兼容波普尔证伪思想的语义信息测度和确证度

——涉及逻辑概率，归纳推理，证伪，确证，乌鸦悖论，语义信息论，统计学习

$$\text{语义信息量}=\log\frac{\text{命题真值}}{\text{谓词逻辑概率}} \qquad \text{确证度}=\frac{\text{正例比例}-\text{反例比例}}{\text{上面两个较大者}}$$

鲁晨光

个人主页: http://survivor99.com

复旦哲学系“科学+哲学”系列讲座第五期2019.11.12

主持人：张志林

自我介绍

- 1955年生于安徽和县。当过插队知青、马钢工人，毕业于南京航空航天大学(77级). 教过计算机（长沙大学副教授），当过长沙市青年学术带头人，赴加拿大访问学者，北师大数学系访问学者（汪培庄教授指导）。目前退休，兼做辽宁工程技术大学智能工程和数学学院客座教授。
- 研究兴趣是：语义信息论，统计学习，色觉的数学和哲学, 美学和进化论，投资组合。专著有：《广义信息论》，《投资组合的熵理论和信息价值》，《色觉奥妙和哲学基本问题》，《美感奥妙和需求进化》。于《光学学报》,《通信学报》,《现代哲学》,《Int. J. of General System》，《Information》等期刊上发表论文多篇。最近主要工作是把语义信息论用到统计学习和归纳推理。

本讲座背后的故事I

第十五卷 总 87 期 自 然 辩 证 法 通 讯 Vol. 15, Sum № 87
1 9 9 3 第 5 期 JOURNAL OF DIALECTICS OF NATURE № 5, 1993

真理、逼真性和实在论*

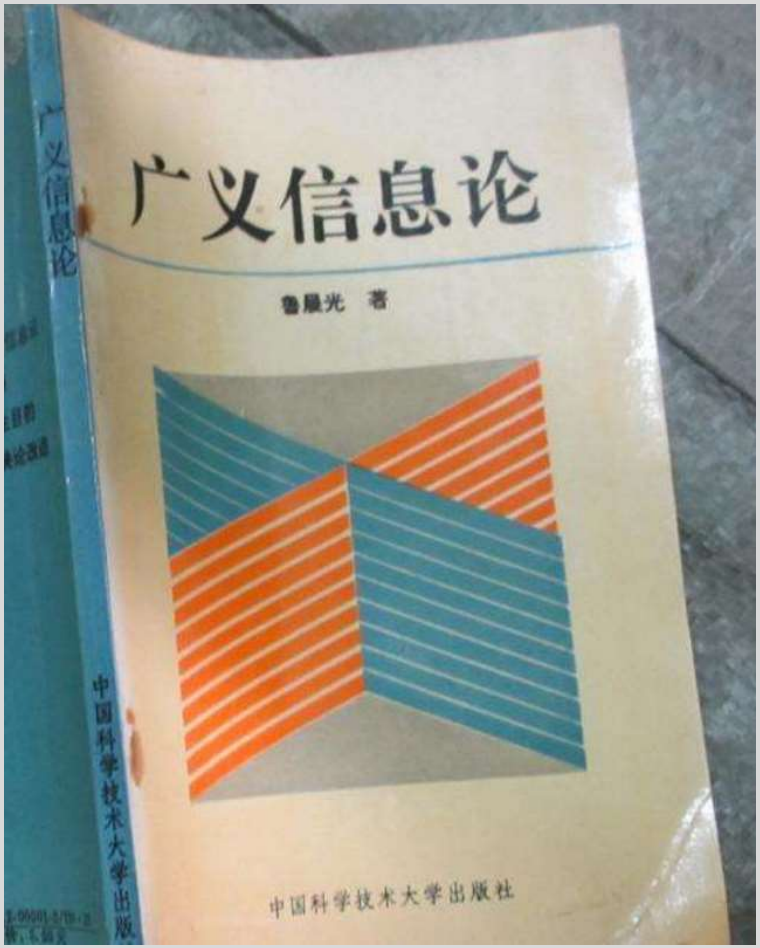
张 志 林

本文有选择地分析了一些科学哲学家对真理、逼真性和实在论的探讨。分析结果表明，他们的探讨存在着逻辑上的循环、理论上的不严谨或应用上的困难。因此，本文基本上是批判性的。至于怎样提出建设性的主张，乃是笔者希望今后从事的一个研究课题。

- 1993年，这篇文章让我认识了张志林教授。其中讨论了Popper等人用逼真度解释科学进步和实在论，但是怎样解释假设和事实符合，解释一个理论比另一个理论逼真？该文认为：已有理论在逻辑上还存在问题，有待改进。

本讲座背后的故事2

- 也是1993年，我发表了专著《广义信息论》——在北师大做访问学者期间写的，。
- 看了张志林的文章，觉得文章非常好。我想我可以用逼真度解释语义信息公式，同时可以从信息论角度解释模糊假设和事实符合。于是写信给他，不久欣闻他的文章获得金岳霖奖。当时我们都30几岁。
- 我今天报告的主要是证伪和确证的技术问题，关于证伪和归纳的哲学问题，我还要请教张志林教授和其他老师。
  - 有网页版全文：http://survivor99.com/



背景I：证实和证伪之争

- 逻辑经验主义认为：科学假设需要通过经验证实，否则没有意义；
- Popper指出：全称假设的证实是不可能的。科学假设的意义在于在逻辑上容易证伪而实际上经得起检验。
- 一致结论：不完全正确的假设可以通过经验得到确证，即证据可以提高假设的确证度。现代归纳问题变成确证度计算问题。
- Carnap和Popper都提出确证度公式：
- Carnap的确证度公式比较有名。最初用条件逻辑概率P(h|e),后来用P(h,e)-P(h)P(e)表示确证度，在0和1之间变化。
- 但是现在更多人关心大前提if-then语句的确证，相应的确证度在-1和1之间变化。

背景II：语义信息公式

- 香农（1948）提出互信息公式：

$$I(X;Y)=\sum_j P(y_j)\sum_i P(x_i\,|\,y_j)\log\frac{P(x_i|y_j)}{P(x_i)}=\sum_j\sum_i P(x_i)P(y_j\,|\,x_i)\log\frac{P(y_j\,|\,x_i)}{P(y_j)}$$

$$=H(X)-H(X\,|\,Y)=H(Y)-H(Y\,|\,X)$$

据此，假设yj提供关于事件xi的信息是

$$I(x_i;y_j)=\log\frac{P(x_i|y_j)}{P(x_i)}$$

- 卡尔纳普和Bar-Hillel（1951）提出语义信息公式：

$$I(p)=\log(1/mp),$$

这个公式反映了波普尔思想， 逻辑概率越小，语义信息量越大。但是这个公式不能表达对错。说错了信息如何?

- 波普尔（1963）检验严厉性公式：

$$S=\log\frac{P(e_i|h_j,b)}{P(e_i,b)}$$

（上下加了b表示背景）。但是P(ei|hj,b）不是逻辑概率，也不太好反映事实检验。

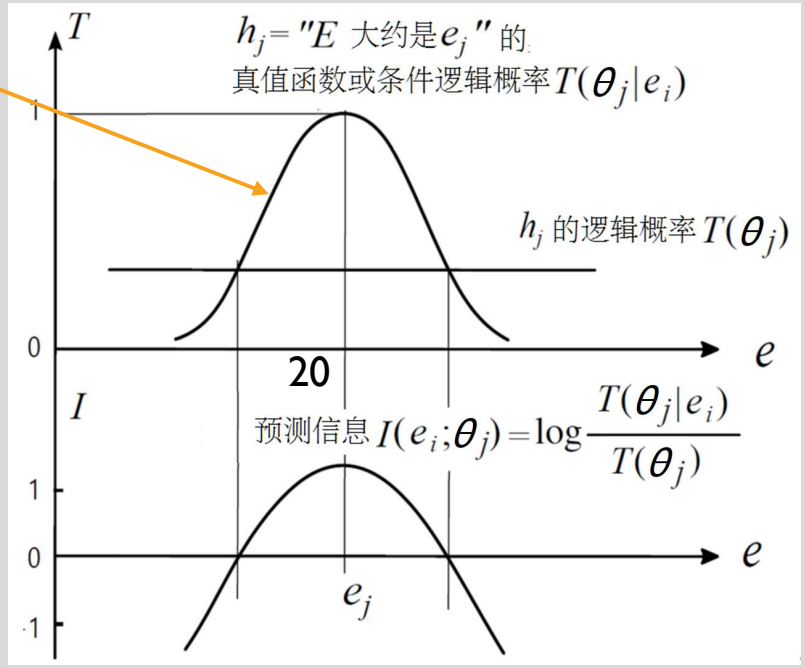
我们需要能反映对错的语义信息公式。

我的语义信息公式

●我的改进：

$$I(e_i;\theta_j)=\log\frac{T(\theta_j|e_i)}{T(\theta_j)}=\log\frac{\text{命题真值}}{\text{谓词逻辑概率}}=\log(1/\text{mp})\text{-相对偏差}$$

T：真值或逻辑概率；θj：是使假设hj为真的E组成的模糊集合， 比如谓词“e大约20岁”，其真值函数如图。



●性质：

1：如果来人ei真的20岁时真值最大为1，偏差越大，真值越小，信息越小甚至是负的。

2：逻辑概率越小，信息的绝对值越大。

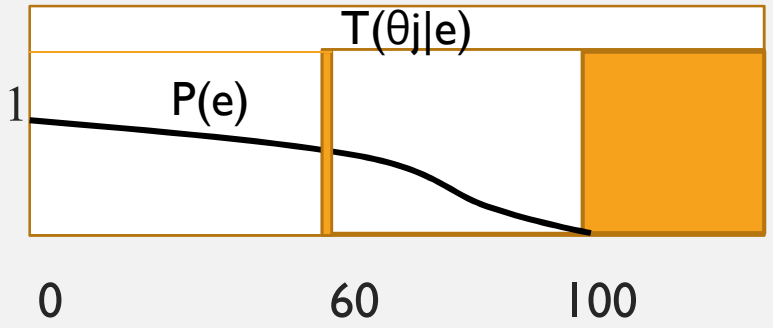
永真句的两条线重合，信息量是log(1/1)=0.

