

刑事证明标准的高概率模型

杜文静

摘要: 在刑事司法证明中, 为了实现司法审判的数字化, 有必要对刑事证明标准进行量化描述。法律概率主义通过有罪概率和概率阈值构建了“高概率裁决模型”。根据这一模型, 当有罪概率是准确的高概率时, 裁决应当认定被告有罪。但是, 该模型受到法律传统主义的强烈批判, 认为用高概率来描述刑事证明标准在道德上不可接受且会产生推理悖论, 而我们将通过对刑事证明标准的重构为其进行辩护。刑事证明标准可细分为叙事性、稳定性、完全性和充分性四要素。当审判满足叙事性、稳定性和完全性时, 有罪概率的准确性得以确保。如果此时的有罪概率还满足充分性, 则可以认为审判满足刑事证明标准, 并应认定被告有罪。因此, 一旦有罪概率的准确性得到保障, 刑事证明标准可以化归为高概率。高概率裁决模型具有正当性, 能消解传统主义的批判, 促使审判实现核心功能。

关键词: 刑事证明标准; 有罪概率; 概率阈值; 高概率裁决模型; 错判风险

中图分类号: B81 **文献标识码:** A

1 引言

数字法治建设包含两个方面, 一是数字世界法律化, 二是法律世界数字化。本文关注后者, 聚焦在刑事审判中, 如何实现事实认定的自动化和智能化。事实认定包含举证、质证和认证三个程序过程, 其中举证和质证由控辩双方在法庭上完成, 而认证是在法庭辩论结束之后由事实认定者依据规则、经验和逻辑通过心证来完成, 其目的是评价证据的相关性、可采性和证明力, 并做出事实裁决。我们所说的事实认定智能化是指认证的智能化。要实现事实认定智能化需要满足两个条件, 首先, 综合案情和全案的可采证据, 量化表示所指控的犯罪事实发生的可能性; 其次, 量化表征刑事证明标准。法律概率主义为完成这两个条件提供了框架。法律概率主义包含两个主要观点。第一, 量化主张。通过对案件所有证据的全面考量, 被告犯罪事实成立的可能性可以借助概率来量化表示。第二, 阈值主张。设定一个适当数值, 如 0.9、0.95 或 0.99, 作为有罪事实裁决的标准, 这一数

收稿日期: 2024-02-29

作者信息: 杜文静 华东政法大学文伯书院
2178@ecupl.edu.cn

基金项目: 国家社科基金一般项目“司法证明的逻辑研究”(23BZX127)。

值被视为概率阈值。当犯罪事实的概率超过阈值时,应当做出有罪裁决,反之,则裁决无罪。([13])犯罪事实成立的概率称为有罪概率,超过阈值的概率称为高概率。因此,阈值主张也可以表述为:当准确的有罪概率是高概率时,则认定被告有罪,否则宣告释放。尽管追求量化主张的准确性,即运用概率推理和智能工具来准确计算有罪概率也同样重要,但我们在此主要关注阈值主张的辩护,为实现智能审判提供必要的理论基础。

2 概率阈值理论

概率与刑事审判的联姻是非常自然的过程。概率论的一些先行者对法律诉讼的兴趣不亚于对机会游戏的兴趣,如雅可比·伯努利(J. Bernoulli)在其1713年的著作《猜想的艺术》中讨论了刑事定罪的要求,并提出一个引人深思的问题:“是否0.99的概率就已经足以作为定罪的依据,或者我们需要更高的0.999概率才能确保公正”。([1],第38页)这可能是最早将刑事证明标准等同于高概率的提议。然而,刑事证明标准是否确实可以用高概率来量化描述?或者更具体地说,在公正的司法审判中,是否可以“仅仅”依据有罪概率足够高而裁定被告有罪?对此,法学界内部形成两种截然不同的观点。法律概率主义者认为这是可以的。刑事审判本质上是概率问题,证据不能“确定性”地证明有罪。在他们看来,如果充分高的有罪概率都不足以定罪,那还有什么可以胜任呢?与此相反,法律传统主义者对此表示反对。他们质疑罪责是否可以量化,并强调即便可以量化,仅凭罪责的高概率也不足以定罪,刑事审判不应该把别人的生命当作赌博。这场围绕概率主义和传统主义的学术争议直至今日仍在持续。

2.1 高概率裁决模型

概率主义者认为,概率和统计方法能够提高我们对法律证据和诉讼的理解,帮助事实认定者理清证据并获得更加准确的事实裁决。例如,卡普兰(J. Kaplan)提议将决策理论应用于刑事事实认定过程,不仅获得量化刑事证明标准的计算方法,还为法律体系的运作注入新见解。卡普兰的理论建立在两点之上:首先,经济决策中的许多问题与法律决策在微妙性和复杂性上不相上下,决策理论在经济领域的显著成就表明其在法律领域的潜在适用性。尽管可用的量化数据在法律问题上可能难以获取,但通过对不同要素之间的抽象数量关系进行思考,至少能促进对相关法律概念的理解。其次,司法审判充满不确定性让我们无法“绝对确信”被告有罪,无论是作伪证或错误陈述的目击证人,或是可能因深层心理冲动而认罪的被告,抑或是可能由一系列不太可能的巧合所产生的确凿证据,都为审判过程带来了不确定性。再精巧的审判流程也无法完全排除这种不确定性,错判是司

