论文层面采用非线性领域归一化方法得到的非线性领域归一化引文计数不能相加或取均值

王兴[a],[1]

[a]中国山西省太原市030006，山西财经大学信息学院

[1]通讯作者。电子邮箱：wangxing@sjtu.edu.cn

摘要学术界有一个非常重要的问题尚未得到足够重视，即使用非线性领域归一化方法得到的论文层面非线性领域归一化引文计数不能相加或取均值。遗憾的是，学术文献中存在许多将单篇论文的非线性标准化引文计数相加或取均值的情况，这表明非线性领域归一化方法长期以来在学术界遭到误用。本文开展了以下两项研究工作。首先，我们从数学理论分析的角度，分析了为什么单篇论文的非线性标准化引文计数不能相加或取均值，并对分析中的关键步骤给出了数学证明。其次，我们将现有的主要领域归一化方法系统地划分为线性领域归一化方法和非线性领域归一化方法。上述两项研究工作为今后正确使用领域归一化方法提供了理论依据，有助于避免非线性数据持续遭到误用。此外，由于我们的数学证明适用于整个实数域中的所有非线性数据，因此本研究对整个数据与信息科学领域也具有重要意义。

关键词 领域归一化方法；非线性归一化方法；非线性变换；等距变换

引言

引文计数是科学界中一项重要且常用的定量指标，被用作衡量论文研究影响力的替代指标，涉及在不同聚合层级上（例如个体学者、研究团队、机构、期刊和国家层面）开展研究影响力评估。此类评估的结果往往在决定科研经费的授予、学者聘用，乃至科研机构的命运方面发挥重要作用。因此，基于引文分析的研究影响力评估应以最精确且无偏的方式进行（Radicchi 等，2008）。

然而，基于引文分析评估研究影响力的一个潜在问题是，由于不同学科的引文实践存在固有差异，不同领域论文的原始引文计数无法直接比较。对原始引文计数进行此类比较会产生偏差，可能妨碍对研究影响力的公平评价。例如，数学领域论文的引文计数通常低于分子生物学领域论文。

鉴于此，有必要采用领域归一化来尽量缩小或消除学科差异，以确保能够公平比较不同学科的引文影响力[1]。领域归一化方法是一种旨在消除引文计数中学科差异的数学变换。归一化后，归一化引文计数应不受学科影响，从而可以通过比较归一化引文计数，自然而然地实现不同学科间引文影响力的比较。

迄今为止，随着信息计量学的发展，各种领域归一化方法相继提出。根据归一化过程是否属于线性变换，这些领域归一化方法可分为线性领域归一化方法和非线性领域归一化方法（Zhang 等，2015）。例如，科研评价实践中一种常用的方法——基于均值的方法，即将一篇论文的原始引文数除以同一学科所有论文的平均引文计数，是一种典型的线性领域归一化方法；而另一种非常流行的归一化方法——百分位排名法，即将论文的原始引文计数映射到其在引文分布中的相对位置，则是一种典型的非线性领域归一化方法。

学术界和评价实践界有一个尚未得到充分关注的极其重要的问题，即，使用非线性领域归一化方法获得的论文层面的非线性领域归一化引文计数无法相加或取平均值。实际的科研评价通常在聚合层级（例如机构、期刊和国家层面）进行，这要求将单篇论文的归一化引文计数相加或取平均值，以衡量聚合层级的标准化引文表现。然而，在学术

文献或评价实践界中，可以发现许多将单篇论文的非线性标准化引文计数相加或取平均值的案例（例如，Bornmann，2020；Bornmann 和 Williams，2020；Lundberg，2007年；Maflahi 和 Thelwall，2021年；Schniedermann，2021年；Thelwall，2017年、2020年；Thelwall 和 Maflahi，2020年；Wu 等，2020年），这表明非线性领域归一化方法长期以来一直被误用。第二节将分析为何单篇论文的非线性领域归一化引文计数无法相加或取平均值，这也是本文的重点。

此外，我们在第三节中根据归一化过程是否属于线性变换，对现有常用的领域归一化方法进行梳理和总结，并据此将这些归一化方法分为线性领域归一化方法和非线性领域归一化方法。最后，我们在“结论与讨论”一节中给出总结性意见，并在补充信息中讨论使用领域归一化方法时面临的一个两难问题，其中涉及归一化效果的检验方法以及领域归一化的理想目标。

为什么对单篇论文进行领域归一化时，所用的归一化方法会产生不同的结果?

什么是线性领域归一化方法，什么是非线性领域归一化方法?

我们首先定义线性领域归一化方法和非线性领域归一化方法。

在同一领域中，如果论文的原始引文数$y = kx + b$（k 为非零常数，b为常数）与该论文的归一化引文计数$y,$之间存在 x这样的映射关系，那么我们称y 是$x,$的线性变换，并将这种变换称为线性领域归一化方法。否则，我们称y是$x,$的非线性变换，相应的变换过程称为非线性领域归一化方法。

例如，常用的基于均值的方法（见公式 (1)）和 z 分数法（见公式 (2)）都是典型的线性归一化方法:

$$y = \frac{x}{m} = \frac{1}{m} \cdot x \quad , \tag{1}$$

$$y = \frac{x - m}{sd} = \frac{1}{sd} \cdot x - \frac{m}{sd} \quad . \tag{2}$$

在公式 (1) 中， x 和 y 分别表示论文的原始引文计数和归一化引文计数，m 表示与给定论文具有相同学科、出版年份和文献类型的论文的平均引文计数。事实上，基于均值的方法是一般线性方法 $y = kx + b,$ 的一个特例，其中缩放项 k =

1/*m*，截距项 $b = 0$。类似地，公式 (2) 中 x、y 和 m 的含义与公式 (1) 中相同，而 sd 表示与给定论文具有相同学科、出版年份和文献类型的论文的引文计数的标准差；z 分数法也是一般线性方法 $y = kx + b$,的一个特例，其中缩放项 $k = 1/sd$ ，截距项 $b = -m/sd$。

广受欢迎的百分位排名法是一种典型的非线性领域归一化方法。单篇论文原始引文数y 对应的百分位排名得分x，定义为：与给定论文具有相同学科、出版年份和文献类型的论文中，有y%的论文的引用次数 不高于（或等于）该给定论文的原始引文数x^2。例如，假设某篇论文的原始引文数为30。如果与该论文具有相同学科、出版年份和文献类型的论文中，有80%的论文的引文数不高于（或等于）30，那么该论文的百分位排名得分就是80。在百分位排名法下，原始引文数x到百分位排名得分y的映射通常无法形成$y = kx + b$这样的线性关系；因此，百分位排名法是一种非线性领域归一化方法。

“等间隔”测量：判断单篇论文的归一化引文计数是否可以相加的标准

人们有时容易忽略的一点是，加法要求进行等距测量，也就是说，数据的每个测量单位都必须等距。例如，测试分数为1分和2分之间的距离（即差值），与2分和3分之间的距离相同。数据测量单位之间等距，是对数据进行加法运算的前提。若忽略这一重要前提，就会得出错误的结果。下面来看一个虚构示例。

图1中有两个方框，方框1和方框2。我们可以用人的手的长度作为测量单位，来测量这两个方框的长度。有两种测量方法。第一种方法是用不同人的手组成一把尺子（即图1中的尺子A），来测量这两个方框的长度。由于不同人的手长度不同，因此这里的测量单位（即每只手的长度）不是等距的。按照这种测量方法，方框1的长度是3个“手长”，方框2的长度是9个“手长”，这会让我们得出这样的结论：三个方框1的长度之和等于方框2的长度，因为“3个手长+ 3 个手长 + 3 个手长 = 9 个手长”。

