

# Application of Bayesian Data Analyses in Psychological Research

Xu Shen

Psychology Department, School of Social Science, Tsinghua University, Beijing  
Email: [shenxu2011@gmail.com](mailto:shenxu2011@gmail.com)

Received: Nov. 14<sup>th</sup>, 2013; revised: Nov. 22<sup>nd</sup>, 2013; accepted: Dec. 3<sup>rd</sup>, 2013

Copyright © 2014 Xu Shen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for Hans and the owner of the intellectual property Xu Shen. All Copyright © 2014 are guarded by law and by Hans as a guardian.

**Abstract:** In recent years, the Bayesian Data Analysis method has increasingly received attention from psychological researchers. This method allows them to directly and objectively estimate the probability of a research hypothesis by deriving a posterior probability from the prior distribution and their own research data, or to use the Bayesian factor to directly compare which of two hypotheses can better explain their data. In this literature review paper, we summarized several major strengths of this inference method and its deficiencies. We used examples to illustrate how it can be used in psychological research and summarized some software and textbooks that psychological researchers can use to learn the Bayesian data analysis method. This article aims to introduce a supplementary method into psychological research.

**Keywords:** Bayesian Data Analysis; Bayes Factor; Hypothesis Testing; Posterior Probability

## 贝叶斯推论统计在心理学研究中的应用

沈 序

清华大学社科学院心理系, 北京  
Email: [shenxu2011@gmail.com](mailto:shenxu2011@gmail.com)

收稿日期: 2013年11月14日; 修回日期: 2013年11月22日; 录用日期: 2013年12月3日

**摘 要:** 近年来贝叶斯统计学越来越受到心理学界的关注。该方法的基本逻辑是综合先验信息和实验结果得出一个后验概率, 令研究者可以直接地、客观地检验研究假设, 或使用贝叶斯因子比较哪种研究假设能更好解释实验数据。本文总结了这种方法的优势和劣势, 举例说明如何在心理学研究中应用它, 并介绍了可以使用的软件和教材, 以此作为对当前心理学研究中统计方法的一种补充。

**关键词:** 贝叶斯数据分析; 贝叶斯因子; 假设检验; 后验概率

### 1. 引言

心理学实验数据分析中的核心问题之一是如何根据实验结果做出合理的统计推断。在理论发展中, 统计推断衍生出了两套体系: 频率学派与贝叶斯学派(陈希孺, 1999), 因此产生了虚无假设检验方法与贝

叶斯统计方法这两种不同的推论统计方法。

虚无假设检验(Null Hypothesis Significance Test)是一种证伪方法, 不能直接检验研究者关于实验效应的研究假设。研究者需要设立一个关于实验效应不存在的虚无假设, 用  $H_0$  表示, 而把自己的研究假设作为备择假设, 然后通过对抽样所得的数据进行统计学

计算,进而判断是否拒绝虚无假设。如果能够拒绝虚无假设,则不能拒绝备择假设;如果不能拒绝虚无假设,则只能拒绝备择假设。虚无假设检验方法在心理学界广为使用,Cumming等(2007)对1998年到2006年4月心理学界中的10种主要期刊的统计方法进行抽样分析发现,虚无假设检验的使用率大于96%。但是,近年来这种统计方法的缺陷和不足也引起越来越多的质疑(仲晓波,王希尧和万荣根,2008;甘怡群等,2005)。

