基于语义信息论的证伪和确证公式

鲁晨光 lcguang@foxmail.com

摘 要：在证伪和证实之间的长期争论之后，Carnap 和 Popper 达成一致结论：对于不确定全称假设，确证是必要的. Popper 肯定：一个假设的逻辑概率越小（在逻辑上越容易被证伪），则语义信息量越大，如果它经得起经验的检验. Carnap 和 Bar-Hillel 提出的语义信息公式 *I*=log(1/逻辑概率)能反映逻辑概率越小信息量越大，但是不能反映经验检验或证伪. Carnap 和 Popper 都提出过确证公式，但是没有被普遍接受. 现在确证公式有很多，没有一个被普遍接受. 基于语义信息研究，作者提出新的语义信息公式 *I*=log(真值/逻辑概率), 它就能反映 Popper 的证伪思想，因为真值越大就表示假设更能经得起检验. 对于一个不太确定的全称假设，我们可以把它看成是一个全称假设和一个永真句的线性组合，永真句的比例就是不信度 *b*’。优化 *b*’使平均语义信息量达最大，就得到优化的不信度或否证度 *b*’*. 确证度 *b**可以通过公式 *b*’*=1-|*b**|得到. *b**具有 HS 对称性——具有 *b**的假设等价于其否定假设具有-*b**. 和流行的多种确证度相比，*b**的不同之处是：1)和统计学中置信水平及检验中的似然比兼容; 2)能用于概率预测，反映预测质量; 3)能克服乌鸦悖论; 4)它肯定较小的反例比例比较大的正例比例更重要，因而和 Popper 的证伪思想兼容.

关键词：证伪, 归纳逻辑，确证度，语义信息，乌鸦悖论，医学检验，置信水平.

中国图书分类号：O21, O23

Formulas for Falsification and Confirmation Based on A Semantic Information Theory

LU Chenguang[1]

(Intelligence Engneering and Mathematics Institute, Liaoning Technical University, 123000)

- **Abstract:** After long arguments between positivism and falsificationism, Carnap and Popper reached a consistent conclusion: for uncertainly universal hypotheses, confirmation is necessary. Popper affirms that the less the logical probability of a hypothesis is (the easier it is falsifiable), the more the semantic information it conveys is if it can survive experimental tests. Carnap and Bar-Hillel proposed a semantic information formula *I*=log(1/logical_probability), which can ensure that the less the logical probability is, the more the information is. However, it cannot express experimental tests or falsification. Both Carnap and Popper proposed confirmation measures, which are not widely accepted. Now there have been many confirmation measures, none of which is widely accepted. Based on studies on the semantic information theory, the author proposed an improved semantic information formula *I*=log(true_value/logical_probability), which can reflect Popper’s falsification thought because the larger true value means that the hypothesis can survive test better. An uncertainly universal hypothesis can be regarded as the linear combination of a universal hypothesis and a tautology. The ratio of the tautology is the degree of disbelief. Optimizing *b*’ for maximum average semantic information, we can obtain the optimized degree of disbelief or the disconfirmation measure *b*’*. Then we can obtain the conformation measure *b** using *b*’*=1-|*b**|. This conformation measure *b** has the symmetry: hypothesis *h* with *b** is equivalent to its negation *h*’ with -*b**. In comparison with the popular confirmation measures, *b** has distinctions：1) *b** is compatible with the confidence level in statistics and the likelihood ratio in medical tests; 2) it can be used for probability predictions and reflect predictive quality; 3) it can be used to clarify the Raven Paradox；4)it emphasizes that the less counterexample ratio is more important than the larger positive example ratio, and hence is compatible with Popper's falsification thought.

Keyword：falsification, inductive logic, degree of confirmation, semantic information, Raven Paradox, medical tests, confidence level.

1. 引言

证伪主义和逻辑经验主义之争的结果是，Carnap(逻辑经验主义的代表)放弃全称假设的证实，转向不确定全称假设的确证[1-4]；同时波普尔也承认确证不确定全称假设是重要的. 于是归纳问题现在转变为确证问题，或确证度计算问题. 两位大师都为语义信息公式(用于证伪)和确证公式做了很多努力. 相比之下，卡尔纳普的公式更加出名. 虽然Shannon信息论[5]在工程上取得极大成功，但是它不能度量语义信息，主要是因为它不考虑假设或预言的真假对错，从而不使用逻辑概率.卡尔纳普曾用条件逻辑概率定义确证度([4] ，p98): $c(e,h)=p(h,e)/p(e)$ (其中 h 是假设，e 是证据，p 是逻辑概率)；后来他用相关性定义确证度[6] ：$c(e,h)=P(h,e)-P(e)P(h)$. 卡尔纳普和Bar-Hillel还提出语义信息公式[7] $I(p)=\log(1/m_p)$. 其中 p 是一假设，m_p 是假设的逻辑概率. 波普尔早就提出([8]，p96)：逻辑概率越小， 信息量越大，如果假设经得起经验的检验. Carnap和Bar-Hillel的信息公式反映了波普尔的思想——逻辑概率越小，信息量越大；但是没有反映波普尔的假设检验思想——说对了信息量才大. 虽然波普尔没有提出语义信息公式，但是Popper提出检验严厉性公式([9], p526)：$s=P(e|h,b)/P(e|b)$ 和 $s=\log P(e|h,b)/P(e|b)$(其中 b 是背景知识. 注意：条件概率 $P(e|h,b)$在原文中写法是 $P(e,hb)$). 因为 b 总是有的，可以忽略. 忽略之后这个公式就变成Shannon互信息公式的对数部分. 这个公式的不足是: $P(e)$和 $P(e|h)$看来是事件的统计概率而不是逻辑概率. Popper也提出确证度公式([9] , p388), 因为比较复杂且不容易理解，所以影响很小.