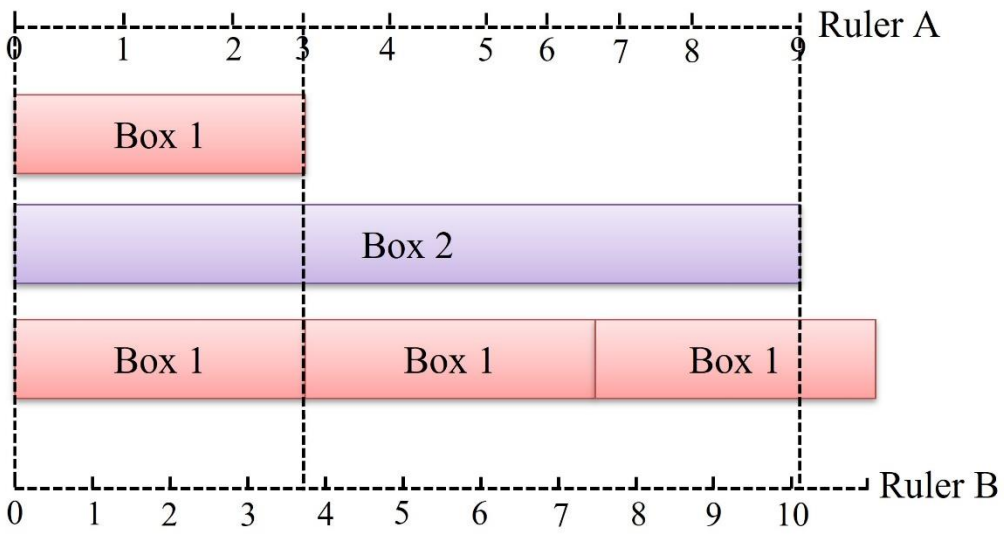

图1用于解释为何加法要求等距测量的虚构示例 注：每个数字之间虚线“------”的长度表示人手的长度，即测量单位。

然而，如图1所示，实际上，三个方框1的长度之和大于方框2的长度。显然，按照第一种测量方法得出的结论是错误的。与第一种测量方法不同，第二种测量方法用同一个人的一只手组成一把尺子（即图1中的尺子B），来测量这两个方框的长度。这样一来，手的长度（即测量单位）是固定的；也就是说，测量单位是等距的。显然，采用第二种科学的测量方法，我们可以得出正确的测量结论，这与实际情况相符。

至此，通过上述示例可以很容易理解，为何加法要求等距测量。当数据的测量单位并非等距时，数据不能相加；否则，我们将得出错误的结论。

同样，在同一领域中，单篇论文的原始引文计数中的每个引文计数彼此等距（事实上，这里的每个引文计数相当于上述示例中的测量单位——手的长度）；因此，同一领域中的原始引文计数可以相加或求平均。然而，如果我们希望在实施领域归一化后，将单篇论文的归一化引文计数相加或求平均，就必须同样确保同一领域中这些归一化引文计数的每个测量单位彼此等距，这意味着领域归一化必须是线性变换；我们将在下一节解释其中的原因。

非线性标准化引文计数不再等距

线性领域归一化方法的归一化过程是一种等距变换，不会改变各引文计数的等距性质（事实上，

线性归一化变换与等距变换互为充分必要条件；我们已在附录A中证明了这一结论）。在此，我们使用了一个虚构数据集，其中包含某一领域内引文计数各不相同的52篇论文，以通俗易懂的方式说明这一点（见表1）。

表 1用于计算基于均值的方法和百分位排名法所得归一化引文计数的论文数据集（52篇论文）

引文计数	归一化论文	$k = 1/m$	m-得分	排名i	百分位<style id='1'> 排名得分
0	9	0.07	0.00	9	17.31
1	8	0.07	0.07	17	32.69
2	6	0.07	0.14	23	44.23
3	4	0.07	0.21	27	51.92
4	4	0.07	0.28	31	59.62
5	2	0.07	0.35	33	63.46
6	2	0.07	0.42	35	67.31
7	1	0.07	0.49	36	69.23
8	1	0.07	0.56	37	71.15
9	1	0.07	0.63	38	73.08
10	2	0.07	0.70	40	76.92
12	1	0.07	0.84	41	78.85
15	1	0.07	1.05	42	80.77
16	1	0.07	1.12	43	82.69
20	2	0.07	1.40	45	86.54
25	1	0.07	1.74	46	88.46
38	1	0.07	2.65	47	90.38
42	1	0.07	2.93	48	92.31
43	1	0.07	3.00	49	94.23
80	1	0.07	5.58	50	96.15
120	1	0.07	8.38	51	98.08
200	1	0.07	13.96	52	100.00

注：k 表示一般线性方法中的缩放项 $y = kx + b$；m表示52篇论文原始引文计数的均值；m得分表示使用基于均值的方法得到的归一化引文计数；i表示原始引文计数的排名；百分位排名得分表示使用百分位排名法得到的归一化引文计数[3]。

在这个虚构示例中，我们使用基于均值的方法（一种典型的线性领域归一化方法）对该领域52篇论文的原始引文计数进行归一化，得到这些论文的归一化引文计数（即m分数）（见表1第四列）。实施基于均值的方法之前，原始引文计数中的每个引文计数（即测量单位）之间的距离相等。实施基于均值的方法之后，归一化引文计数中的各个测量单位之间的距离仍然相等：只是原始引文计数中的每个引文计数都按$k = 1/m$缩放，因此归一化后

的引文计数之间的距离经此缩放后都变为$1/m$，在本例中即为0.07（见图2）。

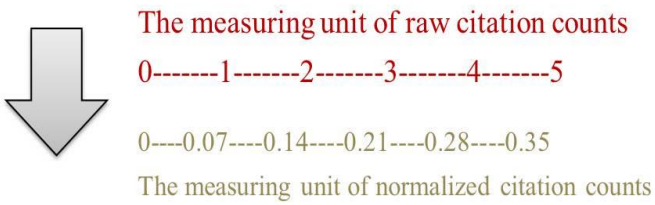

图2从原始引文计数到归一化引文计数的等距变换示意图

其他线性领域归一化方法与上述基于均值的方法具有相似的变换原理：通过线性变换得到的归一化引文计数的每个测量单位仍然是等距的，且该距离等于缩放项k（见附录A）。 因此，单篇论文的线性归一化引文计数可以相加或取平均值。

然而，非线性领域归一化方法的归一化过程并非等距变换；因此，非线性标准化引文计数不再等距，因而无法相加或取平均值。这里，我们使用了与上述相同的另一个虚构数据集示例来说明这一点。