法审判不可避免的系统性风险。犯罪证据越充分，被告的有罪概率越大，却无法达到完全确定。但事实认定者必须在这种不确定性状态下做出裁决，要么认定被告有罪，要么宣告无罪。决策理论正是在不确定状态下进行推理和决策的理性理论，指引事实认定者做出理性裁决。尽管事实认定者的决策过程经常受到一些甚至他们自己都未知的非理性因素影响，但事实裁决过程应该要以一种理性的方式运行。（[6]）

决策理论应用于刑事事实认定涉及两个核心概念：有罪概率和负效用。在司法审判中，当所有证据被出示、检查和交叉询问后，法官或陪审员在权衡全部可采证据后必须做出事实裁决。这一过程涉及两个重要步骤：首先，事实认定者需要评估被告犯罪事实成立的可能性，即在考量所有定罪证据后，被告的有罪概率，记为 $P(G|E)$ ，其中 G 表示命题“被告有罪”， E 表示全案所有的定罪证据。其次，在刑事事实裁决中，事实认定者的决策包含两种行动：认定被告有罪，或宣告被告无罪。这两种行动会产生两类错误后果，一是错误定罪，将无辜的被告认定为有罪；二是错误释放，对有罪的被告宣告无罪释放。错误定罪和错误释放是司法审判不可避免的两类错判风险。因此，事实认定者还需要考虑做出错误事实裁决所带来的后果或代价，这种错判的后果称为负效用。负效用量化表达了这两类错判后果的严重程度，记 $D(CI)$ 表示错误定罪的负效用， $D(AG)$ 表示错误释放的负效用，其中， CI 、 AG 分别是 *convict innocent* 和 *acquit guilt* 的缩写，且 $D(CI)$ 和 $D(AG)$ 均为正数。在经济领域，效用和负效用通常用交易行为中金钱的收益和损失来测量。但在法律中，效用和负效用难以用这样一个简单的数量指标来表示。庆幸的是，“由于概率和效用都是个人主义概念，不同人计算概率或效用的差异主要是程度上的，而不是本质上的，因此将决策理论应用于法律决策过程至少在理论上并非不可能。”（[6]）

在获得有罪概率和负效用之后，就可以计算一个行动的期望负效用，它等于概率乘以相应的负效用，测量的是如果采取这个行动，未来预期会带来的平均负效用。于是，错误定罪的期望负效用等于： $[1 - P(G|E)] \times D(CI)$ ；错误释放的期望负效用等于： $P(G|E) \times D(AG)$ 。决策理论告诉我们，面对一个刑事案件，如果错误定罪的期望负效用“小于”错误释放的期望负效用，那么就应当认定被告有罪。因此，为了做出有罪事实裁决，被告的有罪概率 $P(G|E)$ 必须满足： $[1 - P(G|E)] \times D(CI) < P(G|E) \times D(AG)$ 。经过简单的数学演算和变形，有罪概率 $P(G|E)$ 满足的不等式可用“负效用比”来表示：

$$P(G|E) > 1/[1 + D(AG)/D(CI)],$$

其中， $D(AG)/D(CI)$ 是负效用比。为了下文引用方便，同时强调这个不等式的重要性，我们将它取名为“裁决不等式”。

一方面，裁决不等式的右边给出了计算概率阈值（记为 t ）的方法：

$$t = 1/[1 + D(AG)/D(CI)].$$

因为负效应都是正数,所以 t 是一个大于0且小于1的数值。 t 本身不是概率,但由于它的取值范围与概率的取值范围相同,且人们通常将有罪概率与它比较大小,因而称为概率阈值。当一个概率数值大于阈值时,则称它为高概率。在具体的案件中,犯罪事实的概率 $P(G|E)$ 是否成为高概率,不取决于它的绝对大小,而取决于它与阈值的相对大小。例如,假设 $P(G|E) = 0.93$,如果这个案件的阈值等于0.9,那么它是高概率;但如果该案件的阈值是0.95,那么它不是高概率。

另一方面,裁决不等式也给出了刑事案件事实裁决的概率模型,不妨称之为“高概率裁决模型”:仅当有罪概率大于概率阈值时,才能做出有罪事实认定;否则宣告无罪释放。具体来说,依据高概率模型进行事实裁决分为三个步骤。第一,在法庭辩论结束后,事实认定者评估控辩双方提交给法庭的所有可采证据,依据证据推理的概率方法准确地计算出被告的有罪概率 $P(G|E)$ 。第二,结合案情,事实认定者权衡案件的利害关系,量化评估负效用比,从而确定概率阈值 t 。第三,比较有罪概率和概率阈值的大小,如果有罪概率大于阈值即 $P(G|E)$ 是高概率,那么认定被告有罪,否则认定被告无罪。

2.2 概率阈值的计算和功能

有罪概率和概率阈值的计算是两个相互独立过程,前者涉及证据评价和证据推理的方法,主要是从认识论层面来确定被告有罪的可能性;后者涉及对利害关系的权衡,主要是从道德、社会和政治等非认识论层面来确定负效用比,从而计算出概率阈值。两者都具有主观性,但不同之处在于,有罪概率量化的是事实认定者对证据在认识论层面上主观判断,而概率阈值量化的是事实认定者对证据在道德和社会等非认识论层面的主观判断。在事实认定的认证阶段,主观性是不可避免的,决策理论不能消除主观性,而是以一种量化的方式显性表达事实认定者的主观判断和假设,这样有利于揭示事实认定者心证的犯错点和怀疑源,有利于他人审查和评价心证。因此,在这个意义上,决策理论可以促进事实认定的准确性。

