逻辑学研究　2026 年第 2 期，44–59
文章编号：1674-3202(2026)-02-0044-16

概率确证测度的评估与选择

杜文静

摘　要：贝叶斯方法主张运用概率论的术语来描述证据对假说的确证度。为了量化这种确证度，学术界提出了多种不同的概率确证测度方案。通过典型案例分析和适当性条件这两个维度，能够深入探讨这些确证测度方案的合理性。其目的是排除不合理的方案，并突出两类具有显著比较优势的测度策略。在此基础上，可以提炼出选择确证测度的指导性原则。该原则强调，在具体应用实践中，应根据特定的探究语境选择合适的测度方案，并详细阐述其相较于其他方案的优势，从而为科学决策提供坚实的理论支撑和实践指导。

关键词：确证；确证度；概率确证测度

中图分类号：B81　　　　**文献标识码**：A

贝叶斯主义认为，确证并非仅仅是全有或全无的定性概念，而是有程度之分。（[5]，第 398 页）证据对假说的支持程度被称为确证度，为其确证假说或理论提供了更精细的分析手段。通过确证度的数值，研究者不仅能判定证据何时确证假说，更能评估证据在何种程度上确证假说。确证测度作为计算确证度的数学公式，具有双重效能：一方面，每个确证测度都有一个中立值，通过比较计算所得的确证度与该中立值，可判断证据是否确证假说；另一方面，确证度的数值比较能够直接反映确证强弱并据此排序。例如：证据 E_1 对假说 H 的确证强于证据 E_2 对 H 的确证；证据 E 对假说 H_1 的确证强于它对假说 H_2 的确证。基于概率论框架，学界根据不同的数学直觉，发展出多种概率确证测度。因此，从众多确证测度中筛选出合理且有效的测度，显得尤为重要。

本文致力于分析并评估各种确证测度的合理性，旨在排除不合理的测度，并筛选出具有比较优势的测度方案。首先，介绍确证测度的基本构造，并着重分析九个互不排序等价的测度。随后，结合典型案例和适当性条件，从两个不同视角对确证测度进行评估。最后，提出确证测度选择的实践指导建议，为具体应用情境提供参考依据。

收稿日期：2025-06-16
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）。

1　确证测度的构造

“在科学哲学、形式认识论及相关领域，‘确证’已跃升为一个核心技术术语。从广义上讲，确证与证据如何影响假说的可信度有关，这一问题对科学探究、医学诊断和法律论证等领域中的人类推理至关重要。”（[3]，第 650 页）确证理论意在探究证据对假说产生的影响，描述证据对假说的支持关系，其核心任务是：当证据不逻辑蕴涵假说时，阐明证据确证或支持假说的含义。（[9]，第 146 页）具体而言，确证理论的研究内容包含两方面：第一，定性的确证理论，探讨证据确证假说的条件；第二，定量的确证理论，通过引入“确证度”的概念，不仅回答证据何时确证假说，还要回答证据在何种程度上确证假说。（[17]，第 7 页）本文聚焦于定量的确证理论，研究核心为确证度的计算方法和途径。

1.1　确证测度的多元化

贝叶斯确证理论为量化确证度提供了框架，通过构建概率确证测度使计算确证度成为可能。为深入阐释贝叶斯确证理论，有必要引入一些基本概念与符号。在此框架下，评估证据 E 对假说 H 的确证度，是相对于一个特定的概率分布 P 进行的，而该概率分布 P 是基于一个明确的命题语言 L 构建的。在任何具体的应用领域或探究语境中，需首先明确讨论的范围，厘清相关的背景信息以及所关注的命题，进而构建出一个适宜的命题语言 L。它由原子命题构成，并通过五种命题联结词（否定、合取、析取、实质蕴涵、实质等值）生成所有逻辑上可能的复合命题。（[16]，第 26 页）命题语言 L 界定了交流与对话的范围，证据 E 和假说 H 均为该语言中的命题。概率分布 P 是定义在命题语言 L 上的函数，它为 L 中的每一个命题分配一个实数值，并满足概率论的基本公理。概率分布的作用在于，它能够将一个具体探究语境中的所有相关背景信息以及主体的认知观点，以量化的形式进行编码（[9]，第 149 页），从而直观地反映出主体对每个命题的信念度。

在逻辑学中，非矛盾式且非永真式的命题被称为可能性命题。贝叶斯确证理论的基本原则包含两个方面：首先，所有证据和假说均为可能性命题。若概率分布 P 对命题语言 L 中的每个可能性命题 α 指派的概率满足 $0 < P(\alpha) < 1$，则称 P 是正则概率分布。其次，所采纳的概率分布必须是正则的。（[16]，第 99–100 页）正则概率分布之所以重要，是因为它确保了证据和假说被赋予合理的概率值，从而支持确证度的计算。假设证据 E 和假说 H 是命题语言 L 中的两个可能性命题，P 是建立在 L 上的一个正则概率分布，那么，先验概率 $P(H)$ 反映的是基于现有的背景信息 B，主体对假说 H 的信念度；后验概率 $P(H|E)$ 表示的是，仅将 E 加入 B 之后，在同时考虑 B 和 E 时，主体对假说 H 的信念度。概率 $P(E)$ 表示的是基于现有的背景信息 B，主体预测证据 E 出现的可能性；似然度 $P(E|H)$ 表示

仅将 H 加入 B 之后，在同时考虑 B 和 H 时，主体预测证据 E 出现的可能性；两者测量的都是证据的可预测性。由于背景信息和主体的认知观点都被封装在概率分布中，因此，为了叙述方便，在下文中我们将背景信息视为一种默认，不再单独提及。

基于上述术语和符号，我们进一步引入贝叶斯确证理论。卡尔纳普（R. Carnap）主张，可借助概率理论构建一种令人满意的确证理论。卡尔纳普为此提出了两种不同的确证理论。（[1]，第 xvi 页）第一种是坚固性的确证，其定义为：证据 E 确证假说 H 当且仅当 $P(H|E)$ 很高，即 $P(H|E)$ 要大于某个阈值 k（$0.5 < k < 1$）。第二种是坚固性递增的确证，其定义为：证据 E 确证假说 H 当且仅当 $P(H|E) > P(H)$；证据 E 否证假说 H 当且仅当 $P(H|E) < P(H)$；证据 E 与假说 H 不相干当且仅当 $P(H|E) = P(H)$。前者亦称为绝对确证，后者则称为递增确证。（[3]，第 651 页）

