

# Bayesian Evaluation of Inequality and Equality Constrained Hypotheses for Contingency Tables

Irene Klugkist, Olav Laudy, and Herbert Hoijtink

Department of Methods and Statistics, Utrecht University\*

September 7, 2009

## Abstract

In this paper a Bayesian model selection approach is introduced that can select the best of a set of inequality and equality constrained hypotheses for contingency tables. The hypotheses are presented in terms of cell probabilities allowing researchers to test (in)equality constrained hypotheses in a format that is directly related to the data. The proposed method is investigated by several simulation studies and shows good performance. Software that allows researchers to apply the Bayesian approach to their own data is also provided.

*Keywords:* Bayes factor, Contingency Table, Inequality Constrained Hypothesis, Posterior Model Probability

---

\*Irene Klugkist, Olav Laudy and Herbert Hoijtink, Department of Methods and Statistics, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands. E-mail: I.Klugkist@uu.nl, O.Laudy@fss.uu.nl, H.Hoijtink@uu.nl. This research was supported by a grant from the Netherlands Organization for Scientific Research: NWO-VICI-453-05-002.

## Introduction

Advances in Bayesian methodology in recent years have resulted in a great expansion of applications of Bayesian statistics in a wide variety of fields. Bayesian papers now make up a substantial percentage of the papers published in the top statistical journals. However, to understand and apply Bayesian statistical models, a fair amount of technical knowledge is required. In this paper, we discuss Bayesian model selection in the context of contingency tables and (in)equality constrained hypotheses at a non-technical level. The statistical background is presented in two appendices. The approach presented in this paper fits in the steadily increasing attention for Bayesian evaluation of inequality constrained models. See, for example, Hoijtink, Klugkist and Boelen, 2008; Myung, Karabatsos and Iverson (2005); Klugkist, Laudy and Hoijtink (2005); Karabatsos and Sheu (2004).

In psychological research, theories or expectations are often expressed in the form of expected orderings of cell probabilities. To illustrate this and to demonstrate the translation of expectations in inequality constrained hypotheses we will provide two examples. Lessard and Moretti (1998) investigated the relationship between attachment patterns and suicidal ideation. Quality of attachment was assessed using Bartholomew's two-dimensional model of attachment (Bartholomew, 1990), leading to the cross-tabulation of Self Image (positive/negative), Other Image (positive/negative) and Suicidal Ideation (yes/no). Expectations about these data are represented by the following hypotheses:

$H_1$  Respondents with a positive self image have a lower probability of suicidal ideation compared to respondents with a negative self image

$H_2$  Respondents with a positive other image have a lower probability of suicidal ideation compared to respondents with a negative other image

$H_3$  Both effects occur and they add up, that is, the probability of suicidal ideation is lowest for respondents in the positive-positive cell, highest in the negative-negative cell and moderate (in-between) in the other two (positive-negative) cells

These expectations can be translated into hypotheses that are formulated in terms of inequality constraints on (functions of) cell probabilities. Using the standard notation for contingency tables with three variables with respective indexes  $i = 1, \dots, I$ ,  $j = 1, \dots, J$  and  $k = 1, \dots, K$ , the unknown probabilities of a response in cell  $i, j, k$  are denoted  $\pi_{ijk}$ . The notations used in this example are provided in Table 1.

The first hypothesis expresses an expectation about marginal probabilities since the variable Other Image is not included in  $H_1$ . Therefore, summations are made over the categories positive and negative for Other Image. To abbreviate notation, summation over a third variable is denoted

Table 1: Notations for the Suicidal Ideation Example

Self Image	Other Image	Suicidal Ideation	
		yes	no
pos	pos	$\pi_{111}$	$\pi_{112}$
	neg	$\pi_{121}$	$\pi_{122}$
neg	pos	$\pi_{211}$	$\pi_{212}$
	neg	$\pi_{221}$	$\pi_{222}$

by a '+'-sign in the subscripts, e.g.  $\pi_{1+1} = \pi_{111} + \pi_{121}$ , for the cross classification of positive Self Image and the response yes for the variable Suicidal Ideation. Furthermore, the hypothesis states that for positive Self Image the probability of answering yes (i.e.,  $\pi_{1+1}/\pi_{1++}$ ) is smaller than the probability of answering yes for negative Self Image (i.e.,  $\pi_{2+1}/\pi_{2++}$ ). Note that the notations in the denominators are used to denote further summation over the yes and no categories of Suicidal Ideation.

In a similar fashion, the other two hypotheses are translated into inequality constrained hypotheses, leading to:

$$\begin{aligned}
 H_1 : \quad & \frac{\pi_{1+1}}{\pi_{1++}} < \frac{\pi_{2+1}}{\pi_{2++}}, \\
 H_2 : \quad & \frac{\pi_{+11}}{\pi_{+1+}} < \frac{\pi_{+21}}{\pi_{+2+}}, \\
 H_3 : \quad & \frac{\pi_{111}}{\pi_{11+}} < \left\{ \frac{\pi_{121}}{\pi_{12+}}, \frac{\pi_{211}}{\pi_{21+}} \right\} < \frac{\pi_{221}}{\pi_{22+}}.
 \end{aligned}$$

The brackets in  $H_3$  are used to denote that the first term is restricted to be smaller than both terms in the brackets, and, both terms in the brackets are restricted to be smaller than the last term. Evaluation of these hypotheses is not straightforward with standard approaches. In the discussion of this paper the possibilities and limitations of existing classical methods will be discussed. Especially, the mutual comparison of non-nested hypotheses is problematic in a non-Bayesian framework. With the Bayesian model selection approach presented in this paper, however, the support for each hypothesis under consideration can be mutually compared, providing the information which of the three hypotheses receives most support.

A second example, cross-classifies Number of Siblings with Happiness. The notations for this example are presented in Table 2. Vermunt (1991) investigated whether respondents with more siblings are happier. Although not investigated in the original research, one can easily imagine a competing hypothesis, like, for instance, an increase in Happiness up to a certain number of siblings, let's say 5, and a decrease afterwards (i.e., 'it is nice to have siblings but one can also have too many'). These expectations can be translated in inequality constrained hypotheses in several ways. As an example,

Table 2: Notations for the Happiness with Siblings Example

		Number of Siblings				
		0-1	2-3	4-5	6-7	8+
Happiness	not too happy	$\pi_{11}$	$\pi_{12}$	$\pi_{13}$	$\pi_{14}$	$\pi_{15}$
	pretty happy	$\pi_{21}$	$\pi_{22}$	$\pi_{23}$	$\pi_{24}$	$\pi_{25}$
	very happy	$\pi_{31}$	$\pi_{32}$	$\pi_{33}$	$\pi_{34}$	$\pi_{35}$

we will present hypotheses in terms of local odds ratios. A local odds ratio  $\theta_{ij} = (\pi_{ij}\pi_{i+1,j+1})/(\pi_{i+1,j}\pi_{i,j+1})$ , so, for instance,  $\theta_{11} = (\pi_{11}*\pi_{22})/(\pi_{21}*\pi_{12})$  and  $\theta_{24} = (\pi_{24}*\pi_{35})/(\pi_{34}*\pi_{25})$ . Under the hypothesis of Vermunt ( $H_1$ ), all 8 local odds ratios are expected to be larger than one (positive association between Number of Siblings and Happiness). In the alternative hypothesis some local odds ratios are expected to be larger, and some smaller than 1. This leads to the following formulation of hypotheses:

$$H_1 : \{\theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{21}, \theta_{22}, \theta_{23}, \theta_{24}\} > 1,$$

$$H_2 : \{\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}\} > 1 > \{\theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}\}.$$

Again, the – non-nested – hypotheses can be mutually compared with the Bayesian model selection approach presented in this paper.

These two examples show that expectations of researchers are often stated in terms of *inequality constraints*, e.g. a set of cell probabilities – or a function of cell probabilities, like a (local) odds ratio – is hypothesized to be ordered. In general, hypotheses can be formulated by imposing constraints on cell probabilities using one or more of the operators  $+$ ,  $-$ ,  $*$ ,  $/$ ,  $=$ ,  $>$ ,  $<$  and, 0 and 1. Throughout the paper, several examples are provided and form an illustration of the use of (combinations of) these operators to constrain cell probabilities according to certain theoretical expectations.

In a Bayesian framework, it is very natural to include inequality constraints imposed on model parameters as prior knowledge. However, also some non-Bayesian approaches exist for order constrained hypotheses. An overview of the developments in the classical framework is provided in the discussion. The main limitation of these approaches is that non-nested hypotheses can not be mutually compared, while this is no problem using Bayesian model selection as proposed in this paper.

In the next sections, the Bayesian model selection approach for inequality and equality constrained hypotheses in contingency table analysis is introduced using two relatively simple examples based on a  $2 \times 2$  and a  $2 \times 2 \times 2$  contingency table. This is followed by simulation studies, based on  $2 \times 2 \times 2$  tables, conducted to evaluate the behavior of the proposed method as well as the sensitivity to specification of the prior distribution. This is followed

Table 3: Cross-classification of Internal Assets and Being Sent from Class

Internal Assets	Sent from Class	
	yes	no
low	220	1060
high	96	609

by a more elaborate illustration dealing with delinquency data in a  $4 \times 4 \times 4$  table. The paper is concluded with a discussion.

## Bayesian Model Selection

In this section, Bayesian model selection is introduced in four subsections. In the first part, we will explain the basic ingredients of any Bayesian analysis: likelihood, prior and posterior. For the explanation a simple example dealing with a  $2 \times 2$  table will be used and (in)equality constraints are not yet incorporated. The next part specifically deals with the inclusion of *inequality* constraints and the evaluation of constrained hypotheses. Hypotheses containing *equality* constraints are discussed separately in the third subsection. The section is concluded with the introduction of posterior model probabilities and their interpretation.

### Likelihood, Prior and Posterior

Consider another, simple, example from Nash and Bowen (2002), for which the data are displayed in Table 3. The research question is whether students with low internal assets are more frequently sent from class than students with high internal assets. Data are collected using the School Success Profile (SSP) by Bowen, Richman, Brewster and Bowen (1998). The SSP is a self-administered instrument designed for students in grades 6 to 12. The internal assets index is a dichotomized composite of 10 items that assesses the adolescent’s perception of his or her strength and resources (health, exercises, or involvement in sports). A dichotomous measure was created using a students response to a single SSP item asking whether, during the previous 30 days, the student had been sent from class due to his or her behavior.

Let  $f_{ij}$  denote the observed frequency for cell  $i, j$ . For example, in Table 3,  $f_{11} = 220$ . Let  $\pi_{ij}$  denote the unknown probabilities of a response in cell  $i, j$ , that is,  $\pi_{ij}$  is the probability that an observation sampled at random from the population of interest falls into cell  $i, j$ . Finally, let  $\mathbf{f}$  denote the four frequencies in Table 3 and  $\boldsymbol{\pi}$  the four corresponding probabilities.

The first ingredient is the likelihood of  $\boldsymbol{\pi}$  given  $\boldsymbol{f}$  for a model in which it is assumed that the total sample size is fixed:

$$L(\boldsymbol{\pi}|\boldsymbol{f}) = \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{f_{ij}} \quad (1)$$

with the restriction that  $\sum_i \sum_j \pi_{ij} = 1$ . The likelihood is proportional to the density of the data with the role of data and parameters reversed. It summarizes the information with respect to  $\boldsymbol{\pi}$  that is available in the data. For example, the likelihood that  $\pi_{11} = 220/1985 = .11$ ,  $\pi_{12} = 1060/1985 = .53$ ,  $\pi_{21} = 96/1985 = .05$ , and  $\pi_{22} = 609/1985 = .31$  is rather high because the probabilities are equal to the sample probabilities (the total size of the sample in Table 3 is 1985). The likelihood of  $\pi_{11} = .6$ ,  $\pi_{12} = .1$ ,  $\pi_{21} = .2$ , and  $\pi_{22} = .1$  is much smaller because the probabilities given are not at all supported by the data. Stated otherwise, the likelihood tells us which values of  $\boldsymbol{\pi}$  are more and less supported by the data.