现在确证度公式已有很多. Tentori等人于2007发表的文章［10］包括了几个比较著名的确证度公式(其中 e 是正例，e_0 是反例，h_0 是 h 的否定)：

$d(e,h)=P(h|e)-P(h)$ (Eells, 1982; Jeffrey,1992)

$r(e,h)=\log[P(h|e)/P(h)]$ (Keynes, 1921; Horwich, 1982)

$n(e,h)=P(e|h)-P(e|h_0)$ (Nozik,1981)

$l(e,h)=\log[P(e|h)/P(e|h_0)]$ (Good, 1984)

$c(e,h)=P(h,e)-P(e)P(h)$ (Carnap,1962)

$k(e,h)=[P(e|h)-P(e|h_0)]/[P(e|h)+P(e|h_0)]$ (Kemenv and Oppenheim,1952)

$s(e,h)=P(h|e)-P(h|e_0)$ (Christensen,1999)

但是至今没有一个公式是公认为合理的；争论还在继续. 然而，有一个比较公认的结论是：确证度就是归纳支持程度［1］，应在-1和1之间变化. 据此，不相关的证据带来的确证度是0，支持假设的证据带来正的确证度，支持其否定假设的证据带来负的确证度. Ellery 和 Fitelson［11］因此于2000年提出：确证度公式要能保证假设对称性(Hypothesis Symmetry, 简称HS), 若假设 h 的确证度是-c, 则 h_0 的确证度就是正 c.比如：h="所有乌鸦是白的"，其确证度 c=-1；则它等价于说"所有乌鸦不是白的"，其确证度是 1.可以说检验一个确证度好坏，HS是标准之一. 上面公式中，$n(e,h)$, $k(e,h)$，和 $s(e,h)$具有HS.

笔者认为，统计理论中的置信水平(confidence level [12]，缩写为CL)和确证度含义类似，反映假设在多大程度上可信. 公式是

$$CL=\text{反例比例}/(\text{正例比例}+\text{反例比例})=P(h|e)/[P(h|e)+P(h|e_0)].$$

CL 在0和1之间变化. 确证度是否随置信水平单调增加，这也应该是检验确证度的标准之一.

另外，根据确证度高的假设预测事件发生的确定性应该更高(不确定性或后验熵更小)，这也应该是检验确证度的一个标准. *CL* 就能反映概率预测的确定性或预测质量. 比如 *CL*=1 时，反例比例是 0，正例比例是 1，因而预测质量最好；*CL*=0.5 时，正反例比例一样，对预测没有任何帮助. *CL*=0 时，正例比例是 0，反例比例是 1. 这是预测反了，质量最差.

Hempel 提出著名的乌鸦悖论[13]. 按照数理逻辑，假设 h_1=“若 s_1 则 s_2”和 h_2=“非 s_2 则非 s_1”等价，即“所有乌鸦是黑的”和“不是黑的就不是乌鸦”等价，所以支持 h_2 的证据也支持 h_1. 这是结论 I. 但是，根据结论 I，一根白粉笔支持“不是黑的就不是乌鸦”，从而也支持“所有乌鸦是黑的”. 而根据常识，白粉笔和“所有乌鸦是黑的”不相关——这是结论 II. 所以两个结论之间存在悖论. 一个确证度能否克服乌鸦悖论，应该也是好坏标准之一.

所以我们希望有一个确证度， 1)具有 HS 对称性；2)和置信水平 CL 兼容；3)能反映概率预测质量；4)能克服乌鸦悖论.

在 Carnap 和 Bar-Hillel 之后，也有其他人提出语义信息公式，比如钟义信的语义信息公式[14]和 Floridi 的语义信息公式[15,16]. 但是这些公式都没能很好反映假设检验——即事实和假设不符时信息如何减少. 比如假设 h_1=“张三大约 20 岁”或“张三是年轻人”， 张三是实际年龄是 15 岁， 20 岁，25 岁，30 岁，或 60 岁时，信息量应该不同. 张三真的 20 岁时，h_1 提供的信息量应该最大，其次是 25 岁,15 岁. 张三 60 岁时，h_1 提供的信息应该是负的. 所以我们希望有一个语义信息公式使得信息量在逻辑概率减小时增大，也能在偏差增大时减小.

