

# 司法鉴定意见的概率表述体系 ——法庭统计学理论与实践展望

徐嘉烨

(南开大学 经济学院 数量经济研究所,天津 300071)

**摘要:** 司法鉴定意见的表述需要同时具备科学性、有效性和易于理解的特征,这也是司法鉴定标准化建设的客观需求。由于具备前述特征,因此以似然比为基本形式的概率表述体系逐渐成为量化证据价值的方法和鉴定意见的主要表述方式。从理论和实践两个方面出发,辨析了法庭统计学中有关概率表述的关键概念,阐述了概率表述方式在法庭科学实践中的科学有效性和国际上的标准化成果,并详细介绍了司法鉴定意见的概率表述体系,最后总结了概率表述体系在实践中的主要应用难点和技术难点,并提出值得探索的研究方向。

**关键词:** 司法鉴定;法庭统计学;证据评价;概率表述;似然比;贝叶斯分析

中图分类号: DF794;O211 文献标志码: B doi: 10.3969/j.issn.1671-2072.2023.06.015

文章编号: 1671-2072-(2023)6-0096-05

司法鉴定意见是证据评价全流程的最终步骤,其以科学严谨的语言描述证据的价值和证明力度。因此,出于科学性、有效性、方便理解且易于标准化的考虑,以及与目前法庭科学中普遍采用的统计分析方法相配合的需求,以似然比(likelihood ratio, LR)为主要形式的概率表述体系,逐渐成为证据价值量化陈述的必要方式,在实践中已经被许多国家和地区认可为司法鉴定意见的标准表述形式<sup>[1]</sup>。同时,在我国司法鉴定公信力建设方面,司法鉴定的合规管理、标准化建设,以及依靠科技手段健全司法鉴定管理体制,既是我国近年来司法实践中的迫切需求,也是理论界的研究热点专题<sup>[2-3]</sup>。因此,从鉴定标准和鉴定制度的科学性出发,以 LR 为主要形式的概率表述体系,是值得纳入我国司法鉴定技术标准的。

## 1 法庭统计学中的概率表述

从发展历程来看,法庭科学中的物证鉴定方法离不开多学科的各自发展与交叉应用,其中,证据价值评估、不确定性量化及最终的鉴定意见表述,依

赖于法庭统计学<sup>①</sup>的理论支持。自 1977 年 LINDLEY<sup>[4]</sup>在 *Biometrika* 杂志上发表第一篇关于法庭科学统计方法的论文以来,法庭统计学作为统计学在法庭科学/司法鉴定领域的重要应用分支,一直以刑事案件物证的量化分析、价值评估和结论表述为核心研究对象,为提升鉴定有效性和可靠性持续提供并逐渐完善统计模型和推理方法,支持法庭科学的制度与标准建设<sup>[5]</sup>。

### 1.1 贝叶斯因子与 LR

与统计学在其他应用领域表述“可能性”的方式不同,法庭科学中的概率以“赔率”<sup>②</sup>的形式出现,于是,当把贝叶斯定理以“赔率”形式表示以后,便得到了证据价值的关键度量——贝叶斯因子(Bayes factor, BF),在特定条件下也被称作 LR。尽管在实践中 BF 和 LR 常常被混用,从业者也习惯将这两种度量统称为 LR,然而在理论上却有区分这两种度量的必要,因为两者在定义、计算方法和性质上都有所不同。首先,因为假设的概率模型中含有未知参数,所以在计算 BF(或 LR)之前需要先进行参数估

收稿日期:2022-11-28

作者简介:徐嘉烨(1987—),女,讲师,博士,主要从事统计学与应用概率论研究。E-mail: jiayexu@nankai.edu.cn

①法庭统计学(forensic statistics),也译作法证统计学、刑事技术统计学。

②赔率(odds),也译作胜算、几率,由于概率论起源于博彩问题,统计学沿用了“赔率”这一术语。

计,当采用频度方法估计参数时,得出的度量被称为LR,当采用贝叶斯方法先推断整个后验分布进而再推断参数,之后得出的度量则是BF。其次,OMMEN等<sup>[9]</sup>严格证明了应用于“同一认定”问题的BF和LR从理论上并不等价,并且LR是BF的极限形式,且这一极限性质在“共同(未知)来源”和“特定来源”这两种情景假设下均成立。以上差异导致了由两者量化得出的证据价值不同,相应的,鉴定意见表述在内容(数值)上也会存在差异,这是本文辨析BF和LR概念及理论差异的原因。在后文中,出于对业界称呼习惯的尊重,除非确有区分讨论的必要,否则BF和LR将统一表述为LR。