贝叶斯统计方法则认为,在进行统计推断问题时,除了使用抽样所得的样本数据之外,也必须考虑抽样前就有的先验分布。也就是说,研究者在进行推论统计时应首先获得关于研究对象的先验概率,再根据抽样所得的实验结果校正先验概率而得到后验概率,从而直接检验研究假设。贝叶斯统计方法是建立在贝叶斯法则的基础上的。根据Fienberg(2006)的总结,贝叶斯法则是英国牧师贝叶斯(Thomas Bayes, 1702~1763)在研究二项分布的参数问题时发现的,并在他去世后由他的朋友普赖斯(Richard Price)整理成文并发表。学界一般认为,拉普拉斯(Pierre-Simon Laplace)在贝叶斯去世十二年后独立地得出了与贝叶斯相同的结论并有所扩展。尽管贝叶斯统计在十八世纪就已正式地登上统计学的舞台,但是因为它所依赖的先验分布受到研究者主观决定的影响,因此直到20世纪30年代随着统计学界对“逆概率”(inverse probability)和先验分布问题的研究进展,以及贝叶斯统计在如序列分析中的应用,贝叶斯统计才得到较为广泛的重视。到20世纪80年代,贝叶斯统计开始在统计学领域占有越来越重要的地位,并在农学、天文学、生物学、生态学、动物学、气象学等科学领域得到应用。在心理学实验中应用贝叶斯统计方法最早至少可以追溯到1963年Edwards, Lindman和Savage的工作。在过去的十年内,贝叶斯统计分析方法因为它在模型比较和多重比较中的优势和它在逻辑上的直观性受到了心理学界更多关注(Dixon, 2003; Lee, 2008; Myung, Foster & Browne, 2000; Rouder, Speckman, Sun, Morey, & Iverson, 2009; 仲晓波, 黄希尧和万荣根, 2008)。另外, Kass和Raftery(1995)则提出了将贝叶斯因子这种比值形式用来作为测量数据对实验效应支持强度的一种标准方法。

本文介绍了贝叶斯统计方法的逻辑与基本计算

过程,并结合具体的研究实例,说明贝叶斯统计方法在心理学研究中的实现以及它如何可以避免p值检验方法存在的不足。最后,本文还介绍了贝叶斯统计在未来的发展趋势及其对心理学研究方法论革新的深远影响。

## 2. 贝叶斯推论统计检验

尽管心理学研究中的推论统计至今很大程度上还依赖于虚无假设检验,一些心理学家已经注意到了这种方法的缺陷并提出了两种修正的方案。第一种是改良方案,即在报告p值的基础上,在统计结果中增加置信区间、标准误、效应量的报告(Cohen, 1994)。这种方案提出十几年之后,Cumming等(2007)对心理学界中的10种主要期刊进行了抽样,试图分析出传统统计方法在心理学界的应用是否有所变化。他们对这10种期刊均从1998年、2003至2004年,2005年至2006年4月这三个时间段中出版的40篇文献进行分析,结果发现在这3个时期中虚无假设检验的使用率分别为97.8%、97.7%和96.9%,但是对于置信区间的使用仍仅有3.7%、9.2%和10.6%。这些数据表明,改良方案并未得到足够的重视和广泛的实施。第二种方案,则是使用贝叶斯统计方法。

### 2.1. 贝叶斯推论统计的基本逻辑

传统的假设检验与贝叶斯推论统计最大的区别在于对“不确定性”这个概念的理解。前者认为实验得到怎样的数据 $X$ 是不确定的,为随机变量,而研究者感兴趣的参数 $\theta$ 则为固定值。后者则认为我们实验观察得到的数据是确定的,而由于我们知识的缺乏或对某个现象的不了解,感兴趣的参数 $\theta$ 是不确定的,为随机变量。

在贝叶斯统计的框架中,研究者用先验分布 $p(\theta)$ 来概括前人和自己对某个感兴趣参数已有的知识。在先验模型下结合实验得到的数据根据似然原理(似然函数包含了样本中的所有信息)研究者得到似然函数 $p(x/\theta)$ 。以简单的抛硬币问题为例,如果研究者想研究硬币是否为均质的(正反面朝上的概率相等)做了 $Z$ 次抛硬币实验,有 $N$ 次朝上,则在这个问题中的似然函数 $p(x/\theta)$ 即为将实验得到的观察值 $x_1, x_2, \dots, x_n$ 代入伯努利分布的概率密度函数,即