●意义： 该公式反映事实检验，也反映逼真度——不仅正确而且精确（外延小）时，信息量就大。

澄清逻辑概率——和外延及信源P(E)的关系

- **逻辑概率解释**：命题有真值，谓词（即命题函数）有逻辑概率，即谓词被判断为真的概率。
- 假定有一万个有不同年龄的人走过来，让门卫来判断是否“来人e是老年人”。如果有2000人被判定为老年人，那么“老年人”的逻辑概率就是0.2.
- **逻辑概率计算**：不仅取决于外延大小，还取决于事实e的先验概率分布。计算公式是：

$$T(\theta_j) = \sum_i P(e_i) T(\theta_j | e_i)$$



- 其中T(θj|e)是谓词的真值函数或者说隶属函数，反映hj的外延。
- 卡尔纳普等人使用的逻辑概率中，老年人和非老年人逻辑概率一样，是0.5。我们的逻辑概率还取决于P（e），老年人的逻辑概率和群体寿命有关。
- 比如，“百岁以上老人”比“60岁的老人”外延更大，但是逻辑概率更小，因为超过百岁的人很少。

澄清逻辑概率2——区分逻辑概率和选择概率

- 逻辑概率——假设被判断为真的概率。
- 统计概率——即假设被选择的概率。
- 假设一万个有不同年龄的人走过来，有四个标签：
- “小孩”，“年轻人”，“成年人”，“老年人”
- 让门卫分类，即为每个人选一个标签。 如果贴上“老年人”标签的人数是1000个， 则统计概率是0.1. 它小于逻辑概率0.2. 为什么？因为有的老年人被贴上了“成年人”标签。
- 同样， 这样得到的“成年人”的概率比如（0.5）也是统计概率，它小于成年人的逻辑概率。因为老年人也是成年人，成年人的逻辑概率应该等于0.5+0.1=0.6.
- 最极端的例子：一个永真句“来人可能是老人，也可能不是”，其逻辑概率是1，但是它被选择的概率通常是0.
- 流行的确证度理论的一个严重缺陷是：混淆了逻辑概率和统计概率！因为都用P。而我用T表示逻辑概率。

语义信息和概率预测的关系

根据语义做概率预测： 以前的贝叶斯公式中只有概率（逻辑的或统计的不分），我提出：可以把真值函数带进贝叶斯公式做概率预测，似然函数：

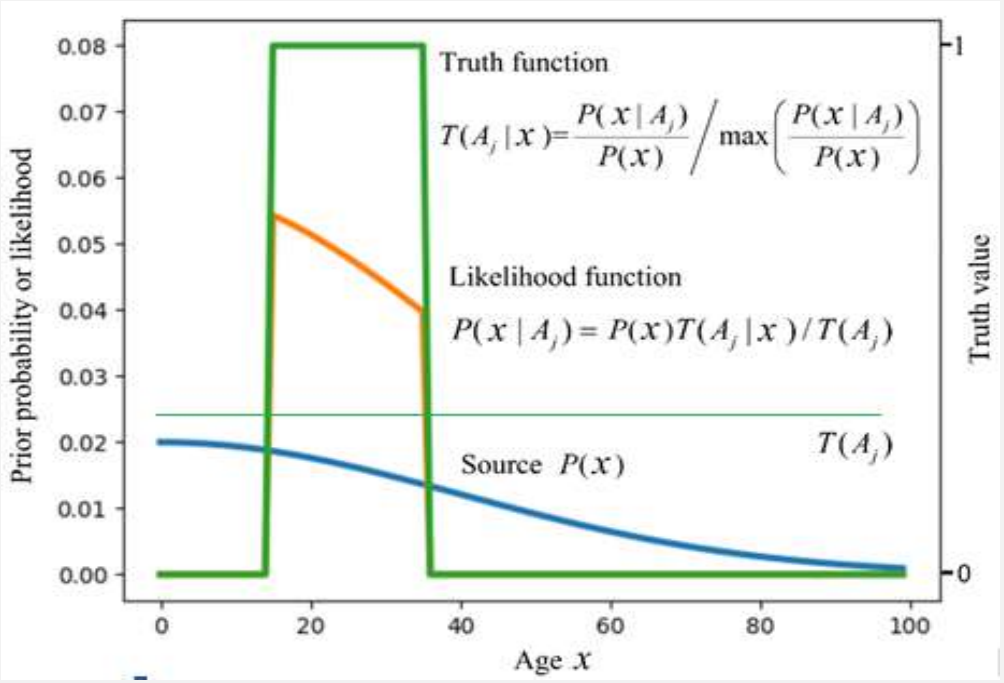
$$P(e|\theta_j)=P(e)T(\theta_j|e)/T(\theta_j),$$

$$T(\theta_j)=\sum_i P(e_i)T(\theta_j|e_i)$$

再把似然函数P(e|θj)带进语义信息公式，得到：

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

右边类似于波普尔的检验严厉性公式。可以说，语义信息公式就是逼真度公式，也是预测检验公式——反映假设是否经得起严厉检验。



从样本或样本分布学习概念外延

- 统计学习中一个带标签样本包含很多数据对，一个关于年龄的样本：

{(10，“小孩”), (20, “年轻人”), (70，“老年人”)，(40，“中年人”)，(19，“成年人”)，(60，“成年人”)，…}

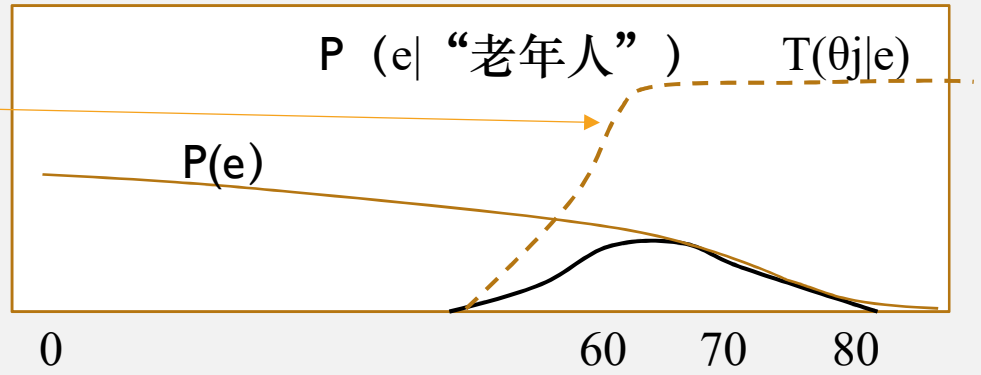
样本分布可用联合概率分布表示P(e,h), 其中e是年龄，h是标签。

- 一个带有老年人标签的子样本

{(60, “老年人”)，(70，“老年人”)，…}={60,70,…| “老年人”}

子样本分布就是条件概率分布：P(e|hj)=P(e|”老年人”).

从语义信息论和统计学习看，
求平均信息，或训练学习函数，
要用样本分布！归纳或确证，
也要用样本分布！

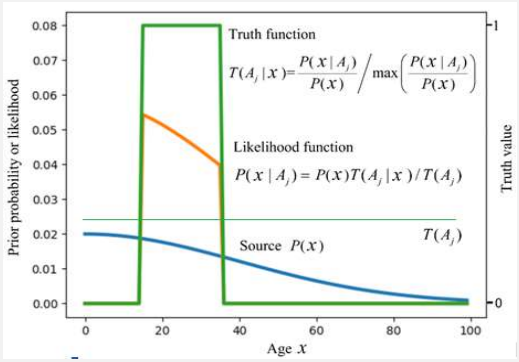


平均语义信息公式

对 $I(e_i;\theta_j)$求平均就得到hj提供的平均语义信息：

样本分布（反映事实）　　似然函数（反映预测）

$$I(E;\theta_j)=\sum_i P(e_i\mid h_j)\log\frac{P(e_i\mid\theta_j)}{P(e_i)}=\sum_i P(e_i\mid h_j)\log\frac{T(\theta_j\mid e_i)}{T(\theta_j)}$$



当集合清析时，在T(θj|e)=0的地方有一个ei使得P(ei|hj) ≠0, 则平均语义信息是-∞. 这正反映Popper的观点：对于全称假设，有一个反例就可以证伪一个理论。

但是对于不确定假设——集合是模糊的， 比如θj={年轻人}，反例增加会减少信息，未必能全部否定假设。

模糊假设可以减少风险。

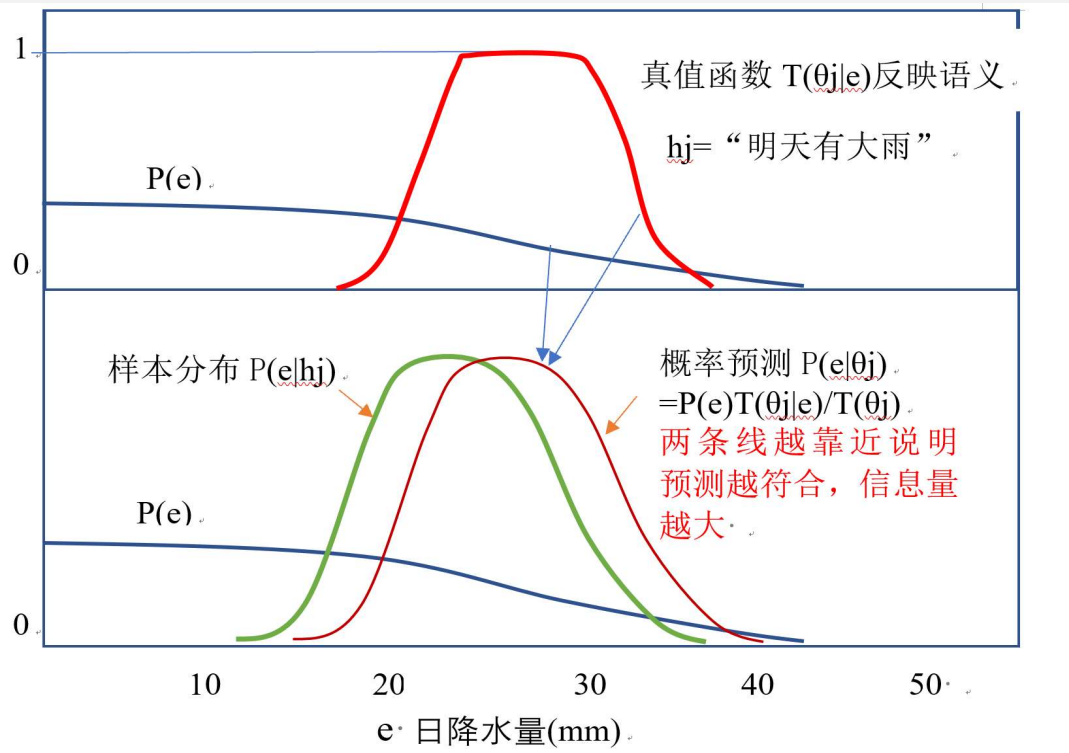
模糊预测和事实符合问题

清晰判断的符合：事实不超过外延，而且外延尽可能小。

模糊判断的符合：当P(e|θj)=P(e|hj)或T(θj|e)∝P(hj|e)时，平均语义信息量最大。这时可以说：判断和事实符合。

$$I(X;\theta_j)=\sum_i P(e_i\,|\,h_j)\log\frac{P(e_i|\theta_j)}{P(e_i)}=\sum_i P(e_i\,|\,h_j)\log\frac{T(\theta_j\,|\,e_i)}{T(\theta_j)}$$

