

概率的不同表征对贝叶斯推理的影响的研究

杨莉 胡竹菁

(江西师范大学 心理学系,南昌 330022)

摘 要 :该研究以贝叶斯推算、判断为实验任务 ,探讨了三种概率表征(百分比、自然频率、概率词) \ 三种概率规则认识水平对贝叶斯推理的影响 ,结果表明 :1) 概率词表征的贝叶斯推算实质是判断过程。2) 在三种概率表征中 ,频率表征更适合于正确完成贝叶斯推理 ,但存在学科背景的差异。3) 贝叶斯推理使用的策略有正确规则、错误规则、命题知识、概率简捷式、直觉和个人意愿六种。4) 贝叶斯判断的正确率高于贝叶斯推算 ,个体在贝叶斯判断上使用的正确规则多于贝叶斯推算过程。

关键词 :贝叶斯推理 ;概率表征 ;推理策略

中图分类号 :B842.5 文献标识码 :A 文章编号 :1003 - 518X(2007)02 - 0032 - 04

1 问题的提出

贝叶斯推理是根据某事件 H 的伴随现象 D 的信息 ,推断出该事件的发生情况的过程。事件 H 的发生情况有两种 :1) 事件 H 发生的可能性大小 ,即 $P(H|D)$;2) 事件 H 是否发生。通常把前一种称为贝叶斯推算 ,后一种称为贝叶斯判断。心理学界对贝叶斯问题的探索一般是推算过程 ,企图验证人类的推算过程是否遵循贝叶斯规则 ;或者在不知道贝叶斯规则的情况下 ,如何进行推算 ,其结论有什么共同之处。贝叶斯规则为 :由 $H, - H$ 所构成的完全事件 ,观察到 $P(D|H), P(D|-H)$,从而得出 $P(H|D)$ 。其中 $P(H), P(-H)$ 为基础率 , $P(D|H)$ 为击中率 , $P(D|-H)$ 为误报率 , $P(H|D)$ 为后验概率。贝叶斯定理的表达形式之一 :

$$P(H|D) = \frac{P(H)P(D|H)}{[P(H)P(D|H) + P(-H)P(D|-H)]}$$

Ward Edwards 和他的同事^[1]首次用实验来检验人类推理是否遵循贝叶斯定理 ,他的结论是 :一般地 ,人们给了基础率足够的重视 ,推理结论还是与贝叶斯定理的计算结果相称的。Kahneman 和 Tversky^[2,3]则由工程师和出租车实验得出了相反的结论 ,人们在推理过程中是忽视基础率的。Eddy^[4]的实验也得出相同的结论 ,人们是忽视基础率的。

有学者认为^[5] ,尽管百分比概率信息在现代世界是有用的 ,但是它们不是人们进化而得到的 ;只有在自然频率表征时 ,基础率的信息才能够被更合适、

更有效地得到利用。所以 ,相对于百分比表征 ,个体在自然频率表征上能更好地解决贝叶斯问题。Gigerenzer(1995)^[6]的实验结果表明 ,使用自然频率表征后 ,正确率从百分比表征的 16% 上升到 46%。

按照进化的思想 ,进化程度越高的表征 ,其隐含的信息量越大 ,加工过程越趋于自动化 ,需要个体的努力程度越低^[7] ;虽然概率的表征系统有哪几种并没有得到严格的论证 ,但是目前大家认同的概率的表征形式常用的是概率词、频率、百分比。频率、百分比是数字表达的概率 ,概率词是言语表达的概率 ,数字表达的概率比言语概率更加精确 ,在预测和比较方面容易操作 ,且拥有一套发展很好的计算系统^[8]。人们对贝叶斯推理过程研究的关注的重点也是数字型的估量 ,几乎没有言语估量的报告。这里选择研究不确定性的言语估量 ,作为一种人们评定事件发生可能性大小的可选择的手段之一 ,补充了贝叶斯推理的研究。

虽然大多数学者认为言语和数字概率间可以转换 ,但两者的概率推理过程却不一样。Windschit^[9]证实数字概率倾向于进行分析和基于规则的推理 ,而言语概率允许更多的联想和直觉思维 ,因此数字和言语概率可能反映了不同的推理系统 ,受到情景的不同方面的影响 ,研究试图探讨不同概率表征对贝叶斯推理(推算、判断)的影响。