The second ingredient is the prior distribution. The prior distribution summarizes the information with respect to  $\boldsymbol{\pi}$  that is available before observing the data:

$$Pr(\boldsymbol{\pi} \mid \boldsymbol{\alpha}, H_u) \propto \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{\alpha_{ij}-1}. \quad (2)$$

Note that this is the prior distribution for the unconstrained hypothesis  $H_u : \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$ , where the comma's are used to denote that no constraints are imposed on the four cell probabilities. As can be seen the mathematical form of the prior distribution is equal to the mathematical form of the likelihood. The equivalence of the mathematical form is called "conjugate". Conjugate prior distributions are a common and convenient choice because (as will be shown below) they render simple posterior distributions that can easily be processed. As elaborated in Albert and Gupta (1983), there are other ways to specify prior distributions for contingency tables. However, as will be shown later in the paper, Bayesian evaluation of constrained hypotheses based on the conjugate prior (2) performs adequately.

The main difference between the prior distribution and the likelihood is the use of  $\alpha_{ij}$  instead of  $f_{ij}$ . Where  $f_{ij}$  is the observed frequency of cell  $i, j$ ,  $\alpha_{ij}$  can be seen as the number of persons in cell  $i, j$  before observing the data, that is, the prior information with respect to  $\pi_{ij}$  before observing the data. In statistics the prior distribution (2) is known as the Dirichlet distribution (Gelman, Carlin, Stern and Rubin, 2004, pp. 576-577). It is a statistical convention to specify this distribution using  $\alpha_{ij} - 1$ . A common choice is to choose  $\boldsymbol{\alpha} = [1, 1, 1, 1]$ , that is, before observing the data our prior knowledge is equivalent to 1 person in each cell of the contingency table. With this choice each  $\boldsymbol{\pi}$  for which  $\sum_i \sum_j \pi_{ij} = 1$  has the same prior density  $Pr(\boldsymbol{\pi} \mid \boldsymbol{\alpha}) \propto 1$ , that is, each  $\boldsymbol{\pi}$  is a priori equally likely. As will

be shown in the section dealing with sensitivity, using  $\boldsymbol{\alpha} = \mathbf{1}$  has good properties if the goal is to select the best of a set of (in)equality constrained hypotheses.

The third ingredient is the posterior distribution. It combines the information with respect to  $\boldsymbol{\pi}$  available in the data and the prior distribution:

$$P(\boldsymbol{\pi}|\mathbf{f}, \boldsymbol{\alpha}, H_u) \propto \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{f_{ij} + \alpha_{ij} - 1}. \quad (3)$$

As can be seen, the total sample size per cell is the sum of the observed frequencies  $f_{ij}$  and the prior frequencies  $\alpha_{ij}$ . The higher the value of  $P(\boldsymbol{\pi}|\mathbf{f})$  the more the information in the data combined with the prior information support the value of  $\boldsymbol{\pi}$  at hand. The posterior distribution (3) illustrates two of the features that distinguishes Bayesian analyses from classical analyses: the use of prior information, and, the fact that  $\boldsymbol{\pi}$  is considered to be random.

The focus of this paper is on selection of the best of a set of constrained hypotheses, hence, parameter estimates and confidence intervals will not be presented. However, they are easily obtained from a sample from the posterior distribution (Hojtink, 2000).

## Inequality Constrained Hypotheses

In the previous section we introduced the use of the conjugate prior containing hardly any prior information (by setting  $\boldsymbol{\alpha} = \mathbf{1}$ ) and an equal amount of prior support for each cell (since  $\alpha_{11} = \dots = \alpha_{IJ}$  for all  $i, j$ ). So far, no constraints have been imposed on the model parameters and therefore this prior distribution is called the unconstrained, or also, the encompassing prior. This is the prior used for what we call the unconstrained hypothesis or unconstrained model  $H_u : \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$ .

Nash and Bowen (2002) hypothesized that students with low internal assets are more frequently sent from class than students with high internal assets. This expectation can be formalized by the inequality constrained hypothesis:  $H_1 : \pi_{11}/(\pi_{11} + \pi_{12}) > \pi_{21}/(\pi_{21} + \pi_{22})$ . Imposing the inequality constraints on (a function of) the cell probabilities  $\pi_{ij}$  allows researchers to test hypotheses in a format that is directly related to the data. It is also common to think about cross-classified data in terms of odds ratios, where the odds ratio  $\theta = (\pi_{11} * \pi_{22})/(\pi_{12} * \pi_{21})$  can be used to specify, for instance, the expectation of a positive association in a  $2 \times 2$  table, that is:  $H_1 : \theta > 1$ . Note that an odds ratio is also a function of cell probabilities. In the remainder of the paper both notations (odds ratios and cell probabilities) will be used for the formulation of hypotheses. In the software provided with this paper one always has to formulate hypotheses in terms of (functions of) cell probabilities.

At this point, it is important to stress that this  $2 \times 2$  illustration is too small to demonstrate the benefits and flexibility of the Bayesian approach,

since for such a simple example and the constrained hypothesis  $H_1$  also other, classical, approaches could be used. It is included, however, to explain the Bayesian procedure in a simple set-up. In the sections that follow, more elaborate examples will be provided, dealing with  $2 \times 2 \times 2$  and  $4 \times 4 \times 4$  contingency tables. In the discussion of this paper an overview of alternative methods for constrained contingency tables is provided.

Bayesian model selection requires a prior parameter distribution under each model. For inequality constrained hypotheses we use the fact that a hypothesis formulated using inequality constraints is nested in the unconstrained model. Compared to the unconstrained model  $H_u$ , hypothesis  $H_1$  puts additional, *a priori*, information on the four model parameters. The assumed ordering is incorporated in the prior distribution  $Pr(\boldsymbol{\pi}|\boldsymbol{\alpha}, H_u)$  by truncation of the parameter space, that is, parts that are not in agreement with the constraints of the hypothesis are given a prior density of zero. This leads to the prior

$$Pr(\boldsymbol{\pi}|\boldsymbol{\alpha}, H_t) = \frac{Pr(\boldsymbol{\pi}|\boldsymbol{\alpha}, H_u)I_{H_t}}{\int Pr(\boldsymbol{\pi}|\boldsymbol{\alpha}, H_u)I_{H_t}d\boldsymbol{\pi}}, \quad (4)$$

for any inequality constrained hypothesis  $H_t$ . The indicator function  $I_{H_t}$  equals one if  $\boldsymbol{\pi}$  is in agreement with the constraints, and zero otherwise. This shows that the prior distribution of the constrained hypothesis has the same shape as the unconstrained prior (i.e., it is proportional) in the area in agreement with the constraints. The denominator is required to make the constrained prior a proper density function that integrates to one. Stated differently, the integral is the normalizing constant that is required to compensate for the part that is truncated (i.e., where the density is set to zero according to the inequality constraints).

In a similar fashion, the resulting posterior distribution of a constrained hypothesis  $H_t$  is proportional to the unconstrained posterior for parameter values in agreement with  $H_t$  and zero otherwise. Again, a normalizing constant is required to compensate for the truncated part of the unconstrained posterior, leading to:

$$P(\boldsymbol{\pi}|\mathbf{f}, \boldsymbol{\alpha}, H_t) = \frac{P(\boldsymbol{\pi}|\mathbf{f}, \boldsymbol{\alpha}, H_u)I_{H_t}}{\int P(\boldsymbol{\pi}|\mathbf{f}, \boldsymbol{\alpha}, H_u)I_{H_t}d\boldsymbol{\pi}}. \quad (5)$$

The two normalizing constants play an important role in the model selection procedure and will from now on be denoted  $c_t$  and  $d_t$  for prior and posterior, respectively.

To evaluate inequality constrained hypotheses, the Bayesian model selection criterion used is the Bayes factor (Kass, 1993; Kass and Raftery, 1995). A Bayes factor is the ratio of two marginal likelihoods, where the marginal likelihood of a hypothesis is a Bayesian measure of the degree of support in the data for that hypothesis. The term *marginal* is used because it is the

likelihood of the data integrated over the prior distribution, that is:

$$m(\mathbf{f}|H) = \int L(\boldsymbol{\pi}|\mathbf{f})Pr(\boldsymbol{\pi}|\boldsymbol{\alpha}, H)d\boldsymbol{\pi},$$

for any hypothesis  $H$ . For a general comprehensive introduction to Bayesian statistics including Bayesian model selection, we refer to Congdon (2001) and Lee (1997). For estimation of the marginal likelihood, see also, for instance, Chib, 1995; Gelfand and Smith, 1990; Verdinelli and Wasserman, 1995.

For the evaluation of inequality constrained hypotheses, however, we use an alternative formulation of the marginal likelihood (Chib, 1995):

$$m(\mathbf{f}|H) = L(\boldsymbol{\pi}|\mathbf{f})Pr(\boldsymbol{\pi}|\boldsymbol{\alpha}, H)/P(\boldsymbol{\pi}|\mathbf{f}, \boldsymbol{\alpha}, H).$$

Using this notation, the Bayes factor of any inequality constrained hypothesis  $H_t$  with the unconstrained  $H_u$  can be written as:

$$BF_{tu} = \frac{L(\boldsymbol{\pi}|\mathbf{f})Pr(\boldsymbol{\pi}|\boldsymbol{\alpha}, H_t)/P(\boldsymbol{\pi}|\mathbf{f}, \boldsymbol{\alpha}, H_t)}{L(\boldsymbol{\pi}|\mathbf{f})Pr(\boldsymbol{\pi}|\boldsymbol{\alpha}, H_u)/P(\boldsymbol{\pi}|\mathbf{f}, \boldsymbol{\alpha}, H_u)} = \frac{c_t}{d_t}. \quad (6)$$

Due to the nesting of  $H_t$  in  $H_u$  and the previously presented formulation of prior parameter distributions as truncated versions of the prior for  $H_u$ , this equation can be simplified into the ratio of the two normalizing constants  $c_t$  and  $d_t$ : For any value  $\boldsymbol{\pi}^*$  that belongs to  $H_t$ , the likelihood terms cancel, and the priors and posteriors cancel up to the two normalizing constants. This can be seen by inserting (4) and (5) in (6); a more formal derivation is provided in Appendix A (and see also Hoijtink, Klugkist and Boelen, 2008; Klugkist, Kato and Hoijtink, 2005).

To get an estimate of  $BF_{tu}$ , all that is needed are estimates of  $c_t$  and  $d_t$ . Since the normalizing constants are needed to compensate for the truncated part of prior and posterior distribution, they are the inverse of the proportions that are truncated. If, for instance, 50% of the unconstrained prior distribution is not in agreement with the constraints of a hypothesis  $H_t$ —and the density in this area thus set to zero in the constrained prior—the prior density in the remaining area must be multiplied by 2, that is,  $c_t = 2$ . Stated differently, the proportion of the unconstrained prior in agreement with the constraints of  $H_t$  provides the value  $c_t^{-1}$ . Similarly, the proportion of the unconstrained posterior in agreement with the inequality constraints is equal to  $d_t^{-1}$ . These two proportions are easily estimated by sampling from both unconstrained prior and posterior and counting the number of iterations in agreement with  $H_t$ . The sampling procedure is not elaborated here but can be found in Appendix B.

## Equality Constrained Hypotheses

For models formulated using equality constraints,  $c_t^{-1}$  and  $d_t^{-1}$  can *not* be interpreted as the respective proportions of the unconstrained prior and posterior distribution in agreement with the constraints, because there is a zero

probability that two or more probabilities  $\pi_{ij}$  are exactly equal. However, as elaborated in Dickey (1971), Gunel and Dickey (1974) and Verdinelli and Wasserman (1995), the Bayes factor for an equality constrained versus unconstrained hypothesis is still a ratio of a posterior and a prior component. Considering a simple equality hypothesis stating  $H_1 : \theta = 1$ , the Bayes factor comparing  $H_1$  with a hypothesis without any constraint on  $\theta$  (denoted as,  $H_u : \theta$ ), is the ratio of the posterior and prior density evaluated at  $\theta = 1$  in the unconstrained posterior and prior distribution, respectively. The interested reader is referred to the last paragraph of Appendix A for the derivation of this result.