天气预报

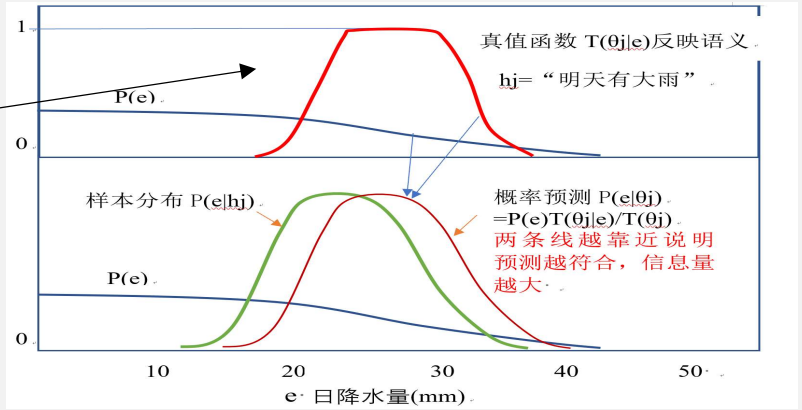


怎样用样本分布优化真值函数——寻找预测符合的外延

- 统计学习中，假设我们原先不知道一个概念（比如“年轻人”）的外延，通过样本分布，我们可以得到优化的真值函数（模糊外延）：

$$T^*(\theta_j \mid e) = \arg\max_{T(\theta_j|e)} \sum_i P(e_i \mid h_j) \log \frac{T(\theta_j \mid e_i)}{T(\theta_j)}$$

- 当样本很大，样本分布连续时，我们可以直接得到优化的真值函数（使图中两条线重合）：



$$T^*(\theta_j \mid e) = [P(e \mid h_j) / P(e)] / \max[P(e \mid h_j) / P(e)]$$

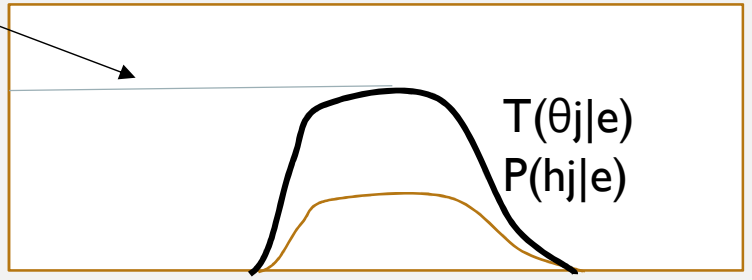
$$= P(h_j \mid e) / \max(P(h_j \mid e))$$

- 上面公式和前面的公式

$$P(e \mid \theta_j) = P(e)T(\theta_j \mid e) / T(\theta_j),$$

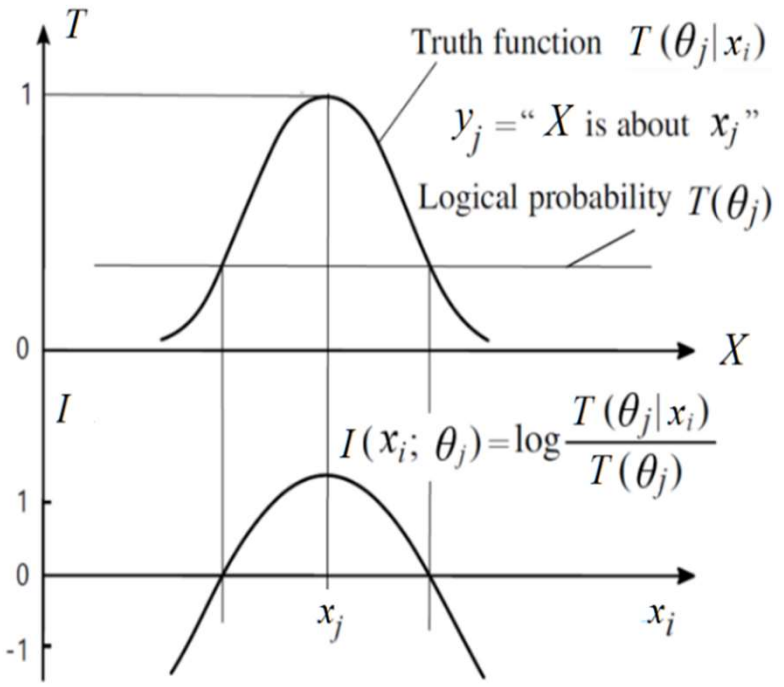
- $$T(\theta_j) = \sum_i P(e_i)T(\theta_j \mid e_i)$$

- 都是新的贝叶斯公式——是我的发现。



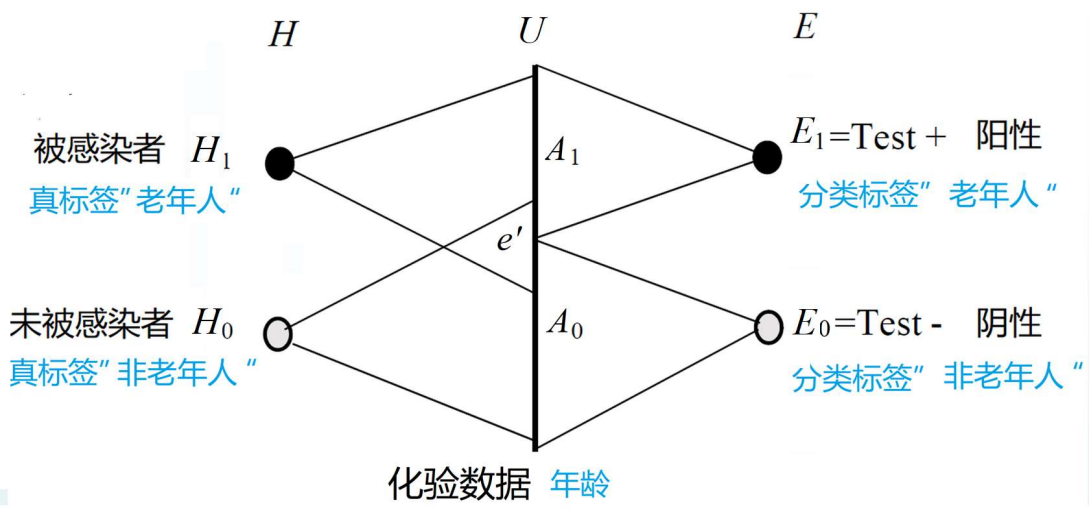
总结：语义信息公式如何兼容波普尔思想

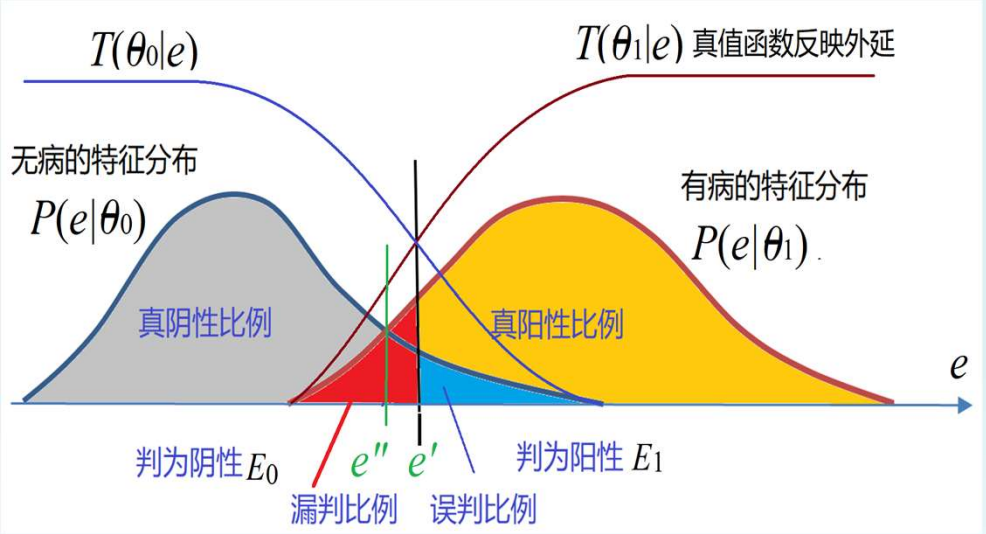
- 符合波普尔思想：
- 1）科学理论的价值在于提供信息；
- 2）逻辑概率越小，如果经得起检验，
- 信息量越大；
- 3) 永真句信息都是0;
- 4）一个反例可以证伪一个全称假设；
- 5）对于模糊假设——波普尔没有明
- 确结论；新公式给出具体结论。
- 结论：语义信息公式就是逼真度公式，
- 也反映检验严厉性，特别适合模糊假设。