根据概率阈值的计算公式可知,当负效用比越大时,概率阈值越小;当负效用比越小时,概率阈值越大。从社会道德层面看,在刑事审判中,一个基本的价值共识是:错误定罪带来的负效用要远比错误释放产生的负效用大,即 $D(CI)$ 要比 $D(AG)$ 大很多。因此,负效用比 $D(AG)/D(CI)$ 很小,概率阈值很大。这也说明了为何刑事证明标准要求很高的原因,法官或陪审员必须被说服到“排除合理怀疑”的程度,才能认定被告有罪。当刑事案件有疑点时,法庭宁愿以错误释放一个有罪之人为代价来减少认定一个无罪之人犯罪的错误。“十个罪犯逃脱,总比一个无辜者受苦好”([14]),这是英国法学家布莱克斯通(W. Blackstone)的名言,10:1的比率被称为“布莱克斯通比率”。如果按照布莱克斯通比率来量化负

效应比值,那么错判一个无辜者的负效用要 10 倍于错误释放一个罪犯的负效用,甚至更大。用符号表示,即为 $D(AG)/D(CI) < 1/10$,代入裁决不等式,则可计算出: $P(G|E) > 1/1.1 \approx 0.91$ 。所以,在布莱克斯通比率的假设下,概率阈值约等于 0.91,仅当有罪概率大于 0.91 时,才能认定被告有罪,否则宣告无罪。

布莱克斯通比率只是量化负效用比的一种方法,基于不同的道德和社会等因素的考量,将得到不同的负效用比,从而确定不同的概率阈值。例如,当错判一个无辜者的负效用 9 倍于错误释放一个罪犯的负效用时,则 $D(AG)/D(CI) = 1/9$,概率阈值等于 0.9;当错判一个无辜者的负效用是错误释放一个罪犯负效用的 19 倍时,则 $D(AG)/D(CI) = 1/19$,概率阈值为 0.95;当错判一个无辜者的负效用 99 倍于错误释放一个罪犯的负效用时,则 $D(AG)/D(CI) = 1/99$,概率阈值等于 0.99。“在不同的案件中,我们通常没有必要假定概率阈值是固定不变的。”“概率阈值可能会根据特定案件的利害关系而发生变化。”〔[12]〕在刑事案件中,由于不同案件可能存在不同程度的利害关系,因此存在变化的概率阈值,但这个阈值一般被认为至少是 0.9。

概率阈值的功能是对审判中两类错判风险的权衡和分配。具体来说,如果有罪概率 $P(G|E)$ 大于阈值 0.95,则认定被告有罪,此时错误定罪的可能性为 $1 - P(G|E)$,即错误定罪率小于 0.05。如果 $P(G|E)$ 小于 0.95,那么认定被告无罪,此时错误释放的可能性为 $P(G|E)$,即错误释放率小于 0.95。这说明,对于某个具体的概率阈值 t ,错误定罪率小于 $1 - t$,错误释放率小于 t ,即最大错误定罪率与最大错误释放率之比等于 $(1 - t) : t$ 。改变阈值 t 的大小,就是在对两类错判风险之间进行权衡和分配。阈值变大,则错误定罪率减小,错误释放率增高;阈值变小,则错误定罪率增高,错误释放率减小。¹所以,概率阈值不仅设定了错误定罪率,也设定了错误释放率,它调节的是这两类错判风险之间的分配。在事实认定过程中,绝对的确定性很难获得,犯错不可避免。在总错判风险确定的情况下,较大的概率阈值让错误定罪率减少,但同时会让错误释放率增大,这种将总错判风险分配在两类错判风险之间的方式符合刑事审判证明责任和道德的要求。无罪推定和谁主张谁举证的原则,以及错误定罪带来的坏处远大于错误释放的道德假设,都说明错误定罪率要远小于错误释放率,这也解释了为何在刑事审判中概率阈值通常要大于 0.9。

¹当最大错误定罪率与最大错误释放率之比 $(1 - t) : t = 1 : 10$ 时,可以计算出 $t = 10/11 \approx 0.91$ 。它与根据布莱克斯通比率计算出的概率阈值相同,这说明计算概率阈值存在两种方法,一是通过评估负效用来确定阈值,二是通过设定最大错误定罪率与最大错误释放率之比来确定阈值。两种方案计算出的结果是一致的。

3 高概率的批判

概率阈值理论遭到法律传统主义者的诸多批判,批判主要来自两个方面。第一,犯罪事实发生的可能性很难用概率来准确地量化表示。例如,富兰克林(J. Franklin)认为,在逻辑上人们很难“将刑事案件中不同种类的复杂证据结合起来”形成有罪概率,概率作为一种“部分蕴含”,它的值不能被唯一确定。([5])第二,即使概率可以量化表示犯罪的可能性,但高概率也不足以定罪。例如,纽曼(J. O. Newman)批判道,“像概率、赔率和机会这样的词语,陪审员并不能清晰地明白它们的含义。事实上,我也不清楚。”“最终,当陪审员被告知为了判定有罪需要达到 95% 的概率时,我不知道他们会怎么想。”([9], 第 267–268 页)前者是对法律概率主义中量化主张的批判,后者是对阈值主张的批判。本文讨论后者,即刑事证明标准是否可以转化为高概率?为止,我们不妨承认前者,即假设犯罪事实发生的可能性可以用有罪概率来准确地表达,事实认定者能有效地运用概率理论计算有罪概率。实际上,芬顿(N. Fenton)和尼尔(M. Neil)的著作系统地介绍了如何借助贝叶斯网络模型和智能软件工具辅助事实认定者准确地计算有罪概率。([4], 第 523–552 页)