Our approach to equality constrained hypotheses is based on an approximating hypothesis, where equality constraints are replaced by *about equality* constraints. For instance,  $H_1$  is approximated by  $H_{1*} : \theta \approx 1$ , evaluated using  $H_{1*} : |\theta - 1| < \delta_0$  for a small value  $\delta_0$ . According to Berger and Delampady (1987) a Bayes factor based on  $\theta = 1$  is indistinguishable from the Bayes factor based on  $|\theta - 1| < \delta_0$  if  $\delta_0$  is small enough. With this approximate hypothesis the dimension of the unconstrained hypothesis is preserved and therefore  $d_t^{-1}$  and  $c_t^{-1}$  can again be interpreted as the proportions of the unconstrained posterior and prior in agreement with the constraints of the hypothesis.

Estimation of these proportions based on samples from unconstrained prior and posterior will, however, be inefficient if  $\delta_0$  is close to zero. Therefore, a stepwise procedure is used that starts with a larger value for  $\delta$  and a stepwise decrease of this value until the desired  $\delta_0$  is obtained. Adopting this approach requires several steps in the sampling scheme as well as constrained sampling. All technical details are provided in Appendices A and B.

With a limited number of steps, we can let  $\delta_0 \rightarrow 0$ , providing a very good approximation of the strict equality constraint. Note that other approaches exist that can handle hypotheses specified using equality constraints (see, for example, Verdinelli and Wasserman, 1995). However, ours is currently the only unified approach that can handle hypotheses specified using both inequality and equality constraints, and therefore also hypotheses that contain a mix of equality and inequality constraints. The simulation studies presented later in this paper will furthermore show that the performance of our approach is good.

## Bayes Factors and Posterior Model Probabilities

The Bayes factor is a model selection criterion that combines fit and complexity, that is, it functions as an automatic Ockham's razor. This means that if two hypotheses fit the data equally well, the 'smallest' hypothesis receives more support, where, in this context, smallest refers to the hypothesis with the smallest non-zero density parameter space. Since the value

Table 4: Hypotheses and Results for the Nash and Bowen Data

Hypothesis	$BF_{tu}$
$H_1 : \pi_{11}/(\pi_{11} + \pi_{12}) > \pi_{21}/(\pi_{21} + \pi_{22})$	1.96
$H_2 : \pi_{11}/(\pi_{11} + \pi_{12}) = \pi_{21}/(\pi_{21} + \pi_{22})$	2.59
$H_u : \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$	

$c_t^{-1}$  reflects the part of the unconstrained prior parameter space that is in agreement with the constraints of  $H_t$ , it can be interpreted as the size or complexity of the constrained hypothesis. Since the posterior support for  $H_t$ , measured by  $d_t^{-1}$ , takes the observed data into account it can be interpreted as a measure of fit. As we saw in (6), the Bayes factor comparing a constrained with the unconstrained hypothesis is the ratio of fit ( $d_t^{-1}$ ) and complexity ( $c_t^{-1}$ ).

As an example, consider  $H_1 : \theta_1 > \theta_2$  versus the hypothesis without constraints  $H_u : \theta_1, \theta_2$ , where  $\theta_1$  and  $\theta_2$  represent two odds ratios, and the Bayes factor  $BF_{1u}$ . The value of  $c_1^{-1} = .5$  since half of the parameter space  $(\theta_1, \theta_2)$  has non-zero density under  $H_1$ . For two sample odds ratios (OR) that are equal, exactly half of the posterior will fall in the area  $\theta_1 > \theta_2$ . In this case, there is no support for  $H_1$  but also no clear evidence against it. This is correctly represented by the resulting Bayes factor: since both  $c_1^{-1}$  and  $d_1^{-1}$  are .5, we obtain  $BF_{1u} = 1$ . For sample odds ratios that are not in agreement with  $H_1$  (i.e.,  $OR_1 < OR_2$ , where  $OR_1$  and  $OR_2$  denote the observed odds ratios in the sample) less than half of the unconstrained posterior will fall in the area of  $H_t$  leading to  $BF_{1u} < 1$ . On the other hand, for a sample with  $OR_1 > OR_2$  the fit  $d_1^{-1}$  will be larger than .5 and the resulting  $BF_{1u} > 1$ . The larger the difference between  $OR_1$  and  $OR_2$ , the larger the part of the unconstrained posterior that falls in the  $H_1$  area, resulting in larger values for  $d_1^{-1}$  and  $BF_{1u}$ . This appropriate behavior of the Bayes factor will be further demonstrated in the simulation studies presented in the remainder of this paper.

For the Nash and Bowen data presented in Table 3, two hypotheses are evaluated. The first ( $H_1$ ) states that children with low internal assets are more frequently sent from class than children with high assets, whereas  $H_2$  states that there is no difference between the two groups. In Table 4, each hypothesis is formulated in terms of constrained cell probabilities and the resulting Bayes factors,  $BF_{1u}$  and  $BF_{2u}$ , are presented. The results show that  $H_1$  is 1.96 times as likely as  $H_u$ , for  $H_2$  this value is 2.59. A direct comparison of  $H_1$  with  $H_2$  can be made using:

$$BF_{t't'} = \frac{BF_{tu}}{BF_{t'u}}. \tag{7}$$

This results in  $BF_{21} = 2.59/1.96 = 1.32$ , that is, after observing the data and accounting for both fit and complexity,  $H_2$  is 1.32 times as likely as  $H_1$ .

A set of Bayes factors can also be translated into posterior model probabilities which are numbers on a 0 – 1 scale and provide a measure of the relative support for each hypothesis in a set of two or more hypotheses (see, for example, Lee (1997, pp. 117-138), Kass and Raftery (1995) and Congdon (2002, pp. 469-472)). If it is assumed that a priori each of the hypotheses under investigation is equally likely, the information contained in a set of posterior probabilities is equal to the information in the corresponding set of Bayes factors. If the set of hypotheses of interest consists of the unconstrained model  $H_u$  and several alternative hypotheses  $H_t$  ( $t = 1, \dots, T$ ), the posterior probabilities are computed using:

$$P_{H_t} = \frac{BF_{tu}}{1 + BF_{1u} + \dots + BF_{Tu}}. \quad (8)$$

For the three hypotheses in our small example, the posterior probabilities are  $P_{H_1} = .35$ ,  $P_{H_2} = .47$ , and  $P_{H_u} = .18$ . As can be seen  $H_2$  has a higher posterior model probability than  $H_1$ . However, the evidence in favor of  $H_2$  is not such that  $H_1$  can be disqualified. Both hypotheses are supported more than  $H_u$ , but a definite choice between both hypotheses would be premature. Stated otherwise, after observing the data it is still unclear whether or not there is a positive association between internal assets and being sent from class.

Two important issues remain. First, are the results sensitive to the choice of the prior distribution? Second, how precise are the sampling based estimates? Prior sensitivity will be investigated in a separate section. Sampling errors are addressed here and in Table 5.

The sampling error in the estimates of  $c_t$ ,  $d_t$  and subsequent Bayes factors can be controlled by taking large samples from the prior and posterior distribution, respectively (see Appendix B for the sampling procedure). Table 5 shows the Bayes factors and posterior model probabilities obtained after three runs of our software. As can be seen, with a sample size of 10,000 from prior and posterior for each step in the analysis (see (18) in Appendix A where the step-wise computation of Bayes factors is elaborated) the sampling error in the resulting Bayes factors and posterior model probabilities is negligible for all practical purposes. A similar table presenting three different runs will also be presented in the other examples in this paper.

## An Extension of the Nash and Bowen Example

Another research interest of Nash and Bowen (2002) is whether the relation between Internal Assets and Sent from Class differs between males and females. Table 6 displays the data. Note that summing over gender gives

Table 5: Investigation of Sampling Error

	Run 1		Run 2		Run 3	
	$BF_{tu}$	$P_{H_t}$	$BF_{tu}$	$P_{H_t}$	$BF_{tu}$	$P_{H_t}$
$H_1$	1.96	.35	1.97	.34	1.97	.35
$H_2$	2.59	.47	2.87	.49	2.70	.48
$H_u$		.18		.17		.18

Table 6: Cross-classification of Internal Assets and Being Sent from Class by Gender

Gender	Internal Assets	Sent from Class	
		yes	no
female	low	79	629
	high	18	323
male	low	141	431
	high	78	286

Table 3. To abbreviate notation, in this example the hypotheses are formulated in terms of odds ratios instead of cell probabilities:  $\theta_f$  denotes the relation between Internal Assets and Sent from Class for females,  $\theta_m$  for males. A positive relationship is represented by an odds ratio with a value larger than one. The observed odds ratios are 2.25 and 1.19 for females and males respectively. This means that for both females and males there is a positive relationship between Internal Assets and Sent from Class in the sample.

The hypotheses that will be investigated are presented in Table 7. Note that we do not include all possible combinations of (in)equality constrained hypotheses in the set, but rather follow an approach where the researcher is invited to provide well-considered hypotheses. These hypotheses may be combinations of theory driven predictions, personal reflection and hypotheses from peer researchers with opposite theories about the data. The hypothesis  $H_u$  is included not because it is so interesting in itself, but because it provides a benchmark for the constrained hypotheses: if a constrained hypothesis receives less support than the unconstrained hypotheses, the constraints are not supported by the data.

Table 7 displays the  $BF_{tu}$  for each hypothesis under investigation computed using  $\alpha = 1$ . Like for the previous example (see Table 5) repeated runs of our software show that the sampling error with a sample size of 10,000 for each step is negligible for all practical purposes. For  $H_1$  and  $H_2$  the Bayes factor is smaller than 1. This implies that the constraints in both hypotheses are not supported by the data. For  $H_3$ ,  $H_4$  and  $H_5$  the

Bayes factor is larger than 1. As can be seen  $H_5$  has the largest posterior model probability (.50), stated otherwise, the expectation that the relation between Internal Assets and Sent from Class is stronger for girls than for boys and positive for both gender groups is supported most by the data.

In the next two sections the format of this example will be used in several simulations, in which the behavior of posterior model probabilities, and, the sensitivity with respect to the specification of the prior distribution will be investigated.

## Behavior of the Posterior Model Probability

In this section the effect of sample size and size of the odds ratio on the posterior model probability is explored for the format of the data in Table 6. This provides some insight into the behavior of the posterior model probability when the goal is to select the best of a set of constrained hypotheses. In this section we will again use  $\alpha = \mathbf{1}$ , in the next section this choice will be motivated.

To explore the behavior of the posterior model probability for various sample sizes and odds ratios, data are constructed as follows. For a specific sample size, say 40, the first data set is constructed to have an odds ratio of exactly 1 for groups 1 and 2. As can be seen in Table 8 this implies that each of the eight cells contains a frequency of 5. The odds ratios  $\theta_1$  and  $\theta_2$  for C=1 and C=2, respectively are defined as  $\pi_{111}\pi_{221}/(\pi_{211}\pi_{121})$  and  $\pi_{112}\pi_{222}/(\pi_{212}\pi_{122})$ . To increase the difference between the two odds ratios in the next data set, for C=2, the diagonal cells are increased with 1 and the off-diagonal cells are decreased with 1. In the next data set, the diagonal cells are again increased with 1, and the off-diagonal cell are decreased with 1, etc. The resulting sequence of tables is displayed in Table 8. Following a similar procedure, a sequence of odds ratios smaller than 1 can be constructed via a stepwise subtraction of 1 from the diagonal cells and addition of 1 to the off-diagonal cells. Note that this sequence is not

Table 7: Hypotheses and Results for the Extended Nash and Bowen Example

	Run 1		Run 2		Run 3	
	$BF_{tu}$	$P_{H_t}$	$BF_{tu}$	$P_{H_t}$	$BF_{tu}$	$P_{H_t}$
$H_1: \theta_f = \theta_m = 1$	.39	.03	.56	.04	.42	.03
$H_2: \theta_f = \theta_m$	.24	.02	.20	.01	.26	.01
$H_3: \theta_f > 1, \theta_m > 1$	3.29	.24	3.44	.24	3.45	.25
$H_4: \theta_f > \theta_m$	1.98	.14	1.94	.14	1.97	.14
$H_5: \theta_f > \theta_m > 1$	6.81	.50	6.96	.49	6.66	.48
$H_u: \theta_f, \theta_m$		.07		.07		.07

presented in Table 8. This procedure is followed for sample size 40, 80, 200, 400, 800 and 4000. For the larger sample sizes, the (off) diagonal steps are changed as indicated between brackets 40(1), 80(5), 200(10), 400(25), 800(50) 4000(100).