绝对确证理论存在诸多问题。首先，精确地确定阈值颇具挑战性，且缺乏明确的标准。其次，绝对确证可能面临不相干证据确证假说的困境。例如，考虑以下探究语境：有一种公平的彩票游戏，共有 10000 张彩票，仅有一张彩票能中奖。假设 10000 张彩票已随机售出，李某购买了其中一张。令假说 H 表示“李某的彩票没有中奖”，在李某未打开彩票查看中奖结果之前，我们分析何种证据会确证 H。显然，在现有背景信息下，$P(H)$ 非常高，其值为 0.9999。设证据 E 为“今天刮风”，那么 $P(H|E)$ 同样非常高，其值也是 0.9999。根据绝对确证的定义，E 确证了 H，然而这与直觉相悖，因为 E 和 H 在直觉上是不相干的。因此，本文仅探讨递增确证，并以此为基础来构造概率确证测度。

依据递增确证理论，只要证据能够“增加”人们对假说的信念度或可信度，那么证据被认为确证假说。在这种意义下，确证度实际上等同于信念度的增加程度。因此，每一个用于测量信念度增加程度的计算公式，本质上就构成一种确证测度。基于此，表 1 所列的确证测度显得非常直观且自然。从数学直觉出发，这些测度可以通过作差、作商和对数运算来构造。例如，$C_{P1}(H,E)$ 通过后验概率与先验概率之差来测量信念度的增加程度，即 $C_{P1}(H,E) = P(H|E) - P(H)$，当 $C_{P1}(H,E) > 0$ 时，$P(H|E) > P(H)$，证据 E 确证假说 H，并且 $C_{P1}(H,E)$ 比 0 越大，E 对 H 的确证度就越高。类似地，$C_{P2}(H,E)$ 通过后验概率与先验概率之商来测量信念度的增加程度，即 $C_{P2}(H,E) = P(H|E)/P(H)$，当 $C_{P2}(H,E) > 1$ 时，$P(H|E) > P(H)$，证据 E 确证假说 H，并且 $C_{P2}(H,E)$ 比 1 越大，E 对 H 的确证度就越高。

为了使确证测度的中立值均为 0，我们对 $C_{P2}(H,E)$ 进行对数运算，从而得到测度 $C_{P3}(H,E)$，即 $C_{P3}(H,E) = \log\ C_{P2}(H,E)$。由于对数运算是单调递增的，即 $C_{P2}(H,E)$ 越大 $C_{P3}(H,E)$ 也越大，$C_{P2}(H,E)$ 越小 $C_{P3}(H,E)$ 也越小，因此 $C_{P2}(H,E)$ 和 $C_{P3}(H,E)$ 是两个排序等价的确证测度。从形式上讲，两个确证测

作差	作商	取对数
$C_{P1}(H,E)=P(H\|E)-P(H)$	$C_{P2}(H,E)=P(H\|E)/P(H)$	$C_{P3}(H,E)=\log[P(H\|E)/P(H)]$
$C_{P4}(H,E)=P(H\|E)-P(H\|\neg E)$	$C_{P5}(H,E)=P(H\|E)/P(H\|\neg E)$	$C_{P6}(H,E)=\log[P(H\|E)/P(H\|\neg E)]$
$C_{P7}(H,E)=P(E\|H)-P(E)$	$C_{P8}(H,E)=P(E\|H)/P(E)$	$C_{P9}(H,E)=\log[P(E\|H)/P(E)]$
$C_{P10}(H,E)=P(E\|H)-P(E\|\neg H)$	$C_{P11}(H,E)=P(E\|H)/P(E\|\neg H)$	$C_{P12}(H,E)=\log[P(E\|H)/P(E\|\neg H)]$

表 1: 根据数学直觉获得的 12 个概率确证测度

度排序等价是指，对任意的概率分布 P，对任意的证据 E_1 和 E_2，以及假说 H_1 和 H_2，都有：$C_{P^*}(H_1,E_1)>/=/<C_{P^*}(H_2,E_2)$ 当且仅当 $C_{P^{**}}(H_1,E_1)>/=/<C_{P^{**}}(H_2,E_2)$。换言之，对于“哪对证据与假说之间的确证度更高”的问题，排序等价的测度会得出相同的答案。（[4]，第 231 页）很显然，如果一个确证测度经过单调递增函数运算后得到另一测度，那么这两个确证测度必然是排序等价的。表 1 中第 3 列的确证测度，分别是由第 2 列的确证测度通过对数运算后得到的测度，因此它们分别排序等价。排序等价的一个重要作用在于，它能够缩减需要研究的确证测度的个数。当存在多个彼此排序等价的确证测度时，我们只需研究其中一个即可。例如，在表 1 中，我们可以选择研究第 3 列的测度，而无需再研究第 2 列的确证测度，这样便减少了 4 个确证测度；同样地，也可以选择只研究第 2 列的测度。

需要说明的是，$C_{P1}(H,E)$ 和 $C_{P3}(H,E)$ 并不排序等价。例如，考察以下探究语境，有两对证据和假说，在概率分布 P 下，它们的概率值分别如下：$P(H_1)=0.02$，$P(H_1|E_1)=0.32$；$P(H_2)=0.15$，$P(H_2|E_2)=0.48$。于是，$C_{P1}(H_1,E_1)=0.32-0.02=0.3$，$C_{P1}(H_2,E_2)=0.48-0.15=0.33$；$C_{P3}(H_1,E_1)=\log(0.32/0.02)\approx 1.2$，$C_{P3}(H_2,E_2)=\log(0.48/0.15)\approx 0.5$。这说明，对于测度 $C_{P1}(H,E)$ 而言，E_1 对 H_1 的确证度低于 E_2 对 H_2 的；而对于测度 $C_{P3}(H,E)$ 来说，E_1 对 H_1 的确证度却高于 E_2 对 H_2 的。这两个确证测度对同一比较判断得出了两个相互矛盾的结果，因此它们不排序等价。