在表1中，我们采用百分位排名法这一典型的非线性领域归一化方法，对52篇论文的原始引文计数进行归一化，得到这些论文的归一化引文计数（即百分位排名得分）（见表1第六列）。在实施百分位排名法之前，原始引文计数的每个引文计数（即测量单位）之间的距离相等。实施百分位排名法之后，归一化引文计数的各测量单位之间不再等距。例如，原始引文中0次与1次、3次与4次，以及42次与43次引文之间的距离均为1；然而，它们对应的百分位排名得分之间的距离分别为15.38、7.7和1.92（见表1和图3）。在这种情况下，拥有42次原始引文的论文A与拥有1次原始引文的论文B的引文计数之和，等于拥有43次原始引文的论文C与拥有0次引文的论文D的引文计数之和；然而，上述论文对应的百分位排名得分之和却不再相等（见表1，公式（3）和（4））。此外，再举一个更夸张的例子：4篇各有1次原始引文的论文，其百分位排名得分之和竟然高于一篇有200次引文的论文的百分位排名得分；这在实际科研评价中非常不合理（见表1，公式（5）和（6））。因此，单篇论文的非线性标准化引文计数不能相加或取平均值。

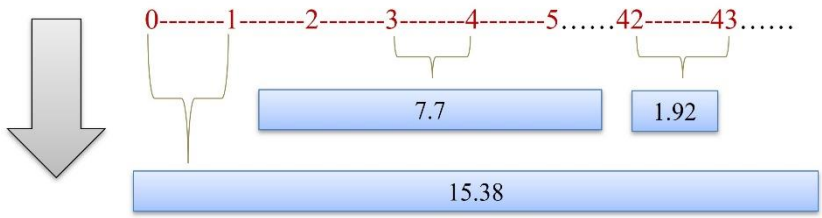

Figure 3 示意图：对归一化引文计数进行非线性归一化变换后，每个测量单位不再等距

进行非线性变换之前：

$$42 \ + \ 1 \ = \ 43 \ + \ 0 \quad , \tag{3}$$

(A) (B) (C) (D)

进行非线性变换之后：

92.31 32.69 94.22 17.31 (A) (B)
(C) (D) . (4)

进行非线性变换之前：

$$1 \ + \ 1 \ + \ 1 \ + \ 1 \ < \ 200 \quad , \tag{5}$$

(E) (F) (G) (H) (I)

执行非线性变换后：

32.69 32.69 32.69 32.69 100 >
(E) (F) (G) (H) (I) . (6)

事实上，非线性领域归一化变换与非等距变换互为充分必要条件（我们已在附录B中证明这一结论）。一方面，如果原始引文计数与其归一化引文计数之间的映射关系是非线性的，那么映射关系图中必定至少存在两个斜率值不同的点：如图4所示，斜率值 $k1$ 和 $k2$ 不同。这里的斜率值实际上就是上文提到的缩放项k，这些缩放项不同意味着缩放并非等距；也就是说，非线性领域归一化变换不是等距变换。另一方面，如果某种领域归一化变换不是等距变换，那么必定存在不同的缩放项k；也就是说，必定存在不同的斜率值，从而导致原始引文计数与其归一化引文计数之间的映射关系呈非线性，如图4所示。

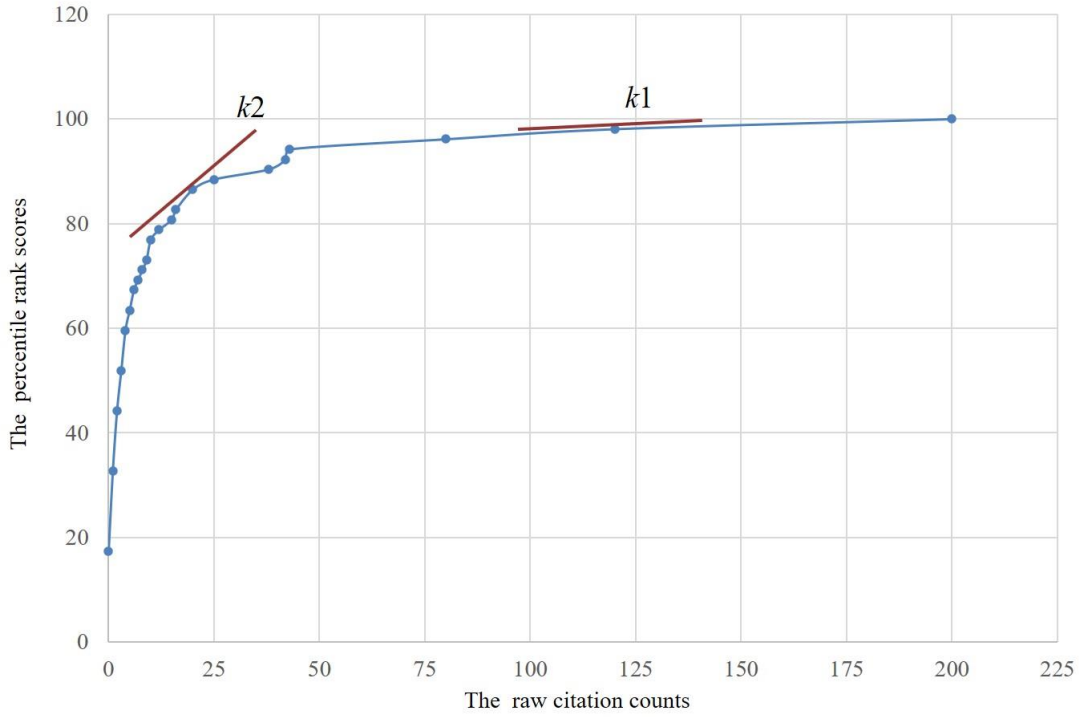

图 452篇论文的原始引文计数与百分位排名得分之间的映射关系

利用一种常用的流行语言和非流行语言模型进行 CoT 思维链和用户级别的参数化微调

在第二节中，我们已经定义了线性领域归一化方法和非线性领域归一化方法。这里，我们梳理并总结现有常用的线性和非线性领域归一化方法。

基于均值的方法、中位数法、z 分数法、基于优化的线性方法（Zhang等，2015）以及相对引文比率（RCR）法（Hutchins等，2016）均属于线性领域归一化方法。在这些领域归一化方法中，同一领域内原始引文数 x 到归一化引文计数y的映射呈线性关系，例如$y = kx + b$。我们汇总了这些线性领域归一化方法中的参数k和 b（见表 2）。

表 2几种主要的线性领域归一化方法概述

信息是什么?	**材料详情及英文描述的空间限制**	***k, b***
基于均值的方法[4]	$y=\frac{x}{m}=\frac{1}{m}\cdot x$ *x* 和*y* 分别表示论文的原始引文计数和归一化引文计数，*m*表示与给定论文属于同一领域的论文的平均引文计数（下同）。	$k=\frac{1}{m},b=0$
基于中位数的方法	$y=\frac{x}{\text{中位数}}=\frac{1}{\text{中位数}}\cdot x$ 中位数是与给定论文属于同一领域的论文的引文计数中位数。	$k=\frac{1}{\text{中位数}},b=0$
z 分数法	$y=\frac{x-m}{sd}=\frac{1}{sd}\cdot x-\frac{m}{sd}$ 标准差是与给定论文属于同一领域的论文的引用次数标准差。	k 标准差 $=\frac{1}{}$，b 标准差 $=-\frac{m}{}$
基于优化的线性方法	该方法首先假设，经过线性方法变换后，所有学科的引文分布曲线都与某条共同参考分布曲线重合 *y* = *kx* + *b*，其中未知参数为*k*和*b*，然后计算此时各学科的*k*、*b* 值（这一过程等同于一个优化问题即寻找归一化效果最佳时*k* 和*b*的值）。确定各学科的参数*k*和*b*后，各学科的归一化方法*y* = *kx* + *b*也就自然确定了。	参见其中的说明左栏
相对引文比（RCR）法	$y=\frac{x}{\text{ECR}}=\frac{1}{\text{ECR}}\cdot x$ 1. 在此方法中，一篇论文所属的领域由该论文所在的共引网络确定。 2. 一篇论文的 *FCR* 值（预期领域引文数）等于该焦点论文所在共引网络中所有论文的期刊引文率的平均值。 3. 使用此方法时，必须先选取一组基准论文，并计算其 *ACR* 值（均值年引文数）和每篇基准论文的 *FCR* 值。然后，通过这些基准论文的 *ACR* 值（因变量）和 FCR 值（自变量）拟合回归曲线（方程）拟合于*ACR*值（因变量）和*FCR*值（自变量）这些基准论文的：ACR FCR= β · +α 4. 在实际评估中，一篇焦点论文的 *ECR* 值（期望被引次数）等于 *ACR* 值。当上述回归方程中的自变量取值为 *FCR* 值。最终焦点论文的归一化引文计数等于该焦点论文的原始引文数除以其期望被引次数——*ECR* 值。	$k=\frac{1}{ECR},b=0$