For each constructed data set, the hypothesis  $H_1 : \theta_1 = \theta_2$  is tested against  $H_u : \theta_1, \theta_2$ . The results are displayed in the first plot in Figure 1. The odds ratio for group 1 is 1, the odds ratio for group 2 is displayed on the x-axis. Note that the figure for the sequence of odds ratios smaller than one is the mirror image of the top plot in Figure 1, that is, the posterior model probabilities for  $H_1$  with  $\theta_2 \in 1, 2.25, 5.44, \dots, 16$  are identical to those for  $H_1$  with  $\theta_2 \in 1/1, 1/2.25, 1/5.44, \dots, 1/16$ .

The figure shows that for  $\theta_1 = \theta_2 = 1$  (i.e.,  $H_1$  is true) the posterior model probability of  $H_1$  increases with increasing sample size. For  $\theta_1 \neq \theta_2$  two things can be observed. First of all, for each sample size it holds that the larger the difference between  $\theta_1$  and  $\theta_2$  the smaller the posterior model probability of  $H_1$ . Secondly, the larger the sample size, the faster the posterior probability of  $H_1$  decreases with an increasing difference between  $\theta_1$  and  $\theta_2$ .

For the same constructed data sets, in the plot at the bottom of Figure 1, the hypothesis  $H_3 : \theta_1 > \theta_2$  is compared with  $H_u : \theta_1, \theta_2$ . As can be seen, the larger  $\theta_2$  the smaller the posterior model probability of  $H_3 : \theta_1 > \theta_2$ . Furthermore, for  $\theta_2$  larger than  $\theta_1 = 1$  (which is not predicted by  $H_3$ ) it holds that the larger the sample size, the smaller the posterior probability of  $H_3$ . For  $\theta_2$  smaller than  $\theta_1 = 1$  (which is predicted by  $H_3$ ) it holds that the larger the sample size, the larger the posterior probability of  $H_3$ . These are all sensible properties and support the usefulness of posterior model probabilities for the selection of the best of a set of constrained hypotheses.

Table 8: Constructed Data for a  $2 \times 2 \times 2$  Contingency Table and  $N = 40$

C	A	B		B		B		B		
		1	2	1	2	1	2	1	2	
1	1	5	5	5	5	5	5	...	5	5
	2	5	5	5	5	5	5	...	5	5
2	1	5	5	6	4	7	3	...	9	1
	2	5	5	4	6	3	7	...	1	9
$\theta_1$		1		1		1		...	1	
$\theta_2$		1		2.25		5.44		...	16	

## Sensitivity to the Prior

In all previous analyses the prior distribution of  $\boldsymbol{\pi}$  was specified using  $\boldsymbol{\alpha} = \mathbf{1}$ . This corresponds to assuming that a-priori (i.e. before observing the data) there is one observation or person in each cell of the contingency table. This implies that a-priori each combination of values of  $\boldsymbol{\pi}$  that sum to 1.0 is equally likely. It is possible to specify other values for  $\boldsymbol{\alpha}$ . Values larger than 1 imply that a-priori values for the elements of  $\boldsymbol{\pi}$  centered around  $1/IJK$  are considered to be more likely, values smaller than 1 imply that the a-priori most likely values are not centered around  $1/IJK$ .

Sensitivity of the posterior model probabilities to the specification of the prior distribution will be demonstrated using the same setup as in the previous section. First of all, prior sensitivity is investigated for  $H_1 : \theta_1 = \theta_2$  versus  $H_u : \theta_1, \theta_2$ .

Figure 2 displays the results for a sample size of  $N = 40$  (top) and 800 (bottom). Note that the line associated with  $\boldsymbol{\alpha} = \mathbf{1}$  is the same as in the first plot of Figure 1. As can be seen the posterior model probability depends on the choice of  $\boldsymbol{\alpha}$ . What is clear from this figure is that values of  $\boldsymbol{\alpha}$  smaller than 1.0 lead to posterior model probabilities that are too small if  $H_1$  is true, that is, if the odds ratio in group 2 is equal to 1. What can also be seen is that the lines for values of  $\boldsymbol{\alpha}$  of 1 and larger are relatively close to each other.

In Figure 3, the sensitivity of the posterior model probability of  $H_3 : \theta_2 > \theta_1$  versus  $H_u : \theta_1, \theta_2$  is evaluated for both  $N = 40$  and  $N = 800$ . As can be seen the sensitivity of the posterior model probability for the inequality constrained hypothesis to the choice of  $\boldsymbol{\alpha}$  is negligible. Note that for  $N = 800$  the lines for the five different  $\alpha$  values are on top of each other, for  $N = 40$  the lines are very close.

For two reasons we recommend to use  $\boldsymbol{\alpha} = \mathbf{1}$ . As is illustrated by Figure 2,  $\boldsymbol{\alpha} = \mathbf{1}$  has a nice trade of in size between the situation in which  $H_1$  is true (in which case it should be relatively large) and the situation in which  $H_1$  is not true (in which case it should be relatively small). The choice between  $\boldsymbol{\alpha}$ -values of 1 and larger is relatively arbitrary because the differences in performance are rather small. Values of  $\boldsymbol{\alpha}$  smaller than 1 lead to a rather weak performance if  $H_1$  is true (the posterior model probabilities are too small). Furthermore, the value 1 is preferred because it is the largest value of  $\boldsymbol{\alpha}$  for which a so called reference prior is obtained (Lee, 1997, pp. 83-86). A nice property of reference priors is that they are uninformative in the sense that they express no preference for specific combinations of values of  $\boldsymbol{\pi}$ .

To further illustrate prior sensitivity we will repeat the evaluation of the hypotheses stated for the data in the  $2 \times 2 \times 2$  table form Nash and Bowen displayed in Table 7 for three different values of  $\boldsymbol{\alpha}$ . The results are presented in Table 9. As can be seen, there are small differences in posterior model

Table 9: Sensitivity of Posterior Model Probabilities

Hypothesis	$\alpha = .5$	$\alpha = 1$	$\alpha = 2$
$H_0: \theta_1 = \theta_2 = 1$	.08	.03	.03
$H_1: \theta_1 = \theta_2$	.00	.02	.03
$H_2: \theta_1 > 1, \theta_2 > 1$	.23	.24	.24
$H_3: \theta_1 > \theta_2$	.14	.14	.14
$H_4: \theta_1 > \theta_2 > 1$	.48	.50	.49
$H_u: \theta_1, \theta_2$	.07	.07	.07

probabilities for the hypotheses involving equality constraints, and hardly any differences for the hypotheses involving inequality constraints. This is a rather common phenomenon: although there may be differences in posterior model probabilities due to sensitivity to the prior, these differences are often small and not of influence on the interpretation of the results. Whatever value of  $\alpha$  would be chosen from Table 9, the conclusion would be the same:  $H_4$  is supported most by the data.

### Illustration: The Effect of Prior Dispositions on Current Disposition for Young Delinquents

In this section an example concerning a  $4 \times 4 \times 4$  contingency table will be elaborated. In addition to the unconstrained hypothesis, two competing hypotheses will be translated into constraints on sums of cell probabilities. Subsequently, the best of the hypotheses will be selected.

Matarazzo, Carrington and Hiscott (2001) investigate the effect of prior dispositions on the current disposition of young delinquents. Their study uses data from the Canadian youth court survey in 1993-1994. The unit of analysis is the “case” which is operationalized as all the charges pertaining to an offender which were disposed of (sentenced) at the same court hearing. In order to study the effect of prior dispositions, the data were limited to the 16,636 cases involving young offenders with at least two previous cases that reached disposition. The data are displayed in Table 10.

Throughout this section, we will label the three dispositions “First”, “Second” and “Third” to denote the chronological order of the three dispositions under consideration. It should be noted, however, that the label “First” does not imply that it was the offenders first disposition. “Second disposition” was defined as the most serious disposition in the case with the most recent date of disposition before the date of the third disposition. “First disposition” was defined as the most serious disposition in the case with the most recent date of disposition before the date of second disposition.

Table 10: Cross-classification of Three Subsequent Dispositions

First disposition	Second disposition	Third disposition			
		Other	Probation	Open	Secure
Other	Other	373	294	100	106
Probation		427	507	239	191
Open		75	74	74	80
Secure		55	58	48	101
Other	Probation	379	496	242	168
Probation		612	1399	814	499
Open		87	168	167	171
Secure		77	127	90	172
Other	Open	72	109	168	138
Probation		177	356	696	421
Open		103	207	537	473
Secure		42	82	183	311
Other	Secure	62	78	51	174
Probation		122	220	205	442
Open		85	145	269	576
Secure		125	195	269	1074

There are four types of disposition: Other, a residual category of dispositions including a fine not exceeding \$1000, community service order, order for compensation or restitution; Probation; Open custody; and, Secure custody. A small number (94) of youth court cases involving very serious charges and/or lengthy criminal histories were transferred to ordinary (adult) criminal court and are excluded from this study.

Much research has been done to investigate different mechanisms underlying judicial decision-making in youth court. Matarazzo et al. (2001) further investigate some contradictory results on the relationships between prior and current youth court dispositions. According to the societal-reaction theory, assessments such as dispositions are a type of labelling and any prior label, like a prior disposition, will strongly affect future assessments. This is also referred to as the stabilization theory. One of the specifications of this stabilization theory investigated in the paper is: 'Current youth court dispositions are dependent upon all prior dispositions as opposed to only the most recent prior disposition.' (Matarazzo et al., 2001). This hypothesis inspired us to formulate two inequality constrained hypotheses. We will evaluate and compare:  $H_1$  : the second disposition is often similar to the third disposition, and  $H_2$  : the first disposition is often similar to the third disposition. According to Matarazzo's expectation support should be found for both  $H_1$  and  $H_2$ .

To clarify the hypotheses, two new tables are constructed. Table 10 is in-

Table 11: Turnover Tables for the Data in Table 10

	Third disposition			
	Other	Probation	Open	Secure
Second Disposition				
Other	.33	.33	.16	.17
Probation	.20	.39	.23	.18
Open	.10	.19	.39	.33
Custody	.10	.16	.19	.55
First Disposition				
Other	.29	.32	.19	.19
Probation	.18	.34	.27	.21
Open	.11	.18	.32	.40
Custody	.10	.15	.20	.55

dexed by  $i, j, k$ , where  $i$  indexes the first disposition,  $j$  the second disposition and  $k$  the third disposition. The upper half of Table 11 displays the data summed over the first disposition, and normalized such that the row sum equals 1. The hypothesis  $H_1$  states that the second disposition is likely to be similar to the third disposition. This implies that the diagonal probabilities are larger than the probabilities in the corresponding row and column. This is formalized in Table 12 in hypothesis  $H_1$ . Note that  $\pi_{+11} = \sum_i \pi_{i11}$ , and that  $\pi_{+1+} = \sum_i \sum_k \pi_{i1k}$ , etc. In the lower half of Table 11, the data are summed over the second disposition and normalized such that the row sum equals 1. The hypothesis  $H_2$  states that the first disposition is similar to the third disposition. In terms of probabilities this means that all the diagonal probabilities are restricted to be larger than the probabilities in the corresponding row and column. This is again formalized in Table 12.