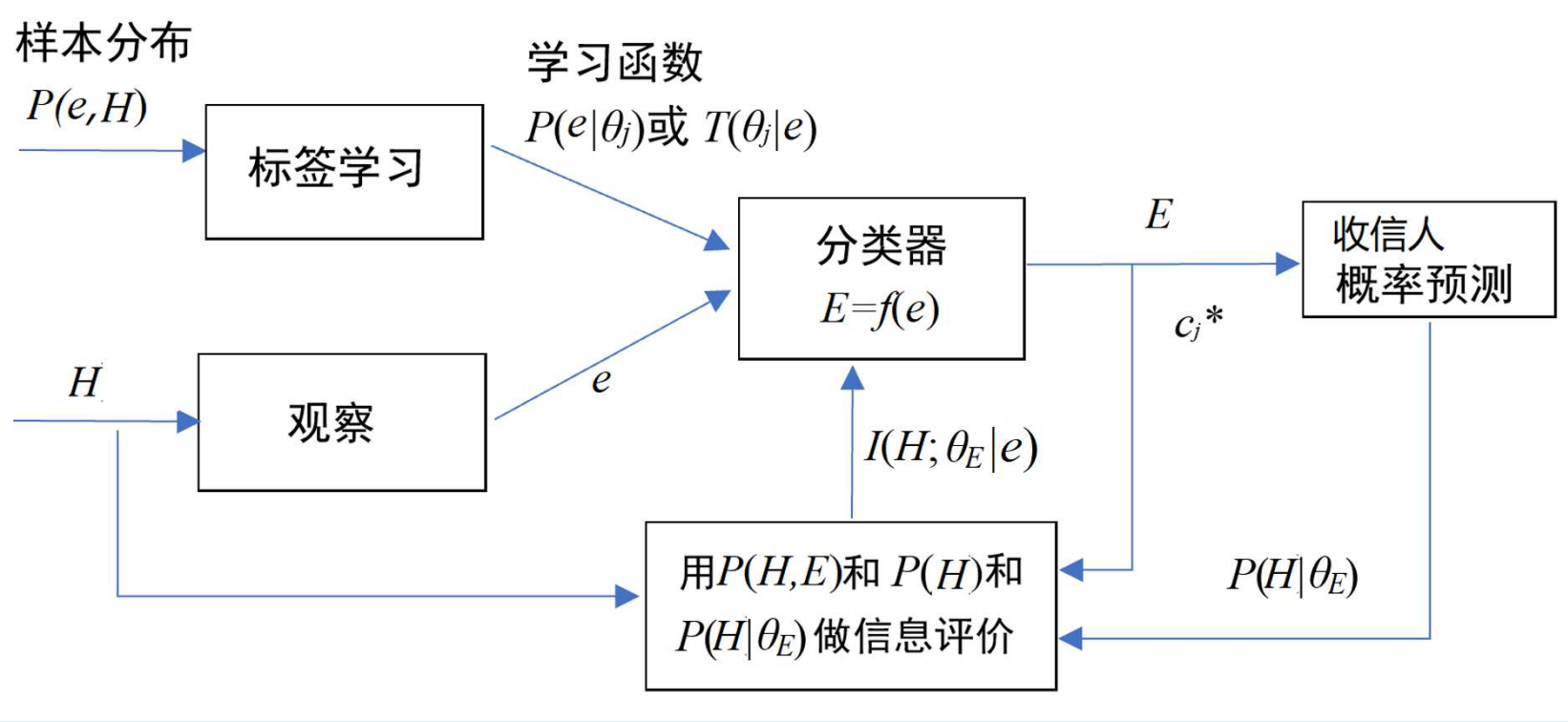
归纳确证：从统计学习模型谈起

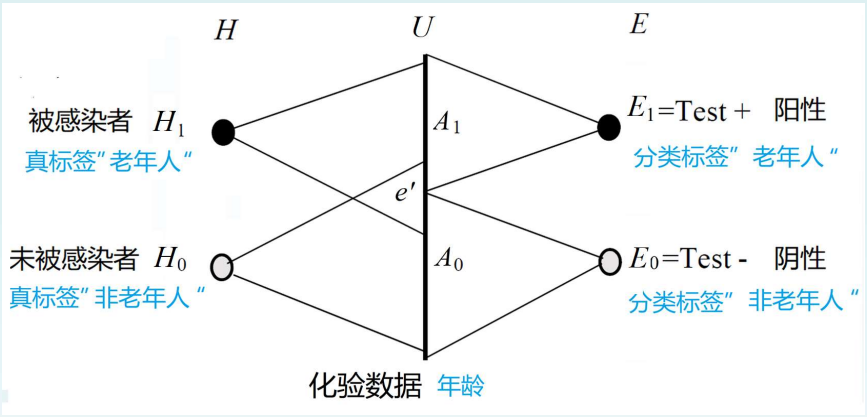
以医学检验和老年人分类为例：根据观察特征，分类可能不准。

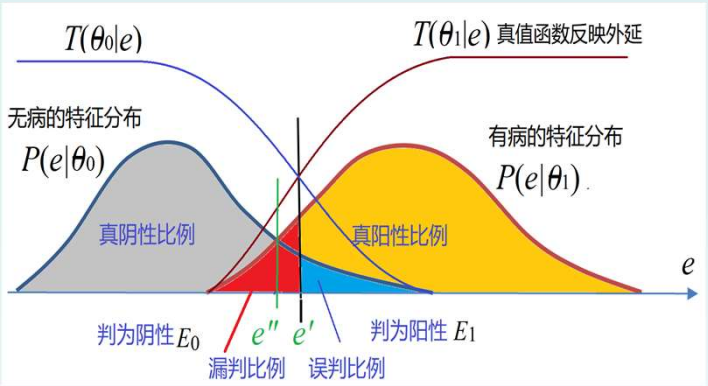




统计学习步骤


样本分布
$P(e,H)$
学习函数
$P(e|\theta_j)$或 $T(\theta_j|e)$
标签学习
分类器
$E=f(e)$
E
收信人
概率预测
c_j*
H
观察
e
$I(H;\theta_E|e)$
用$P(H,E)$和 $P(H)$和
$P(H|\theta_E)$做信息评价
$P(H|\theta_E)$


H
U
E
被感染者 H_1
真标签" 老年人 "
未被感染者 H_0
真标签" 非老年人 "
A_1
e'
A_0
E_1=Test + 阳性
分类标签" 老年人 "
E_0=Test - 阴性
分类标签" 非老年人 "
化验数据 年龄


$T(\theta_0|e)$
$T(\theta_1|e)$ 真值函数反映外延
无病的特征分布
$P(e|\theta_0)$
有病的特征分布
$P(e|\theta_1)$
真阴性比例
真阳性比例
e
判为阴性 E_0
e''
e'
判为阳性 E_1
漏判比例
误判比例

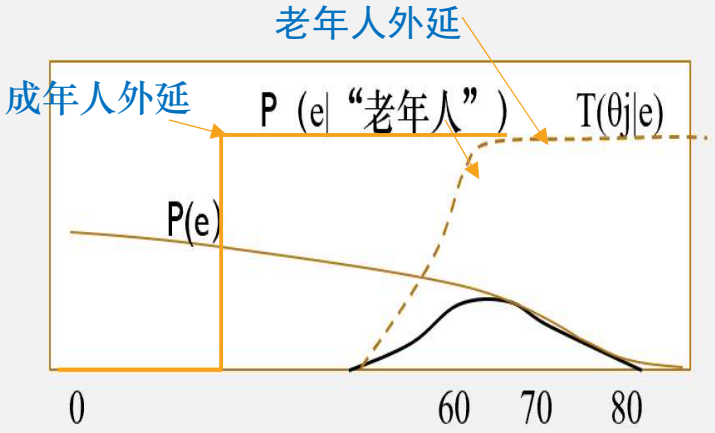
优化真值函数（即隶属函数）——归纳外延

- 1 标签学习：归纳一个全称判断的真值函数，外延或语义。
- 比如归纳谓词“e是成年人”的真值函数。
- 假定一个外国人到中国，不知道“成年人”定义，但是可以归纳出“成年人”的外延。在他看到满18岁，或19, 20，…70…被称为成年人之后，他就有了一个样本，从样本分布知道“满18岁的人是成年人”。如果一个国家对“成年人”或“老年人”没有严格定义，我们也能归纳出模糊外延——用前面的真值函数优化公式：

$$T^*(\theta_j \mid e) = \arg\max_{T(\theta_j|e)} \sum_i P(e_i \mid h_j) \log \frac{T(\theta_j \mid e_i)}{T(\theta_j)}$$

$$T^*(\theta_j \mid e) = [P(e \mid h_j) / P(e)] / \max[P(e \mid h_j) / P(e)]$$

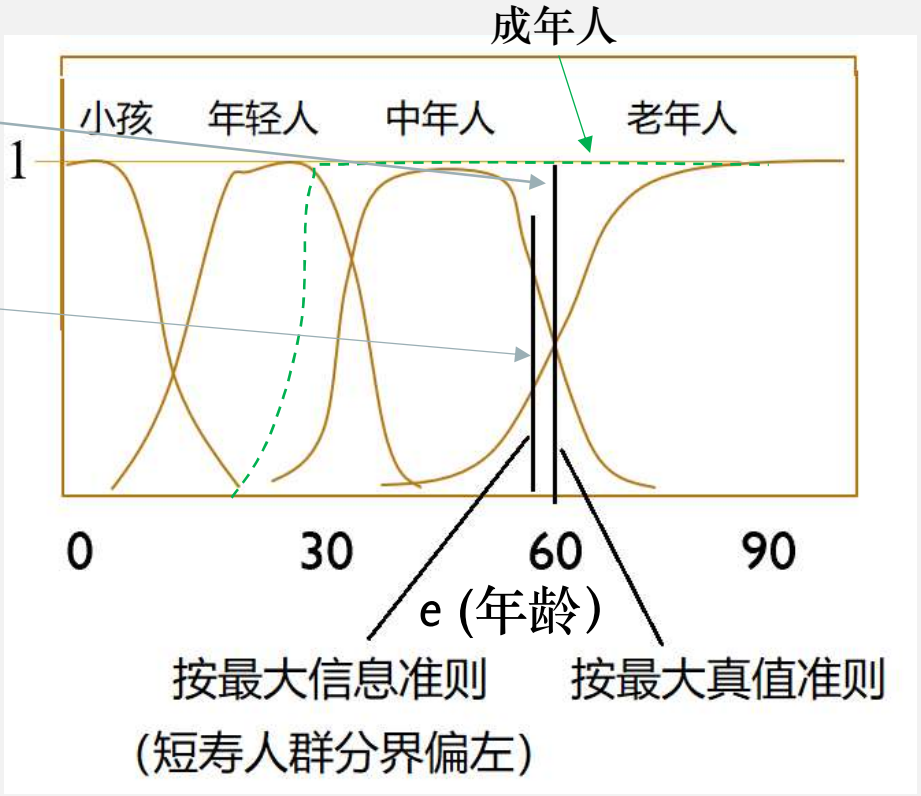
$$= P(h_j \mid e) / \max(P(h_j \mid e))$$