本文作者曾提出广义信息理论[17-19] ——即推广 Shannon 信息论得到的语义信息理论. 现在发现所用语义信息公式可以用于证伪和确证，并且使得两者兼容. 下一节介绍语义信息方法和用于解释证伪的语义信息公式，第三节介绍新的确证公式. 第四节是结论和讨论.

2. 兼容 Popper 证伪思想的语义信息公式

2.1 数学方法

新的数学方法严格区分并同时使用两种概率：统计概率和逻辑概率. 首先定义：

- e: 一个证据点或实例；E 是取值 e 的随机变量，e 或 e_i 是 U={e_1,e_2,…}中的一个. e 发生的频率或频率的极限是统计概率 $P(e)$.
- h: 一个假设或标签； H 是取值 h 的随机变量，h 或 h_j 是 V={h_0, h_1,…}中的一个.
- $P(h_j)$: h_j 被选择的概率，即频率或频率的极限；$P(h_j|e)$ (h_j 不变，e 变)是 h 的转移概率函数，其最大值通常小于 1.
- $T(h)$:h 被判断为真的概率. 它和 $P(h)$的区别比如：设 T_{au} 是永真句，$P(T_{au})$一般说来是 0，因为 T_{au} 极少被选择；而 $T(T_{au})$和 T_{au} 是否被选择无关，总有 $T(T_{au})$=1.
- $P(e|\theta_j)$=$P(e|h_j$ 是真的)是似然函数, θ_j 是 U 中使 h_j 为真的 e 构成的模糊集合.
- $T(h_j|e)$=$T(\theta_j|e)$是 h_j 的模糊真值函数（后面简称为真值函数），取值于[0,1]. 如果它取值于{0,1}, $T(h_j|e)$就是经典的真值函数或集合特征函数. 其最大值通常是 1.

笔者把贝叶斯定理细分称三种[20]. 第一种贝叶斯定理是贝叶斯和拉普拉斯提出的. 设集合 $A, B\in 2^U$，B'是 B 的补集. T(A)=$P(x\in A)$, $T(B)$等同理. 则该定理可以写成公式：

$$T(B|A)=T(A|B)T(B)/T(A),\ T(A)= T(A|B)T(B)+ T(A|B')T(B'), \tag{1}$$

其中 B’是 B 的补集. 注意，其中涉及一个随机变量和一个论域。

第二种贝叶斯定理是 Shannon [8] 和统计学者们使用的：

$$P(e|h_j)=P(e)P(h_j|e)/P(h_j),\ P(h_j)=\sum_i P(e_i)P(h_j|e_i)\,. \tag{2}$$

类似地，也可以对称地求出 $P(h_j|e)$. 注意：其中涉及两个随机变量和两个论域.

我们把贝叶斯定理推广到一个统计概率和一个逻辑概率之间(涉及两个随机变量和两个论域)，得到连接似然函数和真值函数的第三种贝叶斯定理：

$$P(e|\theta_j)=P(e)T(\theta_j|e)/T(\theta_j),\ T(\theta_j)=\sum_i P(e_i)T(\theta_j|e_i)\,, \tag{3}$$

$$T(\theta_j|e)=T(\theta_j)P(e|\theta_j)/P(e)\,. \tag{4}$$

假定 $T(\theta_j|e)$的最大值是 1， 可以推导出

$$T(\theta_j|e)=[P(e|\theta_j)/P(e)]/\max[P(e|\theta_j)/P(e)]\,. \tag{5}$$

其中分母是分子的最大值，可以保证 $T(\theta_j|e)$的最大值是 1.

2.2 改进的语义信息公式

Shannon 互信息公式[8]的对数部分是

$$I(e_i;h_j)=\log\frac{p(e_i|h_j)}{p(e_i)} \tag{6}$$

我们用“h_j 是真的”取代“h_j”， 上面公式就变为

$$I(e_i;\theta_j)=\log\frac{P(e_i|\theta_j)}{P(e_i)}=\log\frac{T(\theta_j|e_i)}{T(\theta_j)}\,. \tag{7}$$

其中利用了第三种贝叶斯定理。假设 h_j=“E 大约是 e_j”的真值函数是没有系数的正态分布(最大值是 1)，则语义信息量随偏差变化如图 1 所示.

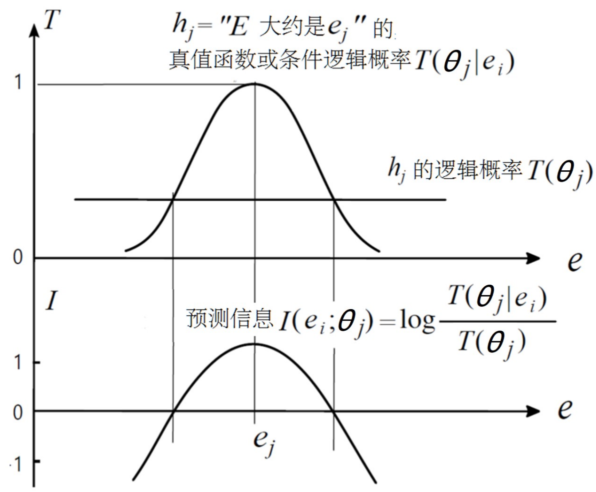

图 1 语义信息量随偏差变化图解

Fig. 1. The amount of semantic information changes with the deviation.