The results are displayed in Table 13 for three different runs of our software (with  $\alpha = 1$ ) to illustrate the effect of sampling to estimate  $c_t$  and  $d_t$  for  $t = 1, 2$ . As can be seen, the effect of sampling is negligible for all practical purposes. The hypothesis that the second disposition is similar to the third disposition is strongly supported by the data (posterior model probability of .993). However, no support is found for the hypothesis stating that the first disposition is similar to the third disposition (posterior model probability of .000). Two conclusions can be drawn: The stabilization hypothesis is supported when the most recent prior disposition is considered. However, the hypothesis of Matarazzo et al. that all prior dispositions have this 'stabilization' effect is not supported.

Table 12: Hypotheses for the Data in Table 10

$H_u$ :	Unconstrained
$H_1$ :	$\pi_{+11}/\pi_{+1+} > \pi_{+1k}/\pi_{+1+}, k \neq 1$
	$\pi_{+11}/\pi_{+1+} > \pi_{+j1}/\pi_{+j+}, j \neq 1$
	$\pi_{+11}/\pi_{+1+} > \pi_{+1k}/\pi_{+1+}, k \neq 2$
	$\pi_{+11}/\pi_{+1+} > \pi_{+j1}/\pi_{+j+}, j \neq 2$
	$\pi_{+11}/\pi_{+1+} > \pi_{+1k}/\pi_{+1+}, k \neq 3$
	$\pi_{+11}/\pi_{+1+} > \pi_{+j1}/\pi_{+j+}, j \neq 3$
	$\pi_{+11}/\pi_{+1+} > \pi_{+1k}/\pi_{+1+}, k \neq 4$
	$\pi_{+11}/\pi_{+1+} > \pi_{+j1}/\pi_{+j+}, j \neq 4$
$H_2$ :	$\pi_{1+1}/\pi_{1++} > \pi_{1+k}/\pi_{1++}, k \neq 1$
	$\pi_{1+1}/\pi_{1++} > \pi_{i+1}/\pi_{i++}, i \neq 1$
	$\pi_{2+2}/\pi_{2++} > \pi_{1+k}/\pi_{1++}, k \neq 2$
	$\pi_{2+2}/\pi_{2++} > \pi_{i+1}/\pi_{i++}, i \neq 2$
	$\pi_{3+3}/\pi_{3++} > \pi_{1+k}/\pi_{1++}, k \neq 3$
	$\pi_{3+3}/\pi_{3++} > \pi_{i+1}/\pi_{i++}, i \neq 3$
	$\pi_{4+4}/\pi_{4++} > \pi_{1+k}/\pi_{1++}, k \neq 4$
	$\pi_{4+4}/\pi_{4++} > \pi_{i+1}/\pi_{i++}, i \neq 4$

Table 13: Bayes Factors and Posterior Model Probabilities for the Hypotheses in Table 12

	Run 1		Run 2		Run 3	
	$BF_{tu}$	Post. Prob	$BF_{tu}$	Post. Prob	$BF_{tu}$	Post. Prob
$H_1$	150.19	.993	153.02	.993	154.61	.994
$H_2$	0.00	.000	0.00	.000	0.00	.000
$H_u$		.007		.007		.006

## Discussion

This paper presented a Bayesian approach that can be used to select the best of a set of (in)equality constrained hypotheses formulated in terms of the cell probabilities of a contingency table. Posterior model probabilities were used to quantify the support in the data for each of the hypotheses under investigation. It was illustrated that posterior model probabilities behave adequately when the goal is to select the best of a set of constrained hypotheses. Bayesian analysis can be sensitive to the specification of the prior distribution. It was illustrated that the sensitivity is almost non-existent if hypotheses formulated using inequality constraints are evaluated. It was also argued that use of a prior distribution with  $\alpha = \mathbf{1}$  has good properties for the evaluation of hypotheses specified using equality constraints.

The analyses described in this paper can be executed using software that can be downloaded from the website of this journal. The input for the soft-

ware consists of a file containing the number of observations in each cell of the contingency table, and, a file in which the restrictions for one or more hypotheses are specified. Running the program will render an output file containing the Bayes factors and posterior model probabilities of the hypotheses under investigation. The software comes with a small manual, and example input and output files that correspond with the examples presented in this paper. As illustrated in this paper, contingency tables consisting of 4 until 64 cells can be handled by this software package. Beyond that there is currently no experience.

The Bayesian approach is not the only option for researchers who want to evaluate a set of hypotheses formulated using equality and inequality constraints. A number of alternatives is shortly discussed.

An overview of hypothesis testing using  $p$ -values when either the null or the alternative hypothesis is specified using inequality constraints can be found in Silvapulle and Sen (2005). Their work is an update and extension of the work described in Barlow, Bartholomew, Bremner and Brunk (1972) and Robertson, Wright and Dykstra (1988). Consider again the hypotheses displayed in Table 12. What can be done with hypothesis testing is test  $H_1$  (which then serves as the null-hypothesis) against  $H_u$ , and test  $H_2$  against  $H_u$ . If the null-hypothesis is not rejected the data support the constraints used to formulate the null-hypothesis. Two issues have to be kept in mind when using null-hypothesis testing: the sample size should be large enough to have sufficient power to reject the null if it is false; and, both of  $H_1$  and  $H_2$  can be evaluated using null-hypothesis testing, but a direct comparison of both is not possible.

Let  $H_0$  denote the counterpart of  $H_1$  in which the inequality constraints are replaced by equality constraints. Hypothesis testing using  $p$ -values can also be used to test  $H_0$  against  $H_1$ . The same holds for  $H_2$  if  $H_0$  is constructed replacing the inequalities in  $H_2$  by equalities. The same issues discussed above apply to this setup in which  $H_1$  and  $H_2$  are the alternative hypotheses. Note that the use of information criteria like AIC (Akaike, 1974) and BIC (Kass and Raftery, 1995) is not an alternative for the fact that null hypothesis significance testing can't be used for a direct comparison of  $H_1$  and  $H_2$ . Both approaches require the number of parameters in a model in order to compute the penalty for model complexity. As to yet is unclear how to quantify the number of parameters in hypotheses specified using inequality constraints.

Gallindo-Garre and Vermunt (2002, 2005) discuss inequality constrained log-linear models. Instead of formulating hypotheses in terms of the cell probabilities of a contingency table, they formulate hypotheses in terms of log-linear parameters. The hypotheses that can be tested are partly different and partly overlapping with the hypotheses that can be tested with the approach of Silvapulle and Sen (2005). Here too, two non-nested hypotheses like  $H_1$  and  $H_2$  cannot directly be compared to each other.

There are two other approaches that can also be used to superimpose a structure on the cell probabilities  $\pi$ . The first approach is the class of so-called row-column (RC) association models (Agresti, Chuang and Kezouh, 1987; Becker, 1989; Anderson, 1996; Gallindo-Garre and Vermunt, 2004; Iliopoulos, Kateri and Ntzoufras, 2007). These models can be used to superimpose different ordinal association structures on contingency tables and to estimate and test the resulting models. The second approach enables a researcher to specify models by means of linear restrictions on the cell probabilities or the logarithm of the cell probabilities (Grizzle, Starmer and Koch (1969), Haber (1985), Haber and Brown (1986)), and to estimate and test the resulting models. Both approaches are interesting in their own right, but can't be used to evaluate hypotheses formulated using inequality constraints. This will be elaborate using three examples.

A core feature of RC-association models for two way tables is that the log odds ratio  $\log \theta_{ij}$  for rows  $i, i + 1$  and columns  $j, j + 1$  is restricted using

$$\log \theta_{ij} = \beta(r_{i+1} - r_i)(c_{j+1} - c_j), \quad (9)$$

where the category parameters  $r$  and  $c$  are estimated from the data, and  $\beta$  is a association parameter that holds uniformly throughout the contingency table. If  $r_{i+1} - r_i = 1$  for  $i = 1, \dots, I - 1$  and  $c_{j+1} - c_j = 1$  for  $j = 1, \dots, J - 1$ , a so-called linear by linear association model is obtained in which  $\beta$  is the log odds ratio that holds uniformly throughout the contingency table (Agresti, 1984, pp. 76-87, 139-141). The hypothesis  $H_0 : \log \theta_{11} = \log \theta_{12} = \log \theta_{13}$ , with respect to three adjacent odds-ratios in a 2 by 4 contingency table, can without problems be evaluated against  $H_u$  : "not  $H_0$ " using a linear by linear association model. However, as far as the authors know, it is not possible to evaluate the hypotheses  $H_1 : \log \theta_1 > \log \theta_2 > \log \theta_3$ ,  $H_2 : \log \theta_1 > 0, \log \theta_2 > 0, \log \theta_3 > 0$  or the combination of both hypotheses using linear by linear or RC-association models. Note that models using linear restrictions on the logarithm of the cell probabilities (Haber, 1985) can also be used to evaluate  $H_0$ . However, provisions for the evaluation of  $H_1$ ,  $H_2$  or the combination of both, are lacking here too.

Consider also the hypothesis  $\pi_1 > \pi_2 > \pi_3 > \pi_4$  for a 1 by 4 contingency table, and the hypothesis  $\frac{\pi_{11}}{(\pi_{11} + \pi_{12})} > \frac{\pi_{21}}{(\pi_{21} + \pi_{22})} > \frac{\pi_{31}}{(\pi_{31} + \pi_{32})} > \frac{\pi_{41}}{(\pi_{41} + \pi_{42})}$  for a 2 by 4 contingency table. Both can be evaluated using the approach described in Silvapulle and Sen (2005) and the approach proposed in this paper, however both hypotheses do not fit in the frame work of RC-association models and can't be evaluated using the approach described in Haber (1985) and Haber and Brown (1986).

Note finally, that both hypotheses in Table 12 can be reformulated as eight odds ratios that are restricted to be larger than one. Evaluation of one odds ratio that is restricted to be larger than one can be achieved using a one sided null hypothesis significance test. However, evaluation of eight odds ratio's restricted to be larger than one constitutes a multivariate one-sided

hypothesis. Multivariate one sided hypotheses can be evaluated using the methods described in Silvapulle and Sen (2005) or the approach proposed in this paper. An advantage of the approach proposed in this paper is that it can be used to directly compare two or more hypotheses that are not nested within each other.

Further study is needed to explore the advantages and disadvantages of inequality constrained hypotheses and hypotheses formulated using ordinal or linear restrictions in different situations. Further study is also required to compare the performance of model selection (either Bayesian or based on information criteria) and null hypothesis significance testing. However, both are beyond the scope of this paper. The interested reader is referred to Dayton (2003) who compares hypotheses testing and model selection in the context of comparisons of means, and Wagenmakers (2007) who gives an evaluation of hypothesis testing and Bayesian model selection.

## References

- Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. New York: John Wiley.
- Agresti, A., Chuang, C. and Kezouh, A. (1987). Order-restricted score parameters in association models for contingency tables. *Journal of the American Statistical Association*, *82*, 619–623.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Albert, J.H. and Gupta, A.K. (1983). Estimation in contingency tables using prior information. *Journal of the Royal Statistical Society, Series B*, *45*, 60–69.
- Anderson, C.J. (1996). The analysis of three way contingency tables by three mode association models. *Psychometrika*, *61*, 465–483.
- Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference Under Order Restrictions*. Wiley: New York.
- Bartholomew, K. (1990). Avoidance of intimacy: An attachment perspective. *Journal of Social and Personal Relationships*, *7*, 147–178.
- Becker, M.P. (1989). Models for the analysis of association in multivariate contingency tables. *Journal of the American Statistical Association*, *84*, 1014–1019.
- Berger, J.O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, *3*, 317–352.