另一方面，从数学上可以证明，下面四个概率不等式是相互等价的：

$$
\begin{aligned}
P(H|E)>P(H) &\Leftrightarrow P(H|E)>P(H|\neg E)\\
&\Leftrightarrow P(E|H)>P(E)\\
&\Leftrightarrow P(E|H)>P(E|\neg H)
\end{aligned}
$$

（[14]，第 400 页）

这些不等式之间的等价性可以通过概率论的基本性质进行推导。这些等价关系表明，当一个不等式成立时，其他三个不等式也必然成立。这种等价性为构造丰富多样的确证测度提供了坚实的数学基础。于是，根据后三个不等式（除了第一个

不等式），按照作差、作商和取对数的运算原理，我们可以自然地得到表 1 中剩余的 9 个确证测度。这些测度通过不同的数学运算方式，从不同的角度刻画了证据对假说的确证度。

此外，在数学上，任何一个概率都对应于一个几率，它们相互唯一确定。例如，概率 $P(\alpha)$ 对应的几率是 $P(\alpha)/P(\neg\alpha)$，即 $P(\alpha)/[1-P(\alpha)]$；条件概率 $P(\alpha|\beta)$ 对应的几率是 $P(\alpha|\beta)/P(\neg\alpha|\beta)$。若概率 $P(\alpha)=0.6$，则几率 $P(\alpha)/P(\neg\alpha)=3/2$；反之，若几率 $P(\alpha)/P(\neg\alpha)=3/7$，则概率 $P(\alpha)=0.3$。概率和几率的对应关系是单调递增的，也就是说，概率越大，对应的几率也越大；几率越大，对应的概率也越大。因此，每个概率不等式都等价于一个几率不等式。于是，上述四个概率不等式都等价于下面四个几率不等式：

$$\begin{aligned}\frac{P(H|E)}{P(\neg H|E)}>\frac{P(H)}{P(\neg H)} &\iff \frac{P(H|E)}{P(\neg H|E)}>\frac{P(H|\neg E)}{P(\neg H|\neg E)}\\ &\iff \frac{P(E|H)}{P(\neg E|H)}>\frac{P(E)}{P(\neg E)}\\ &\iff \frac{P(E|H)}{P(\neg E|H)}>\frac{P(E|\neg H)}{P(\neg E|\neg H)}\end{aligned}$$

换言之，这八个不等式彼此等价。这种等价性进一步丰富了确证测度的构造方法，使得从概率和几率两个角度都可以构建出合理的确证测度。利用这四个几率不等式，并参照表 1 中作差、作商和取对数的运算原理，我们同样可以自然地得到另外 12 个确证测度。尽管本文无意详尽列出这些确证测度的具体计算公式，但阐述这些数学原理的意图在于强调：运用概率方法可以获得丰富多样的确证测度，从而彰显确证测度的多元化特征。

1.2 九个互不排序等价的确证测度

基于数学直觉和基本运算，我们阐述了构造确证测度的一般方法，并列举了其中的 12 个确证测度。实际上，通过类似的数学操作，可衍生出更多的其他确证测度，详见席佩斯（M. Schippers）等人的研究。（[14]，第 405、408 页）然而，鉴于排序等价的测度在确证关系的相对评价上本质相同，确证测度的“多元化”核心应指向存在多种互不排序等价的测度类型。因此，我们聚焦于文献中广泛讨论的 9 个互不排序等价的确证测度，并对其合理性进行评估。（[4]，第 236 页）这 9 个测度的分类标准有两条：第一，它们捕捉了不同的数学直觉，通过不同的数学运算来测量一个数比另一个数增大的程度；第二，它们互不排序等价。

为便于查阅与比较，表 2 汇总了这 9 个测度的计算公式。具体来说，前 4 个确证测度（表 2 第 1 列的前 4 个）分别是：C_{P1}（[1]，第 361 页）、C_{P4}（[2]，第 450 页）、C_{P7}（[11]，第 152 页）和 C_{P10}（[12]，第 252 页），其计算公式与表 1

第 1 列中的公式完全一致。为明确展示本文所列确证测度的个数，后续的第 5 至第 9 个测度的下标序号将延续表 1 中的编号。

$C_{P1}(H,E)=P(H\|E)-P(H)$	$C_{P14}(H,E)=P(E\wedge H)-P(E)P(H)$
$C_{P4}(H,E)=P(H\|E)-P(H\|\neg E)$	$C_{P15}(H,E)=1-[P(\neg H\|E)/P(\neg H)]$
$C_{P7}(H,E)=P(E\|H)-P(E)$	$C_{P16}(H,E)=[P(E\|H)-P(E\|\neg H)]/[P(E\|H)+P(E\|\neg H)]$
$C_{P10}(H,E)=P(E\|H)-P(E\|\neg H)$	$C_{P17}(H,E)=\begin{cases}\frac{P(H\|E)-P(H)}{1-P(H)}, & \text{当}P(H\|E)\geq P(H),\\ \frac{P(H\|E)-P(H)}{P(H)}, & \text{当}P(H\|E)<P(H).\end{cases}$
$C_{P13}(H,E)=[P(H\|E)/P(H)]-1$	

表 2: 9 个常见的互不排序等价的确证测度

在计算确证度和否证度时，C_{P17} 运用两种不同的计算方式，而其他 8 个测度都采用统一的计算公式。因此，表 2 中第 5 至第 9 个互不排序等价的确证测度可划分为两类：一类是基于统一计算公式的测度。第 5 个测度是 C_{P13}（[6]，第 307 页），它由测度 C_{P2} 减 1 得到，目的是为了使中立值为 0。它与前四个测度在数学直觉上均具有明显的直观性，通过差值或商值来衡量一个数相对于另一个数的增加程度。第 6 个测度是 C_{P14}（[1]，第 360 页），它反映了 E 和 H 偏离独立性的程度。当 E 和 H 相互独立时，C_{P14} 达到中立值 0，E 和 H 不相干；当 E 和 H 正向偏离独立性时，$C_{P14}>0$，E 确证 H；当 E 和 H 负偏离独立性时，$C_{P14}<0$，E 否证 H。第 7 个测度是 C_{P15}（[13]，第 129 页），它表达了 E 对 H 的确证度等于 1 减去 E 对 H 的否证程度。第 8 个测度是 C_{P16}（[10]，第 320 页），它是似然比测度 C_{P11} 的单调递增函数，因而与 C_{P11} 排序等价。另一类则是根据确证或否证的具体情境选择计算公式，即第 9 个测度 C_{P17}（[4]，第 233 页），它是一种相对距离的测度。其中，$1-P(H)$ 表示 H 与确定性为真的初始距离，而 $P(H)=P(H)-0$ 表示 H 与确定性为假的初始距离。当 E 确证 H 时，即 $P(H|E)>P(H)$，C_{P17} 反映了 E 使得 H 距离确定性为真的相对缩小程度，换言之，C_{P17} 越大，H 越接近确定性为真。当 E 否证 H 时，即 $P(H|E)<P(H)$，C_{P17} 的绝对值反映了 E 使得 H 距离确定性为假的相当缩小程度，也就是说，C_{P17} 的绝对值越大，H 越靠近确定性为假。