$\prod_{i=1}^N \theta^{x_i} (1-\theta)^{1-x_i} = \theta^z \theta^{N-z}$ 。频率学派的虚无假设检验认为,推论统计应该根据整个样本空间做出;而在贝叶斯推论统计方法中,只有实验以后得到的证据才能用于推断,推断只根据在实验后所收集到的观察值做出,似然函数体现的也正是这一点(Glover & Dixon, 2004)。最后贝叶斯统计学家用实验数据更新自己有关某个参数的先验信念,得到有关参数的后验分布  $p(\theta/x)$ 。后验分布  $p(\theta/x)$  与似然函数  $p(x/\theta)$  和先验分布  $p(\theta)$  的乘积呈正比。后验分布的计算可通过贝叶斯法则完成,以连续性变量的问题为例,后验分布为

$$p(\theta/x) = \frac{p(x/\theta)p(\theta)}{p(x)} = \frac{p(x/\theta)p(\theta)}{\int d_{\theta} p(x/\theta)p(\theta)}$$

(Carlin & Louis, 2008; James & Press, 1989; Kruschke, 2010)。

关于先验分布  $p(\theta)$  的确定原则在学界仍有较大的争议。Stern(2005)提出,一般有三种方法获得先验分布:即以均匀分布为代表的无信息先验分布;研究者结合自身研究经验和知识确定的先验分布;以及为了数学上计算方便保持先验分布和后验分布在形式上一致而将似然函数与先验分布设置成共轭分布。无论采用哪种方法,心理学研究中的先验分布需要根据过去公开发表的、可重复的研究结果确定,经过同行评审,并且被广泛接受(Kruschke, 2010)。另外在贝叶斯统计框架中,传统的假设检验方法其实也可以被视作在实验开始前视感兴趣的参数有等可能取不同的值,即先验分布为无信息的均匀分布的一种贝叶斯统计的特殊形式。如果在操作中研究者使用无信息均匀分布为先验分布,那么贝叶斯推论统计与传统统计在功能上相似。

## 2.2. 贝叶斯推论统计的实现

在贝叶斯统计中,研究者数据分析的结论都可以根据后验分布  $p(\theta/x)$  做出。因此得到的结论更加丰富、全面、系统,而非一个简单的“是否”有影响的二分法论断。研究者可以利用后验概率  $p(\theta/x)$  进行参数估计和区间估计。如果  $\theta$  的分布范围较小,则可能值的确信程度较高;如果  $\theta$  的分布范围较大,则  $\theta$  可能值的确信程度较低。在贝叶斯统计的点估计中,可以根据求得的实验数据的统计量  $\theta(y)$  估计  $\theta$ , 普遍

选取统计量的标准是能使损失函数  $E_{\theta,y}(\theta - \theta(y))^2$  最小,基于该损失函数, Louis 和 Carlin(2008)得出在众数、中位数和平均数三个统计量中,后验分布的平均数为最佳点估计值。研究者使用贝叶斯统计进行区间估计,与虚无假设检验假设检验方法中的置信区间类似。贝叶斯统计使用最高后验密度(Highest posterior density)来表示能够覆盖绝大多数后验分布区域的区间,例如 95% 的区间,这种区间  $C$  可表达为:

$$C = \{\theta \in \Theta : p(\theta/x) \geq k(\alpha)\}$$

且  $k(\alpha)$  是满足  $P(C/y) \geq 1-\alpha$  不等式的最大常数(Brooks, 2003, Kruschke, 2010)。

研究者还可以使用贝叶斯因子(Bayes Factor)比较两个或多个模型或假设对实验数据的解释能力(Jeffreys, 1961; Kass & Raftery, 1995)。贝叶斯因子指在一种假设模型下实验数据出现的概率与另一种假设模型下实验数据出现的概率的比值。以心理学研究中经常需要比较的虚无假设模型  $H_0$  和备择假设模型  $H_A$  为例:比率  $\frac{p(D/H_A)}{p(D/H_0)}$  可以用来衡量哪个模型得到更多

的证据支持(Jeffreys, 1961)。如果贝叶斯因子为 2,则说明在模型下  $H_A$  出现实验结果的可能性是模型  $H_0$  下出现同样实验结果的两倍;贝叶斯因子越大,则实验证据就更强有力地支持模型  $H_A$ (Goodman, 2005)。使用贝叶斯因子令研究者可以更好地权衡两个模型的解释数据的力度,而不仅仅是简单地接受或拒绝某一个模型。