- Bowen, G. L., Richman, J. M., Brewster, A. and Bowen, N. K. (1998). Sense of school coherence, perceptions of danger at school, and teacher support among youth at risk of school failure. *Child and Adolescent Social Work Journal*, 15, 273–286.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90, 1313–1321.
- Congdon, P. (2001). *Bayesian Statistical Modelling*. New York: Wiley.
- Dayton, C.M. (2003). Information criteria for pairwise comparisons. *Psychological Methods*, 8, 61–71.
- Dickey, J. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Statistics*, 42, 204–223.
- Galindo-Garre, F.G., Vermunt, J.K. and Croon, M.A. (2002). Likelihood-ratio tests for order-restricted log-linear models: A comparison of asymptotic and bootstrap methods. *Metodologa de las Ciencias del Comportamiento*, 4, 325–337.
- Galindo-Garre, F.G. and Vermunt, J.K. (2004). The order-restricted association model: two estimation algorithms and issue in testing. *Psychometrika*, 69, 641–654.
- Galindo-Garre, F.G. and Vermunt, J.K. (2005). Testing log-linear models with inequality constraints: a comparison of asymptotic, bootstrap and posterior predictive p-values. *Statistica Neerlandica*, 59, 82–94.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.
- Gelman, A. Carlin B.P., Stern H.S. and Rubin D.B. (2004). *Bayesian Data Analysis*. New York: Chapman and Hall.
- Grizzle, J.E., Starmer, C.F and Koch, G.G. (1969). Analysis of categorical data by linear models. *Biometrics*, 3, 489–504.
- Gunel, E. and Dickey, J. (1974). Bayes factors for independence in contingency tables. *Biometrika*, 61, 545–557.
- Haber, M. (1985). Maximum likelihood methods for linear and log-linear models in categorical data. *Computational Statistics and Data Analysis*, 3, 1–10.
- Haber, M. and Brown, M.B. (1986). Maximum likelihood methods for log-linear models when expected frequencies are subjected to linear constraints. *Journal of the American Statistical Association*, 81, 477–482.

- Hoijtink, H. (2000). Posterior inference in the random intercept model based on samples obtained with Markov chain Monte Carlo methods. *Computational Statistics*, *3*, 315–336.
- Hoijtink, H., Klugkist, I. and Boelen, P.A. (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.
- Iliopoulos, G., Kateri, M. and Ntzoufras, I. (2006). Bayesian estimation of unrestricted and order-restricted association models for a two-way contingency table. *Computational Statistics and Data Analysis*, *51*, 4643–4655.
- Karabatsos, G. and Sheu, C, -F (2004). Order-constrained Bayes inference for dichotomous models of unidimensional nonparametric IRT. *Applied Psychological Measurement*, *28*, 110–125.
- Kass, R.E. (1993). Bayes factors in practice. *The Statistician*, *42*, 551–560.
- Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**: 773–795
- Klugkist, I., Kato, B. and Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, *59*, 57–69.
- Klugkist, I., Laudy, O. and Hoijtink, H. (2005). Inequality Constrained Analysis Of Variance: A Bayesian Approach. *Psychological Methods*, *10*, 477–493.
- Lee, P.M. (1997). *Bayesian Statistics: An Introduction*. London: Arnold.
- Lessard, J.C. and Moretti, M.M. (1998). Suicidal ideation in an adolescent clinical sample: attachment patterns and clinical implications. *Journal of Adolescence*, *21*, 383–395.
- Matarazzo, A., Carrington, P.J. and Hiscott, R.D. (2001). The Effect of Prior Youth Court Dispositions on Current Disposition: An Application of Societal-Reaction Theory. *Journal of Quantitative Criminology*, *17*, 169–200.
- Myung, J.I., Karabatsos, G. and Iverson, G.J. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, *49*, 205–225.
- Nash, J.K. and Bowen, G.L. (2002). Defining and Estimating Risk and Protection: An Illustration from the School Success Profile Child and Adolescent. *Social Work Journal*, *19*, 247–261.
- Robertson T., Wright F.T., Dykstra R.L. (1988). *Order Restricted Statistical Inference*. New York: Wiley.

- Silvapulle, M.J. and Sen, P.K. (2005). *Constrained Statistical Inference: Inequality, Order and Shape Restrictions*. New York: Wiley.
- Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, 90, 614–618
- Vermunt, J.K. (1991). A general class of nonparametric models for ordinal categorical data. *Sociological Methodology*, 29, 187–223.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.

## Appendix A: Derivation of the Bayes Factor for Constrained Hypotheses

The Bayes factor for a constrained hypothesis  $H_t$  with respect to the unconstrained hypothesis  $H_u$  is the ratio of two predictive densities or marginal likelihoods:

$$BF_{tu} = \frac{\int_{\boldsymbol{\pi}} L(\boldsymbol{\pi} | \mathbf{f}) Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_t) d\boldsymbol{\pi}}{\int_{\boldsymbol{\pi}} L(\boldsymbol{\pi} | \mathbf{f}) Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_u) d\boldsymbol{\pi}}. \quad (10)$$

Computation of the Bayes factor can be burdensome (Chib, 1995; Gelfand and Smith, 1990; Verdinelli and Wasserman, 1995), because the integrals involved may not be easy to compute. However, for the purposes of this paper it is convenient to choose the representation used by Chib (1995):

$$BF_{tu} = \frac{L(\boldsymbol{\pi} | \mathbf{f}) Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_t)}{P(\boldsymbol{\pi} | \mathbf{f}, \boldsymbol{\alpha}, H_t)} \bigg/ \frac{L(\boldsymbol{\pi} | \mathbf{f}) Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_u)}{P(\boldsymbol{\pi} | \mathbf{f}, \boldsymbol{\alpha}, H_u)}, \quad (11)$$

for any value of  $\boldsymbol{\pi}$  in agreement with the constraints of  $H_t$ . The prior distribution for a constrained hypothesis can be derived from the prior distribution of the unconstrained hypothesis:

$$Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_t) = \frac{Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_u) I_{H_t}}{\int_{\boldsymbol{\pi}} Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_u) I_{H_t} d\boldsymbol{\pi}}, \quad (12)$$

where  $I_{H_t}$  equals 1 if  $\boldsymbol{\pi}$  is in agreement with the constraints in  $H_t$  and zero otherwise. In the same manner the posterior distribution of a constrained hypothesis is obtained:

$$P(\boldsymbol{\pi} | \mathbf{f}, \boldsymbol{\alpha}, H_t) = \frac{P(\boldsymbol{\pi} | \mathbf{f}, \boldsymbol{\alpha}, H_u) I_{H_t}}{\int_{\boldsymbol{\pi}} P(\boldsymbol{\pi} | \mathbf{f}, \boldsymbol{\alpha}, H_u) I_{H_t} d\boldsymbol{\pi}}. \quad (13)$$

It follows that for  $\boldsymbol{\pi}^* \in H_t$

$$Pr(\boldsymbol{\pi}^* | \boldsymbol{\alpha}, H_t) = c_t^{-1} Pr(\boldsymbol{\pi}^* | \boldsymbol{\alpha}, H_u), \quad (14)$$

where  $c_t^{-1}$  denotes the proportion of  $Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_u)$  in agreement with  $H_t$ , and

$$P(\boldsymbol{\pi}^* | \mathbf{f}, \boldsymbol{\alpha}, H_t) = d_t^{-1} P(\boldsymbol{\pi}^* | \mathbf{f}, \boldsymbol{\alpha}, H_u), \quad (15)$$

where  $d_t^{-1}$  denotes the proportion of  $P(\boldsymbol{\pi} | \mathbf{f}, H_u)$  in agreement with  $H_t$ . Since the likelihoods in (11) cancel against each other, insertion of (14) and (15) renders:

$$BF_{tu} = \frac{c_t}{d_t}. \quad (16)$$

The interested reader is referred to Klugkist, Laudy and Hoijtink (2005) for a similar derivation.

For hypotheses specified using only inequality constraints an estimate of  $c_t$  can be obtained using a sample from  $Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_u)$ . Since the prior

is a Dirichlet distribution with parameters  $\boldsymbol{\alpha}$ , this sample can be obtained using procedure DIR-2 described in Narayanan (1990). This procedure will be elaborated in Appendix B. The proportion of sampled parameter vectors  $\boldsymbol{\pi}$  in agreement with  $H_t$  is an estimate of  $c_t^{-1}$ . The posterior is a Dirichlet distribution with parameters  $\boldsymbol{f} + \boldsymbol{\alpha}$ . The proportion of sampled parameter vectors from  $P(\boldsymbol{\pi} \mid \boldsymbol{f}, \boldsymbol{\alpha}, H_u)$  in agreement with  $H_t$  is an estimate of  $d_t^{-1}$ . If hypotheses are specified using equality constraints in addition to inequality constraints, these estimators can not be used because  $c_t$  and  $d_t$  are not defined for constraints like  $\pi_{11} = \pi_{21}$ . The algorithm for estimation of  $BF_{tu}$  presented below can handle both equality and inequality constraints.

Let the inequality constraints in hypothesis  $t$  be denoted by  $t^*$ . The Bayes factor of a hypothesis containing both equality and inequality constraints versus the unconstrained hypothesis can then be written as:

$$BF_{tu} = BF_{t^*,u} \times BF_{t,t^*}. \quad (17)$$

Note that  $BF_{t^*,u}$  can be computed using the estimates of  $c_t$  and  $d_t$  presented above. The computation of  $BF_{t,t^*}$  will be discussed below and in Appendix B.

To deal with equality constraints like  $\pi_{11} = \pi_{21}$  these will be replaced with about equality constraints like  $|\pi_{11} - \pi_{21}| < \delta_0$ . According to Berger and Delampady (1987) a Bayes factor based on  $\pi_{11} = \pi_{21}$  is indistinguishable from the Bayes factor based on  $|\pi_{11} - \pi_{21}| < \delta_0$  if  $\delta_0$  is small enough. Along the same lines as the derivation in the beginning of this section it can be shown that  $BF_{t,t^*} = c_{t,t^*}/d_{t,t^*}$ , where  $c_{t,t^*}^{-1}$  is the proportion of the constrained (by the inequality constraints of hypothesis  $t$ ) prior distribution in agreement with all the constraints of hypothesis  $t$ , and  $d_{t,t^*}^{-1}$  the corresponding proportion from the constrained posterior distribution. If  $\delta_0$  is very small straightforward estimates of  $c_{t,t^*}$  and  $d_{t,t^*}$  will be very inefficient. However, the following approach solves this problem. Define  $BF_{tu}$  as

$$BF_{tu} = BF_{t^*,u} \times BF_{t_{\delta_R},t^*} \times BF_{t_{\delta_{R-1}},t_{\delta_R}} \times \dots \times BF_{t_{\delta_0},t_{\delta_1}} = \frac{c_{t^*}}{d_{t^*}} \times \frac{c_{t_{\delta_R},t^*}}{d_{t_{\delta_R},t^*}} \times \frac{c_{t_{\delta_{R-1}},t_{\delta_R}}}{d_{t_{\delta_{R-1}},t_{\delta_R}}} \times \dots \times \frac{c_{\delta_0,\delta_1}}{d_{\delta_0,\delta_1}}, \quad (18)$$

where  $\delta_{r-1} = \delta_r/2$  for  $r = 1, \dots, R$ , and  $\delta_R = .2$ . Note that  $c_{t_{\delta_{R-1}},t_{\delta_R}}^{-1}$  denotes the proportion of the constrained (by the inequality constraints of hypothesis  $t$  and about equality constraints with bound  $\delta_R$ ) prior distribution in agreement with a hypothesis with the same inequality constraints, but equality constraints with bound  $\delta_{R-1}$ , and  $d_{t_{\delta_{R-1}},t_{\delta_R}}^{-1}$  the corresponding proportion for the posterior distribution. If  $R$  is large enough,  $BF_{t_{\delta_0},t_{\delta_1}} \rightarrow 1$  which can easily be checked during the computation of (18). The  $c_{t_{\delta_{R-1}},t_{\delta_R}}$ 's and  $d_{t_{\delta_{R-1}},t_{\delta_R}}$ 's needed for the computation of (18) can be obtained using the constrained Gibbs sampler presented in Appendix B.