值得注意的是，近年来在生物医学界颇受欢迎的 RCR方法，可视为一种特殊的线性领域归一化方法。在这种方法中，论文所属的领域由该论文的共引网络决定；因此，现实中两篇论文很难拥有完全相同的共引网络。换言之，现实中通常不存在两篇或更多论文属于同一领域的情况：每篇论文都有其独有的领域。因此，在这种方法下，每篇论文都有其独有的期望被引次数（*ECR* 值）和相应的独有缩放项 k (k=1/*ECR*)。

百分位排名法、引文 z 分数法（Lundberg，2007年）、归一化对数引文分数（NLCS）法（Thelwall，2017年）和逆向工程法（Radicchi 和 Castellano，2012a年）都是典型的非线性领域归一化方法（见表 3）。在这些领域归一化方法下，同一领域中从原始引文数 x 到归一化引文计数 y 的映射无法形成诸如 $y = kx + b$ 这样的线性关系。

Table 3 若干主要非线性领域归一化方法概述

Nai Tf Ti tefd	材料详情及英文描述的空间限制
百分位排名法	一篇论文的原始引文数 y 的百分位排名得分 x 定义为：有 y %的论文，其引文计数与该论文的与给定论文处于同一领域的论文，其原始引文数小于或等于给定论文的原始引文数 x 。
引文 z 分数法	$$y = \frac{\ln(1+x) - \mu_{[\ln]}}{\sigma_{[\ln]}}$$ x 和 y 分别表示论文的原始引文数和归一化引文计数（下同），$\mu_{\ln}$ 和 $\sigma_{\ln}$ 分别表示均值和标准差分别表示对数转换后的引文计数（加一）分布的标准差（下同）。这种方法本质上是一种对数 z 分数法。
归一化对数引文 Score（NLCS）方法	$$y = \frac{\ln(x+1)}{\mu_{[\ln]}}$$ 该方法类似于上述引文 z 分数法，本质上是一种基于对数均值的方法。
逆向工程法	$$x = \lambda(y)^{\alpha}$$ 不同领域引文分布中位于相同分布位置的论文应具有相同的引文影响力。基于这一思想，该方法首先利用不同领域所有论文的引文构建一个共同参考分布。当将某一领域中的原始引文数 x 映射为归一化引文计数 y 时，首先计算 x 在该领域原始引文分布中的位置 P；然后，在共同参考分布中找到位于相同位置是 x 的归一化引文计数。上述映射过程在大多数学科中可以近似描述为非线性幂律函数，如上所示，其中 α 和 λ 是领域相关常数。这种方法旨在从以下角度积极实现归一化的理想目标（即，不同学科的所有归一化引文分布都与同一个分布一致）。

对于百分位排名法，按照定义，百分位排名本身呈矩形分布，而原始引文计数的分布始终呈非矩形分布。将原始引文计数映射为百分位排名分数，会改变原始引文计数非矩形分布的形状，因此这一映射过程并非等距变换。根据本文第二节，非等距变换与非线性归一化互为充分必要条件；因此，百分位排名法是一种非线性领域归一化方法。

如表 3 所示，引文 z 分数法和 NLCS 法均采用对数变换。对数变换属于非等距变换，而对数函数的图像表明，对数变换显然是一种非线性变换。因此，引文 z 分数法和 NLCS 法均为非线性领域归一化方法。

如表 3 所示，对于逆向工程法，在大多数学科中，原始引文数 x 与归一化引文计数 y 之间的映射关系可用非线性幂律函数描述；因此，逆向工程法显然是一种非线性领域归一化方法。

此外，还有三种领域归一化方法难以确定其属于线性还是非线性方法，分别是施引侧归一化方法（Leydesdorff & Opthof，2010；Waltman & van Eck，2013；Zitt & Small，2008）、结合施引侧归一化与引文百分位的方法（Bornmann，2020），以及汇率法（Crespo 等，2013）（见表 4）。

表 4难以归类为线性或非线性归一化方法的其他领域归一化方法概要

信息是什么?	材料详情及英文描述的空间限制
施引侧归一化方法	该方法的主要思想是，每个引文均按施引出版物或施引期刊中的参考文献数量进行归一化；该方法的一项显著优势是无需领域分类系统。该方法的三种具体方法如下，其中 SNCS 是“来源归一化引文得分”的缩写。 $SNCS1=\sum_{i=1}^{c}\frac{1}{a_i}$，　$SNCS2=\sum_{i=1}^{c}\frac{1}{r_i}$，　$SNCS3=\sum_{i=1}^{c}\frac{1}{p_i r_i}$， 其中，$c$为焦点论文的引文计数（下同），$a_i$ 为与第i篇施引出版物在同一期刊和同一出版年份发表的出版物中的活跃参考文献数量均值，r_i 为第i篇施引出版物中的活跃参考文献数量，而p_i 为在发表在与第i篇施引出版物属于同一期刊且出版年份相同。
结合施引侧标准化和引文百分位的方法	此方法首先通过施引侧方法（如 SNCS3）对原始引文计数进行归一化，然后根据每篇论文的 SNCS3 得分，为其赋予一个相应的百分位排名得分。

Table 4 （续）

Nam o 在.....中	材料详情及英文描述的空间限制
汇率法	这种归一化方法的理念，类似于按照汇率将某个国家的货币兑换成通用国际货币，而该方法中的交换率类似于某国货币与通用国际货币之间的汇率。 货币。 **原始汇率与对应的归一化指标如下：** $$e_f(\pi)=\frac{\mu_f^{\pi}}{\mu^{\pi}},\quad y=\frac{x}{e_f(\pi)}\quad (k=\frac{1}{e_f(\pi)},\ b=0).$$ 对于每个学科f，此方法首先根据引文f领域内的论文划分为n个大小为 的分位区间 N_f/n 这些论文的计数（ 表示学科f中的论文总数；表示某个分位区间，并且π $\pi=1,2,\cdots,n$； 有 N_f/n 每个分位区间中有 篇学科f的论文）。类似地，所有学科的全部论文都被划分为n个分位 区间，每个区间大小为 ，划分依据是这些论文的引文计数。表示中论文μ_f^{π}的平均引文计数 分位区间 中的论文平均引文计数 。 表示引文分布中分位区间 的论文平均引文计数 π 由所有学科的所有论文构成， 是分位区间内论文的交换率 所属领域f（相同的 π 如下）。x是属于领域 分位区间内的一篇焦点论文的原始引文数f，而y是标准化引文 计数。 **Th 贸易融资和应收账款融资银团贷款 的条款如下：** $$e_f=\frac{1}{(\pi^{M}-\pi_m)}\sum_{\pi}e_f(\pi),\quad y=\frac{x}{e_f}\quad (k=\frac{1}{e_f},\ b=0).$$ π^{M} π_m 是特殊区间的上（下）边界值，且。，其等于该特殊区间内每个分位区间的原始交换率的算术平均值 ，是属于领域f的所有论文的交换率。x是属于领域f的一篇焦点论文的原始引文数，y是该焦点论文的归一化引文计数。