2　确证测度的评估

上述 9 个确证测度是基于不同的数学直觉构建的，彼此之间互不排序等价。因此，在对同一案例分析时，可能会产生相互冲突的确证结论。鉴于此，对确证测度的合理性进行评估显得尤为重要。现在将结合典型案例和适当性条件，从两

个不同视角对这些确证测度进行评价。

2.1 基于典型案例的评估

为评估确证测度的合理性，我们通过构建一系列典型案例来进行检验。具体而言，首先需创设一个明确的探究语境，确定相关的命题语言，并为其指派一个符合直觉且合理的概率分布。其次，在该命题语言中选择若干感兴趣的证据和假说。然后，一方面依据确证直觉形成一种确证判断；另一方面，运用确证测度计算相应的确证度，从而获得另一种确证判断。最终，对比这两种确证判断。若两者相一致，则该确证测度可视为合理的；若两者相矛盾，则该确证测度可能存在问题。现在，我们将通过两个案例来说明五个确证测度的不合理性。

案例 1. *考虑投掷一枚均匀的六面骰子。在这种语境下，命题语言中的命题是关于骰子可能出现点数的陈述。一个自然的直觉概率分布是给每个可能出现的点数指派相同的概率，即出现点数 i（$i = 1, 2, 3, 4, 5, 6$）的概率均为 $1/6$。现在，我们感兴趣的证据 E 为：出现的点数是集合 $\{2, 3, 5\}$ 中的某个数，感兴趣的假说 H_1 为：出现的点数是 2；H_2 为：出现的点数是 3 或 5。根据直觉，证据 E 应该既确证 H_1 又确证 H_2，但 E 对 H_2 的确证度应该大于 E 对 H_1 的确证度。*

然而，通过概率运算，我们发现测度 C_{P7} 和 C_{P13} 与上述直觉判断相矛盾，其他 7 个测度都符合上述直觉。根据测度 C_{P7}，$C_{P7}(H_1, E) = P(E|H_1) - P(E) = 1 - 1/2 = 1/2$，$C_{P7}(H_2, E) = P(E|H_2) - P(E) = 1 - 1/2 = 1/2$，这说明 E 对 H_1 和 E 对 H_2 的确证度相同。根据测度 C_{P13}，$C_{P13}(H_1, E) = P(H_1|E)/P(H_1) - 1 = (1/3) \div (1/6) - 1 = 1$，$C_{P13}(H_2, E) = P(H_2|E)/P(H_2) - 1 = (2/3) \div (1/3) - 1 = 1$，这说明 E 对 H_1 也和 E 对 H_2 的确证度相同。由此可见，测度 C_{P7} 和 C_{P13} 是不合理的。

案例 2. *考察这样一种情景：一个盒子中装有 100 个小球，其中 90 个是白色的，2 个是黑色的，8 个是红色的。这些小球除了颜色不同外，其他方面均相同。盒子是密闭且不透光的，人们无法从外部看到每个小球的颜色。假定我们将从盒子中随机取出一个小球，并探究或预测取出小球的颜色。在这种探究语境下，命题语言中的命题是关于取出小球颜色的陈述。一种自然的概率分布为：取出的小球是白色的概率为 90/100、取出的小球是黑色的概率为 2/100、取出的小球是红色的概率为 8/100。现在，我们感兴趣的假说 H 是：取出的小球不是红色的；感兴趣的证据 E_1 是：取出的小球是白色的，E_2 是：取出的小球是黑色的。根据直觉，证据 E_1 和 E_2 都应当确证假说 H，并且具有相同的确证度。*

以测度 C_{P4} 为例，其计算结果显示，对于假说 H，证据 E_1 的确证度大于证据 E_2，即根据测度 C_{P4}，$C_{P4}(H, E_1) = 1 - 2/10 = 8/10$，$C_{P4}(H, E_2) = 1 - 90/98 =$

8/98。然而，这与我们的直觉判断相悖，因为直觉上，证据 E_1 和证据 E_2 对假说 H 的确证程度应当是相同的。同样地，测度 C_{P7}、C_{P10} 和 C_{P14} 也给出了证据 E_1 对假说 H 的确证度大于证据 E_2。这些不一致性表明，测度 C_{P4}、C_{P7}、C_{P10} 和 C_{P14} 在评估确证度时可能存在问题，因此它们是不合理的。

2.2　基于适当性条件的评估

适当性条件是评价确证测度合理性的另一种方法。这些条件是基于确证直觉提出的一般性原则，被视为合理的确证测度要满足的必要条件。当确证测度满足某个适当性条件时，它有可能是合理的；反之，若不满足，则应被视为不合理。在文献中存在许多适当性条件（[3]，第 654、658 页），其中最具典型意义的条件有以下八个，它们从不同侧面刻画了合理确证测度应具备的性质。

第一，分类原则。定量确证理论是定性确证理论的推广，它要求确证度不仅能判断证据是否确证假说，还能反映确证的程度。因此，一个合理的确证测度应当具备分类的能力，这意味着它应设定一个中立值作为基准。当确证度大于中立值时，证据就确证假说；当确证度小于中立值时，证据否证假说；当确证度等于中立值时，证据与假说不相干。在这个适当性条件的审视下，上述 9 个确证测度具有同等优势，它们都设定了中立值，且中立值均为 0。