当 $T(\theta_j|e)$总是 1 时，上式就变为 Carnap-Bar-Hillel 语义信息公式. 式(7)能保证：

1） 逻辑概率越小， 信息量越大；

2） 偏差越小， 信息量越大；偏差太大时，信息量是负的(反映证伪)；

3） 永真句和矛盾句(逻辑概率和真值都是 0)的信息量是 0.

可见分子用真值函数 $T(\theta_j|e)$而不是 1 就可以反映波普尔的假设检验和证伪思想.

Popper 把 h 的逼真度定义为 h 的真内容减去其假内容，即 $Vs(h)=Ct_T(h)-Ct_F(h)$. 设 h_j 的真值是预测偏差的函数，是没有系数的高斯分布 $T(\theta_j|x)=\exp[-(x-x_j)_2/(2\sigma^2)]$，则上面公式变为

$$I(e_i;\theta_j)=\log[1/T(\theta_j)]-(x-x_j)^2/(2\sigma^2). \tag{8}$$

其中 $\log[1/T(\theta_j)]$反映真内容；$(x-x_j)_2/(2\sigma^2)$是相对偏差，反映假内容。所以 $I(e_i;h_j)$也可以看做是逼真度。张志林[21]曾指出, 按 Popper 的逼真性定义，并不是任意两个假设的逼真性是可比较的。笔者建议：用 $\log[1/T(\theta_j)]$即 Carnap-Bar-Hillel 的语义信息定义检验严厉性，用 $I(e_i;h_j)$定义逼真度。这样，任意两个假设的逼真度都可以相互比较。

对 $I(e_i;\theta_j)$ 求平均，就得到平均语义信息公式：

$$I(E;\theta_j)=\sum_i P(e_i|h_j)\log\frac{T(\theta_j|e_i)}{T(\theta_j)}. \tag{9}$$

其中 $P(e|h_j)$是样本分布。这个公式可以保证，当集合清晰时，即 $T(\theta_j|e)\in\{0,1\}$时, 如果有一个反例，则平均信息就是负无穷大. 这正反映了波普尔的思想——一个反例就可以证伪一个全称假设. 但是，如果假设是不确定或模糊的，即使有反例存在，适当调整反例的真值也可以得到正的语义信息.

对 $I(E;\theta_j)$求平均就得到语义互信息。可以证明当 $P(e|\theta_j)=P(e|h_j)$ 或 $T(\theta_j|e)\propto P(h_j|e)$，语义互信息达到其最大值——Shannon 互信息。

下面以医学检验为例说明如何优化反例的真值或不信度，从而得到确证度。

3. 确证度推导及检验

3.1 语义信道和不信度优化

信任度(Degree of Belief)有时用来指主观概率，有时用来指我们对关系比如“若 A 则 B”的相信程度. 本文仅指后者，它是主观的. 确证度则是归纳出的或样本支持的信任度[1] ，或优化的信任度，它是客观的. 不信度 b'和信任度 b 的关系是

$$b'=1-|b|. \tag{10}$$

不信度 b'在 0 和 1 之间变化， 而信任度 b 在-1 和 1 之间变化.

常识告诉我们， 对于医学诊断、天气预报、股评，…，我们不能不信，也不能全信. 对于医学诊断，我们的信任度要高点；而对股评家的话信任度要低很多；对于骗子，我们的信任度是负的. 语义信息分析表明，根据预测的正例反例比例适当调整对预测的信任度，可以增加我们接收到的信息. 下面我们以医学检验为例说明检验和二分类(根据观察特征二分类)的 Shannon 信道，语义信道，和不信度.

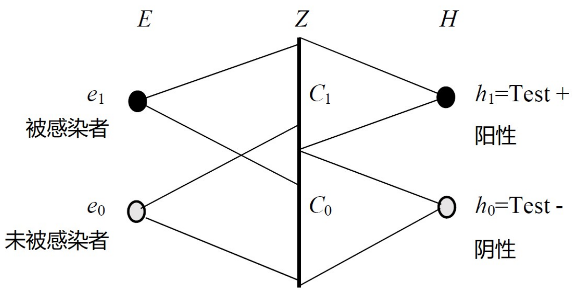

图 2 医学检验图解。检验可以被抽象为 2X2 有噪声 Shannon 信道。Z 是化验数据。

Figure 2. Illustrating the medical test. The test can be abstracted as a 2×2 Shannon nosy channel. Z is a laboratory datum.