线性领域归一化方法和非线性领域归一化方法的概念以领域分类系统为前提。在执行归一化变换之前，同一领域内的原始引文计数具有可比性且可以相加。执行线性领域归一化变换后，同一领域内的归一化引文计数仍然具有可比性且可以相加（此时，不同领域的归一化引文计数当然也具有可比性且可以相加，因为已经完成了领域归一化变换）。根据上述第二节的分析，对原始引文计数执行非线性领域归一化变换后，同一领域内的归一化引文计数将无法再相加。就施引侧归一化方法而言，该方法本身没有领域分类的概念，归一化过程也不需要领域分类系统。因此，该方法本身不属于区分线性和非线性领域归一化方法的概念范畴；我们很难将该方法归类为线性或非线性领域归一化方法。

结合施引侧归一化与引文百分位的方法，本质上是一种施引侧归一化方法。不过，为了便于人们更好地理解 SNCS3 等施引侧方法下的归一化引文计数，该方法最终会根据每篇论文的 SNCS3 得分，为其赋予相应的百分位排名得分。由于该方法本质上是一种施引侧归一化方法，因此它也没有领域分类系统这一概念，其归一化过程也不需要领域分类系统。因此，该方法也不属于区分线性与非线性领域归一化方法的概念范畴。

从原始……的角度来看，汇率法是一种非线性领域归一化方法

该方法的定义，以及相关原因，可以通过分析原始交换率的公式来解释（见表 4）。不同分位区间内论文的平均引文计数π 在同一领域内（即交换率的分子）各不相同，例如，。同样，不同分位区间内论文的平均引文计数π 在由所有学科的全部论文构成的引文分布中（即交换率的分母）也各不相同，例如，。因此，不同分位区间的交换率π 在同一领域内很可能不会完全相同，例如，即。也就是说，缩放项 对于不同分位区间论文的原始引文计数

$$\mu^{1} \neq \mu^{200}$$

$$\frac{\mu}{\mu} \neq \frac{\mu}{\mu} \qquad (1)\ (200)_{f}$$

$$千 = \frac{1\ (\ )_{f}}{\pi}$$

区间π 在同一领域内有所不同（例如，），这表明这种缩放并非等距；因此，该方法是一种非线性领域归一化方法。

然而，可能是出于计算便利等原因，Crespo 等选择了某个特殊区间 [ ] 例如 [706, 998] ，以重新定义交换率：，（见表 4），该交换率等于特殊区间π 内每个分位区间 [ ] 的原始交换率的算术平均值，并且被用作属于领域 f 的所有论文的统一交换率。在这种情况下，同一领域内所有论文原始引文计数的缩放项 为一个统一值，等于 ；因此，这种缩放是等距的，这一变换过程自然是线性的。

$$\frac{1}{e_f}$$

因此，总而言之，对于汇率法，该方法属于线性还是非线性领域归一化方法，必须根据具体的算法定义来确定。

结论与讨论

本文开展了以下两项研究工作，它们也是本文的两项学术贡献。第一，从数学理论分析的角度，我们分析了为什么使用非线性领域归一化方法得到的单篇论文非线性标准化引文计数不能相加或求平均；此前学界尚未对此进行研究。数据能够相加的前提是其测量单位等距；然而，非线性领域归一化方法的归一化过程并非等距变换（针对非线性领域归一化方法并非等距变换这一关键步骤，我们给出了数学证明）；因此，非线性标准化引文计数的测量单位不再等距。所以，单篇论文的非线性标准化引文计数不能相加或求平均。第二，我们将现有的主要领域归一化方法系统地划分为线性领域归一化方法和非线性领域归一化方法；此前学界也未进行过这种分类。

上述两项研究工作为未来正确使用

科研评价中的领域归一化方法提供了进一步的理论依据，避免学术界和评价实践界继续误用非线性领域归一化方法。

此外，我们在附录中给出的定理1和定理2的证明，旨在探讨的不仅是线性（非线性）领域归一化变换与等距（非等距）变换之间的具体关系，还包括更一般意义上，整个实数域上的线性（非线性）变换与等距（非等距）变换之间的关系。因此，我们实际上证明了，任何实数域中的所有非线性数据（例如其他研究主题中的百分位排名分数）都不能相加或求平均。这一结论对整个数据科学和信息科学领域都具有重要意义。

事实上，长期以来，不仅非线性领域归一化方法遭到误用，数据科学和信息科学领域其他研究主题中的其他非线性数据（例如百分位排名分数和对数分数）也一直被误用，因为这些非线性数据同样被相加或求平均。即使在一些信息科学顶级期刊中，也经常出现误用非线性数据的情况（例如，Jackson 等，2019；Safon 和 Docampo，2020；Zheng 等，2022）。因此，本文的研究不仅对领域归一化、科研评价、科学学和信息计量学领域的理论与实践具有重要意义，也对整个数据科学和信息科学领域具有重要意义。

结语

[1] 在实际科研评价中，跨领域比较引文影响力以及相应的领域归一化问题往往不可避免，也常常是必要的，因为科研评价通常是在一定的汇总层面（例如机构、期刊和国家层面）进行的。机构等受评对象很少只涉及单一学科，通常涵盖多个研究领域（Bornmann，2020；Radicchi 和 Castellano，2012b；Rafols 等，2012）。此外，不同领域的学者争夺相同资源（例如学术职位或学术奖项），也涉及这种“苹果与橘子”式的比较难题（Radicchi 和 Castellano，2012b）。

[2] 计算百分比排名分数有多种具体方法，例如 *P*100、*P*100’、 *CP-IN* 和 *CP-EX* 方法。有关这些具体方法，请参阅题为《引文影响力百分位测度的评估及改进建议》的论文（Bornmann 和 Williams，2020）。

[3] 此处计算百分位排名分数的具体方法是*CP-IN*方法（Bornmann 和 Williams，2020）。

[4] 在本节和本表中，我们仅梳理基于论文层面的领域归一化方法，因此，诸如新皇冠指标（Waltman 等，2011）和 WCNCI 指标（Wang 和 Zhang，2020）等汇总层面上的一些基于均值的归一化指标未单独列出。同样，其他汇总层面上的领域归一化指标，例如*I*