## 2.3. 贝叶斯推论统计的优势

与虚无假设检验的方法相比,贝叶斯统计方法在逻辑上更加严密。第一,贝叶斯统计方法用计算得到的后验概率  $p(H_0/X)$  表示基于实验数据和结果  $X$  的前提下  $H_0$  为真的概率,即直观地说明了虚无假设为真的可能性,而不是像传统的假设检验方法那样估计了如果  $H_0$  为真,重复无数次实验出现观察到的实验结果的可能性。

第二,贝叶斯统计中可以使用贝叶斯因子  $\frac{p(D/M1)}{p(D/M2)}$  表示虚无假设条件下实验结果出现的概率与备择假设条件下实验结果出现的概率的比值,直观地说明了虚无假设和备择假设这两种模型中究竟是哪一种能更好地解释实验数据,而不是简单地得出



“接受”或“拒绝”的结论。

第三，贝叶斯统计分析中的虚无假设是备择假设的一个特例，即  $H_0$  是  $H_a$ ,  $a = 0$  时的一个特例，所以研究者可以更加自由地使用贝叶斯统计来比较两个不同的模型，这两个模型可以类似于虚无假设检验中的虚无假设和备择假设，也可以类似于虚无假设检验中的两个备择假设。例如，某研究者在采集了 20 个观察值后发现自变量对因变量存在主效应并想弄清楚这种效应是以怎样的形式作用，就可以对自变量和因变量进行线性回归和二次曲线回归，然后通过计算贝叶斯因子比较这两个模型中谁能更好地解释数据：

$$BF = \frac{p(X/\hat{\theta}_2)}{p(X/\hat{\theta}_1)} = \left( \frac{1-R_1^2}{1-R_2^2} \right)^{\frac{n}{2}}$$

其中  $R_1$  和  $R_2$  代表两种不同回归模型得到的残差(residual) (Glover 和 Dixon, 2004)。

贝叶斯推论统计在实际操作中能更好地降低因为实验者的主观意愿和偏见比如研究者认知的系统性偏差、缺失值处理方法、抽样方法、多重比较校正方法和实验停止标准等影响客观实验结果的强度。

首先，人类在对不确定事件判断时存在的系统偏差之一是对有关先验概率或者基率信息(baseline rate)不敏感(Kahneman 和 Tversky, 1982)，而贝叶斯统计方法中先验分布的确定过程帮助研究者克服这一偏差。其次，贝叶斯统计分析的结果受到似然函数和先验分布影响，而似然函数中的数据都是来自实验已经观察到的数据。这令贝叶斯统计分析不会受到实验者抽样方法和实验停止标准等主观意愿的影响。而虚无假设检验的推断要根据观察到的数据和未观察到的样本空间做出，而样本空间的大小则会可能受到研究者的操控。第三，在使用虚无假设检验进行多重比较时，研究者使用的不同的校正方法可能得到不同的结论，使结论受到研究者的主观影响；在贝叶斯统计过程中，由先验分布和似然函数决定的后验分布涵盖了有关数据的所有信息，研究者可以根据自己的研究兴趣自由地从不同角度对各类实验条件进行比较分析，比较的次数多也不会增加犯第 I 类错误的可能性。最后，贝叶斯推论统计对实验缺失值的处理也更加完善，在该框架下，数据被分为两种类型，未观察到的参数和观察到的数据。遵循以上逻辑，实验中的缺失值可以

被视作一种未被观察到的参数值体现在似然函数中。具体操作可使用数据增强(data augmentation)的算法。鉴于篇幅原因，本文不展开介绍，有兴趣的读者可参考 Tanner 和 Wong(1987)关于如何在贝叶斯统计产生的后验分布中使用该方法的综述文章。

### 3. 贝叶斯统计方法在心理学研究中的应用

为了使读者对贝叶斯统计有更直观的理解，本文举两个具有代表性的例子说明贝叶斯统计方法在心理学研究中的应用，尤其是关于贝叶斯统计方法和虚无假设检验方法可能对相同的实验结果做出不同的解释，其中一个 Bem(2011)关于人类是否具有心灵感应能力的研究，另一个是作者正在进行的研究。

