Bayesian Evaluation of Inequality Constrained Hypotheses

Xin Gu[1], Joris Mulder [2], Herbert Hoijtink[1]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] Utrecht University, Department of Methods and Statistics, Utrecht, The Netherlands
`x.gu@uu.nl; h.hoijtink@uu.nl`

[2] Tilburg University, Department of Methods and Statistics, Tilburg, The Netherlands
`j.mulder3@uvt.nl`

Abstract

Bayesian evaluation of inequality constrained hypotheses enables researchers to investigate their expectations with respect to the structure among model parameters. This study proposes an approximate Bayes procedure that can be used for the selection of the best of a set of inequality constrained hypotheses based on the Bayes factor in a very general class of statistical models. A software package BIG is provided such that researchers can use it for the analysis of their own data. Execution of BIG renders the Bayes factor that can determine the support in the data for each candidate hypothesis.
Keywords: Bayes Factor, Inequality Constrained Hypotheses, Normal Approximations.

1 Introduction

Bayesian evaluation of inequality constrained hypotheses has become an attractive alternative for the evaluation of null hypotheses. It allows researchers to focus on the theory or expectation they are interested in and to answer the question: is my theory/expectation supported by the data or not. The theory/expectation will be translated into inequality constrained hypotheses, which can be evaluated by means of Bayes factors. Using the Bayes factor the evidence from the data for a hypothesis against another can be quantified. To evaluate inequality constrained hypotheses in a very general statistical model, this study demonstrates a Bayes procedure in which the posterior distributions of model parameters are approximated by normal distributions.

2 Inequality Constrained Hypotheses

The inequality constrained hypothesis is formally defined as:

$$H_i : \boldsymbol{R}\boldsymbol{\beta} > \boldsymbol{r}, \tag{1}$$

where $\boldsymbol{R}$ is the restriction matrix containing inequality constraints, and $\boldsymbol{\beta}$ and $\boldsymbol{r}$ denote the structural parameter vector and constant vector in H_i, respectively. The inequality constrained hypothesis H_i is often compared to an alternative hypothesis H_a or another inequality constrained hypothesis $H_{i'}$. These hypotheses can be evaluated using Bayes factors, which will be elaborated in the next section.

3 Bayes Factor

The Bayes factor is a Bayesian hypothesis testing criterion. The Bayes factor of an inequality constrained hypothesis H_i against an alternative hypothesis H_a can be represented as the ratio of the posterior and prior probability that the inequality constraints hold [1]

$$BF_{ia} = f_i/c_i \tag{2}$$

where c_i is the proportion of the prior distribution in agrement with H_i, and f_i is the proportion of the posterior distribution in agreement with H_i. It can be seen from equation (2) that the Bayes factor of one hypothesis H_i against another hypothesis $H_{i'}$ can be obtained as $BF_{ii'} = BF_{ia}/BF_{i'a}$. The Bayes factor BF_{12} of H_1 against H_2 quantifies the relative evidence in the data in favor of H_1 against H_2. For example, if $BF_{12} = 10$, this implies that there is 10 times more evidence for H_1 than for H_2.

4 Prior and Posterior Distributions

In order to estimate the complexity c_i and fit f_i, prior and posterior distributions for the parameter $\boldsymbol{\beta}$ have to be specified. For the evaluation of inequality constrained hypotheses in general statistical models, noninformative normal prior distributions and normally approximated posterior distribution are specified. Specifically, the prior distribution is a normal distribution with a mean vector of $\mathbf{0}$ and a diagonal covariance matrix in which the diagonal elements are infinity, i.e., $\pi(\boldsymbol{\beta}) = N(\mathbf{0}, \boldsymbol{\Sigma}_\infty)$. The posterior distribution is approximated by a normal distribution $\pi(\boldsymbol{\beta}|\boldsymbol{X}) \approx N(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_\beta)$, where $\hat{\boldsymbol{\beta}}$ is the estimates of structural parameters, and $\boldsymbol{\Sigma}_\beta$ is their covariance matrix. The estimates and covariance matrix can be obtained using various softwares such as Mplus and OpenBUGS.

5 Computation of the Bayes Factor

As was shown in equation (2), the Bayes factor is the ratio of fit and complexity. The complexity and fit can be estimated by sampling form the prior and posterior distributions, respectively. This can be achieved using the Gibbs sampler for the multivariate normal distribution. Take hypothesis $H_2 : \beta_1 > \beta_2$ for example. If the posterior distribution is

$$\pi(\boldsymbol{\beta}|\boldsymbol{X}) = N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \right), \tag{3}$$

then the fit of H_2 is around $f_2 = 0.33$, which can be seen from Figure 1. After estimating the complexity and fit using Gibbs sampler, the Bayes factor can be directly obtained. Based on this algorism, a software package `BIG` is developed to compute the Bayes factor as long as the estimates and covariance matrix of the parameters are obtained.

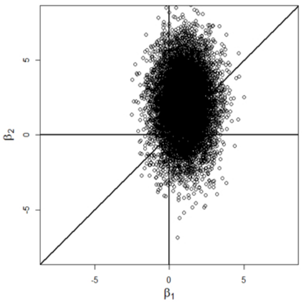

Figure 1: Sample from posterior distribution of β_1 and β_2

References

[1] Hoijtink, H. (2012). *Informative Hypotheses: Theory and Practice for Behavioral and Social Science*. Chapman and Hall/CRC, Boca Raton, FL.