指标（Leydesdorff 和 Bornmann，2011），也未在本节及后续表格中单独列出。

参考资料

Bornmann, L.（2020）。如何对文献计量学中的引文影响力进行标准化？一种结合施引侧标准化与引文百分位数的新方法。《定量科学研究》，*1* (4), 1553-1569。Bornmann, L., & Williams, R.（2020）。引文影响力百分位测度的评估及改进建议。《科学计量学》，*124*(2), 1457-1478。Crespo, J. A., Li, Y., & Ruiz-Castillo, J.（2013）。测量不同科学领域引文实践差异对引文不平等的影响。《*PLoS ONE*》，*8*(3), e58727。Hutchins, B.I., Yuan, X., Anderson, J.M., & Santangelo, G.M.（2016）。相对引文比（RCR）：一种利用引文率测量文章层面影响力的新指标。《*PLoS Biology*》，*14*(9), e1002541。Jackson, S., Vanteeva, N., & Fearon, C.（2019）。数据泄露严重程度对强制性数据泄露通知函可读性影响的研究：来自美国企业的证据。《美国信息科学与技术学会会刊》，*70*(11), 1277-1289。Leydesdorff, L., & Bornmann, L.（2011）。综合影响力指标与影响因子（影响因子）的比较：一种具有政策启示的替代性研究设计。《美国信息科学学会会刊》，*62*(11), 2133-2146。Leydesdorff, L., & Opthof, T.（2010）。Scopus 的每篇论文来源标准化影响力（来源标准化影响力指标）与基于引文分数计数的期刊影响因子之比较。《美国信息科学与技术学会期刊》，*61*(11), 2365-2369。Lundberg, J.（2007）。提升皇冠指标——引文 z 分数。《信息计量学杂志》，*1*(2), 145-154。Maflahi, N., & Thelwall, M.（2021）。职业生涯较长的本国研究人员能够产生更高的平均引文影响力，但这种影响力不会随时间增加。《定量科学研究》，*2*(2), 560-587。Radicchi, F., & Castellano, C.（2012a年）。通过逆向工程方法抑制引文偏差，揭示引文分布的普适性。《*PLoS ONE*》，*7*(3), e33833。Radicchi, F., & Castellano, C.（2012b）。检验引文指标在不同科学领域间比较中的公平性：以分数引文计数为例。《信息计量学杂志》，*6*(1), 121- 130。Radicchi, F., Fortunato, S., & Castellano, C.（2008）。引文分布的普适性：迈向科学影响力的客观测量。《美国国家科学院院刊》，*105*(45), 17268-17272。Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A.（2012年）。期刊排名如何抑制跨学科研究：创新研究与商业与管理的比较。《研究政策》，*41*(7), 1262-1282。Safon, V., & Docampo, D.（2020）。基于客观科研绩效数据分析声誉偏差对全球大学排名的影响：以上海排名（世界大学学术排名）为例。《科学计量学》，*125*(3), 2199-2227。Schniedermann, A.（2021）。医学研究中系统综述与基于指南的系统综述之比较。《科学计量学》，*126*(12), 9829-9846。Thelwall, M.（2020）。1996—2018年英语国家中七个国家有六个存在女性引文影响力优势

。《美国信息科学与技术学会会刊》，*71*(8), 979-990。Thelwall, M.（2017）。用于科研评价的三个实用的学科标准化替代指标公式。《信息计量学杂志》，*11*(1), 128-151。Thelwall, M., & Maflahi, N.（2020）。不同国家和学科之间的学术合作率与引文关联存在显著差异。《美国信息科学与技术学会会刊》，*71*(8), 968-978。Waltman, L., & van Eck, N. J.（2013）。不同引文影响指标标准化方法的系统实证比较。《信息计量学杂志》，*7*(4), 833-849。Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J.（2011b）。迈向新的皇冠指标：一些理论思考。《信息计量学杂志》，*5*(1), 37-47。Wang, X., & Zhang, Z.（2020）。通过考虑短期与长期引文影响之间的相关性，提高短期引文影响指标的可靠性。《信息计量学杂志》，*14*(2), 101019。Wu, J., Xiao, H., Ren, M., Feng, S., & Ding, Z.（2020）。论文引文标准化指数（PCSI）：一种基于引文定量评价科学论文的新指数。《图书情报工作》，*64*(23), 22-30。（中文）。Zhang, Z., Cheng, Y., & Liu, N. C.（2015）。从优化角度改进基于均值的方法的归一化效果：基于优化的线性方法及其性能。《科学计量学》，*102*(1), 587-607。Zheng, H., Goh, D. H., Lee, E. W. J., Lee, C. S., & Theng, Y.（2022）。理解信息提示对 Twitter 上 COVID-19 信息分享的影响。《美国信息科学与技术学会会刊》，*73*(6), 847-862。Zitt, M., & Small, H.（2008）。通过分数引文加权修正期刊影响因子：受众因子。《美国信息科学与技术学会期刊》，*59*(11), 1856-1860。

附录 A：定理 1 的证明

在本附录中，我们给出定理1的证明。

定理 1。 线性变换与等距变换互为充分必要条件。

证明 我们将证明分为两个步骤。首先，需要证明线性变换必须是等距变换，即线性变换是等距变换。
⇒

设 表示任意一种线性变换方法，其中 x 是
进行线性变换前的值，其定义域为实数域（在本文语境中，指某篇论文的原始引文数），y 是进行线性变换后的值（在本文语境中，指该论文的归一化引文计数），k 是非零常数，b 是常数。

接下来，在自变量x的实数域中任取四个值x_i、x_j、x_m 和x_n，使得

$$x_j - x_i = x_n - x_m \ (x_j > x_i, x_n > x_m). \tag{7}$$

也就是说，我们假设自变量的取值等距（在本文语境中，这意味着在进行线性归一化变换之前，原始引文数中的每个引文计数都是等距的）。

接下来，我们的任务是计算y_j 与y_i之差，得到

$$\begin{aligned} y_j - y_i &= f(x_j) - f(x_i) \\ &= kx_j + b - (kx_i + b) \\ &= k(x_j - x_i). \end{aligned} \tag{8}$$

将方程（7）代入方程（8），得到

$$\begin{aligned} y_j - y_i &= k(x_j - x_i) \\ &= k(x_n - x_m) \\ &= kx_n - kx_m \\ &= (kx_n + b) - (kx_m + b) \\ &= y_n - y_m \end{aligned} \quad . \tag{9}$$

Th该方程 意味着变量的取值仍然 1

在线性变换后等距。因此，我们已经证明线性变换必定是等距变换（类似地，在.....的情境下）

本文中，令上述 x_i、x_j、x_m 和x_n 为原始引文数，满足 x_i、x_j 是相邻项，并且 x_m、x_n 是相邻项，即 ；那么，

=1 ( ) 字母数字母数X X X X X X X X> >

su将该方程代入上述方程（8），可得 k ，

这意味着，在执行线性归一化变换后，归一化引文计数的每个测量单位仍然是等距的，且距离等于 k）。

其次，证明的下一步是证明等距变换必定是线性变换，即等距变换 是 线性变换。 ⇒

设 表示任意一种等距变换方法，其中 x 是

执行等距变换前的值，其定义域为实数域（在本文语境中，指一篇论文的原始引文数）；y 是执行等距变换后的值（在本文语境中，指该论文的归一化引文计数），而f是等距映射。