Bem(2011)试图用 10 个实验和虚无假设检验的统计方法证明，人们拥有某种未卜先知的能力，在实验中这种能力表现为对色情刺激的趋近和对负性情绪刺激的回避。他的虚无假设为  $H_0 =$  不存在预知能力，备择假设  $H_1 =$  存在预知能力。在他的实验 1 中，被试的任务是按键选择幕布图片后面是色情图片还是普通图片，结果发现色情图片的选择正确率为 53.1%， $t(99) = 2.51, p = 0.01$ ，这显著地高于随机水平；其它几个实验也用类似的范式得出了显著的结果。但是，Wagenmakers, Wetzels, Borsboom 和 Maas(2011)指出，Bem(2011)没有考虑“ $H_0 =$  不存在预知能力”的先验概率。理论上，人的未卜先知能力缺乏生理心理证据；现实中，人们并不能预知未来，Wagenmakers 等人以 Bem 实验得到的结果 53.1% 计算赌徒在赌场中一年的预期收益，发现收益极高，数值几乎可以让世界上任何一个赌场经营者破产。基于这两个事实， $p(H_0)$  的先验概率极高，Wagenmakers 等用  $9 \times 10^{-20}$  来表示预知能力存在的概率。即使验证性的实验结果 100% 地证实了实验结果在  $H_1 =$  存在预知能力模型下发生，则根据贝叶斯法则可以得出一个后验概率

$$\begin{aligned} p(H_1/D) &= \frac{p(D/H_1)p(H_1)}{p(D/H_0)p(H_0)+p(D/H_1)p(H_1)} \\ &= \frac{0.95 \times 10^{-20}}{0.05(1-10^{-20})+0.95 \times 10^{-20}} \\ &= 1.9 \times 10^{-19} \end{aligned}$$

计算结果表明当考虑实验结果以后对人类存在预知能力的后验信念比先验概率大，但仍旧是很小的数值，

令我们相信这种超能力几乎不可能存在。Wagenmakers 等人(2011)利用贝叶斯 t 检验方法重新分析 Bem(2011)的结果以后发现,上文提到的实验 1 的贝叶斯因子仅为 1.64; 如果按照贝叶斯因子的大小进行衡量,他的 10 个实验中只有 1 个支持他的结论, 3 个支持相反的结论, 其余的只能为他的结论理论提供传闻性的证据 (anecdotal evidence)(Jeffreys, 1961)。

下文将介绍的第二个例子是作者有关情绪图形的视觉搜索研究中的一个实验(Shen 和 Wan, in preparation)。大量的实验证据证明迅速地判断出刺激物中含有的威胁或恐怖的情绪信息对人类的生存与繁衍至关重要。研究证明, 在情绪面孔的识别中含有更多类似 V 形尖角形状的面孔传递出愤怒信息, 而含有更多圆角的面孔则传递出愉悦信息(Aronoff, 2006)。Larson, Aronoff 和 Stearns(2007)等人的研究也证明在视觉搜索中搜索 V 图形比搜索 ^ 图形更快更准确。基于以上发现, 采用类似的范式对中国被试进行重复实验, 利用传统假设检验框架下成对样本 t 检验发现在  $4 \times 4$  的矩形框中在 ^ 背景下搜索 V 的反应时显著地快于在 V 背景下搜索 ^,  $t(17) = 2.11, p = 0.03$ 。

如果对此实验重新进行贝叶斯统计框架下的 t 检验, 那么就需要计算在虚无假设模型  $H_0: \mu_1 - \mu_2 = 0$  下和备择假设模型下  $H_1: \mu_1 - \mu_2 \neq 0$  出现实验观察到数据的概率的比值, 也称为贝叶斯因子

$$\frac{P(x/\mu_1 - \mu_2 \neq 0)}{P(x/\mu_1 - \mu_2 = 0)}$$

同时贝叶斯统计也可以得到两组实验条件下参数平均反应时的  $\mu_1 - \mu_2$  后验分布。在数据分析中, 采用的参数的先验分布为 JZS 先验分布 (Jeffreys-Zellner-Siow Prior), 关于 JZS 先验分布的特性以及在 t 检验情况下使用它的功能以及优势的相关细节可参见(Rouder, Speckman, Sun, Morey, & Iverson, 2009)的工作。