2 研究方法

2.1 材料和设计

实验采用 2 (推理任务 :推算、判断) $\times 3$ (概率表征 :百分比、频率、概率词) $\times 3$ (学科背景 :专家、理科、文科) 三因素混合设计。其中 ,推理任务和概率表征为被试内变量 ,学科背景为被试间变量。因变量为后验概率的推算和判断的准确率以及推理策略的人次分布。

实验材料共六道贝叶斯推理题。包括 ,推算题三题 ,内容分别为传染病、抑郁症、智力竞赛 ;判断题三题 ,内容分别为水果质量、禽流感、胃癌。概率表征形式有百分比、自然频率、概率词三种。涉及百分比、概率词的题目与经典的 Eddy 的“乳腺癌问题”^[10]的形式相同 ,涉及自然频率的题目与 Gigerenzer 的“传染病问题”^[10]的形式相同。数字型表征的先验概率值满足在给出贝叶斯定理的前提下 ,通过心算就可以完成任务的前提条件。问题示例如下 :

“接受常规检查的超过 40 岁的妇女确实患乳癌的概率是 1%。如果一个妇女患有乳癌 ,那么她在乳房照影法中得到阳性反应的概率是 80%。如果一个妇女没有得乳癌 ,那么她在乳房照影法中得到阳性反应的概率是 9.6%。一位这一年龄组的妇女在常规检查的乳房照影法中呈阳性反应。问 :她确实患乳癌的概率是多少 ?”

“设想一下有一位身处文盲社会中但却经验丰富的老医生。她没有任何书籍 ,也无法进行统计调查 ,只能依赖她的经验来为患者治病。她的患者中有人染有一种以前不为人所知的恶性疾病。幸运的是这个医生发现有一种症状能够表征这种疾病 ,不过这种症征与恶性疾病之间的对应关系不是很确定。在她的行医史中 ,她见过 1000 个患者 ,其中有 10 个人患有这种疾病。在这 10 个人中 ,有 8 个显现出这种症状。另外 990 个没有感染这种疾病的人中 ,有 95 人也表现出这种症状。现在有一个新的患者来了 ,她有这种症状。问 :他确实患有这种疾病的可能性有多大 ?”

2.2 被试

江西师范大学共 125 名大学生。其中 ,学过概率统计 ,了解贝叶斯推理的数学系学生 41 名组成专家组。其余不了解贝叶斯推理的 84 名被试 ,分为文科组 (中文系学生) 44 人 ,理科组 (物理系学生) 40 人。

2.3 实验程序

正式实验中 ,要求被试先在实验室的电脑上完

成同一表征系统下的三道贝叶斯推算题 ,休息五分钟 ,再完成三道贝叶斯判断题。完成贝叶斯推理任务后 ,被试就刚才的题目写下思考过程和答案。

3 结果与分析

剔除没按要求作答的数据 ,有效数据 105 个 ,其中数学 36 人 ,汉语言 34 人 ,物理 35 人。

在贝叶斯推算任务中 ,百分比、频率表征下 ,对专家、文科生、理科生的正确率进行统计 ,结果见表 1。进行 χ^2 检验 ,结果如下 :百分比 , $\chi^2_{(2,105)} = 21.735$, $p < 0.000$;频率 , $\chi^2_{(2,105)} = 15.992$, $p < 0.000$;学科背景在贝叶斯推算题中的正确率差异显著。在用概率词汇表征的推理题回答中 ,根据被试的用词 ,就事件发生的可能性分为 5 个类别 :很大、较大、一般、较小、很小。不同学科背景被试的概率词回答的统计结果见表 2 ,进行 χ^2 检验 , $\chi^2_{(8,105)} = 22.970$, $p < 0.005$,学科背景差异显著。