There is an interesting link between (16) and an estimator of the Bayes factor discussed by Dickey (1971), Gunel and Dickey (1974) and Verdinelli and Wasserman (1995). Their estimator can be used, for example, to compare  $H_0 : \theta = 1$  with  $H_u : \text{not } H_0$ . Like for (16) the derivation of their estimator starts with (11) which after cancelation of the likelihood renders:

$$BF_{1u} = \frac{Pr(\theta | \boldsymbol{\alpha}, H_0)}{P(\theta | \mathbf{f}, \boldsymbol{\alpha}, H_0)} / \frac{Pr(\theta | \boldsymbol{\alpha}, H_u)}{P(\theta | \mathbf{f}, \boldsymbol{\alpha}, H_u)}, \quad (19)$$

where

$$Pr(\theta | \boldsymbol{\alpha}, H_u) = \int_{\{\boldsymbol{\pi}: \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \theta\}} Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_u) d\boldsymbol{\pi}, \quad (20)$$

and

$$P(\theta | \mathbf{f}, \boldsymbol{\alpha}, H_u) = \int_{\{\boldsymbol{\pi}: \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \theta\}} P(\boldsymbol{\pi} | \mathbf{f}, \boldsymbol{\alpha}, H_u) d\boldsymbol{\pi}. \quad (21)$$

Note that  $Pr(\theta | H_0)$  equals 1 if  $\theta = 1$  and 0 otherwise. The same holds for  $P(\theta | \mathbf{f}, H_0)$ . The implication is that for  $\theta = 1$ , that is,  $\theta \in H_0$  (19) reduces to:

$$BF_{1u} = \frac{P(\theta = 1 | \mathbf{f}, \boldsymbol{\alpha}, H_u)}{Pr(\theta = 1 | \boldsymbol{\alpha}, H_u)}. \quad (22)$$

Comparing (16) - the ratio of a posterior and prior proportion with respect to an unconstrained hypothesis - with (22) - the ratio of a posterior and prior density with respect to an unconstrained hypothesis - it can be seen that (16) is an extension of (22) that can handle inequality constraints in addition to equality constraints.

## Appendix B: Estimation of the $c$ 's and $d$ 's

The prior distribution (2) is a Dirichlet distribution:

$$Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_u) \sim Dir(\boldsymbol{\pi} | \boldsymbol{\alpha}). \quad (23)$$

The posterior distribution (3) is also a Dirichlet distribution:

$$P(\boldsymbol{\pi} | \mathbf{f}, \boldsymbol{\alpha}, H_u) \sim Dir(\boldsymbol{\pi} | \mathbf{f} + \boldsymbol{\alpha}). \quad (24)$$

Due to the similarity of prior and posterior distribution, we will only explain how to obtain a sample from the prior distribution. Let the elements of  $\boldsymbol{\pi}$  be denoted by  $\pi_a$  for  $a = 1, \dots, A$  and the corresponding elements of  $\boldsymbol{\alpha}$  by  $\alpha_a$  for  $a = 1, \dots, A$ . We will subsequently explain how to estimate  $c_{t^*}$ ,  $c_{\delta_{t^*}}$ , and  $c_{t_{\delta_{r-1}}, t_{\delta_r}}$  for  $r = R, \dots, 1$ .

We will use procedure DIR-2 from Narayanan (1990) to obtain a sample from the prior distribution:

- Initialize  $z_a = 1$  for  $a = 1, \dots, A$ .

- Step 1. For  $a = 1, \dots, A$  sample  $z_a$  from

$$Pr(z_a | \alpha_a) \sim \text{Gamma}(z_a | \alpha_a, 1). \quad (25)$$

- Step 2. For  $a = 1, \dots, A$  compute

$$\pi_a = \frac{z_a}{\sum_{a=1}^A z_a} \quad (26)$$

- Iterate Steps 1 and 2, delete a burn-in phase of 10,000 iterations, and use the remaining 100,000 iterations to estimate the  $d$ 's of interest.

Using the resulting sample of  $\boldsymbol{\pi}$ 's it is easy to estimate  $c_{t^*}^{-1}$  from (18) by means of the proportion of the sampled  $\boldsymbol{\pi}$ 's in agreement with the inequality constraints of hypothesis  $t$ . Note that for the estimation of  $c_{t^*}^{-1}$  a burn-in is not necessary because the algorithm renders independent samples of  $\boldsymbol{\pi}$ . Note also that this is not the case in the next steps in which the algorithm will be transformed to a Gibbs sampler (Gelman, Carlin, Stern and Rubin, 2004, pp. 287-289; Lee, 2001, pp. 259-268).

The proportion of the prior distribution of  $H_{t^*}$  in agreement with  $H_{t_{\delta_R}}$  is represented by  $c_{\delta_R, t^*}^{-1}$ . To estimate  $c_{\delta_R, t^*}^{-1}$  a sample has to be obtained from

$$Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_{t^*}) \propto \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}) I_{H_{t^*}}, \quad (27)$$

where  $I_{H_{t^*}} = 1$  if  $\boldsymbol{\pi}$  is in agreement with the constraints of  $H_{t^*}$  and zero otherwise. The proportion of the sampled  $\boldsymbol{\pi}$ 's in agreement with  $H_{t_{\delta_R}}$  is the required estimate.

The proportion of the prior distribution of  $H_{t_{\delta_r}}$  in agreement with  $H_{t_{\delta_{r-1}}}$  is represented by  $c_{t_{\delta_{r-1}}, t_{\delta_r}}^{-1}$ . To estimate  $c_{t_{\delta_{r-1}}, t_{\delta_r}}^{-1}$  a sample has to be obtained from

$$Pr(\boldsymbol{\pi} | \boldsymbol{\alpha}, H_{t_{\delta_r}}) \propto \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}) I_{H_{t_{\delta_r}}}, \quad (28)$$

where  $I_{H_{t_{\delta_r}}} = 1$  if  $\boldsymbol{\pi}$  is in agreement with the constraints of  $H_{t_{\delta_r}}$  and zero otherwise. The proportion of the sampled  $\boldsymbol{\pi}$ 's in agreement with  $H_{t_{\delta_{r-1}}}$  is the required estimate.

A sample from (27) and (28) can be obtained using the sampling algorithm described above if Step 1 is modified: For  $a = 1, \dots, A$  sample  $z_a$  from

$$Pr(z_a | \alpha_a, L_a, U_a) \propto \text{Gamma}(z_a | \alpha_a, 1) I_{L_a, U_a}, \quad (29)$$

where  $I_{L_a, U_a} = 1$  if  $z_a$  is a value within the interval with lower bound  $L_a$  and upper bound  $U_a$ , and zero otherwise. Sampling within bounds can be achieved using inverse probability sampling. The interested reader is referred to Gelfand, Smith and Lee (1992). The bounds  $L_a$  and  $U_a$  are determined from the restrictions on the  $\pi$ 's. We will illustrate this by means of five

examples.

Example 1: If  $H_{t^*} : \pi_1 < \pi_2 < \pi_3$  then  $\frac{z_1}{\sum_a z_a} < \frac{z_2}{\sum_a z_a} < \frac{z_3}{\sum_a z_a}$ , that is,  $z_1 < z_2 < z_3$ . The implication is that the lower and upper bound when sampling  $z_1$  are  $-\infty$  and the current value of  $z_2$ , respectively. The lower and upper bound when sampling  $z_2$  are  $z_1$  and  $z_3$ , respectively. The lower and upper bound when sampling  $z_3$  are  $z_2$  and  $\infty$ , respectively.

Example 2: If

$$H_{t^*} : \frac{\pi_1 \pi_2}{\pi_3 \pi_4} > 1, \quad (30)$$

then

$$\frac{\frac{z_1}{\sum_a z_a} \frac{z_2}{\sum_a z_a}}{\frac{z_3}{\sum_a z_a} \frac{z_4}{\sum_a z_a}} > 1, \quad (31)$$

that is,  $\frac{z_1 z_2}{z_3 z_4} > 1$ . The implication is that the lower bound when sampling  $z_1$  is  $\frac{z_3 z_4}{z_2}$  and the upper bound is  $\infty$ . When sampling  $z_3$  the lower bound is  $-\infty$  and the upper bound is  $\frac{z_1 z_2}{z_4}$ . For  $z_2$  and  $z_4$  bounds can be derived analogously.

Example 3: As discussed in Appendix A, the restriction  $\pi_1 - \pi_2 = 0$  can be handled using a sequence of restrictions  $|\pi_1 - \pi_2| < \delta_r$  for  $r = R, \dots, 1$ . Each element of this sequence can be rewritten as

$$\left| \frac{z_1}{z_1 + z_2} - \frac{z_2}{z_1 + z_2} \right| < \delta_r \quad (32)$$

which leads to

$$|z_1 - z_2| < \delta_r (z_1 + z_2). \quad (33)$$

This equation has two implications:

$$z_1 - z_2 < \delta_r (z_1 + z_2), \quad (34)$$

and

$$z_1 - z_2 > -\delta_r (z_1 + z_2). \quad (35)$$

Both equations can be rewritten with the focus on  $z_1$  (similar equations can be obtained for  $z_2$ ):

$$z_1 < z_2 \frac{1 + \delta_r}{1 - \delta_r}, \quad (36)$$

and

$$z_1 > z_2 \frac{1 - \delta_r}{1 + \delta_r}, \quad (37)$$

that is,  $z_1$  should be sampled with upper bound (36) and lower bound (37). Note that  $z_1$  and  $z_2$  are gamma deviates and thus positive numbers. Sampling with these bounds renders a sample from the prior distribution of

$H_{t_{\delta_r}} : |\pi_1 - \pi_2| < \delta_r$ . The proportion of this sample in agreement with  $H_{t_{\delta_{r-1}}} : |\pi_1 - \pi_2| < \delta_{r-1}$  is an estimate of  $c_{t_{\delta_{r-1}}, t_{\delta_r}}^{-1}$ .

Example 4: The hypothesis  $H_t : \frac{\pi_1}{\pi_2} = \frac{\pi_3}{\pi_4} = \frac{\pi_5}{\pi_6}$  can be handled using the sequence of hypotheses

$$H_{t_{\delta_r}} : \left| \frac{\pi_1}{\pi_2} - \frac{\pi_3}{\pi_4} \right| < \delta_r, \left| \frac{\pi_1}{\pi_2} - \frac{\pi_5}{\pi_6} \right| < \delta_r, \left| \frac{\pi_3}{\pi_4} - \frac{\pi_5}{\pi_6} \right| < \delta_r \quad (38)$$

for  $r = R, \dots, 1$ . In terms of  $z$ 's (38) can be written as

$$\left| \frac{z_1}{z_2} - \frac{z_3}{z_4} \right| < \delta_r, \left| \frac{z_1}{z_2} - \frac{z_5}{z_6} \right| < \delta_r, \left| \frac{z_3}{z_4} - \frac{z_5}{z_6} \right| < \delta_r, \quad (39)$$

from which lower and upper bounds for the  $z$ 's can easily be determined.

Example 5: The hypothesis  $H_t : \pi_1 < \pi_2 = \pi_3$  can be replaced by a sequence of hypotheses of the form

$$H_{t_{\delta_r}} : \pi_1 < \{\pi_2, \pi_3\}, |\pi_2 - \pi_3| < \delta_r \quad (40)$$

for  $r = R, \dots, 1$ . Written in terms of  $z$ 's this changes to:

$$z_1 < \{z_2, z_3\}, \left| \frac{z_2}{\sum_{a=1}^3 z_a} - \frac{z_3}{\sum_{a=1}^3 z_a} \right| < \delta_r. \quad (41)$$

Using results from Examples 1 and 3, it can be seen that a sample from the prior distribution of  $H_{t_{\delta_r}} : \pi_1 < \{\pi_2, \pi_3\}, |\pi_2 - \pi_3| < \delta_r$  can be obtained using the bound  $L_1 = -\infty$  and  $U_1 = \min(z_2, z_3)$ ,  $L_2 = \max(z_1, \frac{-\delta_3 z_1 + (1-\delta_r)z_3}{1+\delta_r})$  and  $U_2 = \frac{\delta_r z_1 + (1+\delta_r)z_3}{1-\delta_r}$ , and  $L_3$  and  $U_3$  analogous to  $L_2$  and  $U_2$ .