- 流行的方法用Logistic函数做真值函数，只能用于2分类；而上面公式可以用于多分类。

从多标签分类看正确率准则和信息准则差别

- 为什么一般不用“成年人”？
- 逻辑概率大，提供信息少！
- 所以“真”不是一个好准则，还需要逻辑概率小。
- 后面讨论 3提供确证度；4概率预测。



流行的确证度公式

- 现在确证度公式已有很多.Tentori等人于2007发表的文章
- Tentori K, Crupi V, Bonini N, Osherson D. Comparison of Confirmation Measures [J]. *Cognition* 2007,103:107-119
- 包括了几个著名的确证度公式(其中E_1是正例，E_0是反例，H_0是H_1的否定):
- $d(E_1,H_1)=P(H_1|E_1)-P(H_1)$ (Eells, 1982; Jeffrey,1992)，后验-先验
  $r(E_1,H_1)=\log[P(H_1|E_1)/P(H_1)]$ (Keynes, 1921; Horwich, 1982)，对数
- $n(E_1,H_1)=P(E_1|H_1)-P(E_1|H_0)$ (Nozik,1981)，正例-反例
- $l(E_1,H_1)=\log[P(E_1|H_1)/P(E_1|H_0)]$ (Good, 1984)，对数似然比
- $c(E_1,H_1)=P(H_1,E_1)-P(E_1)P(H_1)$ (Carnap,1962)， 相关性增量
- $k(E_1,H_1)=[P(E_1|H_1)-P(E_1|H_0)]/[P(E_1|H_1)+P(E_1|H_0)]$ (Kemenv and Oppenheim,1952)， 相对（正例-反例），网红确证度
- $s(E_1,H_1)=P(H_1|E_1)-P(H_1|E_0)$ (Christensen,1999)，阳性预测-阴性预测
- 他们的解释：证据e提高了假设H的可信程度P(逻辑概率)，所以E确证了H. 也有统计解释或频率解释：Can Bayesian confirmation measures be useful for rough set decision rules? Salvatore Grecoa,*, Zdzisw Pawlak , Roman Sowinski

流行的确证度公式存在的问题

- 考虑三段论, 存在两种证据:

1. 支持大前提的证据：归纳要提供的样例序列——样本。
2. 支持结论H1的证据： 小前提E1和大前提E1->H1。

三段论：
大前提E1->H1
小前提:E1
结论： H1

一个样例是 (E1，H1)， （E1, H0）, (E0, H1), (E0, H0)

四个中的一个。应该用样例序列做大前提的证据。

- 大多数人混淆了两种证据。用E1做大前提的证据显然不对。

权威Eells和Fitelson用H表示大前提，用 E做证据，E表示四个样例中的一个或若干个。但是一两个样例怎能确证一个大前提？他们辩解说：不是算出确证度， 而是算出确证度的增量。

这也不对。因为：1）增量太大；2）没考虑以前的样例。3）带来新的悖论：(E0, H1)和(E0, H0)既支持E1->H1也支持E1->H0.

也有统计解释。我用统计概率算出逻辑概率和确证度。

虽然解释有争议，可幸的是，检验公式都用4个样例的个数a, b, c, d计算确证度。

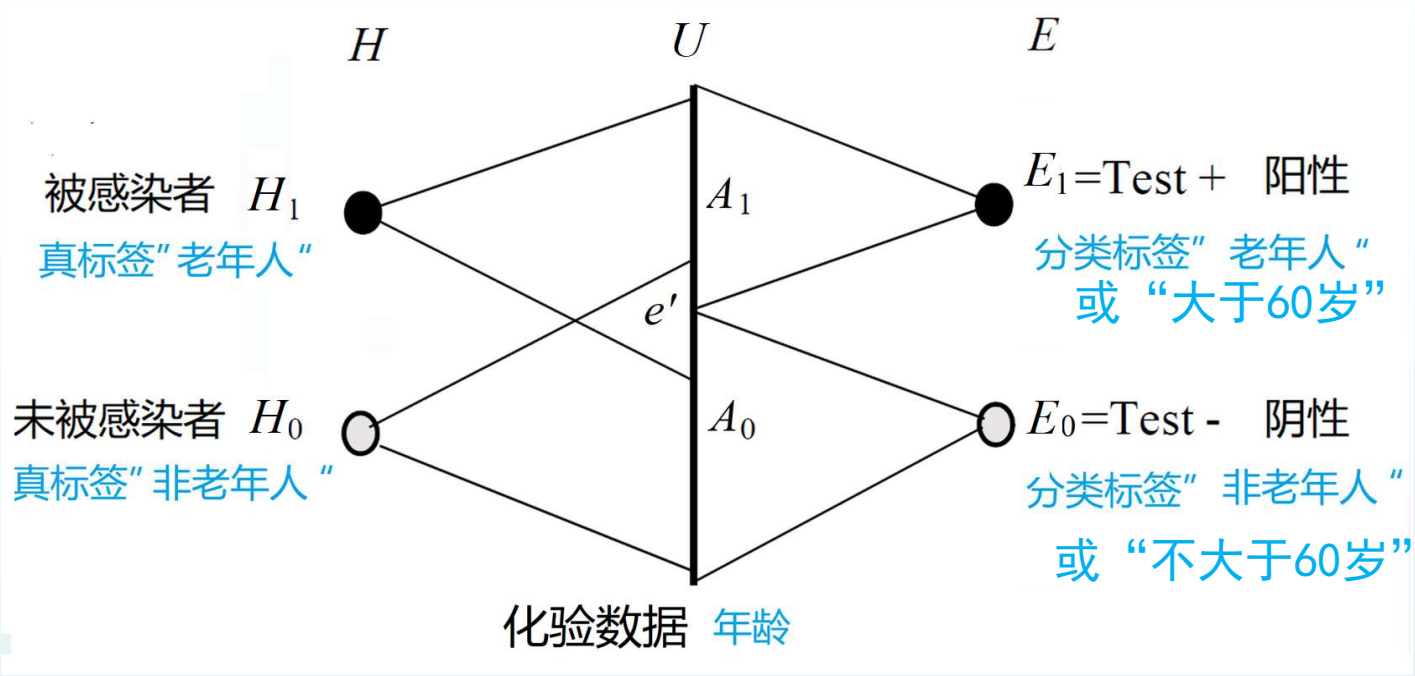
评价确证度的一个标准：4个对称性

- Eells and Fitelson 的重要文章：
- Symmetries and Asymmetries in Evidential Support 证据支持的对称性和不对称性
- http://fitelson.s464.sureserver.com/symmetry.pdf
- 其中肯定HS假设对称性(Hypothesis Symmetry):
- $b^*(E_1 \rightarrow H_1) = -\ b^*(E_1 \rightarrow$ 非 $H_1)$ ——后件否定
- 比如“所有乌鸦是黑的”其确证度为x等价于“所有乌鸦不是黑的”其确证度为-x。
- 同时否定ES和EH对称性：
- ES证据对称性：$b^*(E_1 \rightarrow H_1) = -\ b^*($非$E_1 \rightarrow H_1)$——前件否定;
- EH互换对称性：$b^*(E_1 \rightarrow H_1) = b^*(H_1 \rightarrow E_1)$ ——前后件互换.
- TS（整体对称）：$b^*(E_1 \rightarrow H_1) = -\ b^*($非$E_1 \rightarrow$非$H_1)$——前后件都否定，EC条件;
- 另外还有单调增加标准和归一化标准——[-1,1]。
- 只有上面k满足所有要求，比如满足HS对称性：

$$k(E_1 \rightarrow H_1) = \frac{P(E_1 \mid H_1) - P(E_1 \mid H_0)}{P(E_1 \mid H_1) + P(E_1 \mid H_0)} = -\frac{P(E_1 \mid H_0) - P(E_1 \mid H_1)}{P(E_1 \mid H_0) + P(E_1 \mid H_1)} = -b^*(E_1 \rightarrow H_0).$$

- x　　-x

医学检验中两个大前提



医学检验提供两个大前提：

E1->H1= "如果E=E1(阳性)，则H=H1(有病)"

E0->H0= "如果E=E0(阴性), 则H=H0(没病)"

现在考虑两个大前提的确证度，并且希望算出的确证度能用于有病没病的概率预测。

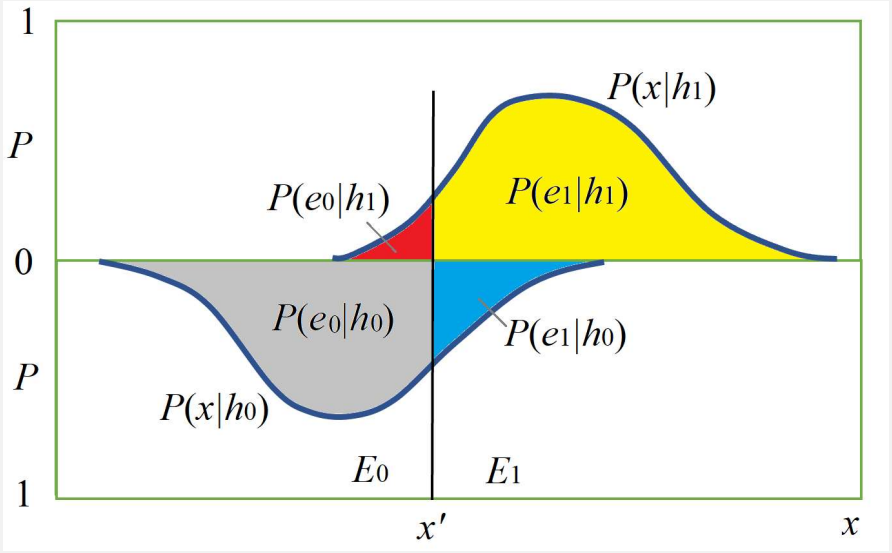
推导确证度I：医学检验的香农信道

表I. 医学检验：敏感性——有病判为阳性的正确率

特异性——无病被判为阴性的正确率；它们构成香农信道$P(E|H)$

	真没病h_0	真有病好h_1
阳性E_1	漏报比例=I-$P(E_1\|H_1)$=$P(E_0\|H_1)$	敏感性=$P(E_1\|H_1)$ (黄色区域)
阴性E_0	特异性=$P(E_0\|H_0)$	误报比例=I-$P(E_0\|H_0)$=$P(E_1\|H_0)$