表 1 数字表征的贝叶斯题项后验概率的准确人数

表征	数学	中文	物理
百分比	14	0	3
频率	16	1	10

表 2 概率词汇表征的贝叶斯题结论的类别的人次

表征	类别	数学	中文	物理
概率词	很大	8	22	11
	较大	12	8	9
	一般	1	1	5
	较小	9	0	6
	很小	6	3	4

对被试在完成贝叶斯推算题所使用的策略分类 ,有 5 种 :1)正确规则 :被试根据这类规则能够得出正确的结论 ;2)错误规则 :有明确的计算规则 ,能够完成一系列的计算 ,但不符合贝叶斯规则的原理 ;3)概率简捷式 :个体不经过正式的规则运算得出结论 ,而是利用概率知识 ,完成粗略的估算 ;4)日常经验 :被试根据生活经验中现有的结论 ,直接从大脑的表征中提取内容 ,得出结论 ;5)直觉 :被试没有任何推导过程 ,也没有根据经验的判断过程 ,纯粹凭感觉猜测。在不同概率表征的贝叶斯题中 ,每一类策略的具体内容有一定的差别。在百分比表征中 ,不同专业被试所有策略进行了 χ^2 检验 , $\chi^2_{(8,105)} = 44.551$, $p < 0.000$,差异极度显著 ;在频率表征中 , $\chi^2_{(8,105)} = 54.784$, $p < 0.000$,差异极度显著 ;在概率词表征中 , $\chi^2_{(4,105)} = 9.627$, $p < 0.05$ 。

在贝叶斯判断中,按个体认为概率均值小于 27.95% 的事件判断为不会发生,三道判断题根据题目得出的后验概率均小于 27.95%,其中概率词是通过与百分比的换算而得出,故正确答案是判断为“不是”。不同学科背景的判断结果见表 3,进行 χ^2 检验:百分比, $\chi^2_{(2,105)} = 9.364, p < 0.01$, 差异显著; 频率, $\chi^2_{(2,105)} = 9.333, p < 0.01$, 差异显著; 在频率词表征上差异不显著。

表 3 贝叶斯判断题答案为“不是”的人次分布

表征	数学	物理	中文
百分比	25	12	15
频率	31	20	19
频率词	19	24	27

对被试在完成贝叶斯判断题所使用的策略分类,除了在贝叶斯推算题中出现的 5 种策略外,在胃癌问题的判断上,还出现了“个人意愿”策略。对不同学科背景的被试所用策略进行 χ^2 检验,结果为:百分比, $\chi^2_{(8,105)} = 40.825, p < 0.000$, 差异极度显著; 频率, $\chi^2_{(6,105)} = 13.085, p < 0.05$, 差异显著; 概率词, $\chi^2_{(8,105)} = 40.551, p < 0.000$, 差异极度显著。

4 讨论

4.1 不同学科背景下的个体完成贝叶斯推理的特征分析

不同学科背景的被试在三种表征下贝叶斯推算的正确率差异显著。其中,百分比表征下,了解贝叶斯规则的数学专业被试正确回答人数最高,有 14 人,物理专业 3 人,中文专业无人正确回答; 频率表征下,数学专业有 16 人,物理有 10 人,中文 1 人; 概率词汇表征下,按数字和概率词转换关系,回答“较小”答案属于正确答案,数学专业有 9 人,物理专业有 6 人,中文专业无人回答正确。频率表征更利于贝叶斯问题的解决,如果按照进化论的观点(2003),频率表征之所有优越于其他表征,是因为长期进化的缘故,个体对它更具有适应性,那么中文系和物理系被试在解决频率表征的贝叶斯题时所表现出来的优势应该是一样的,但是事实上,中文系被试正确回答的人数远远不如物理系多。

从被试选择的答案来看,在贝叶斯判断中,回答正确的人数为:百分比有 52 人,频率有 70 人,概率词有 70 人,而在贝叶斯推算中,回答正确的人数为:

百分比有 17 人,频率有 27 人,概率词有 15 人回答正确。在两种任务性质下,被试做出正确反应的差距很大,也就是说,也许被试无法正确估算贝叶斯的后验概率,但却能做出正确的判断。是否可以这样认为,贝叶斯判断更接近现实生活,也许人们不需要了解贝叶斯定理,也能在实践中很大程度上做出正确的反应。

概率词表征的贝叶斯推算题,被试的回答实质上是基于 5 个层次的判断,应该属于判断反应,而非推算。

4.2 不同学科背景下的个体完成贝叶斯推理的策略分析

贝叶斯推算和判断过程中,用到六种策略:正确规则、错误规则、概率简捷式、命题知识、直觉和个人愿望。正确规则在贝叶斯推算任务中使用情况为:百分比表征,数学专业被试使用的有 18 人,一般来说,只有经过专业训练的个体在百分比表征中才能使用正确的规则,在频率表征中,有 34 人使用正确规则,其中,没有学过贝叶斯定理的被试有 12 人使用,物理专业 11 人,中文 1 人。理科学生在统计规则的认识上可能高于文科学生,即使不了解贝叶斯定理,个体具有一定的统计知识,也能使用正确规则。而在判断过程中,百分比 30 人,频率 55 人使用正确规则,远远多于贝叶斯推算的人数,说明在完成贝叶斯判断任务时,个体更容易使用正确规则。