The following constrained hypotheses can be handled using the procedure described in Appendix A and B:

- Hypotheses that are specified using one or more equality constraints as illustrated in Examples 3 and 4 as long as each constraint is of the form:

$$\left| \frac{\prod_{a=1}^A \pi_a}{\prod_{b=1}^B \pi_b} - \frac{\prod_{c=1}^C \pi_c}{\prod_{d=1}^D \pi_d} \right| < \delta_r, \quad (42)$$

with  $A = B$  and  $C = D$ . For restrictions of this type the term  $\sum_a z_a$  - in Example 3  $z_1 + z_2$ , in Example 4 the term has already cancelled in (39) - can easily be handled, which enables sampling of  $\boldsymbol{\pi}$  under restrictions by means of the  $z$ 's.

- Hypotheses with one or more arbitrary constraints on  $\boldsymbol{\pi}$  as long as equality constraints are not used to specify these hypotheses. For hypotheses without equality constraints only  $d_t$  and  $c_t$  have to be estimated.

- Hypotheses in which equality constraints (as specified under the first bullet) are combined with inequality constraints can be handled as long as each inequality constraint is of the form:

$$\prod_{a=1}^A \pi_a > \prod_{b=1}^B \pi_b, \quad (43)$$

with  $A = B$ . For restrictions of this type the  $\sum_a z_a$  as appears in the examples cancels, and consequently the  $z_a$  based sampling algorithm can easily be used to obtain a sample of  $\boldsymbol{\pi}$  in accordance with the constraints of the hypothesis at hand.

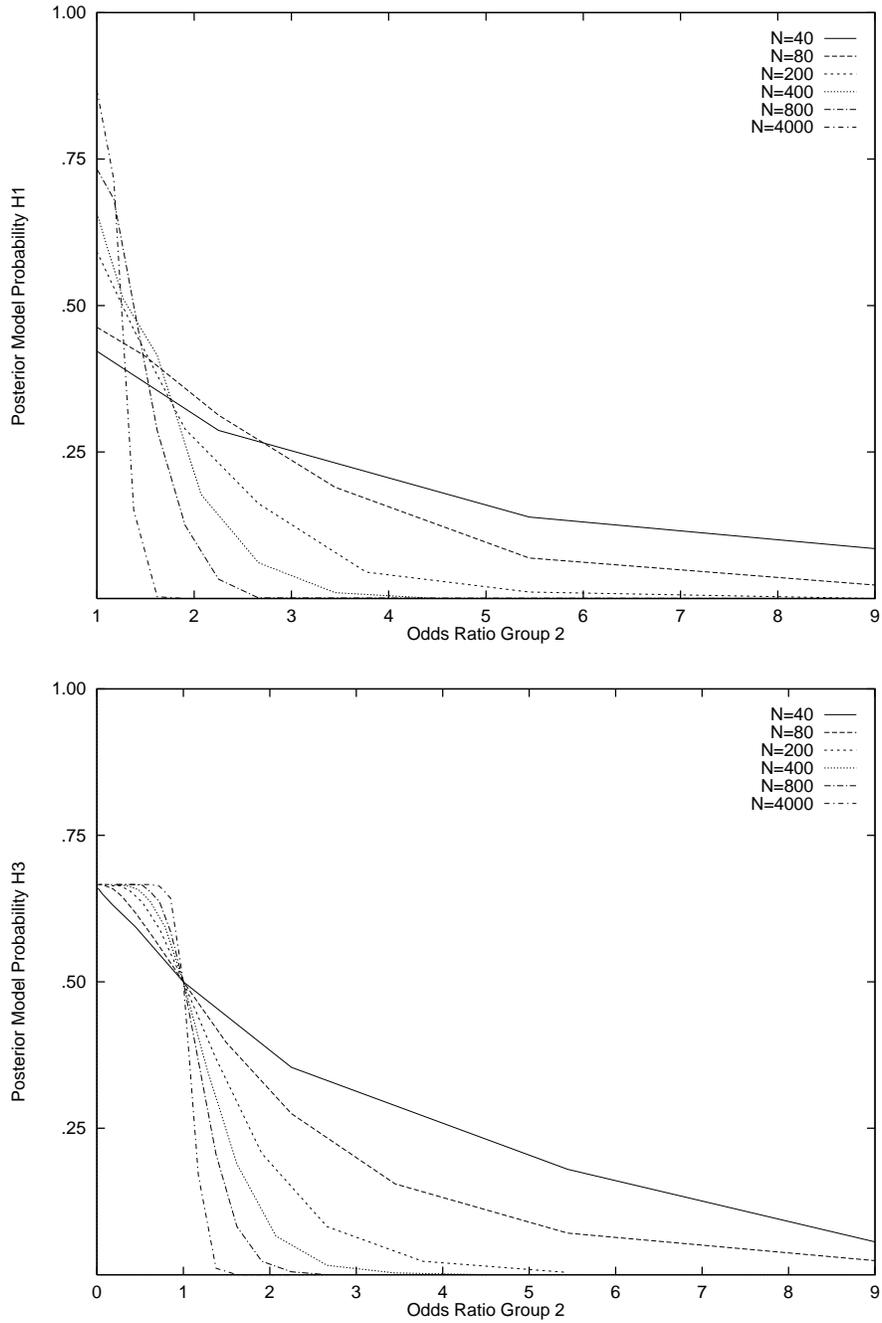


Figure 1: Posterior model probabilities for  $H_1 : \theta_1 = \theta_2$  if it is compared with  $H_u : \theta_1, \theta_2$  (top) and for  $H_3 : \theta_1 > \theta_2$  if it is compared with  $H_u : \theta_1, \theta_2$  (bottom). The value of  $\theta_1 = 1$  and the values of  $\theta_2$  are given along the x-axis.

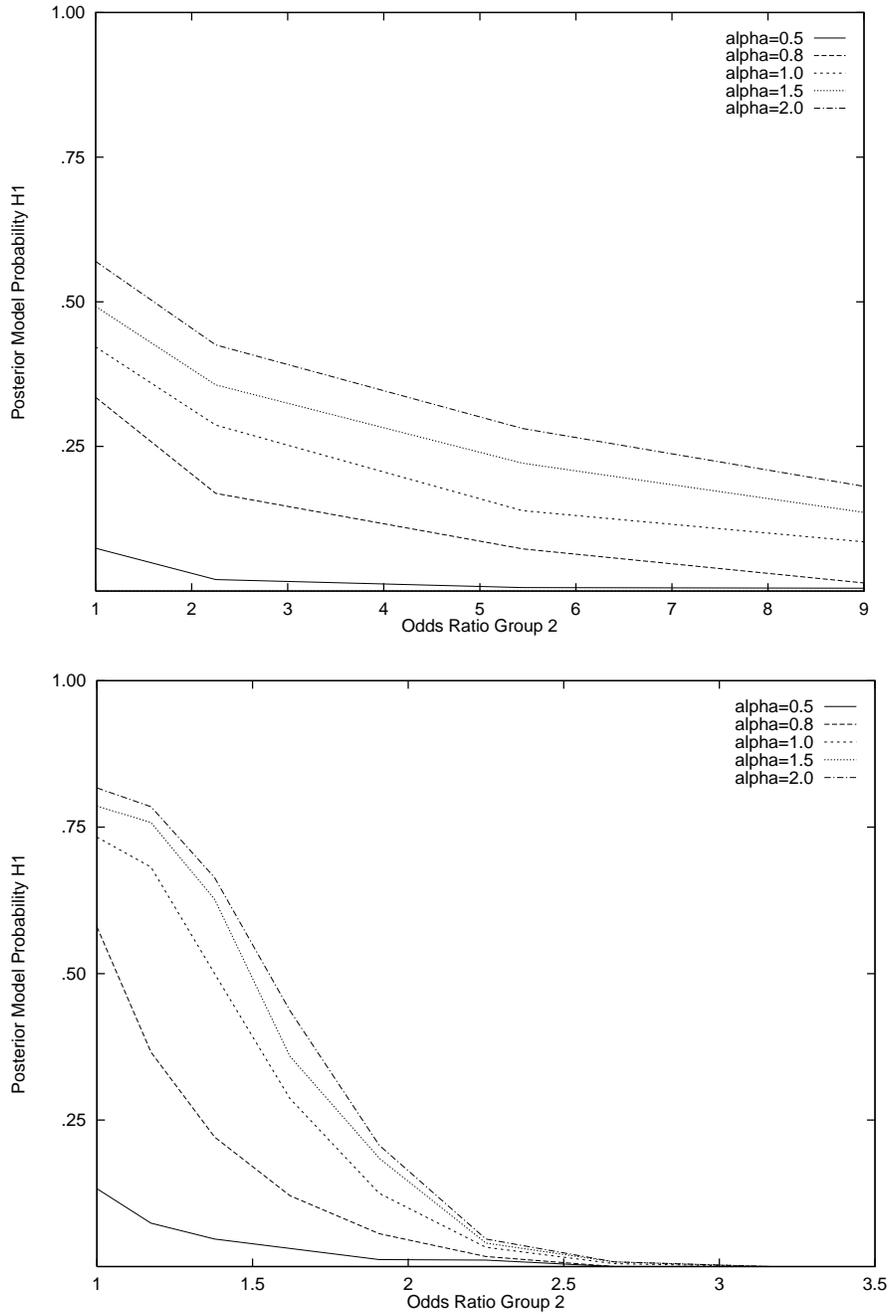


Figure 2: Sensitivity of posterior model probabilities for  $H_1 : \theta_1 = \theta_2$  if it is compared with  $H_u : \theta_1, \theta_2$  for  $N = 40$  (top) and  $N = 800$  (bottom). The value of  $\theta_1 = 1$  and the values of  $\theta_2$  are given along the x-axis. For various choices of  $\alpha$  the posterior model probability of  $H_1$  is displayed on the y-axis.

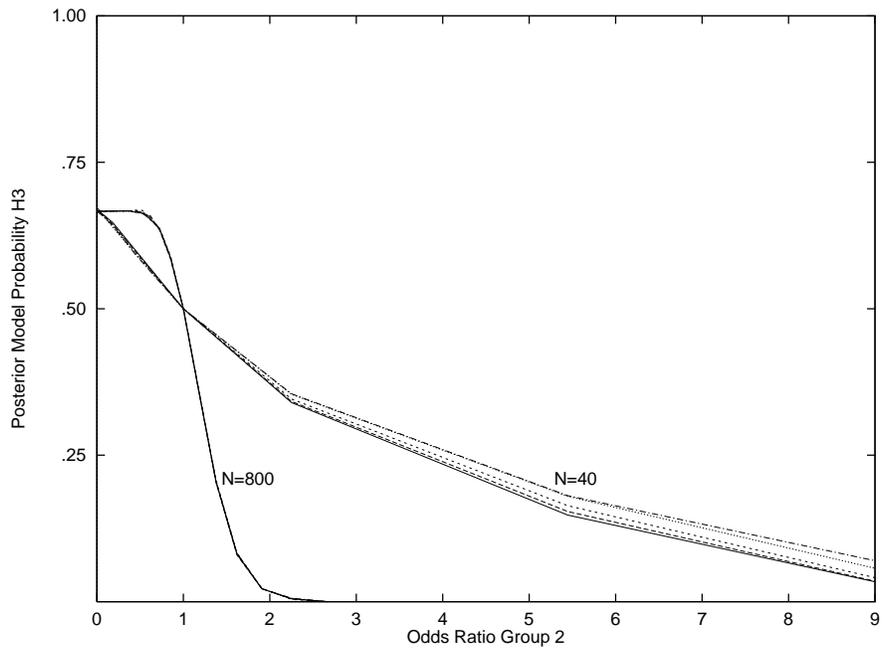


Figure 3: Sensitivity of posterior model probabilities for  $H_3 : \theta_1 > \theta_2$  if it is compared with  $H_u : \theta_1, \theta_2$  for  $N = 40$  and  $N = 800$ . The value of  $\theta_1 = 1$  and the values of  $\theta_2$  are given along the x-axis. For various choices of  $\alpha$  (0.5, 0.8, 1.0, 1.5, 2.0) the posterior model probability of  $H_3$  is displayed on the y-axis.