使用 Morey 和 Rouder(2013)开发的贝叶斯因子软件包(BayesFactor)进行检验分析发现贝叶斯因子为 2.385, 根据 Jeffreys(1961)给出的证据属性与贝叶斯因子的对应表格, 2.385 的贝叶斯因子仅能为备择假设模型即在 ^ 背景下搜索 V 的反应时显著地快于在 V 背景下搜索 ^ 提供传闻性证据(anecdotal evidence), 而非重要证据(substantial evidence)或者强有力证据(strong evidence)。利用该软件包笔者也可以计算两组

实验条件下的参数实验效应量  $d = \frac{\mu_1 - \mu_2}{\sigma}$  的后验分布的概率密度函数, 如图 1 所示。

#### 4. 贝叶斯统计方法的评价与前景展望

贝叶斯统计的兴起在未来的心理学发展中可能称为方法论上的一个重要革新。总体而言, 贝叶斯统计从四个方面弥补了虚无假设假设检验的不足。第一, 虚无假设检验方法只能证伪拒绝虚无假设却不能直接提供有关备择假设在多大程度上成立的信息, 而贝叶斯统计方法在比较不同模型的基础上能给出量化的评价标准。第二, 科学理论的建立依赖于可重复的实验, 但是虚无假设的显著性水平检验不利于实验的可重复性。Miller(2009)认为在频率学派的理论框架中实验得到的一些点估计值如平均数或者标准差只能对实验结果的重复概率(replication probability)提供非常少量的信息, 实验重复的概率对后续研究者几乎是不可知的。相反由于在贝叶斯统计中, 研究者得到的是整个后验概率的分布, 他可以将原先实验得到的后验分布作为下一个重复实验的先验分布, 并从这个先验分布中模拟地产生出新的数值。通过比较这些模拟产生出的新数值与原先实验得到的数据, 研究者可以估计实验重复的概率(Kruschke, 2010)。第三, 虚无假设检验中的 p 值仅针对虚无假设下观察到的数据出现的概率, 并且虚无假设常常是关于实验效应为零的假设, 而贝叶斯统计方法中的贝叶斯因子令研究者可以

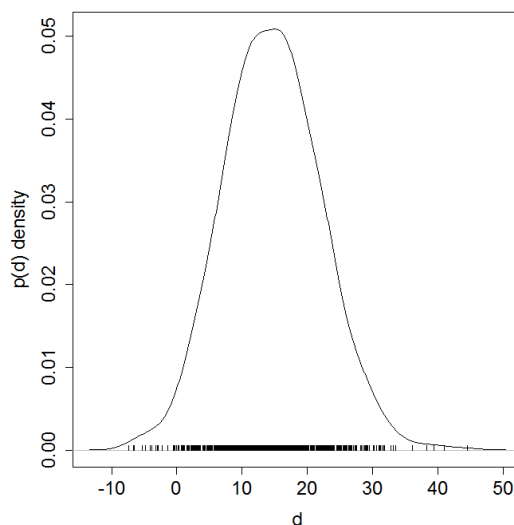


Figure 1. Probability density function of effect size  
图 1. 实验效应量的概率密度函数

自由地比较不同的模型,包括比较虚无假设和备择假设,或是比较两个不同的备择假设。第四,虚无假设检验容易受到实验设计、停止实验的准则、实验中的数据缺失以及实验者选择的多重比较的方法等主观意愿的影响,而贝叶斯统计则对这些因素不敏感。需要强调的是,虽然本文中总结了虚无假设检验的缺陷和贝叶斯统计分析的优势,我们并不认为应摒弃虚无统计检验方法,这种方法在心理学研究中的重要性是不容置疑的。在实际研究的过程中,是选择虚无假设检验的方法,还是贝叶斯统计方法,应联系实际情况,各取所长。在参数模型确定且样本量不大的情况下,贝叶斯方法比虚无假设的方法更具有逻辑上的严密性和操作上的客观性。但是在非参数模型下,先验分布不容易设定,虚无假设检验的方法仍然具有强大的优势。我们建议,在条件允许的情况下,研究者可以对自己的研究数据同时采用虚无假设检验和贝叶斯统计分析,相互验证后得出更为完整、客观的结论。