考虑到道德的因素,却伯(L. H. Tribe)主张审判的目标不应该是仅凭高概率来定罪,而是努力寻求“尽可能接近人类所能达到的确定性”。([11])由此看出,尽管这种确定性不等同于绝对的确定性,但却伯认为事实裁决也不能用高概率作为评判标准。从认识论的角度出发,科恩(J. L. Cohen)提出了一系列所谓的悖论来说明传统概率理论不能恰当地解释法律推理,逃票者悖论便是其中之一。([3], 第 74 页)同样地,汤姆森(J. J. Thomson)提出的蓝色公交车悖论也反映了相似的问题。([10])基于社会和政治的考量,尼森(C. R. Nesson)构建了一个“囚犯”案例,展示了刑事证明标准不能用概率阈值来描述。([8])由于本文旨在探讨刑事证明标准,因此重点分析囚犯案。为了更好地说明问题,我们对囚犯案例做了细微调整。

一段监控录像显示,在一个封闭的庭院内,1000 名身着统一囚服的囚犯与一名穿着制服的狱警同在一处。突然,狱警意外绊倒,此刻囚犯们围拢过来,争吵声四起。在混乱中,一名囚犯挤出人群,寻得院子角落里的一个小棚子,并悄然藏身其间。随后,剩余 999 名囚犯扑向倒在地上的狱警将其杀害。谋杀发生后,原先隐藏起来的囚犯从棚子里走出,重新回到其他囚犯之间。在这段录像中,999 名囚犯都故意参与杀戮,只有一人是无辜的。但这些囚犯的穿着一样,体型也没有显著差异,更不用说面部特征几乎难以分辨,因而仅凭这段录像无法辨认出谁是无辜者。

假定这些囚犯潜逃后,只有一个名叫甲某的囚犯被捕并被起诉,但除了监控

录像中的信息，并没有其他的证据。这样的情况下，似乎可以计算出“甲某参与杀害狱警”或“甲某有罪”的概率高达 0.999。依据这样的高概率裁决模型，应当认定甲某有罪。但在法律实践中，这种基于概率的判决几乎是不被接受的。如尼森所指出，类似囚犯情节的案子应该被驳回，更不应该交给陪审团，因为陪审团不可能给甲某定罪。威尔斯在他的心理学实验中评估了有经验的法官、商学院学生和心理学院学生，他们在面对与囚犯案例相似的情境时的反应。他发现绝大多数参与者在“仅”根据高概率定罪时会感到不安。〔16〕这导致一个棘手的困境。一方面，如果根据高概率裁决模型，应当认定甲某有罪；另一方面，如果根据法律实践中的直觉，应该宣告甲某无罪。显然，法律实践直觉不会出错，造成僵局的原因只能是高概率。因此，在传统主义者看来，囚犯案例对概率主义提出了的严重挑战，高概率裁决模型不能适用于刑事事实认定。

为了进一步澄清导致高概率不足以定罪的原因，传统主义者纷纷转向分析纯统计证据的缺陷。所谓纯统计证据，是指在一个案件中，如果所有的定罪证据只有统计数据或主要依赖统计数据，则称这些定罪证据为“纯”统计证据。〔1〕，第 104 页）在囚犯案例中，唯一的证据就是一项统计数据，即在 1000 名囚犯中，有 999 人故意杀害狱警，只有 1 人是无辜的。这种“999:1”的数据就是典型的纯统计证据。传统主义者认为，纯统计证据在本性上存在某种缺陷，即使依据纯统计证据计算出的有罪概率是高概率如 0.999，它也不足以定罪。首先，纯统计证据缺乏个体性或具体性。定罪证据应当针对性地指向被告或犯罪中的事实，而不应该指向一个碰巧包含被告在内的群体。人们不应当因为别人的行为而被定罪。定罪证据要与犯罪事实具有因果联系。在囚犯情节中，尽管有 999:1 的统计证据，但这一数据既不能针对性地指向甲某，将甲某与其他人区分开来，也无法证明“甲某有罪”的因果联系。其次，在刑事审判中，被告应当被视作唯一的、具有自由权的主体，仅依赖群体行为频率的统计证据来定罪，违背了这一道德要求。再次，若纯统计证据成为定罪的基础，那么将会传递一个社会信息，即被告被定罪不是根据他实际所为，而是因为群体的统计规律。这可能导致公民认识到，作为个体做什么或不做什么并不重要，这就会破坏定罪的威慑作用。最后，由于纯统计证据缺乏个体性，因而无法对其进行交叉询问。〔2〕

纯统计证据在本性上是否真的存在缺陷？这是一个具有争议的问题，迪贝罗就对证据的个体性就提出了异议。〔1〕，第 18 页）但这不是本文的重点，我们要讨论的是“犯罪事实的高概率是否足以定罪？”传统主义者从认识论、道德、社会和法律程序的不同层面分析了纯统计证据的缺陷，我们认同这些分析的合理性。刑事审判是复杂的程序，在权衡和评估证据时需要考虑诸多因素，不能仅考虑高概率就做出裁决。但是，这并不意味着高概率裁决模型存在谬误，更不代表只基于纯统计证据获得的高概率就不足以定罪，正如在某些案件中，仅凭 DNA 证据获