## 1.2 LR表述的科学性与有效性

针对基于LR的证据价值度量,GOOD<sup>[7]</sup>于1991年给出了数学证明,CRUPI等<sup>[8]</sup>于2013年完成了LR度量的公理化。至此,这一度量和表述形式的科学性便有了理论保障。经过多年实证研究和实践探索,LR已成为国际上广泛认可的科学有效的证据评价方法和鉴定表述形式。例如:1996年,美国国家科学研究委员会(National Research Council, NRC)在《法庭DNA证据评估》报告中,提出使用LR进行DNA鉴定意见的表述<sup>[9]</sup>,之后又在2004版和2009版NRC报告中,提出了在其他证据类别上的LR法的应用<sup>[10-11]</sup>;欧洲法庭科学研究所联盟(European Network of Forensic Science Institutes, ENFSI)在《ENFSI法庭科学评价报告指南(2015版)》中也规定了LR方法的实施细则<sup>[1]</sup>;英国法庭科学管理局(UK Forensic Science Regulator)在其2017年版的技术准则中推荐使用LR表述法<sup>[12]</sup>。

## 2 概率表述体系的主要内容

本节将详细介绍司法鉴定意见概率表述体系的主要组成部分,以单项证据的鉴定意见为例,逐一介绍以“赔率”表示的贝叶斯定理、LR、LR数值型结论与描述型结论的对应关系,以及基于相似度分数的LR。

### 2.1 单项证据的统计假设与LR

首先将待评价的证据记作 $E, E=(E_x, E_y)$ ,其中 $E_x$ 表示从犯罪现场采集的证据, $E_y$ 表示采集自嫌疑人的证据。在这里, $E$ 可以代表多种类别的证据,比如DNA图谱,玻璃碎片中的化学成分,潜在指纹、

足迹、笔迹等以形态特征为主的证据。接下来,是在假设(或称“主张”)的语境下进行证据评价。需要说明的是,此处仅讨论出于“同一认定”目的的假设,即“来源级别的主张”而非“犯罪级别的主张”所对应的统计假设。先将控方主张和辩方主张分别记作假设 $H_p$ 和假设 $H_d$ ( $H_p$ 和 $H_d$ 满足互斥性和完备性),即 $H_p: E_x, E_y$ 具有相同来源; $H_d: E_x, E_y$ 具有不同来源。之后将贝叶斯定理写成“赔率”形式并化简,可得:

$$\frac{P(H_p|E, I)}{P(H_d|E, I)} = \frac{P(E|H_p, I)}{P(E|H_d, I)} \frac{P(H_p|I)}{P(H_d|I)} \quad (1)$$

其中, $I$ 表示可控的附加背景信息,例如物证采集所用的工具和材料。

上式中的 $\frac{P(E|H_p, I)}{P(E|H_d, I)}$ ,也常简写作 $\frac{P(E|H_p)}{P(E|H_d)}$ ,记

作LR,即为LR的定义式<sup>[4]</sup>。 $\frac{P(H_p|E, I)}{P(H_d|E, I)}$ 和 $\frac{P(H_p|I)}{P(H_d|I)}$