如果要在心理学研究中使用贝叶斯统计,有两个问题需要引起注意。一方面,贝叶斯统计的计算过程比较复杂,尤其实际研究中产生的后验分布可能非常繁琐,这使针对后验分布的一系列计算很难进行。但是,近几十年来统计软件 R 的普及、BUGS 工具包的开发以及马尔科夫链蒙特卡罗算法(Markov chain Monte Carlo(MCMC))的不断发展比较有效地解决了这一难题。马尔科夫链蒙特卡罗算法是指一种从后验分布中进行抽样,从而求得待估参数近似值的算法。R 软件是免费开源的,并且它可以自由地引用其他功能强大的软件,能够实现蒙特卡罗算法的 OpenBUGS 工具包就是其中之一。蒙特卡罗算法的核心是从任意复杂的后验分布(比如贝叶斯推论统计的后验分布)中产生一系列观察值,然后通过计算这些观察值来高度近似研究者感兴趣的实际量(Smith 和 Roberts, 1993)。BUGS 工具包最早发明于 1996 年,该工具包中的语言使得用户能够在构建完统计模型,即输入先验分布与似然函数后,自动通过蒙特卡罗算法计算后验分布中研究者感兴趣的统计量。仍以抛硬币为例,如果希望结合实际抛硬币的实验数据来计算正面朝上的概率,则 BUGS 语言在 R 平台上的实现方法就是:搭建统计模型,设立先验分布为 Beta 分布,似然函数为伯努利概率密度函数,随后导入实际的实验数据。关于在 R

软件中进行贝叶斯统计的方法和具体步骤,可以参考 Ntzoufras(2011)出版的 *Bayesian Modeling Using WinBUGs*,这是一本兼具统计原理和计算机语言实现方法的高质量教材。Kruschke(2010)所著的 *Doing Bayesian Data Analysis: A Tutorial with R and BUGs* 同样是较好地能够帮助初学者迅速领略到贝叶斯统计的思想核心的高质量教材,另外由于该书的作者是心理学家,因此教材中的许多内容紧扣心理学研究,对如何用贝叶斯统计实现方差分析、t 检验、计算统计效力、确定样本数量都做出了详细的介绍。

另一方面,贝叶斯推论统计中确定先验分布是容易受到质疑的地方。许多批评认为先验分布的确定是个主观性太强的过程。但是笔者想强调的是先验分布并不是随意确定的,而应该受到同行的检验与评审。如果有学者提出反对意见,那么也可以在怀疑者认为更可信的先验分布下重新分析,检验的过程也较为简单,只需要在搭建统计模型时重新设置先验分布的概率密度函数与决定该函数的关键参数,其余计算过程照旧。除此之外,先验分布确定的过程也有利于科学界知识的不断积累,研究者可以更好地把自己的研究与过往与自己相似的研究联系起来,广泛吸收已有的成果与成熟的结论使自己的分析更加深刻,这个过程也能促进科学知识的传承。

当然,贝叶斯统计本身的一些问题也制约着它的发展。第一,如何较为客观地确定先验分布这一问题一直没有得到很好地解决。第二,后验分布概率密度函数的计算较为复杂。然而随着关于对各种先验分布研究的丰富和越来越多的统计工具包的普及,贝叶斯统计的应用也必将越来越广泛。事实也证明,近年来在美国许多以研究为导向型的经济学、生物学、社会学和心理学研究生项目中,项目会至少开设一门关于贝叶斯统计的课程,比如印第安纳大学伯明顿分校(Indiana University Bloomington)心理系在 2013 年开设了贝叶斯统计的入门课程、密歇根大学安娜堡分校(University of Michigan, Ann Arbor)在 2012 年起也开始定期举办关于贝叶斯统计的工作坊。在教学过程中,Wetzels, Grasman 和 Wagenmakers(2012)等人总结认为如何教会学生用一个默认的先验分布以及如何使用 R 语言进行贝叶斯统计的实践是两个较为核心的问题。而学习者只要掌握基础的概率论知识以及 R 软件

