
A Statisticians View on Bayesian Evaluation of Informative Hypotheses

Jay I. Myung¹, George Karabatsos², and Geoffrey J. Iverson³

¹ Department of Psychology, Ohio State University, 1835 Neil Avenue, Columbus, OH 43210 USA myung.1@osu.edu

² College of Education, University of Illinois-Chicago, 1040 W. Harrison Street, Chicago, IL 60607 USA georgek@uic.edu

³ Department of Cognitive Sciences, University of California at Irvine, 3151 Social Science Plaza, Irvine, CA 92697 USA giverson@uci.edu

[Chapter to appear in H. Hoijtink, I. Klugkist, & P. Boelen (eds., 2008), *Bayesian Evaluation of Informative Hypotheses*. Springer, Berlin.]

1 Introduction

Theory testing lies at the heart of the scientific process. This is especially true in psychology, where typically, multiple theories are advanced to explain a given psychological phenomenon, such as a mental disorder or a perceptual process. It is therefore important to have a rigorous methodology available for the psychologist to evaluate the validity and viability of such theories, or models for that matter. However, it may be argued that the current practice of theory testing is not entirely satisfactory. Most often, data modelling and analysis are carried out with methods of null hypothesis significant testing (NHST). Problems with and deficiencies of NHST as a theory testing methodology have been well documented and widely discussed in the field, especially in the past few years (e.g., [42]). The reader is directed to Chapter 9 of this volume for illuminating discussions of the issues. Below we highlight some of the main problems of NHST.

First of all, NHST does not allow one to address directly the questions s/he wants to answer: how does information in the data modify his or her initial beliefs about the underlying processes?; and how likely is it that a given theory or hypothesis provides an explanation for the data?, i.e., one would like to compute $Prob(hypothesis|data)$. Instead, the decision as to whether one should retain or reject a hypothesis is based on the probability of observing the current data given the assumption that the hypothesis is correct, that is, $Prob(data|hypothesis)$. The two probabilities, $Prob(data|hypothesis)$ and $Prob(hypothesis|data)$, are generally not equal to each other, and may even

differ from each other by large amounts. Second, NHST is often conducted in a manner that it is the null hypothesis that is put to the test, not the hypothesis the researcher would like to test. The latter hypothesis called the alternative hypothesis does not get attended to unless the null hypothesis has been examined and rejected subsequently. In other words, there is an imbalance in weighing the null and alternative hypotheses against each other as an explanation of the data. Third, many NHST tests are loaded with simplifying assumptions, such as normality, linearity, and equality of variances, that are often violated by real-world data. Finally, the p-value, the yardstick of NHST, is prone to misuse and misinterpretation, and this occurs more often than one might suspect, and is, in fact, commonplace (see, e.g., Chapter 9). For example, a p-value is often misinterpreted as an evidence measure of the probability that the null hypothesis is true.

These methodological problems go beyond just NHST and are intrinsic to any frequentist methodology. Consequently, they represent limitations and challenges for the frequentist approach to statistical inference. Are there any alternatives to NHST? Fortunately, there is one, namely, the Bayesian approach to statistical inference, which is free of the problems we discussed above. Unlike NHST, in Bayesian inference, (1) one directly computes the probability of hypothesis given data $Prob(hypothesis|data)$; (2) two or more hypotheses are evaluated by weighing them *equally*; (3) any realistic set of assumptions about the underlying processes can easily be incorporated into a Bayesian model; and (4) interpretations of Bayesian results are intuitive and straightforward. What is apparent from the other chapters of the current volume (see the evaluation given in Chapter 5) is the fact that it is much more straightforward to pose and test order-restricted hypotheses with order constraints within the Bayesian framework, compared to the frequentist NHST approach to testing such hypotheses. By definition, an order-restricted hypothesis is a hypothesis where a set of parameters are consistent with a particular order relation, and will be called an ‘informed hypothesis’ in this chapter, so as to be consistent with the terminology used in the other chapters.

A purpose of this chapter is to present a review of recent efforts to develop Bayesian tools for evaluating order-constrained hypotheses for psychological data. In so doing, we provide our own critiques on some of the chapters in this volume, discussing their strengths and weaknesses. Another purpose of writing this chapter is to present an example application of hierarchical Bayesian modelling for analyzing data with a structure that is ideal for an Analysis of Variance (ANOVA) and to compare performance of several Bayesian model comparison criteria proposed and discussed throughout the current volume. We begin by reviewing the literature on Bayesian order-restricted inference.

2 Bayesian Order-restricted Inference

Order-restricted models, that is, models with parameters subject to a set of order constraints, have long been considered in frequentist statistics. Isotonic regression exemplifies this approach, the theoretical foundations of which are summarized in [2], [35] and [40]. It seems appropriate, then, to include a brief description of the frequentist approach to order-restricted inference, before discussing a Bayesian alternative.

For the purposes of testing order-restricted hypotheses, the isotonic regression model leads to a special kind of likelihood-ratio test. Specifically, the test statistic in isotonic regression is the log-likelihood ratio of the maximum likelihood estimate of a reduced model with equal means to that of a full model with certain order constraints imposed on its means. Note that the former model is nested within the latter one. The sampling distribution of the test statistic is then sought under the null hypothesis that all means are equal against the alternative hypothesis that the means satisfy the order constraints. This turns out, however, to be a major hurdle to the method's widespread application in practice; there is no easy-to-compute, general solution for finding the sampling distribution for given forms of order constraints, unless the constraints belong to one of a few simplified forms.⁴ Even if one is able to derive the desired sampling distribution, given the fact that isotonic regression is a null hypothesis testing significance test (NHST), the problems associated with the use of NHST and p-value for model evaluation are still at issue, as discussed at length in Chapter 9 and as critiqued by Kato and Hoijtink [23] who commented "Even though a great deal of frequentist literature exists on order-restricted parameter problems, most of the attention is focused on estimation and hypothesis testing [as opposed to model evaluation and comparison]" (p. 1).