则分别被称作后验比率和先验比率。由定义式可知,LR是在两种假设分别成立的条件下获得证据 $E$ 的可能性之比,更具体地解读可知,LR沟通了先验比率和后验比率,量化了由于新证据的出现,(用以衡量对 $H_p$ 和 $H_d$ 的相对支持度的)先验比率向其相应的后验比率的移动程度。更重要的是,由于LR表示的是由新证据带来的对主张 $H_p$ 的相对支持力度的“增量”,因此在度量证据证明力和客观程度方面具有极大优势。正因其优势,LR很早便被应用在DNA证据的评价上,并且从目前的发展程度来看,DNA证据评价的LR计算方法发展得最为成熟,与其他类型的证据相比,其算法标准化程度也最高,比如美国的DNA综合指标系统(combined DNA index system, CODIS),有关DNA图谱LR的具体计算细节,可以参考文献[13]。

### 2.2 数值型结论与描述型结论之间的对应

计算出LR数值之后,在给出司法鉴定意见时,通常还需要同时给出该数值型结论所对应的描述型结论,又称作证据的“支持程度”,此处的支持程度是指相比于 $H_d$ ,证据对 $H_p$ 的(相对)支持程度。构建数值型结论和描述型结论(即“支持程度”)的意义在于,方便那些对概率统计知识了解不充分的公众理解LR的含义,即数值对主张(假设)的相对支持力度。为方便读者对描述型结论量表形成直观

印象,本文引用了《ENFSI 法庭科学评价报告指南(2015 版)》中的一个示例<sup>[1]</sup>;表 1 给出了可能存在的不同 LR 数值区间所对应的 7 种支持程度。必须说明的是,这种对应关系需要在大量的实证研究并充分结合专家意见之后才能确定下来,事实上,描述型结论量表的制定并不像数值型 LR 结论那样在业界和学界形成共识和标准,对于区间划分、程度等级数量、描述用语等问题的研究和讨论仍在继续<sup>[14-15]</sup>。

表 1 ENFSI 准则中的示例;LR 数值区间与(相对)支持程度的对应关系

支持程度	LR 数值区间
不支持	$1 < LR \leq 2$
弱	$2 < LR \leq 10$
一般	$10 < LR \leq 100$
较强	$100 < LR \leq 10^4$
强	$10^3 < LR \leq 10^4$
很强	$10^4 < LR \leq 10^6$
极强	$LR > 10^6$

### 2.3 基于相似度分数的 LR

当 LR 遇到概率模型中参数的高维推断问题时,通常采用“基于相似度分数的似然比”(score-based likelihood ratios, SLRs)来趋近 LR、近似量化证据价值,SLR 的定义式为:

$$SLR = \frac{f(\Delta(E_x, E_y) | H_p, I)}{f(\Delta(E_x, E_y) | H_d, I)} \quad (2)$$

其中,  $\Delta(E_x, E_y)$  表示不同来源证据异同程度的相似度分数。这种以 SLR 为表述形式的近似 LR 法,避免了对于模型的概率分布假设,转而将重点放在了量化相似度上,其来源于机器学习中应对大数据问题的思路,也是高维推断方法在解决证据价值近似量化问题上的成功应用。在目前的法庭科学研究和实践中,SLRs 方法及其相应的 SLR 表述形式主要应用于以形态特征为主的证据评价。例如,JOHNSON 等<sup>[16]</sup>于 2021 年发布的笔迹鉴定算法“基于随机森林相似度分数的 LR 法”(RF-SLRs),其鉴定意见就表述为 SLR 的形式。需要补充说明的是,在 SLR 的表达式中,为了与前述基于特征的普通

LR 法相区分,采用  $f(\cdot)$  表示以非参数方法估计出的概率密度函数。

### 3 前景与挑战

近十年来,我国学者在司法鉴定意见的概率化表述问题上做了丰富的探索和论述:从特定某一类证据的鉴定意见表述到整个表述体系的构建设想,都为我国司法鉴定意见概率表述的科学化、标准化和制度化做了充分的理论准备<sup>[17-19]</sup>。该领域未来的研究还需多方面的努力,恰如法庭科学作为交叉学科,其发展离不开各领域的理论和技术迭代。仅就法庭统计学而言,在以 LR 形式为基础的概率表述体系框架确定下来之后,未来的应用难点和技术难点主要集中在以下方面。