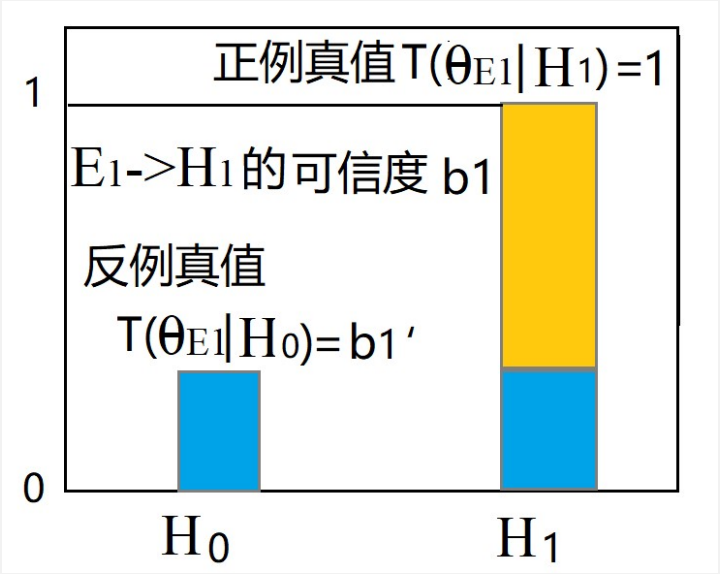
- 四个区域面积比例
- 就代表上面四个概率。
- 对于EI->HI, 正例是
- 黄色区域；反例是
- 蓝色区域。



推导2：语义信道和推理的可信度

E1的模糊真值函数看作是清晰可信部分和不可信部分的组合：

$T(E_1|H)= b_1' + b_1 T(A_{z1}|H)$ （H是变量）



医学检验的语义信道

	真没病 H0	真有病 H_1
阳性E_1	$T(E_0|H_1)=b_0'$ 反例真值	$T(E_1|H_1)=1$ 正例真值
阴性E_0	$T(E_0|H_0)=1$ 正例真值	$T(E_1|H_0)=b_1'$ 反例真值

定义：信任度：主观的，在-1和1之间变化。知道随机乱说，信任度是0；知道故意说谎，可信度是负的。

确证度=样本支持的信任度， 是客观的。

推导3：优化的反例真值就是否证度

可以证明， 当语义贝叶斯预测和条件概率分布相等时， 信息量最大。 令 $P(H|\theta_{E1})=P(H|E1)$

- 我们得到阳性的优化的不信度

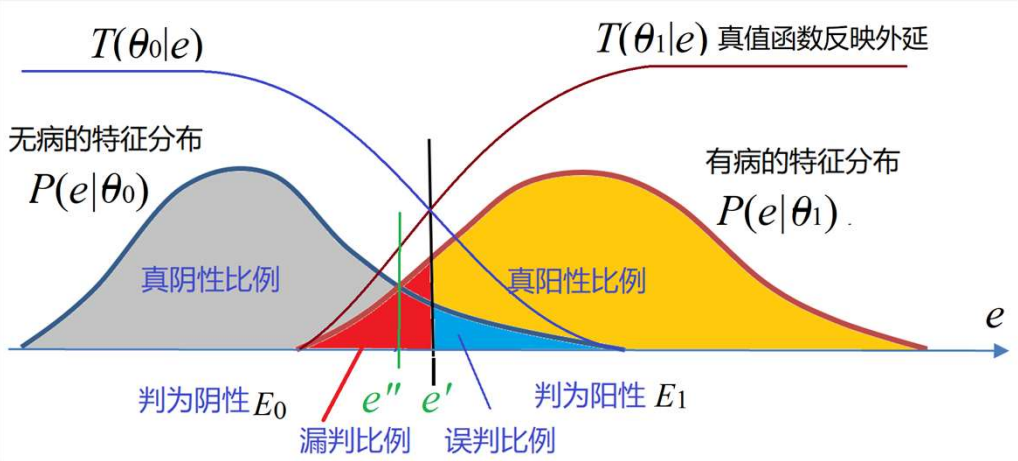
  $b_1'^*= P(E_1|H_0)/P(E_1|H_1);$

- 大前提EI->HI(如果“阳性则有
- 病”） 的确证度:

  $b_1^*=1-b1'^*$

  $= [P(E_1|H_1)-P(E_1|H_0)]/ P(E_1|H_1).$

  =(正例比例-反例比例)/正例比例



黄色占右侧比例就是确证度

推导4：更一般的信道确证度b*
——考虑负的确证度

- 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

$$b_1^* = [P(H_1|E_1)-P(H_1|E0)]/\ P(H_1|E_0).$$ 分母不同

- 综合上面两个公式， 我们得到：

$$b_1^* = \frac{P(E_1|H_1)-P(E_1|H_0)}{\max[P(E_1|H_1),P(E_1|H_0)]} = \frac{\text{正例比例-反例比例}}{\text{上面两个较大者}}$$

$$k = \frac{P(E_1|H_1)-P(E_1|H_0)}{\max[P(E_1|H_1),P(E_1|H_0)]} = \frac{\text{正例比例-反例比例}}{\text{上面两个相加}}$$

上面确证度：优化的真值函数中有用部分比例。它和信源无关，只反映信道特性，所以我们称b*是信道确证度。

反向推理的确证度

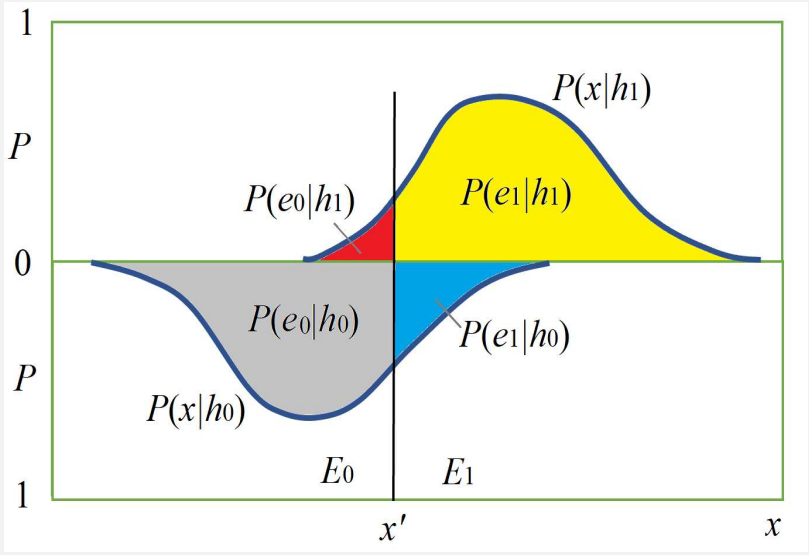
解释确证度c*(H1->E1)（反向推理）：

在图中，上面都是纵向推理，E在前面。现在大前提反过来（要横向推理），H在前：

H1->E1=“如果谁被称为老年人，则他大于60岁”。

用清晰集合解释模糊集合，反例是红色。

$$b_1^* = \frac{P(E_1 \mid H_1) - P(E_0 \mid H_1)}{\max(P(E_1 \mid H_1), P(E_0 \mid H_1))}$$



确证度b*可用于概率预测

是阳性就一定有病？未必！

确证度可以用于有病的概率预测：

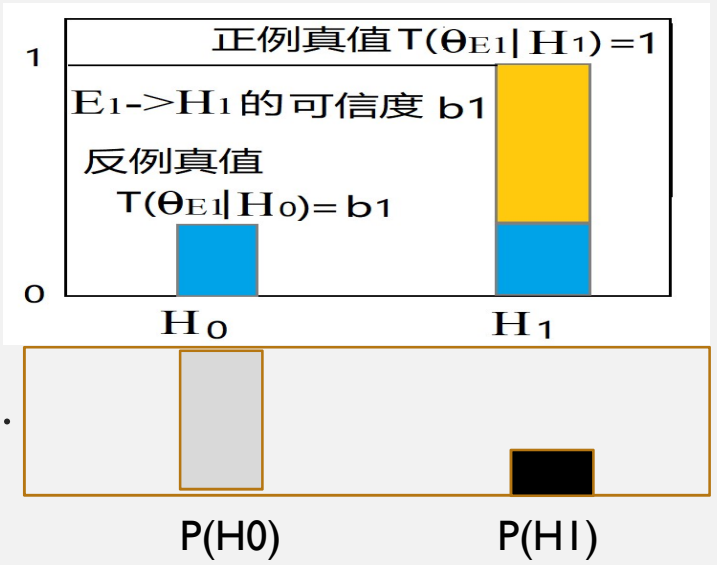
$P(H1|E1)=P(H1)/T(\theta_{E1})$

$=P(H1)/[P(H1)+b1'^{*}P(H0)]$

比如对于一种HIV检验，

敏感性P(E1|H1)= 0.917；特异性P(E0|H0)=0.999.

则$b1'^{*}$=0.0011.



当基础概率（被感染的先验概率）是

P(H1)=0.0001（不太可能人群），P(H1）= 0.002，（普通），P(H1)=0.1(高危）时，当检验为阳性时，预测被HIV感染的概率分别是

$P(H1|\theta_{E1}) = 0.084,\ 0.65,\ 0.99$。

用上面公式用经典的贝叶斯公式，算出结果是一样的，这是因为优化的真值函数正比于转移概率函数，即：$T^{*}(\theta_{z1}|H)\propto P(E1|H)$

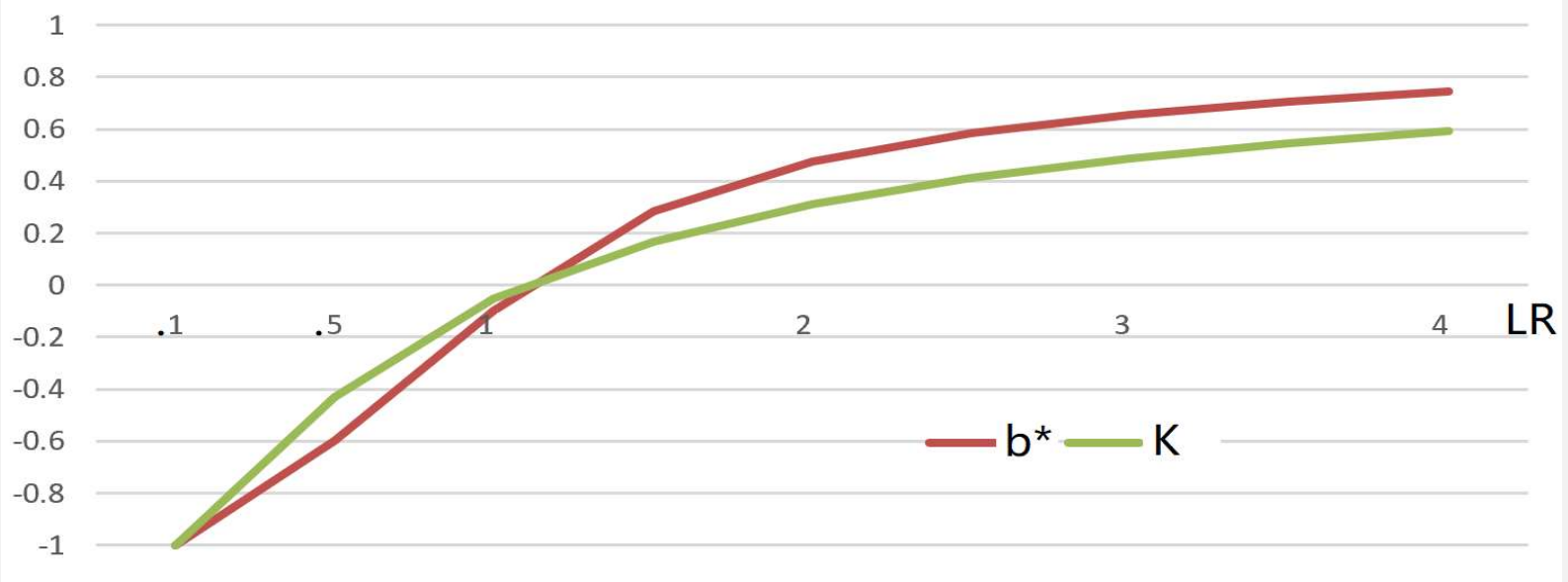
用否证度预测的优点是：1)信源P(H)改变后，否正度$b_1'^{*}$仍然有用；

2)记忆$T^{*}(\theta_{z1}|H)$比记忆$P(E_1|H)$简单——只记一个数$b_1'^{*}$.