的操作方法, 结合实际的研究问题, 经过一定时间的训练, 掌握贝叶斯统计也并非十分耗时耗力的事。

## 致谢

作者感谢宛小昂博士对本文初稿给出的宝贵建议。

## 项目基金

清华大学人文社科振兴基金; 清华大学文化传承创新基金项目; 清华大学亚洲研究中心 2012 青年项目支持。

## 参考文献 (References)

- 陈希孺(1998). 数理统计学小史. *数理统计与管理*, 2 期, 60-64.
- 甘怡群等(2005). *心理与行为科学统计*. 北京: 北京大学出版社.
- 吕佳, 乔克林(2010). 浅谈假设检验中的 P 值. *科学技术与工程*, 10 期, 1671-1815.
- 仲晓波, 黄希尧, 万荣根(2008). 心理学中对假设检验一些批评的分析. *心理科学*, 4 期, 1010-1013.
- Aronoff, J. (2006). How we recognize angry and happy emotion in people, places, and things. *Cross-Cultural Research*, 40, 83-105.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407.
- Brooks, S. P. (2003). Bayesian computation: A statistical revolution. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361, 2681-2697.
- Carlin, B. P., & Louis, T. A. (2008). Bayesian methods for data analysis (Vol. 78). Chapman & Hall/CRC.
- Cohen, J. (1994). The earth is round (P-less-than.05). *American Psychologist*, 49, 997-1003.
- Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., Lo, J., McMenamin, N., & Wilson, S. (2007). Statistical reform in psychology: Is anything changing? *Psychological Science*, 18, 230-232.
- Dixon, P. (2003). The p-value fallacy and how to avoid it. *Canadian Journal of Experimental Psychology*, 57, 189-202.
- Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242.
- Fienberg, S. E. (2006). When did Bayesian inference become "Bayesian". *Bayesian Analysis*, 1, 1-40.
- Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review*, 11, 791-806.
- Goodman, S. N. (2005). Introduction to Bayesian methods I: Measuring the strength of evidence. *Clinical Trials*, 2, 282-290.
- Jeffreys, H. (1961). *Theory of probability*. Oxford: Oxford University

- Press.
- Kahneman, D., Slovic, P., & Tversky, A., Eds. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.
- Kruschke, J. (2010). *Doing Bayesian data analysis: A tutorial introduction with R and BUGS*. Academic Press.
- Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 658-676.
- Larson, C. L., Aronoff, J., & Stearns, J. J. (2007). The shape of threat: Simple geometric forms evoke rapid and sustained capture of attention. *Emotion*, 7, 526-534.
- Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, 15, 1-15.
- Lindley, D. V. (1965). Introduction to probability and statistics from bayesian viewpoint. Part 2 inference. CUP Archive.
- Lindley, D. V. (1975). The future of statistics: A Bayesian 21st century. *Advances in Applied Probability*, 7, 106-115.
- Miller, J. (2009). What is the probability of replicating a statistically significant effect? *Psychonomic Bulletin & Review*, 16, 617-640.
- Morey, R. D., Rouder, J. N., & Morey, M. R. D. (2013). Package "Bayes factor". <ftp://ftp.openjournals.org/pub/cran/web/packages/BayesFactor/BayesFactor.pdf>
- Mr. Bayes, & Price, M. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philosophical Transactions*, (1683-1775), 370-418.
- Myung, I. J., Forster, M., & Browne, M. W. (2000). Special issue on model selection. *Journal of Mathematical Psychology*, 44, 1-2.
- Ntzoufras, I. (2011). Bayesian modeling using WinBUGS (Vol. 698). Wiley.
- Press, S. J., & Press, J. S. (1989). *Bayesian statistics: Principles, models, and applications*. New York: Wiley.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225-237.
- Shen, X., & Wan, X. Visual search for threatening and pleasant shapes.
- Smith, A. F., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B (Methodological)*, 3-23.
- Stern, H. (2005). Bayesian statistics for experimental scientists: ANOVA examples.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-540.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & Van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology*, 100, 426-432.
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology an empirical comparison using 855 t tests. *Perspectives on Psychological Science*, 6, 291-298.
- Wetzels, R., & Wagenmakers, E. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 66, 104-111.