#### 3.1 在多重证据分析上的应用难点

前文中以单项证据为例,阐述了司法鉴定意见的概率表述体系,然而现实中的复杂应用情景,使得鉴定意见的使用者在面对多重证据<sup>③</sup>的案件时,需要分析判断各个证据之间以及各个证据的鉴定意见之间的相关关系,以便确定综合 LR。为了简明且不失一般性,此处以二重证据为例,阐述多重证据分析应用中综合 LR 计算可能遇到的复杂情景:

当新证据  $E_2$  出现,在对现有的多重证据进行整体评价时,需要考虑所有证据的联合分布,此时的综合 LR 公式为:

$$LR = \frac{P(E_1, E_2 | H_p, I)}{P(E_1, E_2 | H_d, I)} \quad (3)$$

当各证据之间具有条件独立性时,联合分布概率可简化为边际概率之积,由此,多重证据的 LR 计算问题就分解为各个单项证据的 LR 分别计算的问题,此时化简后的综合 LR 公式为:

$$LR = \frac{P(E_1 | H_p, I) P(E_2 | H_p, I)}{P(E_1 | H_d, I) P(E_2 | H_d, I)} \quad (4)$$

上式可以推广到一般多重证据(三件及以上)的实践中:对于某个具体案件,当各证据之间具有科学合理的独立假设时,当前案件所涉及的全部证据的综合 LR 等于所有证据各自被单独鉴定时得出的边际 LR 之积。然而,当证据之间没有条件独立性时,比如某两件证据是协同证据的情况,联合分布

③多重证据包括同类别多项证据和多类别证据。



问题会变得十分复杂,此时需要先分析两件证据之间的关系,进而确定条件概率,这种情形的综合 LR 公式为:

$$LR = \frac{P(E_1|H_p, I)P(E_2|E_1, H_p, I)}{P(E_1|H_d, I)P(E_2|E_1, H_d, I)} \quad (5)$$

此种应用场景的难点在于:一是复杂的条件概率模型难以确定,并且模型假设具有很大的主观性;二是当前案件所涉及证据的各自单独司法鉴定意见中的 LR 不能简单相乘,因为  $P(E_i|H_p, I)$  中没有包含  $E_j$  的条件性信息<sup>④</sup>。由此可见,在具体案件中,未来可预见的以概率表述的司法鉴定意见的应用难点来源于证据之间的复杂关系,需要专门的证据分析方法来应对这一情形。当然,法庭统计学中也有用来辅助分析证据关系的量化方法,比如贝叶斯网络,其作为一种基于有向无环图(directed acyclic graph, DAG)的概率模型,通过画图来厘清并展示各个证据之间的关系,利用马尔科夫假设将模型的联合分布概率进行因子化解,再根据专家意见假定出必要的证据之间的条件概率,并计算出供参考的综合 LR。由于本文专注于概率表述体系而非证据分析的数量方法,故不在此处介绍模型细节,感兴趣的读者可以参考 TARONI 等<sup>[20]</sup>有关应用于法庭科学的贝叶斯网络的研究。

### 3.2 法庭统计学方法上的技术难点

(1)开发应用于以形态特征为主的证据鉴定的统计推断方法。与 DNA 证据中成熟的鉴定方法不同,以形态特征为主的证据(如笔迹、足迹等)的同一认定目前仍然主要依靠专家意见,具有很强的主观性。因此,以 SLRs 为基础结合现代统计学方法,开发专门的算法及(开源)程序,与建立专门的数据库用以训练模型估计总体参数,应处于同等优先级。

(2)发展用于多重证据分析的统计方法。上述介绍过的贝叶斯网络正是目前受法庭统计学界关注度较高的一种分析多重证据方法,但是在应用中还有很多技术细节有待完善,距离其成为多重证据分析的标准方法,尚需进行大量情境模拟实验和模拟数据分析,以及利用案例数据库进行实际数据建模分析。另外,本文并不推崇以神经网络模型为代