医学检验可以看成一个 2X2 有噪声 Shannon 信道(见图 2). 相应实例集合是 $U=\{e_0, e_1\}$，其中 e_0 表示未感染者，e_1 表示感染者；相应标签集合是 $V=\{h_0, h_1\}$；h_0 表示阴性，h_1 表示阳性. 图 2 中 $Z\in C$ 表示观察数据. h_0 和 h_1 是根据 Z 选择的标签或预测.

医学检验中，条件概率 $P(h_1|e_1)$叫做敏感性(sensitivity)，$P(h_0|e_0)$叫做特异性(specificity) [22]. 敏感性和特异性构成 Shannon 信道，如表 1 所示.

表 1. 医学检验的敏感性和特异性构成香农信道 $P(H|E)$

Table 2 Sensitivity and specificity of medical tests form a Shannon's channel $P(H|E)$

	真涨或真有病 e_1	真跌或真没病 e_0
$P(h_1\|E)$	敏感性=$P(h_1\|e_1)$	1-特异性=1-$P(h_0\|e_0)$=$P(h_1\|e_0)$
$P(h_0\|E)$	1-敏感性=1-$P(h_1\|e_1)$=$P(h_0\|e_1)$	特异性=$P(h_0\|e_0)$

设检验绝对可信时，四个真值是：$T(A_1|e_1)=T(A_0|e_0)=1$, $T(A_1|e_0)=T(A_0|e_1)=0$ (A_0 和 A_1 表示清晰集合). 如果可能有错，集合变为模糊集合，优化的真值函数应在 0 和 1 之间. 设一个假设 h_j 分为可信部分和不可信部分，可信比例是 b_j， 则其真值函数是

$$T(\theta_j|E)= b_j' +b_jT(A_j|E), \quad j=0,1. \tag{11}$$

其中 $b_j'=1-|b_j'|$是 h_j 中含有永真句(无意义)的比例，也是反例的真值，可谓不信度. 我们称一组真值函数为一个语义信道，则相应的语义信道如表 2 所示.

表 2 两种不信度构成医学检验的语义信道

Table 2. Two no-confidence levels of a medical test form a semantic channel

h	**感染者 e_1**	**未感染者 e_0**
阳性 h_1	$T(\theta_1\|e_1)=1$	$T(\theta_1\|e_0)=b_1'$
阴性 h_0	$T(\theta_0\|e_1)=b_0'$	$T(\theta_0\|e_0)=1$

我们用 $h_1^{b1'}$表示假设“检验阳性，其不信度是 b_1'”，把第一行数据代入公式(9)得到

$$I(E;h_1^{b_1'})=P(e_0|h_1)\log\frac{b_1'}{b_1'P(e_0)+P(e_1)}+P(e_1|h_1)\log\frac{1}{b_1'P(e_0)+P(e_1)}. \tag{12}$$

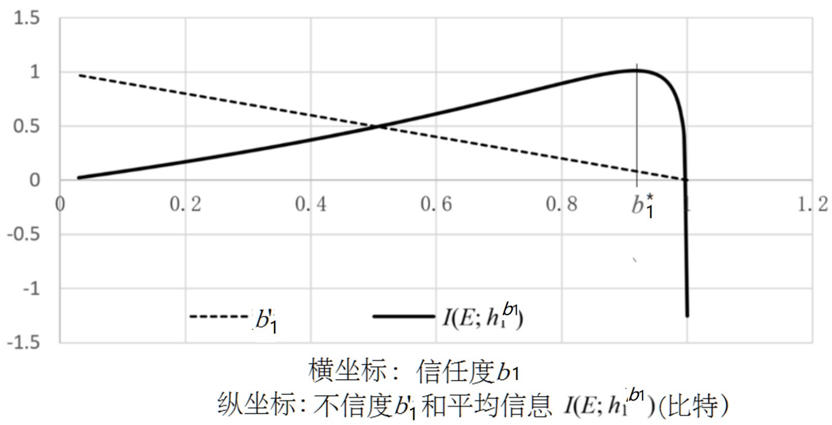

图 3. 平均信息 $I(E; h_1^{b1'})$随不信度 b_1'变化(假定 $P(e_1)$=0.2; $P(e_1|h_1)$=0.75）

Figure 3. Average semantic information $I(E; h_1^{b1'})$ changes with b_1', where $P(e_1)$=0.2; $P(e_1|h_1)$=0.75.

通过改变 b_1'最大化 $I(E; h_1^{b1'})$，即令 $\mathrm{d}I/\mathrm{d}b_1'=0$，我们得到优化的不信度

$$b_1'^* = P(h_1|e_0)/P(h_1|e_1); \tag{13}$$

同理可得

$$b_0'^* = P(h_0|e_1)/P(h_0|e_0). \tag{14}$$

这时候语义信息量等于 Kullback-Leibler 信息量[23]．我们也可以通过 $P(e_1|\theta_1)=P(e_1|h_1)$得到上面结论．令

$$\frac{P(e_1)}{b_1' P(e_0)+P(e_1)} = \frac{P(e_1)P(h_1|e_1)}{P(e_1)P(h_1|e_1)+P(e_0)P(h_1|e_0)}, \tag{15}$$

求解上式中的 b_1'就得到公式（13）中的 $b_1'^*$.