得的高概率也能定罪，如强奸案。传统主义者的这些批判，最多只能表明在类似囚犯情节的案例中高概率不足以定罪。那么，在何种条件下得到的高概率足以定罪呢？为了回答这一问题，并回应传统主义者的批判，下文将论述司法审判的两大核心目标，并阐述高概率裁决模型的逻辑运作，从而为其进行辩护，并为实现智能审判提供理论基础。

4 司法审判的两大核心目标

在事实认定过程中，缘于人类认知能力的局限和现实条件的限制，事实认定者无法获得对事实真相的绝对确信，错判成为司法审判中不可避免的系统性风险。法庭在做出事实裁决时总是承担着一定的错判风险，不论是错误定罪还是错误释放。因此，司法审判的两大核心目标是：一是竭尽所能地减少总错判风险；二是在总错判风险无法进一步降低时，以一种合理的方式在控辩双方之间分配总错判风险。

为了减少总错判风险，法庭需要提高审判的辨识力。([1]，第 58 页) 辨识力反映的是在审判过程中准确地区分无辜被告和有罪被告的能力。在一个辨识力很高的审判体系中，大多数无辜者的有罪概率会在较小值附近波动，如 0.1 附近；而对于大多数犯罪者，有罪概率则应在较大值附近波动，如 0.95 附近。若这里的较小值越小，较大值越大，则说明审判的辨识力就越高。质言之，在一个高辨识力的审判体系中，无辜者的有罪概率会被计算得很小，被认定为无罪；而犯罪者的有罪概率会被计算得很大，被认定为有罪。相反，在一个低辨识力的审判体系中，无辜者的有罪概率会被评估得较高，犯罪者的有罪概率也会被评估得较低，这样直接导致了错判的增加。因此，提高审判的辨识力是减少总错判风险的关键。

如何提高审判的辨识力？这依赖于法庭在处理证据时的规则和措施。如果法庭在审理证据时，“单方面”地偏向任何一方，那将会降低审判的辨识力。不论是仅向控方或辩方提供更多的司法资源，还是仅向控方或辩方赋予更多传唤证人的权利，这些偏颇的证据措施都可能导致有罪证据和无罪证据的不平衡，从而损害司法公平和正义。为了更好的识别被告有罪还是无罪，法庭应该从“双方面”获得尽可能好的证据。首先，法庭应该给予控方和辩方同等的机会和权利来提交证据。其次，在没有任何预知情况下，证据措施应当符合“不可预测的受益人标准”([1]，第 70 页)，即在收集或分析证据之前，法庭无法确定证据的结果将有利于控方，还是有利于辩方，满足这个标准的证据措施或规则就是双方面的。例如，在刑事审判中，引入 DNA 证据就是双方面的，符合不可预测的受益人标准。对于犯罪现场收集的 DNA，其鉴定的结果有两种可能，一是与被告匹配，二是与被告不匹配，前者有利于控方，但后者有利于辩方。在鉴定结果出来之前，不能确

定有利哪一方，但在结果出来之后，却具有很好的区分性。因此，在审判体系中，任何双方面的证据措施都有利于提高审判的辨识力，从而减少总错判风险。叙事性、稳定性和完全性，作为评价证据的三种措施，它们是双方面的，对控辩双方都可能有利，并非单方面的偏向任何一方，既体现了司法的公平和正义，又能提高审判的辨识力。

第一，叙事性。在刑事审判中，叙事性要求控方在法庭论辩中提供一个针对被告的足够详细、完整的犯罪叙事或故事。这个叙事要能回答案前、中、后发生的所有合理问题，例如，犯罪人是谁？他做了什么？为什么要犯罪？何时犯罪？在哪里犯罪？犯罪过程如何？叙事是由一系列事件根据物理因果和意图因果按照时间顺序组织起来的整体。（[17]，第74-81页）叙事由控方随着庭审的进程“逐步”构建。构建叙事是一个不断提出问题和回答问题的动态过程，受到诸多因素影响，例如，被告被控诉的罪名、审判前和审判中获得的证据、一般情况下的常识、以及来自辩方的质疑和反驳。当所有可能自然或合理产生的疑问都得到回答时，最终的叙事才彻底建构完成。一个犯罪叙事越详细，它所断言的事实主张越多，也就意味着它面临辩方审查、质疑和反驳的挑战越多。如果犯罪叙事经受住辩方的严格审查，无疑将对控方有利。反之，如果它被辩方的审查驳倒，控方的主张便会削弱，从而有利于辩方。所以，叙事性要求满足了不可预测的受益人标准，有助于降低少总错判风险。然而，一旦控方未能提供一个详细的犯罪叙事，这就成为引发合理怀疑的首要来源。

第二，稳定性。叙事的稳定性要求法庭为辩方提供有效的机会去审查控方的犯罪叙事，辩方有责任充分利用这个机会去质疑犯罪叙事。如果在辩方的严格审查、交叉询问、反对和挑战下，犯罪叙事成立的可能性没有明显地降低，其结论保持不变，那么就称犯罪叙事具有稳定性。辩方的有效审查是检验犯罪叙事稳定性的途径，它满足不可预测的受益人标准，有助于减少总错判风险。这是因为，在面对辩方的有效审查时，若犯罪叙事具有稳定性，控方的主张将因此得到强化，从而对控方有利；但如果犯罪叙事不具有稳定性，这将有利于辩方。此外，如果在辩方的严格审查下犯罪叙事变得脆弱、不稳定，或由于缺乏资源或程序保障，辩方无法严格审查犯罪叙事，那么这个犯罪叙事就值得怀疑。犯罪叙事缺乏稳定性是产生合理怀疑的第二个来源。