人们使用正确规则将得到准确的结论,个体在不知道正确的规则的情况下,会使用一些跟概率规则有关的策略,这些策略无法得到正确的回答,但却能快速地做出反应,人们把它们称为启发式,研究中,错误规则和概率简捷式就是启发式。

在判断中,概率词表征上用到一种策略:个人意愿,这在贝叶斯推算中也没有出现这种策略。分析贝叶斯推算和判断的六个题项,它们涉及的内容为:感染传染病、智力选拔赛、抑郁症、水果质量、禽流感、胃癌。其中,禽流感、传染病都很容易治愈,智力选拔、抑郁症、水果质量在生活中也不是很重大的事件,只有胃癌才是一种绝症,在被试面临判断此人是否为绝症时,有 17 人认为不希望此人得胃癌,从而判断为不是胃癌,占总体的 16%。在命名这种策略时,可称为“个人意愿”,这也许是“情感启发式”。

5 结论

- 5.1 概率词表征的贝叶斯推算实质是判断过程。
- 5.2 在三种概率表征(百分比、频率、概率词)中,频率表征更适合于正确完成贝叶斯推理,但存在学科背景的差异。
- 5.3 贝叶斯推理的策略分为正确规则、错误规则、命题知识、概率简捷式、直觉和个人意愿。
- 5.4 贝叶斯判断的正确率高于贝叶斯推算,个体在贝叶斯判断上使用的正确规则多于贝叶斯推算过程。

参考文献

1

Edwards W. Conservatism in human information processing. In :B Kleinmuntz ,Ed. Formal representation of human judgment. New York :Wiley ,1968. 17 – 52.

2

Kaheman D ,Tversky A. Subjective probability :A judgment of representativeness. Cognitive Psychology ,1972 3 :430 – 454.

3

Tversky A ,Kahneman D. Evidential impact of base rates. In :D Kahneman ,D. P. Slovic ,A. Tversky ,Eds. Judgment under uncertainty :Heuristics and biases. Cambridge ,England :Cambridge University Press ,1982. 153 – 160.

4

Eddy D M. Probabilistic reasoning in clinical medicine :Problem and opportunities. In :D. Kahneman ,P. Slovic ,A. Tversky ,Eds.

5

Judgment under uncertainty :Heuristics and biases. Cambridge , England : Cambridge University Press ,1982. 249 – 267.

5

Hoffrage U ,Gigerenzer G ,Krauss S ,et al. . Representation facilitates reasoning :what natural frequencies are and what they are not. Cognition 2002 ,78 :47 – 276.

6

Gigerenzer G ,Hoffrage U. How to improve Bayesian reasoning without instruction :Frequency formats. Psychological Review , 1995 ,102 :684 – 704.

7

Sloman S A. The empirical case for two systems of reasoning. In : T. Gilovich ,et al. ,eds. Heuristics and Biases :The Psychology of Intuitive Judgment ,Cambridge University Press ,2002. 397 – 398.

8

Hardman D ,Macchi L. Thinking :psychological perspectives on reasoning ,judgment and decision making. John Wiley ,Sons Ltd , 2003. 128.

9

Windschitl P D. The binary additivity of subjective probability does not indicate the binary complementarity of perceived certainty. Organizational Behavior and Human Decision Processes , 2000 81 :195 – 225.

10

胡竹菁 ,朱丽萍. 人类推理的心理学研究. 北京 :高等教育出版社 2007. 204 ,209.

Effects of Different Probability Representation on Bayesian Reasoning

Yang Li Hu Zhuqing

(Institute of Education , Jiangxi Normal University , Nanchang 330022)

Abstract :The reasoning tasks were Bayesian calculation and judgment to explore the effects of different probability representation , understanding probability rule on the Bayesian inference. The results showed : 1) Process of verbal probability format on Bayesian reasoning was judgment. 2) Frequency format was better than other formats , but there was difference among participants of different subject. 3) Strategies of Bayesian reasoning included right rule , error rule , proposition , probability heuristics , intuition , desire. 4) Judgment was better than calculation , right rules of judgment were more than of calculation.

Key words :Bayesian reasoning ;probability representation ;strategy of reasoning