当 $P(h_1|e_1) \geqslant P(h_1|e_0)$时，从公式(10)和(13)，我们得到

$$b_1^* = [P(h_1|e_1)-P(h_1|e_0)]/ P(h_1|e_1). \tag{16}$$

如果反例比例超过正例比例，即 $P(h_1|e_1)<P(h_1|e_0)$时，我们得到

$$b_1^* = [P(h_1|e_1)-P(h_1|e_0)]/ P(h_1|e_0). \tag{17}$$

综合上面两个公式， 我们得到

$$b_1^* = \frac{P(h_1|e_1)-P(h_1|e_0)}{\max[P(h_1|e_1),P(h_1|e_0)]} = \frac{CL_1-CL_1'}{\max[CL_1,CL_1']} = \frac{\text{正例比例-反例比例}}{\text{上面两个较大者}}. \tag{18}$$

其中 max[]表示其中两个数较大者；CL_1 是 h_1 的置信水平并且 $CL_1= P(h_1|e_1)/[P(h_1|e_1)+P(h_1|e_0)]$；$CL_1'=1-CL_1$.

3.2 检验新的确证度

现在我们用前面四个标准检查新的确证度 b_1^*．检查 HS 就是证明把式(18)中的 h_1 改成 h_0 时，确证度就改变符号．下式是证明．用 h_0 取代 h_1，则正例变成反例，反例变成正例，于是得到：

$$b_1^* = \frac{P(h_0|e_1)-P(h_0|e_0)}{\text{上两个较大者}} = -\frac{P(h_0|e_0)-P(h_0|e_1)}{\text{上两个较大者}} = -b_0^*. \tag{19}$$

我们再看 b^*(b_1^*或 b_0^*)和 CL(CL_1 或 CL_0)的关系．图 4 显示了 b^*随 CL 的变化.

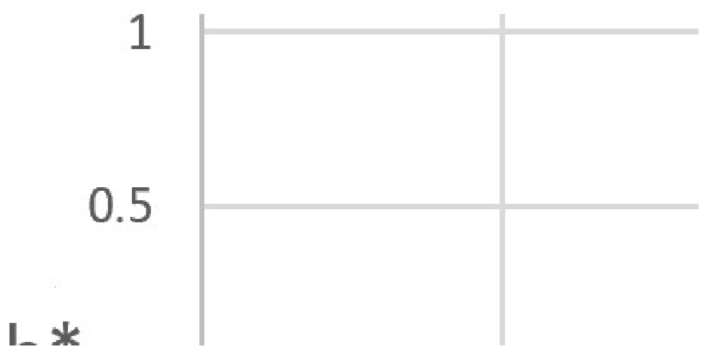

图 4 确证度 b^*随置信水平的变化.
Figure 4. The degree of confirmation b^* changes with confidence level CL.

可见 b^*能和 CL 一样反映检验的好坏程度. 不同的是 b^*从-1 到 1 变化，能更明确反映证据支持假设 h 还是它的否定.

医学检验中用似然比[22] $LR_1= P(h_1|e_1)/P(h_1|e_0)$表示检验阳性的可靠性. 它和否证度 $b_1'^*$之间有简单关系(b_1^*大于 0 时)：

$$LR_1= P(h_1|e_1)/P(h_1|e_0)=1/b_1'^*=1/(1-b_1^*). \tag{20}$$

可见确证度和似然比也有一一对应关系.

我们再看 b_1^*或 $b_1'^*$如何用于概率预测.

用转移概率函数 $P(h_1|e)$和 $P(e)$预测 e_1 发生的概率 $P(e_1|h_1)$如公式(15)右边所示， 用优化的真值函数 $T^*(\theta_1|e)$和 $P(e)$预测 e_1 发生的概率如公式(15)左边所示. 两者是等价的. 可见真值函数更容易理解和记忆.

比如对于某种 HIV 病毒检验[1]，敏感性是 0.917，特异性是 0.999. 当 $P(e_1)$分别是 0.0001, 0.002 (普通人群)和 0.1 (高危人群)时，根据阳性结果预测感染 HIV 的概率 $P(e_1|\theta_1)=P(e_1|h_1)$ 分别是 0.084, 0.65 和 0.99.

值得注意的是，阳性 h_1 的确证度 b_1^*或置信水平 CL_1 的大小主要取决于阴性的正确率即特异性 $P(h_0|e_0)$的大小(大则反例少)，而阴性 h_0 的确证度 b_0^*或置信水平 CL_0 的大小主要取决于阳性的正确率即敏感性 $P(h_1|e_1)$的大小. 医学界很早就知道置信水平的这一性质. 确证度 b^*与置信水平 CL 类似.