表的深度学习方法,因为法庭科学的具体问题更侧重于统计模型的可解释性,而这一点恰恰是深度学习这类“黑盒子”方法的短板。

(3)数值型结论与专家意见类型的科学对应。上述提到在整个概率表述体系中,除了 LR 的数值型结论之外,还应有与之对应的描述型结论。为了使统计方法真正成为与专家意见相辅相成的科学工具,未来需要通过实证研究来确定在不同证据评价中的数值结论与专家意见之间的一一对应关系。

#### 参考文献:

- [1] ENFSI. Guideline for evaluative reporting in forensic science[S/OL]. [2022-10-09]. <http://enfsi.eu/documents/forensic-guidelines/>.
- [2] 霍宪丹.让科技为正义说话——从建立到健全统一司法鉴定管理体制的改革探索与若干思考[J].中国司法鉴定, 2020(5): 17-24.
- [3] 王旭,陈军.合规管理与标准化:科学证据时代的司法鉴定公信力建设[J].中国司法鉴定, 2021(6): 1-9.
- [4] LINDLEY D V. A problem in forensic science[J]. Biometrika, 1977, 64(2): 207-213.
- [5] BANKS D, KAFADAR K, KAYE D, et al. Handbook of forensic statistics[M]. New York, USA: Chapman and Hall/CRC, 2021.
- [6] OMMEN D M, CHRISTOPHER P S. A problem in forensic science highlighting the differences between the Bayes factor and likelihood ratio[J]. Statistical Science, 2021, 36(3): 344-359.
- [7] GOOD I J. Weight of evidence and the Bayesian likelihood ratio[C]// AITKEN S. The use of statistics in forensic science. Chichester: Ellis Horwood, 1991: 85-106.
- [8] CRUPI V, CHATER N, TENTORI K. New axioms for probability and likelihood ratio measures[J]. The British Journal for the Philosophy of Science, 2013, 64(1):189-204.
- [9] National Research Council. NRC II-The evaluation of forensic DNA evidence[M]. Washington, D.C., USA: National Academy Press, 1996.
- [10] National Research Council Committee Report. Forensic analysis: weighing bullet lead evidence[M]. Washington, D.C., USA: National Academies Press, 2004.
- [11] National Research Council. Strengthening forensic science in the United States: A path forward[M]. Washington, D.C., USA: National Academy Press, 2009.

<sup>④</sup>本文使用了适用于多重证据的数学记号:假设  $n$  是当前案件涉及的证据数量,  $i, j \in \{1, 2, \dots, n\}$ , 对于任一  $i \neq j$  的非独立证据,两者各自单独的鉴定意见中的 LR 不能简单相乘。

- [12] FSR. Forensic Science Regulator: technical guidance[S/OL]. [2022-10-09].<https://www.gov.uk/government/collections/forensic-science-regulator-technical-guidance>.
- [13] BUTLER J M. Advanced topics in forensic DNA typing: interpretation[M]. San Diego, USA: Academic Press, 2015.
- [14] NORDGAARD A, ANSELL R, DROTZ W, *et al*. Scale of conclusions for the value of evidence[J]. *Law, Probability and Risk*, 2012, 11(1): 1-24.
- [15] NORDGAARD A, RASMUSSEN B. The likelihood ratio as value of evidence—more than a question of numbers[J]. *Law, Probability and Risk*, 2012, 11(4): 303-315.
- [16] JOHNSON M Q, OMMEN D M. Handwriting identification using random forests and score-based likelihood ratios[J]. *Statistical Analysis and Data Mining*, 2022, 15(3): 357-375.
- [17] 郭华.司法鉴定结果的表述体系与建构——基于贝叶斯定理在鉴定中的展开[J].*中国司法鉴定*,2020(3): 1-7.
- [18] 胡卫平.指纹鉴定标准及鉴定意见概率化研究[J].*证据科学*,2012, 20(4): 480-488.
- [19] 张翠玲.论法庭证据评估及鉴定意见表述[J].*中国人民公安大学学报(自然科学版)*,2017(1): 41-45.
- [20] TARONI F, BIEDERMANN A, BOZZA S, *et al*. Bayesian networks for probabilistic inference and decision analysis in forensic science[M]. 2nd ed. West Sussex, UK: John Wiley & Sons, 2014.

(本文编辑:卢启萌)