接下来，如图5所示，在自变量x的实数域内任取三个相邻值x_i、x_j 和x_m，使得

$$x_m - x_j = x_j - x_i \quad (x_m > x_j > x_i). \tag{10}$$

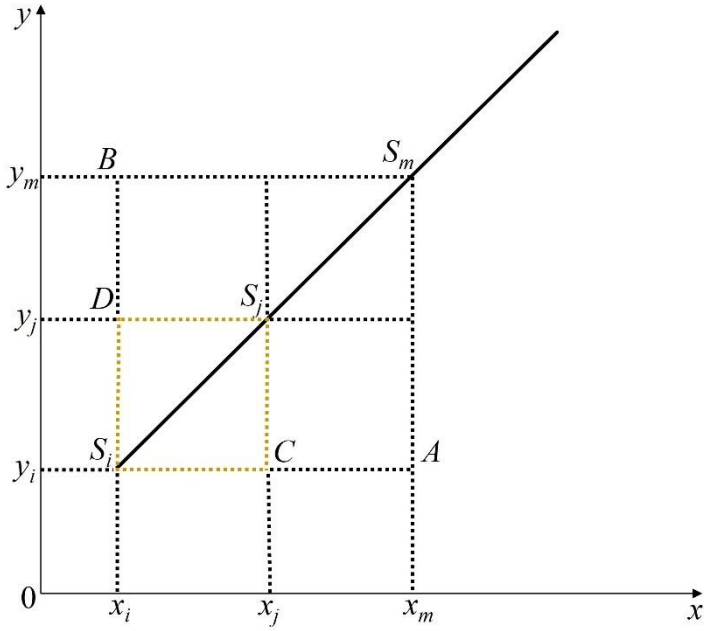

图5 用于证明“等距变换必为线性变换”的第一个示意图

也就是说，我们首先假设自变量的相邻取值之间的距离相等（在本文语境中，这意味着在进行归一化变换之前，原始引文数中每个引文计数之间的距离相等）。由于$f(x)$是等距变换，因此

$$y_m - y_j = y_j - y_i \tag{11}$$

（在本文语境中，这意味着进行归一化变换后，归一化引文计数的每个测量单位之间仍然等距；S_j是矩形S_iAS_mB的中心），并且它必须是端点S_i和S_j之间的线段，也必须是端点S_j和S_m之间的线段（见图5）。

如果端点S_i和S_j之间的距离必须为线段这一断言不成立，并且如果此时我们还希望函数$f(x)$仍然是等距变换，那么函数$f(x)$的像必须经过

矩形中心S_iCS_jD，这就导致 和 x x x x = j − i

$y_j - y_1 = y_1 - y_i$。端点之间的距离S_i和S_1以及两者之间的距离

端点S_1和S_j可以是任意形状的曲线（见图6）。

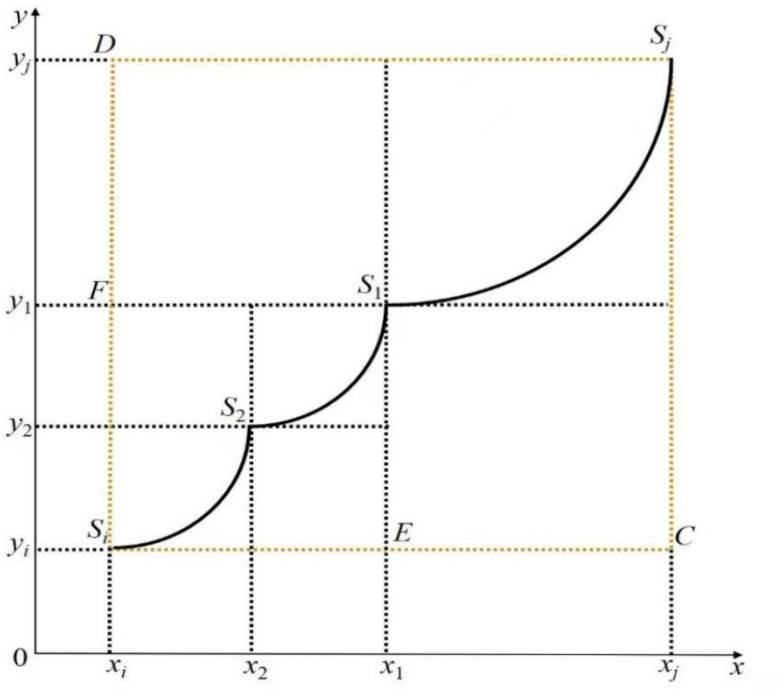

Figure 6 用于证明“等距变换必定是线性变换”的第二幅示意图

同样，如果我们想确保端点S_i和S_1之间的距离不是一条线段，并确保函数$f(x)$仍然是等距的

此时的变换中，函数$f(x)$的像必须经过矩形中心S_iES_1F，因此可得且。

端点S_i 与S_2 之间的距离，以及端点S_2 与 S_1之间的距离，可以是任意形状的曲线（见图6）。

按照上述相同的步骤，必定存在一个值 x_{n+1} ，使得 ，从而得到 ，即函数 f(x) 的像不经过矩形 S_iGS_nH 的矩形中心 S'_{n+1}（见图7）。这与函数 f(x) 是等距变换这一事实相矛盾。因此，端点 S_i 与 S_n 之间的距离必定是一条线段。同理，可以证明端点 S_n 与 S_{n-1} 之间的距离必定是一条线段。由于 S_n 是矩形 $S_iIS_{n-1}J$ 的矩形中心，因此端点 S_i、S_n 和 S_{n-1}都在线段上。

基于相同的逻辑，我们可以通过逐步逆向推导证明，端点 S_i 与S_m 之间的距离必定是一条线段，且端点S_i、S_j 和S_m 都在线段上。

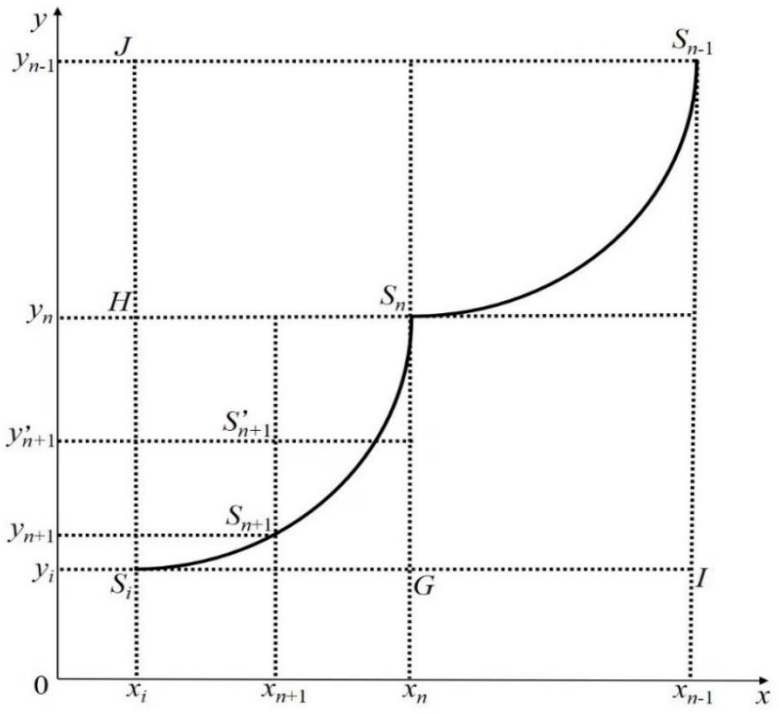

Figure 7 用于帮助证明“等距变换必须是线性变换”的第三幅示意图

由于自变量x的实数域上存在许多相邻点，例如x_i、x_j 和x_m，由上述证明过程类推可得

存在许多线段，例如 S_iS_j、S_jS_m，以及 S_iS_m；也就是说，线段S_iS_m 可以无限延长，因此它是一条直线。由此我们证明了等距变换$f(x)$必定是线性变换。