第二，后验概率原则。给定一个命题语言 L 及其上的概率分布 P，E_1 和 E_2 是 L 中的证据，H 是 L 中的假说，则 $C_P(H,E_1) > / = / < C_P(H,E_2)$ 当且仅当 $P(H|E_1) > / = / < P(H|E_2)$。对于“同一个”假说 H，后验概率原则捕捉的直觉是：证据的确证度与后验概率或后验信念度应具有共变性，即确证度随后验概率的增大而增大，后验概率也随确证度的增大而增大。从数学上可以证明，测度 C_{P4} 和 C_{P14} 不满足后验概率原则，而其他 7 个测度均满足这个原则。因此，在后验概率原则下，测度 C_{P4} 和 C_{P14} 表现出劣势，这也进一步彰显了它们的不合理性；相比之下，其他 7 个测度表现出明显的优势。

第三，不可交换原则。给定一个命题语言 L 及其上的概率分布 P，E 和 H 分别是 L 中的两个命题，如果 E 确证 H，即 $C_P(H,E) > 0$，那么通常 $C_P(H,E) \neq C_P(E,H)$。[1] 根据这个原则，在确证情形下，E 对 H 的确证度通常不等于 H 对 E 的确证度，E 和 H 在确证测度的计算公式中不能交换位置。例如，在一副不含大小王的 52 张扑克牌中，随机从中抽取一张扑克牌，令 E 表示“这张扑克牌是黑桃 A”，H 表示“这张扑克牌是黑桃”。根据直觉，E 确证 H，H 也确证 E，但 E 对 H 的确证度要明显大于 H 对 E 的确证度，这是因为，E 逻辑蕴涵 H，但 H 不蕴涵 E。同样，在数学上可以证明，测度 C_{P13} 和 C_{P14} 不满足这一不可交换原

[1] 从确证的定性维度来看，根据贝叶斯定理，当 E 确证 H 时，H 也确证 E，即 $C_P(H,E) > 0 \Leftrightarrow C_P(E,H) >$ 0。但从定量维度来看，虽然两个确证度都大于 0，但它们的具体数值通常是不相等的。

则，而其他 7 个测度均满足该原则。在不可交换原则下，测度 C_{P13} 和 C_{P14} 表现出劣势，从而说明这两个测度不合理；而其他 7 个测度表现出优势。

第四，逻辑原则。逻辑原则是指，当证据 E 逻辑蕴涵假说 H 时，确证度 $C_P(H,E)$ 应达到最大值；当 E 反驳 H，即 E 逻辑蕴涵 $\neg H$ 时，$C_P(H,E)$ 应取到最小值。逻辑原则所反映的直觉是：演绎逻辑主要关注逻辑蕴涵或逻辑后承，而归纳逻辑则侧重于归纳推理和确证度的量化。逻辑蕴涵被视为最大的确证，而反驳则被视为最大的否证。（[8]，第 501–502 页）在数学上可以证明，只有测度 C_{P16} 和 C_{P17} 满足逻辑原则，其他 7 个测度都不满足这个原则。所以，在逻辑原则条件下，只有测度 C_{P16} 和 C_{P17} 表现出优势。

第五，对称原则。给定一个命题语言 L 及其上的概率分布 P，E 和 H 分别是 L 中的两个命题，则有：（1）如果 E 确证 H，那么 $C_P(H,E)=C_P(\neg E,\neg H)$；（2）如果 E 否证 H，那么 $C_P(H,E)=C_P(E,H)$。对称原则表达的直觉是：确证作为逻辑蕴涵的推广，应当模拟蕴涵的相关性质。对称原则的第一条模拟的逻辑性质是，E 蕴涵 H 当且仅当 $\neg H$ 蕴涵 $\neg E$；第二条模拟的逻辑性质是，E 反驳 H 当且仅当 H 反驳 E。可以证明，只有测度 C_{P17} 满足对称原则，其他 8 个测度都不符合这个原则。在对称原则下，只有测度 C_{P17} 表现出优势。

第六，模块性原则。给定一个命题语言 L 及其上的概率分布 P，E 和 E^* 是 L 中的两个证据，H 是 L 中的假说，P^* 是在 E^* 为真的条件下 L 上的后验概率分布。如果在概率分布 P 下，$P(E|E^*,H)=P(E|H)$ 和 $P(E|E^*,\neg H)=P(E|\neg H)$，即证据 E 和 E^* 关于 H 条件独立，那么 $C_P(H,E)=C_{P^*}(H,E)$。模块性原则反映的直觉是：概率不相干信息的加入不会改变证据的确证度。可以证明，只有测度 C_{P16} 满足模块性原则，其他 8 个测度都不符合这个原则。如果模块性原则是确证测度需要满足的必要条件，那么只有测度 C_{P16} 表现出优势。

第七，互不相容假说的析取原则。给定一个命题语言 L 及其上的概率分布 P，H_1 和 H_2 是 L 中两个互不相容的假说，即 $P(H_1\wedge H_2)=0$，E 是 L 中的证据，则 $C_P(H_1\vee H_2,E)>/=/<C_P(H_1,E)$ 当且仅当 $P(H_2|E)>/=/<P(H_2)$。这个原则表达的直觉是：当向 H_1 中增加一个互不相容的假说 H_2，形成一个逻辑上更弱的析取命题 $H_1\vee H_2$ 时，当且仅当 E 确证 H_2 时，E 对 $H_1\vee H_2$ 的确证度才会大于 E 对 H_1 的确证度。在扑克牌例子中，如果 E 表示“这张扑克牌是红色的”，H_1 表示“这张扑克牌是方块”，H_2 表示“这张扑克牌是红心”，那么 E 确证 H_1，也确证 H_2，E 对 $H_1\vee H_2$ 的确证度大于 E 对 H_1 的确证度。但如果 H_2 表示“这张扑克牌是黑桃”，那么 E 否证 H_2，E 对 $H_1\vee H_2$ 的确证度小于 E 对 H_1 的确证度。从数学上可以证明，只有测度 C_{P1} 满足互不相容假说的析取原则。若将此原则视为合理确证测度必须满足的必要条件，那么只有测度 C_{P1} 表现出优势。