信道确证度b*和似然比关系

- 医学检验中用似然比LR(likelihood ratio)表示一个检验有多好，阳性似然比是　LR+=P(H1|E1)/P(H1|E0)=正例比例/反例比例=1/b1'*
- 确证度b*和似然比的关系是：
- $$b_1^* = \frac{P(E_1 \mid H_1) - P(E_1 \mid H_0)}{\max(P(E_1 \mid H_1), P(E_1 \mid H_0))} = \frac{LR^+ - 1}{\max[LR^+, 1]}$$

b*和k随LR变化：



用k做概率预测没那么方便。

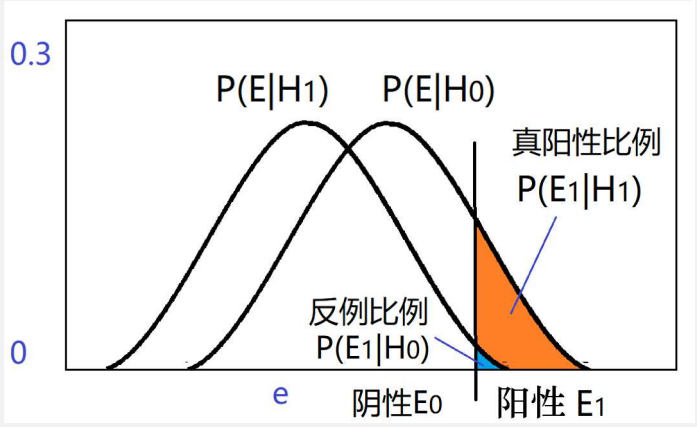
b*和k重视反例比例，兼容Popper思想

- 大多数确证度强调正例比例大重要， 而*b**和*k*强调反例比例 小重要。
- 比如对于E1—>H1

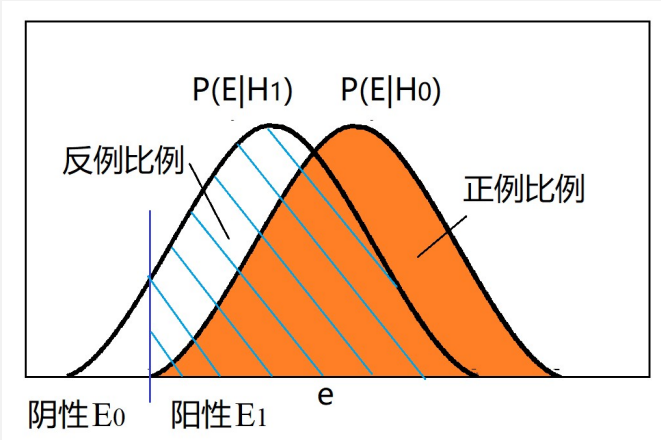
	正例比例	反例比例	其他确证度	确证度b*
例1	$P(E_1\|H_1)=0.1$	$P(E_1\|H_0)=0.01$	较小，比如0.1	b1*=0.9, 较大
例2	$P(E_1\|H_1)=1$	$P(E_1\|H_0)=0.9$	较大， 比如1或0.5	b1*=0.1， 较小

- 例1虽然正例少，反例更少，所以b1*大； 例2虽然正例多，反例也多，所以b1*小.

例1



例2



- 重要结论: 1) 推理阳性->有病的确证度主要取决于无病被判为阴性的正确率P(E0|H0)， 阴性->无病的确证度主要取决于有病被判为阳性的正确率P(E1|H1);
- 2) 反例少比正例多更加重要，所以确证测度 b*兼容 Popper的证伪理论。

信道确证度b*和预测确证度c*——解释乌鸦悖论

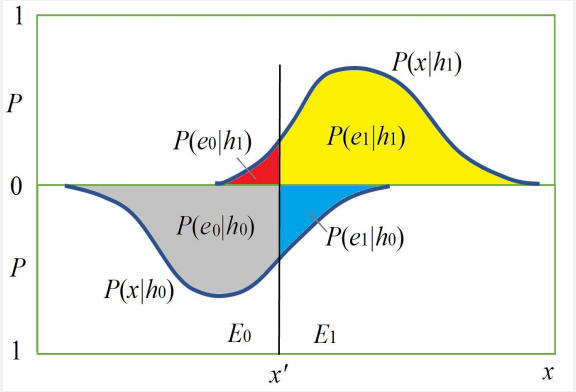
前面讲的确证度反映信道特性——比如反映医学检验手段好坏。
有时候我们还需要反映概率预测好坏的确证度——预测确证度c*:

$$c_1^* = c_1^*(E_1 \to H_1) = \frac{P(H_1 \mid E_1) - P(H_0 \mid E_1)}{\max(P(H_1 \mid E_1), P(H_0 \mid E_1))}$$

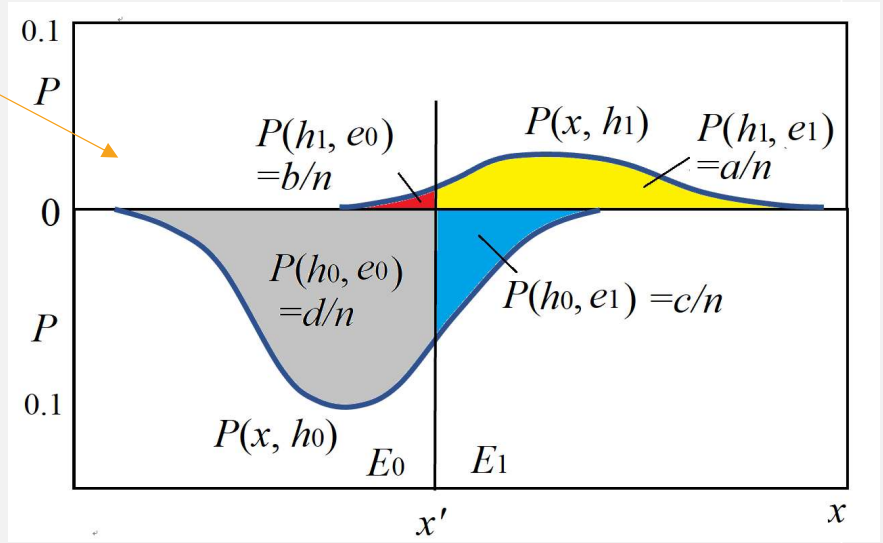
$$= \frac{P(H_1, E_1) - P(H_0, E_1)}{\max(P(H_1, E_1), P(H_0, E_1))} = \frac{\text{正例比例-反例比例}}{\max(\text{正例比例，反例比例})}$$

c1*和b1*不同， 所用的两条曲线P(H1,e)和P(H0,e) 下面积不同. 四个区域代表四个样例数字a,b,c,d.

而前面计算b*的上下面积相同:



P(H1)=P(H0)=0.5时， c*=b*



E1->H1的确证度c*(E1->H1)= （黄-蓝）/黄
H1->E1的确证度c*(H1->E1)=(黄-红)/黄

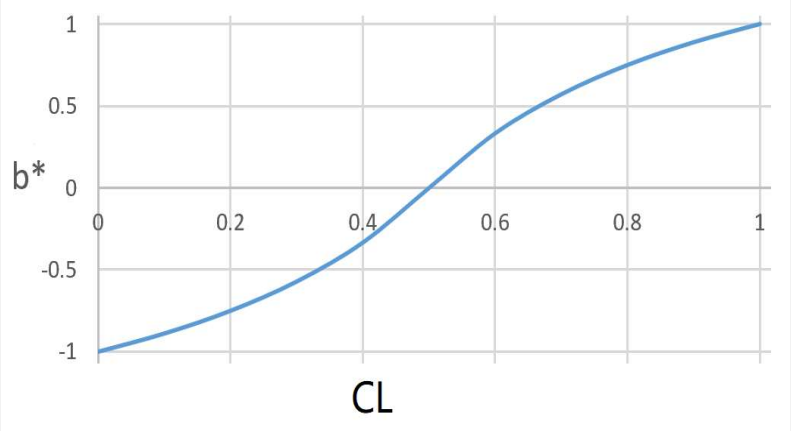
预测确证度c*和置信水平的关系

- 经典统计中用置信水平CL和置信区间表示统计预测质量。
- c*和CL的关系是:

$$c_1^* = c_1^*(E_1 \to H_1) = \frac{P(H_1 \mid E_1) - P(H_0 \mid E_1)}{\max(P(H_1 \mid E_1), P(H_0 \mid E_1))}$$

$$= \frac{2P(H_1 \mid E_1) - 1}{\max(P(H_1 \mid E_1), 1 - P(H_1 \mid E_1))} = \frac{2CL_1 - 1}{\max(CL_1, 1 - CL_1)}$$

- 置信区间就是前面分类的清晰集合——比如"大于60岁"。
- 预测确证度c*和置信水平*CL*的关系如图:



- 用c*做概率预测更加简单:

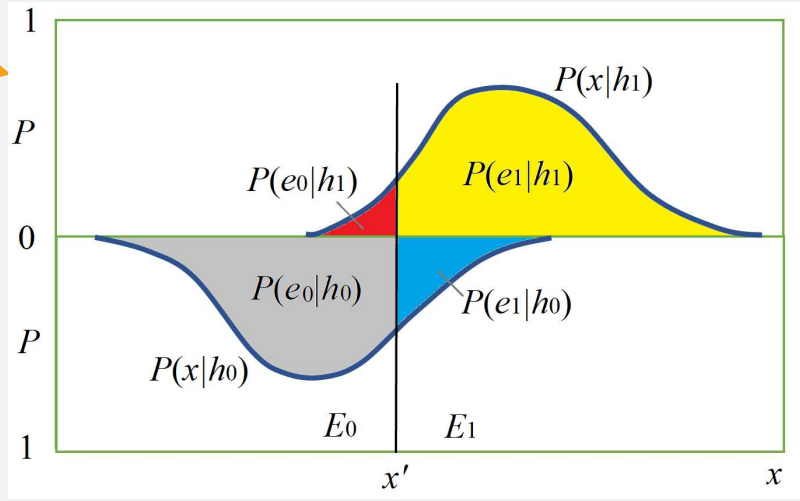
$$P^*(H_1 \mid \theta_{E1}) = 1/(2 - c_1^*)$$

- 比如c1*=0.8, P(H1|θE1)=1/1.2=0.833
- 但是基础概率P(H1）变了,c*就不
- 合适了，要用b*.