定理1的证明至此完成。

附录B：定理2的证明

在本附录中，我们给出定理2的证明。

定理2。非线性变换与非等距变换互为充分必要条件。

证明 由于我们已在附录A中证明线性变换必定是等距变换（即线性变换蕴含等距变换），上述结论的逆否命题也成立，也就是说，非等距变换必定是非线性变换（即非等距变换蕴含非线性变换）。在本文的语境下，非等距变换必定是非线性领域归一化变换。

$\Rightarrow$

类似地，由于我们也已在附录A中证明等距变换必定是线性变换（即等距变换蕴含线性变换），上述结论的逆否命题也成立，也就是说，非线性变换必定是非等距变换（即非线性变换蕴含非等距变换）。在本文的语境下，非线性领域归一化变换必定是非等距变换。

$\Rightarrow$

因此，非线性变换和非等距变换互为充分必要条件。定理2的证明完毕。

支持信息

王星[a],1

[a]山西财经大学信息学院，中国太原 030006

关于线性与非线性领域归一化方法使用困境的讨论

根据本文第二节的分析以及附录中定理1和定理2的证明，我们知道，在聚合层级的科研评价活动中，应选择并使用线性领域归一化方法。然而，关于线性与非线性领域归一化方法的选择和使用，仍存在一个有待讨论的困境，讨论应从归一化效果的检验方法以及领域归一化的理想目标入手。

判断一种领域归一化方法性能的关键标准，是该方法的归一化效果，即该方法能否有效抑制原始引文计数的领域偏差，从而使归一化引文计数不受领域影响。“有效抑制原始引文计数的领域偏差，从而使归一化引文计数不受领域影响，并据此公平比较不同领域论文的研究影响力”，正是领域归一化的理想目标。Bornmann（2018）总结了以下三种用于检验领域归一化方法归一化效果的方法，分别是公平性检验方法（Radicchi 等，2008）、组内相关系数（ICC）检验方法（Leydesdorff 和 Bornmann，2011），以及引文实践差异导致的不平等（IDCP）项检验方法（Crespo 等，2013；Li 和 Ruiz-Castillo，2013）。在这些检验方法中，公平性检验方法最常用，也最容易理解，因此本文仅关注这一方法。

公平性检验方法分为以下两种具体方法。第一种方法称为前 z% 方法，它从数值量化的角度检验 归一化方法的归一化效果。该方法首先根据论文的归一化引文计数，对所有领域的论文进行降序排列，然后选取所有论文中排名前 z% 的论文。接着，我们计算各领域中位于前 z% 的论文所占比例。如果归一化方法公平且有效，那么所有领域的这一比例都应为 z%，因为上述比例不应取决于论文所属的具体领域，而只应与该领域的论文数量有关。各领域的比例越接近 z%，归一化方法就越有效。第二种方法称为概率分布图方法，它从图形角度检验归一化方法的归一化效果。

根据这种检验方法，如果归一化方法公平且有效，那么它应能使不同领域的所有归一化引文分布重合为同一分布。不同领域的归一化引文分布曲线越接近，归一化方法就越有效。

显然，从实现领域归一化理想目标的角度来看（即有效抑制原始引文计数中的领域偏差，从而使归一化引文计数独立于领域，并能够公平比较不同领域论文的研究影响力），百分位排名法是一种理想的归一化方法，因为其定义本身就符合公平性检验方法——前 z% 方法的理念。也就是说，百分位排名法是根据最终的归一化效果，即归一化的最终理想目标而设计的归一化方法。同样，逆向工程法也是一种理想的归一化方法，因为其定义本身也符合公平性检验方法——概率分布图方法的理念。

这种情况将带来关于使用领域归一化方法的两难困境。一方面，百分位排名法和逆向工程法都具有理想的归一化效果，因为它们本身是根据归一化的最终理想目标设计的；尤其是百分位排名法，还具有简单易懂的优点。然而，另一方面，使用作为非线性归一化方法的百分位排名法或逆向工程法，在论文层面获得的归一化引文计数无法相加或取平均。因此，在实际科研评价活动中最常用的汇总层面，百分位排名法和逆向工程法自然无法用于科研评价。

解决这一困境的一种可能折中方案，是采用 Zhang 等（2015）提出的基于优化的线性方法。该方法受到逆向工程法思想的启发，同样是从领域归一化的最终理想目标出发设计的，因此也具有良好的归一化效果。此外，该方法属于线性领域归一化方法，因此可以自然地用于聚合层级的科研评价。

然而，基于优化的线性方法最大的缺点是其算法过程过于复杂，难以应用于实际的科研评价活动中。此外，由于通过线性变换得到的归一化引文分布曲线，其形状变化不如通过非线性变换得到的曲线灵活，因此，基于优化的线性方法得到的归一化引文分布曲线无法与参考分布曲线完全重合（Zhang 等人，2015），所以基于优化的线性方法并未显示出相较于基于均值的方法的明显优势，这也是这种线性归一化方法的另一个主要缺点。尽管如此，基于优化的

线性方法是基于领域归一化的最终理想目标设计的，相较于基于均值的方法所具有的微弱优势恰恰表明，基于均值的方法同样具有良好的归一化效果。

一些学者认为，由于引文分布通常呈偏态，中位数比算术平均数更适合作为集中趋势的度量；基于中位数的方法比基于均值的方法更合理，因此在领域归一化中应使用基于中位数的方法，而非基于均值的方法。然而，评价一种领域归一化方法的表现，应依据该方法的归一化效果，或考察它是否实现了领域归一化的最终理想目标——有效抑制原始引文数的领域偏差，从而使归一化引文计数不受领域影响，而不应依据它是否采用了引文分布的集中趋势。采用更好的集中趋势参数是否能带来更好的归一化效果，并无理论依据。事实上，引文数据的中位数通常小于算术平均值。如果采用基于中位数的方法，引文分布头部的高引文计数就会除以一个比算术平均值相对更小的数值，因此无法有效抑制这些高原始引文数所带来的领域优势。

总之，综合考虑以下因素，包括归一化效果、算法复杂度以及领域归一化方法的应用场景，例如归一化方法是否能够应用于汇总层面，作为一种线性归一化方法，基于均值的方法在当前阶段的实际科研评价活动中可能已经是一种非常出色的归一化方法，并且可能在领域归一化方法使用困境方面实现平衡。

参考文献

Bornmann, L.（2018）。批判理性主义与文献计量学中标准（领域归一化）指标的探索。《信息计量学杂志》，*12*（3），598-604。Radicchi, F.、Fortunato, S. 和 Castellano, C.（2008）。引文分布的普适性：迈向科学影响力的客观测量。《美国国家科学院院刊》，*105*（45），17268-17272。Crespo, J. A.、Li, Y. 和 Ruiz-Castillo, J.（2013）。科学领域间引文实践差异对引文不平等影响的测量。《*PLoS ONE*》，*8*（3），e58727。Leydesdorff, L. 和 Bornmann, L.（2011）。引文分数计数如何影响影响因子：根据科学领域间引文潜力的差异进行归一化。美国信息科学与技术学会期刊，*62*（2），217-229。Li, Y. R. 和 Ruiz-Castillo, J.（2013）。基于不同分类体系的归一化程序比较。《信息计量学杂志》，*7*（4），945-958。Zhang, Z.、Cheng, Y. 和 Liu, N. C.（2015）。 从优化视角改进基于均值的方法的归一化效果：基于优化的线性方法及其性能。

《科学计量学》，*102*（1），587-607。