但是流行的确证度公式[10]都和医学检验的置信水平不兼容或不太兼容. 举个例子，阳性的正确率即敏感性是 0.1，阴性的正确率即特异性是 1. 这意味着，虽然阳性的正确率很小，但是没有反例。这时确证度 b_1^*和 CL_1 都是 1；预测感染的概率 $P(e_1|\theta_1)$也是 1. 但是，按流行的确证度公式(其中 h 是这里的 h_1, e 是这里的 e_1)算出的确证度都很小，因为这些公式更加重视阳性的正确率即 $P(h_1|e_1)$. 而新的确证度公式更加重视阴性的正确率，即认为较小的反例比例比较大的正例比例更重要. 根据阳性预测有病是百分之百正确时，说阳性的确证度很小是不合理的.

再举个例子，阳性正确率 $P(h_1|e_1)$是 0.8， 阴性正确率 $P(y_0|x_0)$只有 0.1，从而反例的比例 $P(h_1|x_0)$是 0.9. 则阳性的正例和反例的比例是 0.8:0.9. 按流行的公式，只有 $s(e,h_1)$和 b_1^*相近——$s(e, h_1)$=0.8-0.9=-0.1 和 b_1^*=(0.8-0.9)/0.9=-0.125 接近, 其他确证度公式大多没有用到反例比例，算出的确证度都比较大. 在反例比例比正例比例还大的情况下，忽视反例比例，说确证度大于 0 是不合理的.

3.3 用新的确证度公式澄清乌鸦悖论

关于乌鸦悖论[2，13] 的两个相互矛盾的结论是：

结论 I：s_1=“是乌鸦就一定是黑的”和 s_2=“非黑的东西就不是乌鸦”等价；

结论 II: 一个非白非乌鸦物体，比如白粉笔，支持 h_2 但是和 h_1 无关.

很多人为了克服乌鸦悖论，肯定两个结论中的一个，否定另一个. 但是下面我们证明，对于不确定假设，这两个结论都不成立.

虽然 s_1 和 s_2 在经典逻辑中等价，但是在归纳逻辑中并不等价. 我们用 n_{ji} 表示 h_j 和 e_i 同时发生的次数. 当 $n=n_{00}+n_{10}+n_{01}+n_{11}$ 很大时，n_{ji}/n 就是联合概率 $P(h_j, e_i)$，j, i=0, 1. 虽然 s_1

[1] http://www.oraquick.com/taking-the-test/understanding-your-results

和 s_2 的反例个数同是 n_{10}，但是 s_1 和 s_2 正例个数 n_{11} 和 n_{00} 并不相同. 两者反例的比例和正例的比例也不同，所以两者的确证度也不同.

相应 s_1 的正例比例是= $P(h_1|e_1)=n_{11}/(n_{11}+n_{01})$, 反例比例是 $P(h_1|e0)= P(h_1|e_0)=n_{10}/(n_{00}+ n_{10})$. s_1 的确证度是

$$b_1^* = 1 - P(h_1/e_0)/P(h_1/e_1) = 1 - \frac{n_{10}}{n_{00}+n_{10}} \bigg/ \frac{n_{11}}{n_{01}+n_{11}}. \tag{21}$$

同理可得 s_2 的确证度：

$$b_2^* = 1 - P(e_0/h_1)/P(e_0/h_0) = 1 - \frac{n_{10}}{n_{11}+n_{10}} \bigg/ \frac{n_{00}}{n_{00}+n_{01}}. \tag{22}$$

我们再看结论 II. 当 n 极大从而 n_{00} 和 n_{10} 也极大时，n_{00} 增加 1 对 s_1 的反例比例 $P(h_1|e_0)$ 影响很小，所以可以认为一个白粉笔和 b_1*无关. 但是它对 b_2*影响也很小. 因为当 n_{00} 很大时，n_{00} 增加 1，对 b_2*影响也很小.

当 n 不大且有限时，比如 e 只能是鸟，并且数目有限，则一个非黑非乌鸦的鸟可以增大特异性比例 $P(h_0|e_0)$，减少反例比例 $P(h_1|e_0)$，从而支持 s_1. 它也支持 s_2. 但是在 n_{10} 小而 n_{01} 大时，n_{00} 增大对 b_1*影响更大.

所以，结论 II 在 n 不大是也是错的. 因为相互矛盾的两个结论都不成立，所以该悖论不存在.

4. 结论和讨论

新的语义信息公式可以写成 I=log(真值/逻辑概率), 它反映了 Popper 的证伪思想，应该就是 Popper 想要的检验严厉性公式或逼真度公式，它能保证逻辑概率越小(在理论上容易被证伪)且真值越大(经得起检验)，则信息量越大.

新的确证度 b*具有 HS 对称性，和统计学中置信水平及似然比兼容，能反映预测质量，且能克服乌鸦悖论. Popper 的证伪思想通俗说来就是：对于全称假设，正例再多对假设影响也不大；有没有反例更加重要. 新的确证度公式最重要的特点就是强调：较小的反例比较大的正例更重要. 所以，新的确证度公式和 Popper 的证伪思想兼容. 证伪和证实(或归纳)历来对立，新的确证度公式可以使两者兼容，这应该是比较理想的结果.

