 需要通过第 2 节的模拟方法确认 $\chi^2_{\gamma|y=a}$ 服从或者近似服从 $n-1$ 阶卡方分布。由图 1 可见, 在本例中统计量 $\chi^2_{\gamma|y=a}$ 服从或者近似服从 3 阶卡方分布。因此, 本例中可应用 $\chi^2_{\gamma|y=a}$ 对 \bar{x} 进行 $n-1$ 阶卡方检验。

依据式(9)及表 2 中的比对结果及不确定度, 卡方统计量 $\chi^2_{\gamma|y=a}$ 为 8.384, 此值大于 3 阶卡方分布的 0.95 分位点 7.815。由此得到, 本例中对 \bar{x} 的卡方检验失败。

5.2 REM 中估计随机影响

在 5.1 节中的比对又加入两家实验室, 测量结果及不确定度见表 3(实验室 5 和实验室 6)。

表 3 比对结果和不确定度 (正态分布且 $n = 6$)
Tab.3 Comparison results and uncertainties (Gaussian distribution and $n = 6$)

结果	实验室1	实验室2	实验室3	实验室4	实验室5	实验室6
x_i	9.5	13.9	7.2	11.6	13.5	8.7
u_i	1.4	2.0	1.6	1.8	0.1	2.5

在 CMM 条件下, 基于表 3 中的不确定度, 得到 $\chi^2_{\gamma|y=a}$ 的模拟密度曲线如图 6 中绿色线所示。由于新加入的两个实验室不确定度与其他实验室不相当, 导致 $\chi^2_{\gamma|y=a}$ 的密度曲线处于 5 阶和 4 阶卡方分布之间。基于式(9)和表 3 中的比对结果, 得到 $\chi^2_{\gamma|y=a}$ 为 12.178, 大于通过模拟曲线得的 0.95 分位点 12.073。因此, 在 CMM 条件下, 对于 \bar{x} 的卡方检验失败。

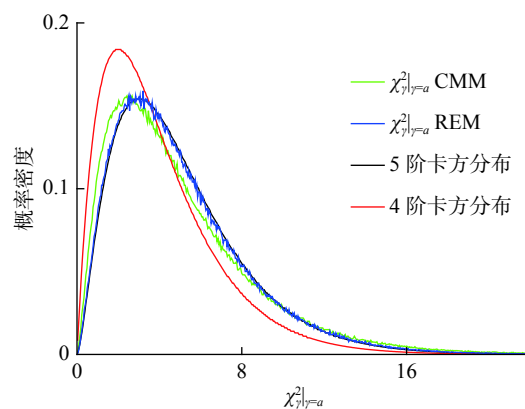


图 6 $\chi^2_{\gamma|y=a}$ 的模拟概率密度曲线 (CMM 和 REM)
Fig.6 Simulated density curves of $\chi^2_{\gamma|y=a}$ (CMM and REM)

在 REM 条件下, 每个实验室加入相等的随机影响 b_i 且 $b_i \sim N(0, \tau_a^2)$, 因此每个实验室的不确定度变为 $u^2(x_i) = \tau_a^2 + u_i^2$ 。以 $u^2(x_i)$ 替代式(9)中的 u_i^2 , 得

到 REM 中的统计量为

$$\chi^2_{\gamma|_{\gamma=a}} = n \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (u_i^2 + \tau_a^2)} \quad (12)$$

由第 2 节 χ^2_{γ} 的性质 2) 可知 $E(\chi^2_{\gamma|_{\gamma=a}}) = n - 1$, 采用矩估计法得到

$$n \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (u_i^2 + \tau_a^2)} = n - 1 \quad (13)$$

得到 τ_a^2 的估计公式为

$$\tau_a^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} - \frac{\sum_{i=1}^n u_i^2}{n} \quad (14)$$

基于式 (14) 及表 3 中的比对结果, 得到 τ_a^2 为 4.311, 并由此得到 $u(x_i) (i = 1 \cdots 6)$ 为 {2.504, 2.883, 2.621, 2.748, 2.079, 3.250}。另外, 基于此不确定度获得 $\chi^2_{\gamma|_{\gamma=a}}$ 的模拟密度曲线如图 6 中蓝色线所示, 可见其与 5 阶卡方分布非常吻合。

6 结论

卡方统计量是实验室比对线性参考值估计分析的重要统计工具, 但传统的卡方统计量只适用于不确定度权重平均值和等不确定度时的算术平均值。本研究针对一般线性参考值提出了新卡方统计量, 并采用理论分析和蒙特卡洛模拟研究了该统计量的性质和分布, 得到:

1) 当权重向量在卡方子空间内取值时, 对应的线性参考值估计的卡方统计量服从或近似服从 $n-1$ 阶卡方分布;

2) 当权重向量在卡方子空间外取值时, 对应的线性参考值估计的卡方统计量的概率密度曲线介于 $n-1$ 阶与 1 阶卡方分布概率密度曲线之间, 具体曲线可通过蒙特卡洛模拟获取;

3) 在任意权重组合时, 对应线性参考值估计的

卡方统计量期望为 $n-1$ 。

故此, 基于 CMM 和 REM, 新提出的卡方统计量可用于所有线性参考值估计的比对数据一致性检验和随机影响方差估计, 为更为广泛的线性参考值估计应用提供了统计分析工具。作为示例, 针对算术平均值给出了新卡方统计量及扩展应用, 为这种常用的线性参考值估计提供了新的统计分析方法。

参考文献

- [1] 原遵东, 王池, 林延东. 确定量值比对参考值的准则 [J]. 仪器仪表学报, 2010, 31(10): 2381-2385.
- [2] 张海云, 吴忠杰, 杭晨哲, 等. MCMC 法确定关键比对参考值及其不确定度 [J]. 计量技术, 2020(7): 3-7, 26.
- [3] Cox M G. The evaluation of key comparison data[J]. *Metrologia*, 2002, 39(6): 589-595.
- [4] Koepke A, Lafarge T, Possolo A, *et al.* Consensus building for interlaboratory studies, key comparisons, and meta-analysis[J]. *Metrologia*, 2017, 54(3): S34-S62.
- [5] Toman B, Possolo A. Laboratory effects models for interlaboratory comparisons[J]. *Accred Qual Assur*, 2009, 14(10): 553-563.
- [6] Elster C, Toman B. Analysis of key comparisons: estimating laboratories' biases by a fixed effects model using Bayesian model averaging[J]. *Metrologia*, 2010, 47(3): 113-119.
- [7] Steele A G, Douglas R J. Extending chi-squared statistics for key comparisons in metrology[J]. *Journal of Computational and Applied Mathematics*, 2006(192): 51-58.
- [8] Rukhin A L. Weighted means statistics in interlaboratory studies[J]. *Metrologia*, 2017, 46(3): 323-331.
- [9] Toman B, Fischer J, Elster C. Alternative analyses of measurements of the Planck constant[J]. *Metrologia*, 2012, 49(4): 567-571.
- [10] Paule R C, Mandel J. Consensus values and weighting factors[J]. *J. Res. Natl Bureau Stand*, 1982, 87(5): 377-385.
- [11] Graybill F A, Deal R B. Combining unbiased estimators[J]. *Biometrics*, 1959(15): 543-550.
- [12] BIPM. CCQM guidance note: estimation of a consensus KCRV and associated degrees of equivalence[DB/OL]. http://www.bipm.org/cc/CCQM/Allowed/19/CCQM13-22_Consensus_KCRV_v10.pdf, 2013: 23-25.

本文编辑: 李润芝