第八，似然度原则。给定一个命题语言 L 及其上的概率分布 P，H_1 和 H_2 是

L 中的两个假说，E 是 L 中的证据，则 $C_P(H_1, E) > / = / < C_P(H_2, E)$ 当且仅当 $P(E|H_1) > / = / < P(E|H_2)$。似然度原则表达的直觉是：$E$ 对 H_1 的确证度大于 E 对 H_2 的确证度，当且仅当 E 在 H_1 的条件下比在 H_2 的条件下更符合预期。通过数学证明，我们发现只有测度 C_{P13} 满足似然度原则。因此，在似然度原则下，其他 8 个测度均表现出劣势，应视为不合理的；只有测度 C_{P13} 表现出优势。

由此可见，各类确证测度各具特色，它们分别满足不同的适当性条件，这也在一定程度上反映出，不同确证测度所捕获的确证直觉存在差异。借助对这些适当性条件的对照分析，我们能够排除那些不合理的测度，进而筛选出具有比较优势的测度。在这 9 个确证测度中，并未有一个测度能全面满足上述 8 个原则，但测度 C_{P16} 和 C_{P17} 满足的原则最多，各自满足了 5 条原则，表现最优。

需要指出的是，运用适当性条件来评价确证测度并非没有局限性。每个适当性条件皆基于特定的直觉提出，而这些直觉的合理性直接决定了条件本身的合理性。若某一适当性条件所依赖的直觉缺乏足够的说服力，则该条件的合理性亦会受到质疑。相应地，将其作为确证测度必须满足的必要条件也值得商榷。在上述八个适当性条件中，似乎仅有前四个原则的直觉基础具有充分的说服力。相较而言，后四个适当性条件虽看似合理，但其直觉基础并不完全令人信服。例如，对称原则的直觉基础在于确证应当模拟逻辑蕴涵的特征。尽管确证可以视为蕴涵的一种扩展，但为何确证必须模拟蕴涵的特征？这一点难以给出令人信服的解释。（[4]，第 241 页）同理，要彻底说服人们相信模块性原则是确证测度必须满足的必要条件，也并非易事。

此外，有些适当性条件自身的合理性是存在争议的。互不相容假说的析取原则在某些情境下显得不够合理。例如，在一副不含大小王的 52 张扑克牌中，随机抽取一张，令 H_1 表示“这张扑克牌是黑桃 A”，H_2 表示“这张扑克牌是红色的”，E 表示“这张扑克牌是人头牌”（其中人头牌是指 J、Q 和 K 这三类扑克牌）。显然，命题 H_1 和 H_2 是互不相容的，即 $P(H_1 \wedge H_2) = 0$；$P(H_2|E) = 6/12 = 0.5$，$P(H_2) = 26/52 = 0.5$，于是 $P(H_2|E) = P(H_2)$。据此，根据互不相容假说的析取原则，可以得到：E 对 $H_1 \vee H_2$ 的确证度等于 E 对 H_1 的确证度，即 $C_P(H_1 \vee H_2, E) = C_P(H_1, E)$。然而，这样的结论似乎不符合直觉，这是因为 E 完全反驳 H_1，但 E 却不能完全反驳 $H_1 \vee H_2$。

似然度原则主要源自似然主义思想。似然主义者主张，比较证据 E 对两个互不相容假说 H_1 和 H_2 的影响，仅需考虑两个似然度 $P(E|H_1)$ 和 $P(E|H_2)$ 的大小。具体而言，如果 $P(E|H_1) > P(E|H_2)$，那么 E 对 H_1 的支持力强于对 H_2 的支持力。如果 $P(E|H_1) = P(E|H_2)$，那么 E 对 H_1 的支持力等于对 H_2 的支持力。如果 $P(E|H_1) < P(E|H_2)$，那么 E 对 H_1 的支持力弱于对 H_2 的支持力。然而，此处的“支持力”与确证度是两个不同的概念。支持力是一种四元关系，涉

及一个证据 E、两个不同的假说（设为 H_1 和 H_2）以及一个概率分布 P，而确证度 $C_P(H,E)$ 是一种三元关系，关涉一个证据 E、一个假说 H 和一个概率分布 P。二者之间并不存在一种直观的、能够直接将四元关系的支持力还原为三元关系的确证度的联系。此外，似然主义者对确证度持怀疑态度，认为其缺乏客观性。他们指出，确证度依赖于先验概率或含义不明确的否定命题，因此认为确证度要么过于主观，要么定义不清晰，甚至无意义。因此，基于似然主义思想提出的似然度原则，并非确证测度必须满足的必要条件。

由此可见，在用适当性条件评估确证测度的合理性时，分类原则、后验概率原则、不可交换原则和逻辑原则的直觉性最强，所占权重最高。如果某个测度违背这四个原则中的一个，那么它将被视为不合理的测度。对称原则和模块性原则的直觉基础一般，所占权重较弱。然而，互不相容假说的析取原则和似然度原则自身的合理性还存在争议。

除了结合典型案例和适当性条件来评价确证测度之外，亦有学者从心理学层面借助实验手段来评估测度的合理性。在心理学领域，合理的确证测度应当能够解释或预测人类在实践中关于确证的判断行为。换言之，根据确证测度所得到的确证判断，应当与人类确证判断的结果相契合。克鲁比（V. Crupi）等人依据心理学实验结果，对表 2 中 9 个确证测度在描述测试者确证判断结果的表现进行了评估。研究发现，描述效果最佳的测度是 C_{P17}，次之为 C_{P16}，而描述效果最差的则是 C_{P1}。（[4]，第 243 页）因此，从心理学角度而言，测度 C_{P16} 和 C_{P17} 显现出比较优势，表现最为出色。

3 确证测度的选择

在对九个常见的确证测度进行分析后，可以观察到 C_{P16} 和 C_{P17} 展现出了最为显著的比较优势。因此，它们在当前学界获得了广泛认可，与之排序等价的测度也同样受到了重视。例如，在司法证明中，似然比测度 C_{P11} 被普遍采纳以测量证据的确证度（[15]，第 24 页），它与 C_{P16} 排序等价，同属于一类测度。在法庭科学中，证据评价国际准则也倡议使用似然比测度作为确证度的标准化度量工具。尽管未来研究可能会发现更多具有比较优势的新测度，但目前的挑战在于如何在实践中选择最适合的测度来测量确证度。