第三，完全性。证据的完全性要求控方提供尽可能多的证据以支持犯罪叙事，是在数量上对定罪证据的要求。在评价控方构建的犯罪叙事时，事实认定者不仅要考虑已经提供的证据，还要考虑应当提供但还未提供的证据。犯罪叙事有一项重要的预测功能，它能促发一个理性人期待控方应该提供什么证据。为了证成其主张，控方必须提供尽可能完全的证据。虽然提供完全的证据是控方的举证责任，但它不是单方面的，而是双方面的。提供完全的证据意味着要寻找更多的证据，这

可能对控方有利,也可能对辩方有利,因为寻找到的证据可能是有罪证据,也可能是无罪证据。因此,完全性也满足不可预测的受益人标准,有助于减少总错判风险。“在某些案件中,陪审员期望听到某些证据,他可能会对无法提供这些证据的案件持严重怀疑态度。”〔7〕因此,定罪证据不完全是产生合理怀疑的第三个来源。

除了上述三个来源,证据缺乏充分性构成合理怀疑的第四个来源。证据的充分性要求有罪概率是高概率。犯罪叙事必须受到证据的充分支持,也就是说,在证据基础上,犯罪叙事成立的可能性要非常高,否则就值得怀疑。证据不充分或有罪概率不是高概率,则成为产生合理怀疑的第四个来源。高概率使定罪更难,它系统性的偏向有利于被告或辩方,不利于控方,这符合司法道德和证据法的要求。但证据充分性是单方面的,不是双方面的,因而不满足不可预测的受益人标准,不能减少总错判风险。

由此可见,在司法审判中,当叙事性、稳定性和完全性的要求被严格遵循时,我们可以认为审判的辨识力已经达到最大化,总错判风险已经降至最低,这时的风险就是审判不可避免的“最小总错判风险”。接下来的任务就是,如何在控辩双方之间合理地分配最小总错判风险。充分性是单方面的,它不具有减少总错判风险的作用,但却具有分配最小总错判风险的功能。在刑事审判中,伦理道德表明错误定罪的危害要远大于错误释放的危害。在分配最小总错判风险时,应该让错误定罪风险更小,错误释放风险较大,比如错误定罪风险与错误释放风险按1:10的比例分配。根据概率阈值的计算原理,我们知道每个概率阈值确定了一种分配最小总错判风险的方案,当概率阈值等于0.9时,错误定罪风险与错误释放风险之比就等于1:9。反之,每种分配最小总错判风险的方案,也确定了一个概率阈值,当错误定罪风险与错误释放风险之比等于1:19时,概率阈值就等于0.95。²因此,充分性较高的概率有助于促进最小总错判风险朝着道德要求的方向在控辩双方之间分配,即让错误定罪风险远小于错误释放风险。

叙事性、稳定性、完全性和充分性,统称为“四要素标准”〔1〕,第68页),它们构成了合理怀疑的来源。违背这四要素中的任何一条标准,都会产生合理怀疑。那么当这四个要素都满足时,审判是否就达到了排除合理怀疑的证明标准呢?这是一个值得进一步讨论的问题,但本文认为答案是肯定的。如果在四要素都满足的情况下,有罪裁决仍不能得以保证,那么有罪裁决的保证就只能是绝对的确定有罪,但绝对确定在现实中不可获得。因此,在四要素标准框架下,当审判满足四个要素时,它就达到了刑事证明标准的要求,事实认定者则应当认定被告有罪。当前三个要素都满足时,审判的辨识力达到了最大化,总错判风险达到了最小状态;充分性则对应着对最小总错判风险的分配方案,让错误定罪风险远小于错误

²这里的错误定罪风险、错误释放风险分别与上文中的错误定罪率、错误释放率含义相同。

释放风险。仅有高概率不能定罪，但在前三个要素都满足之后，高概率就足以定罪，此时刑事事实裁决的依据就转化成了高概率。由此可见，刑事证明标准在减少总错判风险和分配总错判风险上均起到规制作用，全方位约束司法审判，促使审判实现其核心功能。四要素标准能使有罪裁决得到有效保证，刑事证明标准可细化为叙事性、稳定性、完全性和充分性的四要素标准。

实际上，学术界和实务界对到底何为“排除合理怀疑”或“确信无疑”还远未达成共识。当一个法官或律师被问及何为排除合理怀疑时，人们可能听到的答案是：排除任何可能的、似真的、现实的、非人为的或实质性的怀疑。这些回答是促进了还是模糊了人们对排除合理怀疑的直觉理解，存在争议。难怪有法官认为，排除合理怀疑的标准不需要解释，任何对合理怀疑的解释措辞并没有使其含义更加清楚。但这里的四要素标准是对排除合理怀疑的一种细化解构，它不是在为排除合理怀疑给出一个系统的、完全清晰的解释或定义，而是在为达成裁决提供一些实用的和可操作的证据处理措施或手段。