在流行的方法中，统计概率和逻辑概率没有很好区分，更没有同时使用. 本文严格区分并同时使用两者. 度量语义信息应该用逻辑概率(放在 log 右边)，度量确证度应该用统计概率(虽然推导过程也用到逻辑概率)——因为确证度一般取决于大量数据的统计结果. 虽然相应一个例子(e_i, h_j)存在一个真值或条件逻辑概率 $T(\theta_j|e_i)$，但是它取决于语言用法，并不表示 e_i 确证 h_j 的确证度. 我们讨论的确证是对两个假设的不对称的相关性的确证，确证度和数学中的相关(correlation)系数类似，可以有负值. 只是数学中的相关系数是对称的，而确证度一般是不对称的(“若 A 则 B”的确证度和“若 B 则 A 的确证度”不同).

逻辑方法和统计方法一直未能很好统一，笔者以为主要是因为有人坚持概率的统计解释，有人坚持概率的逻辑解释或主观解释. 流行于统计学习的贝叶斯主义只强调主观解释，几乎放弃了逻辑解释. 笔者认为，逻辑概率和统计概率不应是概率的两种互斥解释，而应是并存的两种概率. 要想统一两者，必须先严格区分两者. 虽然两者不同，但是我们可以通过 $P(e|\theta_j)=P(e|h_j)$ 或 $T(\theta_j|e)\propto P(h_j|e)$在统计和逻辑之间架起桥梁. 这一桥梁已经用于统计学习，并且解决了不少难题[20]. 新的语义信息方法不仅有理论意义，也有实践意义. 它在不确定推理中的应用有待进一步研究.

参考文献:

[1] Hawthorne J. Inductive Logic. In Edward N. Z. (ed.), Stanford Encyclopedia of Philosophy. 2004/2012, http://plato.stanford.edu/entries/logic-inductive/ (accessed on May 24,2019)

[2] 顿新国. 归纳悖论研究[M]，北京：人民出版社，2012.

[3] 任晓明,陈晓平等. 决策、博弈与认知——归纳落实的理论与应用[M]. 北京:北京师范大学出版集团，2014.

[4] 熊立文.现代归纳逻辑的发展[M] ，人民出版社，2004.

[5] Shannon, C. E. A mathematical theory of communication [J]. Bell System Technical Journal, 1948, 27:379–429, 623–656.

[6] Carnap, R. Logical Foundations of Probability [M], the University of Chicago Press, Chigago,1962.

[7] Carnap, R. Bar-Hillel, Y.: An Outline of a Theory of Semantic Information[R]. *Tech. Rep.* 1952, No. 247, Research Lab. of Electronics, MIT.

[8] Popper，K.科学发现的逻辑[M]，1935(德文)，2008， 中国美术学院出版社.

[9] Popper K. Conjectures and Refutations[M]. London and New York: Routledge，1963/2002.

[10] Tentori K, Crupi V, Bonini N, Osherson D. Comparison of Confirmation Measures [J]. *Cognition* 2007,103:107-119

[11] Ellery E, Fitelson B. Measuring confirmation and evidence [J]. *Journal of Philosophy* 2000, 97:663–672.

[12] Neyman, J. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society A*. 1937, 236 (767): 333–380.

[13] Hempel, C. G. Studies in the Logic of Confirmation [J]. *Mind*, 1945,54:1–26 and 97–121.

[14] Zhong, Y.X. A theory of semantic information [J]. *China Communications* 2017, 14:1-17.

[15] Floridi, L. Outline of a theory of strongly semantic information [J]. *Minds and Machines*, 2004,14:197-221.

[16] Floridi L. Semantic conceptions of information[J]. in *Stanford Encyclopedia of Philosophy*, Edward N Zalta (ed.), 2005/2015, http://plato.stanford.edu/entries/information-semantic/

[17] 鲁晨光，广义信息论[M].合肥：中国科技大学出版社,1993.

[18] 鲁晨光， 广义熵和广义互信息的编码意义[J]，通信学报，1994, 15, 37-44.

[19] Lu C. (鲁晨光). A generalization of Shannon's information theory[J]. Int. J. of General Systems, 1999，28 (6): 453-490.

[20] Lu, C(鲁晨光). From Bayesian inference to logical Bayesian inference: A new mathematical frame for semantic communication and machine learning. In: *Intelligence Science II*. Proceedings of ICIS2018, pp.11-23; Beijing, China, 2 Oct. 2018; Shi, Z.Z. Ed. Springer International Publishing: Switzerland, 2018.

[21] 张志林，真理、逼真性和实在论，自然辩证法[J]，1993, 87(5),1-8.

[22] Thornbury J R. Fryback, D.G.; Edwards, W. Likelihood ratios as a measure of the diagnostic usefulness of excretory urogram information [J]. *Radiology* 1975, 114, 561–565.

[23] Kullback S, Leibler R. On information and sufficiency[J]. Annals of Mathematical Statistics. 1951，22:79–86.