事实上，在实践中，我们应当依据应用领域或探究语境中的具体需求，选择合适的确证测度，并为所选择的测度提供支持性理由，阐释在这个特定语境中该测度为何优于其他测度。（[7]，第 S372 页）例如，若关注的是证据 E 对假说 H 和 $\neg H$ 的辨识力，则应选择测度 C_{P16}，或与之排序等价的测度 C_{P11} 和 C_{P12}。若期望确证能尽可能地反映逻辑蕴涵的性质，则应选择测度 C_{P17}。在科学探究、医

学诊断、司法证明和法庭科学等多个领域中，推理活动的进行往往需要计算确证度。接下来，我们将以法庭科学家评估科学证据的确证度为例，说明确证测度选择的依据。

在法庭科学领域，法庭科学家在其鉴定报告中常常需要评估科学证据的确证度或证明力。通常情况下，确证度较高的证据，其证据价值也相对更高。因此，警方应当集中力量和资源收集确证度较高的证据，法官也应当着重审查评估确证度较高的证据。例如，DNA、指纹、纤维、玻璃碎片等常见科学证据，对于警方快速锁定嫌疑人并明确侦查方向具有至关重要的作用。以一起刑事案件为例，当警方从犯罪现场收集到一块血迹，并经由法庭科学家分析确认该血迹的 DNA 与张某的 DNA 相匹配时，警方便可以根据这份 DNA 匹配的鉴定意见，将张某暂时列为嫌疑人，并围绕张某展开深入调查，以寻找破案的线索。这是因为，DNA 匹配证据确证了血迹来源于张某，进而可以推断案发时张某在犯罪现场，再进一步推断出张某可能是凶手。然而，这一推论链包含三个推断步骤，每个推断均伴随着不确定性、可错性和可废止性。其中，与 DNA 匹配证据直接相关的命题是：犯罪现场的血迹来源于张某（记为 H）。设 E 表示命题"犯罪现场血迹的 DNA 与张某的 DNA 相匹配"，那么 E 确证 H。现在的问题是：如何测量 E 对 H 的确证度？实际上，在鉴定报告中，法庭科学家不仅要给出 DNA 是否匹配的结论，即形成证据 E，还需提供证据 E 对假说 H 的确证度。那么，专家应如何选择合适的确证测度来测量这一确证度？我们主张，似然比测度 C_{P11}（或其排序等价的测度 C_{P12} 和 C_{P16}）是最佳选择，原因有三：

首先，如前文所述，似然比测度在典型案例和适当性条件两个维度均展现出显著的比较优势。其次，法庭科学家评估确证度的目的在于辨识或区分假说 H 和 $\neg H$，即探究 DNA 匹配证据是更支持血迹来源于张某，还是支持血迹并非来源于张某。似然比测度恰好能够满足这一需求，能有效区分两个对立的假说。最后，法庭科学家的职责在于形成科学证据，并结合其专业知识评估科学证据的价值或确证度。他们无权调阅案件的其他证据，更无需对案件事实发生的可能性进行评判。评估案件事实发生的可能性，是法官或事实调查者的职责范畴。因此，对于证据 E 和假说 H，法庭科学家无需评估假说的先验概率 $P(H)$ 和 $P(\neg H)$，也不需要去计算假说的后验概率 $P(H|E)$。法庭科学家并不擅长且不应涉足先验概率的评估，这超出了其专业范围。即使专家评估先验概率，其结果也可能不准确且难以被接受，因为法律禁止法庭科学家就自己专业领域之外的知识作证或提供专家意见。他们真正需要评估的概率是似然度 $P(E|H)$ 和 $P(E|\neg H)$，这两个概率在法庭科学的专业领域之内。

$P(E|H)$ 表示在犯罪现场的血迹属于张某的条件下，鉴定结果显示犯罪现场血迹的 DNA 与张某的 DNA 相匹配的概率。这一概率反映了法庭科学家在鉴定

DNA 过程中，所采用的鉴定方法与程序的可靠性。若鉴定方法完全可靠，则 $P(E|H)$ 应为 1。然而，科学实验往往并非绝对可靠，总会存在一些不可控的系统性风险。科学实验通常存在两类错误率：一是错误肯定率，即错误地将两个不匹配的 DNA 样本鉴定为匹配的比率；二是错误否定率，即错误地将两个相匹配的 DNA 样本鉴定为不匹配的比率。这两类错误率可通过鉴定机构长期实验的数据由法庭科学家进行评估。于是，$P(E|H)$ 应等于 1 减去错误否定率。例如，若错误否定率为 0.002，则 $P(E|H) = 1 - 0.002 = 0.998$。

$P(E|\neg H)$ 表示在犯罪现场的血迹不属于张某的条件下，鉴定结果显示犯罪现场血迹的 DNA 与张某的 DNA 相匹配的概率。该概率的大小受到两个主要因素的影响。一是 DNA 随机匹配的概率。即便犯罪现场的血迹并非来自张某，张某的 DNA 仍有可能偶然与血迹的 DNA 相匹配，这种偶然匹配的概率由 DNA 随机匹配率决定。随机匹配率可根据 DNA 数据库，由法庭科学专家进行评估。鉴于 DNA 的生物特性，随机匹配率通常极低，例如 0.0000001。二是错误肯定率。在犯罪现场的血迹不属于张某的情况下，鉴定结果通常应显示两个 DNA 不匹配，但由于操作失误，可能会错误地显示两个 DNA 相匹配，这种可能性由错误肯定率决定，例如 0.001。法庭科学需要综合考虑这两个影响因素来确定 $P(E|\neg H)$ 的大小，例如可以取这两者的平均值，于是 $P(E|\neg H) = (0.0000001 + 0.001)/2 = 0.00050005$。