8种确证度总结

- 信道确证度四种:

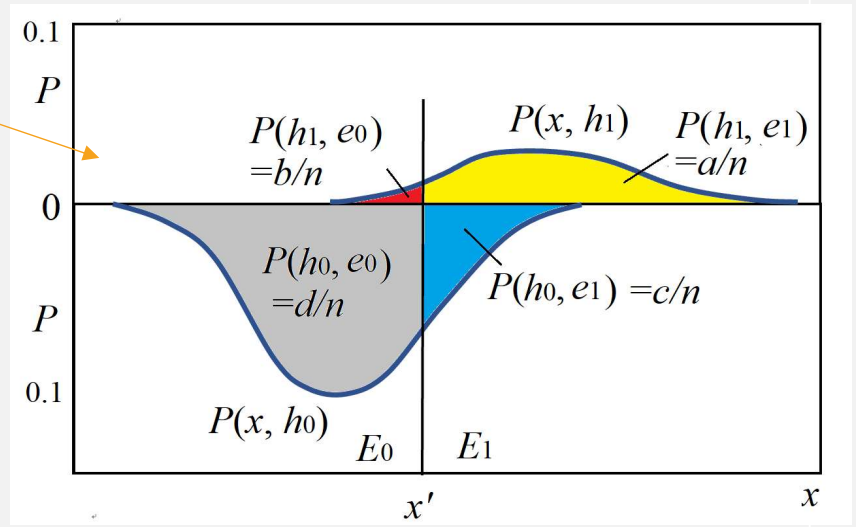
纵向：清晰E->模糊H，2种;
横向：模糊H->清晰E， 2种;



- 加上预测确证度四种，

所有确证度一共8种，都可以写成：

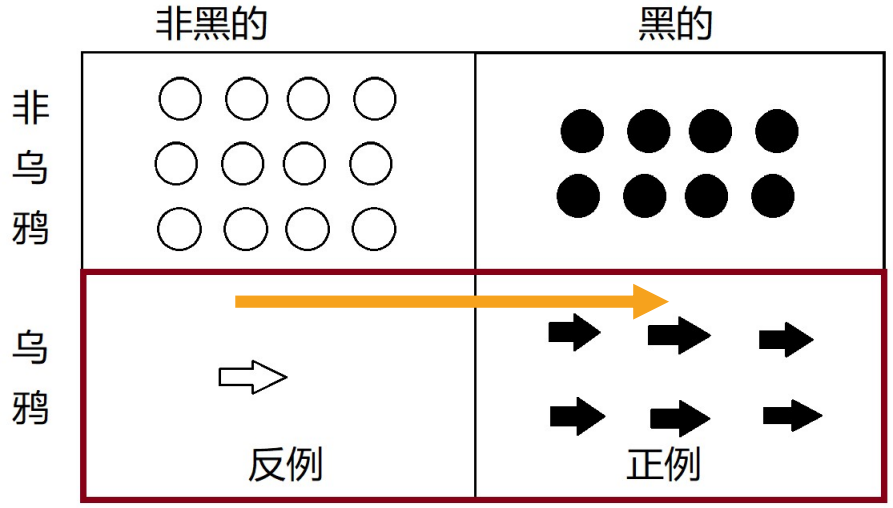
$$\text{确证度} = \frac{\text{正例比例}-\text{反例比例}}{\text{上面较大者}}$$



如果考虑交叉推理, 比如b*(e1->h0)（用HS对称性得到负的）, 一共有8对16种。

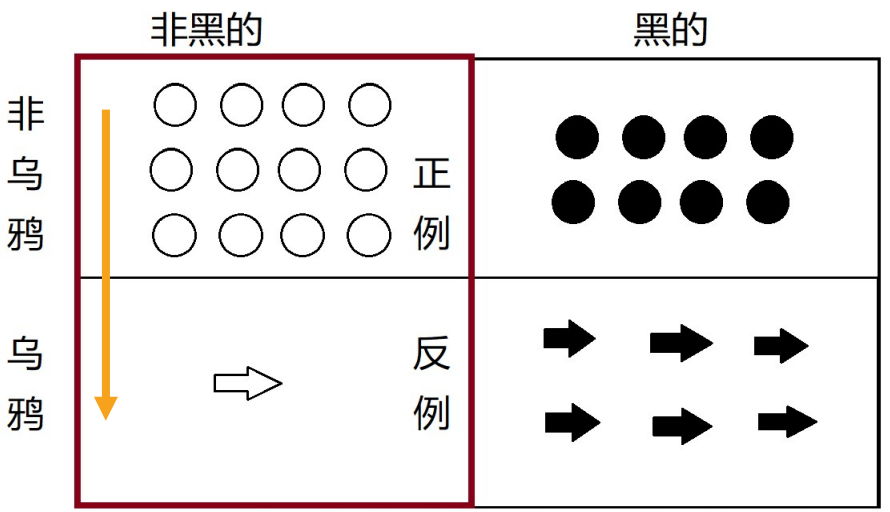
澄清乌鸦悖论1：c*肯定Nicod准则

- 结论1(等价条件EC）：按照经典逻辑 E1->H1 和 H0->E0等价，即“所有乌鸦是黑的”和“所有非黑的东西就不是乌鸦”等价。
- 结论2(Nicod准则中不相关准则NC)：一个非黑的东西，比如一枝白粉笔，支持“所有非黑的东西都不是乌鸦”，但是和“所有乌鸦是黑的”不相关。
- 结论1和结论2矛盾！所以叫乌鸦悖论。为了消除悖论，有人否定结论1，有人否定结论2，还有人否定两者。
- 用预测确证度公式：确证度c*=（正例比例-反例比例）/分子中最大者——肯定NC
- 可见对于模糊逻辑c*(E1->H1) 和c*(H0->E0)不等价，因为正例比例不同。所以结论1错了。不存在悖论。



E1->H1：乌鸦是黑的　　确证度=5/6

增加一个白粉笔O时确证度不变



E0->H0：非黑的就不是乌鸦　　确证度=11/12

增加一个白粉笔O时确证度变为12/13

澄清乌鸦悖论2：c*比k好

- 考虑c*(E1->H1), 设四种例子个数是：

	E0非乌鸦	E1乌鸦
H1黑	b 黑猫	a 黑乌鸦
H0不黑	d 白粉笔	c 白乌鸦

- 乌鸦->黑的

$$k(E_1 \to H_1) = \frac{\frac{a}{a+b} - \frac{c}{c+d}}{\frac{a}{a+b} + \frac{c}{c+d}} = \frac{ad - bc}{ad + bc + 2ac}$$

$$c^*(E_1 -> H_1) = \frac{a - c}{\max(a, c)}$$

- 非黑—>非乌鸦

$$k(H_0 \to E_0) = \frac{\frac{d}{c+d} - \frac{c}{c+d}}{\frac{d}{c+d} + \frac{c}{c+d}} = \frac{d - c}{d + c}$$

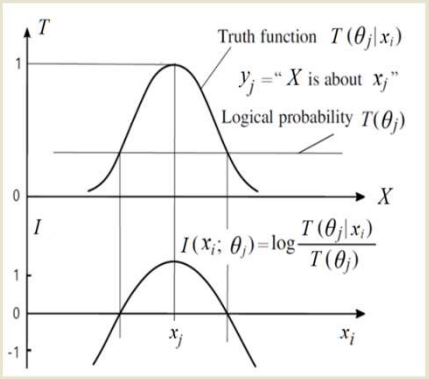
$$c^*(H_0 -> E_0) = \frac{ad - bc}{\max(ad + dc, bc + dc)}$$

- 两个k随d增加，增量不等；只有下面c*随d增加。所以等价条件EC是错的！
按k, NC(Nicod准则)不成立,悖论还在；按c*是Nicod准则成立，不存在悖论！

总结

- 语义信息公式

$$I(e_i;\theta_j)=\log\frac{T(\theta_j|e_i)}{T(\theta_j)}=\log\frac{P(e_i\,|\,\theta_j)}{P(e_i)}=\frac{\text{命题真值}}{\text{谓词逻辑概率}}.$$



也是逼真度公式，体现Popper的检验严厉性。

- 确证度公式：

$$b^*(E_1\to H_1)=\frac{P(E_1\,|\,H_1)-P(E_1\,|\,H_0)}{\max(P(E_1\,|\,H_1),P(E_1\,|\,H_0))}=\frac{LR^+-1}{\max[LR^+,1]}=\frac{\text{正例比例-反例比例}}{\text{分子中两个较大者}}$$

$$c^*(E_1\to H_1)=\frac{P(H_1\,|\,E_1)-P(H_0\,|\,E_1)}{\max(P(H_1\,|\,E_1),P(H_0\,|\,E_1))}=\frac{2CL_1-1}{\max(CL_1,1-CL_1)}=\frac{\text{正例比例-反例比例}}{\text{分子中两个较大者}}$$

- 因为新确证测度重视反例比例，因而也兼容Popper思想；c*还能消除乌鸦悖论。

谢谢听讲，欢迎交流！

更多有关文章，见：http://survivor99.com

（本PPT的PDF版已上载到该网站）