5 高概率裁决的正当性

法律概率主义通过两个核心概念即有罪概率和概率阈值，构建了一个高概率裁决模型，并将刑事证明标准解释为高概率，这一解释与四要素标准是相容、相通的。首先，为了减少总错判风险，必须保障有罪概率计算的准确性，准确的有罪概率是审判辨识力很高的体现。叙事性、稳定性和完全性是获得准确有罪概率的基础和条件，概率理论是准确计算有罪概率的方法。由于人类认知能力的局限性、证据的可错性以及其它诸如时间和成本等条件的限制，审判很难绝对确定地区分有罪和无罪，总是存在某种程度上的不确定。因此，被告犯罪的可能性，可以用概率来衡量。在辨识力很高的审判系统中，对于有罪的被告，其有罪概率应当很高；而对于无辜的被告，其有罪概率应当很低。为了获得准确的有罪概率，就必须提高审判的辨识力。当叙事性、稳定性和完全性都满足时，审判辨识力达到了最大状态，此时再依据概率理论的逻辑方法去计算有罪概率，其准确性就有了保障；反之，有罪概率的准确性值得怀疑。有罪概率的准确性越高，总错判风险就越小；有罪概率不准确，总错判风险就很大。需要强调的是，这里说的是有罪概率之准确性的高低，而不是有罪概率本身在数值上的大小。其次，为了合理地分配总错判风险，必须设定一个适当大的概率阈值。概率阈值与总错判风险的分配方案具有相互确定的对应关系，概率阈值越大，错误定罪风险比错误释放风险就越小。依据高概率做出裁决，有助于促进总错判风险朝着道德要求的方向在控辩双方之间分配。由此可见，有罪概率的准确性对应着叙事性、稳定性和完全性，有罪概率的高要求对应着充分性，刑事证明标准用纯粹的概率语言可以重构为：准

确的高概率。高概率裁决模型具备内在的正当性。

传统主义者之所以反对概率主义，是因为他们对高概率裁决模型存在如下两种误解。第一，忽视准确性要求。高概率裁决模型依据概率阈值理论而得到，但阈值强调的是有罪概率要足够高，没有突出说明有罪概率要足够准确。这就让人容易产生这样的误会：高概率裁决模型只要求有罪概率足够高，从而忽视了准确性要求。实际上，准确性具体体现在概率主义的量化主张中，即犯罪事实成立的可能性可以用有罪概率来准确地量化表示。在阈值理论中，有罪概率可以被准确地计算，是一个潜在的默认假设。因此，高概率裁决模型本质上对有罪概率有两个要求，即准确性和充分性。第二，忽略道德约束。刑事证明标准对审判的规制是全方位的，在认识论层面上，它要求审判发现事实真相；在道德层面上，它要求事实认定者尽全力减少在可能错误定罪时带来的危害。从表面上看来，高概率裁决模型似乎只表达了刑事证明标准在认识论层面上对审判的约束，但实际上它也反映了刑事证明标准的道德要求。这主要体现在两个方面：首先，有罪概率的准确性要求司法审判必须为准确地计算有罪概率提供条件和保障。叙事性、稳定性和完全性是提供这种条件和保障的三项具体证据措施，不仅是认知手段，也兼具道德属性。例如，稳定性要求审判给予辩方充分的权利和机会去审查控方提出的犯罪叙事，辩方也有责任以最有效的方式去质疑和挑战犯罪叙事。只有当犯罪叙事抵挡住了这些反对和质疑，稳定性才得以满足，这既可以为准确地认定事实提供认知保障，又可以减少错误定罪时带来的危害。其次，确定概率阈值完全是基于道德权衡的，通过考量负效用比或两类错误定罪率之比来确定阈值。

在囚犯案例中，基于 999 : 1 的纯统计证据，似乎可得出甲某的有罪概率为 0.999，但这个有罪概率实际上是不准确的。根据高概率裁决模型，准确的高概率才是裁决的依据，仅满足充分性的高概率不能作为裁决依据，因此 0.999 的有罪概率也不能裁定甲某有罪。之所以传统主义者甚至有些概率主义者认为，按照高概率裁决模型，0.999 的有罪概率足以认定甲某有罪，是因为他们忽视了有罪概率的准确性要求。只有在叙事性、稳定性和完全性都被逐一遵循时，审判辨识力才能达到最高或总错判风险达到最小，在此基础上计算有罪概率，其准确性才有保证。在囚犯案例中，叙事性、稳定性和完全性都不满足，所以 0.999 的有罪概率不够准确，不能作为裁决依据。首先，依据 999 : 1 的纯统计证据，控方无法构建一个针对甲某的详细和具体的犯罪叙事。谁实施了谋杀？其犯罪动机是什么？他做了什么事情？他与其他囚犯有什么不同？这些疑问都没有得到回答，无法排除合理怀疑。其次，由于甲某的犯罪叙事不够详细，辩方没有足够的机会去审查叙事，这样导致犯罪叙事没有经历稳定性检验，因而也无法排除合理怀疑。最后，如果甲某真的参与了谋杀，那么在犯罪现场应该还会留下指向甲某的其它证据，但本案例只有一个纯统计证据，不满足完全性要求。由此可见，0.999 的有罪概率不够

准确,或者说,甲某的有罪概率不是 0.999,因而不能认定甲某定罪。韦尔奇也表达了类似的观点,他认为仅根据纯统计证据获得的概率,可能是有罪的先验概率或后验概率,也可能是证据出现的概率。([15])换言之,将 0.999 视为甲某有罪的后验概率是不准确的和不可靠的。