由此可见，在法庭科学的语境中，法庭科学家仅能获取似然度 $P(E|H)$ 和 $P(E|\neg H)$ 的数值，而无权亦无能力去评估先验概率 $P(H)$、$P(\neg H)$ 以及后验概率 $P(H|E)$ 的数值。因此，法庭科学家在衡量证据 E 对假说 H 的确证度时，只能采用似然比测度 C_{P11}，或其排序等价的测度 C_{P12} 和 C_{P16}，而不能使用测度 C_{P17}，因为在 C_{P17} 的计算公式中，需要知晓 $P(H)$ 和 $P(H|E)$ 的具体数值。若将 $P(E|H) = 0.998$ 和 $P(E|\neg H) = 0.00050005$ 代入似然比测度 C_{P11} 中，则可得确证度为：

$$0.98/0.00050005 \approx 1959.8;$$

代入测度 C_{P16} 中，则可得确证度为：

$$(0.998 - 0.00050005)/(0.998 + 0.00050005) \approx 0.999.$$

这些计算结果均表明，证据 E 对假说 H 的确证度极高。正因如此，在案件侦查过程中，一旦警方收集到血迹或毛发等证据，他们会毫不犹豫地将其交给法庭科学家进行进一步鉴定，以期获得 DNA 匹配证据，为快速锁定嫌疑人提供关键线索。

4 结语

确证度作为衡量证据对假说支持程度的关键指标，为确证假说或理论提供了更为精细的分析工具。概率确证测度则是计算确证度的数学公式。本文首先基于递增确证的概念，融合数学直觉，提出了一种发现确证测度的一般性方法，该方法直观且自然。基于此，系统呈现了 17 个确证测度的计算公式，并重点讨论了其中 9 个互不排序等价的确证测度。为深入探究，结合典型案例和适当性条件，从两重维度对这 9 个确证测度的合理性进行全面而细致地评估。通过这一评估，排除了一些不合理的测度，并筛选出两类具有比较优势的测度：C_{P16} 和 C_{P17}，以及分别与它们排序等价的测度。最后，在理论探讨的基础上，研究进一步结合实际应用场景，提出确证测度选择的具体建议。我们主张，实践中应根据具体的探究语境，从合理测度中做出恰当的选择，同时阐明所选测度优于其他测度的原因。例如在法庭科学领域，法庭科学家应优先选用然比测度 C_{P11}，或与它排序等价的测度 C_{P12} 和 C_{P16}，以此测量确证度或证明力。这样的选择不仅符合法庭科学的严谨性要求，亦可增强司法决策的科学性与可靠性。值得注意的是，本文主要通过考察单个证据对假说的确证度来评估确证测度，从而筛选出了两类具有比较优势的测度。一个未来的可能研究方向是，通过考虑多个证据或证据链场景，分析在动态确证过程中确证度的累积误差表现，从而在这两类测度中做进一步的筛选，以期获得唯一的“好”测度。

参考文献

[1] R. Carnap, 1962, *Logical Foundations of Probability*, Chicago, IL: University of Chicago Press.

[2] D. Christensen, 1999, “Measuring confirmation”, *The Journal of Philosophy*, **96(9)**: 437–461.

[3] V. Crupi and K. Tentori, 2016, “Confirmation theory”, in A. Hájek and C. Hitchcock(eds.), *The Oxford Handbook of Probability and Philosophy*, pp. 650–665, Oxford: Oxford University Press.

[4] V. Crupi, K. Tentori and M. Gonzalez, 2007, “On Bayesian measures of evidential support: Theoretical and empirical issues”, *Philosophy of Science*, **74(2)**: 229–252.

[5] I. Douven, 2021, “Tracking confirmation”, *Philosophy of Science*, **88(3)**: 398–414.

[6] H. A. Finch, 1960, “Confirming power of observations metricized for decisions among hypotheses”, *Philosophy of Science*, **27(3–4)**: 293–307, 391–404.

[7] B. Fitelson, 1999, “The plurality of Bayesian measures of confirmation and the problem of measure sensitivity”, *Philosophy of Science*, **66(**Supplement**)**: S362–S378.

[8] B. Fitelson, 2006, “Logical foundations of evidential support”, *Philosophy of Science*, **73(5)**: 500–512.

[9] A. Hájek and J. M. Joyce, 2014, “Confirmation”, in S. Psillos and M. Curd(eds.), *The Routledge Companion to Philosophy of Science*, pp. 146–159, New York: Routledge.

[10] J. Kemeny and P. Oppenheim, 1952, “Degrees of factual support”, *Philosophy of Science*, **19(4)**: 307–324.

[11] H. Mortimer, 1988, *The Logic of Induction*, Paramus, NJ: Prentice Hall.

[12] R. Nozick, 1981, *Philosophical Explanations*, Oxford: Clarendon Press.

[13] L. J. Rips, 2001, “Two kinds of reasoning”, *Psychological Science*, **12(2)**: 129–134.

[14] M. Schippers and J. Koscholke, 2020, “A general framework for probabilistic measures of coherence”, *Studia Logica*, **108(3)**: 395–424.

[15] F. Taroni, P. Garbolino, S. Bozza and C. Aitken, 2021, “The Bayes’ factor: The coherent measure for hypothesis confirmation”, *Law, Probability and Risk*, **21(1)**: 15–36.

[16] M. G. Titelbaum, 2022, *Fundamentals of Bayesian Epistemology*, Oxford: Oxford University Press.

[17] 顿新国,“确证逻辑的研究困境及其解决原则”,江西社会科学,2018 年第 12 期,第 5–11 页。

（责任编辑：映之）

Evaluation and Selection of Probabilistic Measures of Confirmation

Wenjing Du

Abstract

The Bayesian approach advocates employing the terminology of probability theory to characterize the degree of confirmation that evidence provides for a hypothesis. To quantify this degree of confirmation, scholars have developed a range of probabilistic measures designed to quantify this degree of confirmation. This paper seeks to evaluate the rationality of these measures from two key perspectives: typical cases and adequacy conditions. Through this dual lens, the paper aims to eliminate measures that lack rational grounding, while highlighting those that exhibit clear comparative advantages. Moreover, this paper endeavors to summarize the guiding principles for selecting confirmation measures. These principles emphasize that, in specific practical applications, the choice of an appropriate measure should align with the particular investigative context, while also providing a detailed explanation of its advantages over alternative measures. This approach aims to offer robust theoretical support and practical guidance for scientific decision-making.

Wenjing Du　Wenbo College, East China University of Political Science and Law
2178@ecupl.edu.cn