6 结论

在事实认定过程中,由于人类认知能力的局限和各种条件的限制,司法审判不可避免地会犯错。在刑事审判中,存在两类错误:错误定罪和错误释放。因此,司法审判拥有两个核心目标:一是减少总错判风险,二是分配总错判风险。减少总错判风险要求法庭提高审判辨识力,而叙事性、稳定性和完全性则是提高审判辨识力的三项证据措施。它们不仅是保障准确地计算有罪概率的手段,还是产生合理怀疑的来源。为了合理地分配总错判风险,让错误定罪率远小于错误释放率,这就要求证据满足充分性,有罪概率要大于相应的概率阈值,即有罪概率是高概率。高概率具有双重作用,一方面,它说明犯罪叙事得到了充分的证据支持;另一方面,高概率裁决能促进审判的错误定罪率远小于错误释放率,使总错判风险的分配方案符合刑事司法的道德要求。充分性或高概率也是产生合理怀疑的来源。至此,排除合理怀疑的刑事证明标准就可细化为四要素标准,即叙事性、稳定性、完全性和充分性。当四个要素都满足时,事实认定者则应当做出有罪事实裁决。用纯粹的概率语言,刑事证明标准可表达为:准确的高概率,即准确的、大于阈值的有罪概率。仅有高概率不能定罪,但准确的高概率可以作为定罪的依据。当审判满足叙事性、稳定性和完全性要求时,有罪概率的准确性就获得了保障,如果此时有罪概率还满足充分性,即它是高概率,那么就可以认为审判达到了刑事证明标准的要求,应当认定被告有罪。因此,当有罪概率的准确性有了保证之后,刑事证明标准就可以化归为高概率。高概率裁决模型具有正当性,它能消解尼森等传统主义者提出的批判,从而实现智能审判提供理论基础。智能审判的另一个理论基础是依据哪种概率模型来准确地计算有罪概率,这涉及如何运用概率推理和智能工具来执行运算,贝叶斯网络模型或许是一种可行的方案,也是我们要进一步研究的方向。

参考文献

- [1] M. Di Bello, 2013, *Statistics and Probability in Criminal Trials: The Good, the Bad, and the Ugly*, Phd thesis, Stanford University.

- [2] M. Di Bello, 2019, “Trial by statistics: Is a high probability of guilt enough to convict?”, *Mind*, **128(512)**: 1045–1084.
- [3] J. L. Cohen, 1977, *The Probable and the Provable*, New York: Oxford University Press.
- [4] N. Fenton and M. Neil, 2019, *Risk Assessment and Decision Analysis with Bayesian Networks, second edition*, Boca Raton: CRC Press.
- [5] J. Franklin, 2006, “Case comment—United States v. Copeland, 369F. Supp. 2d 365 (E.D.N.Y. 2005): Quantification of the ‘proof beyond reasonable doubt’ standard”, *Law, Probability and Risk*, **5(2)**: 159–165.
- [6] J. Kaplan, 1968, “Decision theory and the factfinding process”, *Stanford Law Review*, **20(6)**: 1065–1092.
- [7] D. H. Kaye, 1986, “Do we need a calculus of weight to understand proof beyond a reasonable doubt?”, *Boston University Law Review*, **66**: 657–672.
- [8] C. R. Nesson, 1979, “Reasonable doubt and permissive inferences: The value of complexity”, *Harvard Law Review*, **92(6)**: 1187–1225.
- [9] J. O. Newman, 2006, “Quantifying the standard of proof beyond a reasonable doubt: A comment on three comments”, *Law, Probability and Risk*, **5**: 267–269.
- [10] J. J. Thomson, 1986, “Liability and individualized evidence”, *Law and Contemporary Problems*, **49(3)**: 199–219.
- [11] L. H. Tribe, 1971, “Trial by mathematics: Precision and ritual in the legal process”, *Harvard Law Review*, **84**: 1329–1393.
- [12] R. Urbaniak, 2018, “Narration in judiciary fact-finding: A probabilistic explication”, *Artificial Intelligence and Law*, **26(4)**: 345–376.
- [13] R. Urbaniak and M. Di Bello, 2021, “Legal probabilism”, in Edward N. Zalta(ed.), *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/fall2021/entries/legal-probabilism/>.
- [14] A. Volokh, 1997, “N guilty men”, *University of Pennsylvania Law Review*, **146(2)**: 173–216.
- [15] J. R. Welch, 2020, “Rebooting the new evidence scholarship”, *The International Journal of Evidence & Proof*, **24(4)**: 351–373.
- [16] G.L. Wells, 1992, “Naked statistical evidence of liability: Is subjective probability enough?”, *Journal of Personality and Social Psychology*, **62(5)**: 739–752.
- [17] 贝克斯 (著), 杜文静、兰磊、周兀 (译), 论证、故事与刑事证据: 一种形式混合理论, 北京: 中国政法大学出版社, 2020 年。

(责任编辑: 映之)

High Probability Model of the Criminal Standard of Proof

Wenjing Du

Abstract

In criminal justice proof, in order to achieve the digitization of judicial trials, it is necessary to quantify the criminal standard of proof. Legal probabilism constructs the “high probability decision model” through the guilty probability and probability threshold. According to this model, when the guilt probability is accurate high probability, the verdict should find the defendant guilty. But, this model has been strongly criticized by legal traditionalists, who argue that it is morally unacceptable to describe the criminal standard of proof with high probability and it will lead to reasoning paradoxes. However, we will defend this model by reconstructing the criminal standard of proof. The criminal standard of proof can be divided into four-fold elements: narrativity, stability, completeness and sufficiency. When a trial satisfies narrativity, stability and completeness, the accuracy of the guilt probability is ensured. If the guilt probability further satisfies sufficiency at this time, then the trial should be considered to meet the criminal standard of proof and the defendant should be found guilty. Therefore, once the accuracy of the guilt probability is guaranteed, the criminal standard of proof can be reduced to a high probability. The high probability decision model has justification and can dispel the criticisms of traditionalism, and can also promote the core function of trials.