As an alternative to the frequentist framework, a Bayesian approach to order-restricted inference was considered in the past (e.g., [39]). However, its application was limited due to the intractability of evaluating the posterior integral. This long-standing difficulty in Bayesian computation has been overcome in the 1990s with the introduction of general purpose sampling algorithms collectively known as *Markov chain Monte Carlo* (MCMC: [9], [13], [34]). With MCMC, *theoretical Bayes* has become *practical Bayes*. In particular, Gelfand, Smith and Lee [10] developed easily implementable MCMC methods for sampling from posterior distributions of model parameters under order-constraints. Since then, a group of quantitative psychologists have

⁴ As an alternative to the isotonic regression likelihood-ratio test, Geyer [12] proposed bootstrap tests in which one computes approximate p values for the likelihood ratio test by simulating the sampling distribution by an *iterated* parametric bootstrap procedure. One problem with the bootstrap, which may be easy to compute, does not have finite sampling properties, and therefore, can give biased estimates of sampling distributions for finite samples [7]. Further, the bootstrap is a frequentist approach that is subject to the problems discussed earlier.

demonstrated the application of the Bayesian framework on a wide range of order-restricted inference problems in psychology, education and economics ([16], [20], [21], [24], [30]). This success prompted Hoijtink and his colleagues to organize the Utrecht Workshop in the summer of 2007, which subsequently led to the publication of the current volume.

2.1 Why Bayesian?

Bayesian inference, at its core, is the process of updating one's initial belief (prior) about the state of a world in light of observations (evidence) with the use of Bayes theorem, thereby forming a new belief (posterior). This way of making inferences is fundamentally different from that of frequentist inference. Among many differences between the two schools of statistics, Bayesian and frequentist, the most notable include the former's interpretation of probability as an individual's degree of belief as opposed to the long-run frequency ratio in the latter and also, the Bayesian view of model parameters as random variables as opposed to fixed but unknown constants in frequentist statistics. For up-to-date and comprehensive treatments of Bayesian methods, the reader is directed to [11] and [33].

Besides such theoretical, and philosophical, differences between the two inference schemes, Bayesian inference offers many pragmatic advantages over its frequentist counterpart, in particular, in the context of evaluating informed hypotheses with parametric order constraints. The advantages may be termed directness of inference, automaticity, power of priors, and finally, ease of computation. Firstly, by directness of inference, we mean that the Bayesian inference process directly addresses the question the researcher wishes to answer, that is, how data modifies his or her belief about initial hypotheses. In contrast, frequentist inferences are based on the probability (i.e., p-value) of obtaining current data or more extreme data under the assumption that the researcher's initial hypothesis is correct, which seems awkward and even confusing. Secondly, Bayesian inference is automatic as there is just *one road* to data analysis: each and every inference problem boils down to finding the posterior from the likelihood function and the prior by applying Bayes theorem. Thirdly, Bayesian statistics allows one to easily incorporate any available relevant information, other than observed data, into the inference process through priors. Being able to incorporate prior information into data modeling, which undoubtedly improves the quality of inferences, is indeed a powerful and uniquely Bayesian idea, with no counterpart in frequentist statistics. This is also one of the reasons Bayesian statistics has gained such popularity in fields dealing with practical problems of real-world significance such as biomedical sciences and engineering disciplines—one cannot afford to disregard potentially useful information which might help save lives or generate millions of dollars! Finally, as mentioned earlier, the recent breakthrough in Bayesian computation makes it routinely possible to make inferences about any given informed hypothesis. The necessary computations for any arbitrary form of

order constraints can be performed via MCMC, as easily as one is running simple simulations on computer.

In what follows, we provide a broad-brush overview of the Bayesian order-restricted inference framework that is described and illustrated in greater detail by various authors of this volume, with a special attention given to the comparative review on the pros and cons of the Bayesian methods discussed in the various chapters.

2.2 The Specifics of The Bayesian Approach

The key idea of the Bayesian approach for testing and evaluating an informed hypothesis is to incorporate the order constraints specified by the hypothesis into the prior distribution. For example, for an informed hypothesis $H : \mu_1 < \mu_2 < \mu_3$ expressed in terms of means μ 's, the order constraint is represented by the following prior for the parameter vector $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3)$

$$p(\boldsymbol{\theta}) = \begin{cases} g(\boldsymbol{\theta}) & \text{if } \mu_1 < \mu_2 < \mu_3 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

for some probability measure function that integrates to 1. Given observed data $\mathbf{y} = (y_1, \dots, y_n)$ and the likelihood $f(\mathbf{y}|\boldsymbol{\theta})$, the posterior is obtained from Bayes rule as

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (2)$$

The posterior distribution in (2) represents a complete summary of information about the parameter $\boldsymbol{\theta}$ and is used to draw specific inferences about it. For instance, we may be interested in finding the posterior mean and Bayesian credible intervals. Each of these measures can be expressed as a posterior expectation. The trouble is that since the normalizing constant $\int f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ in the denominator is commonly intractable for all but the simplest models, the posterior distribution is only known up to a proportionality constant. Even if the posterior is known in analytic form, finding its mean and credible intervals can be challenging. The next best thing, then, beyond knowing the exact expression of the posterior, is to generate a large number of samples that approximate the distribution and to use the samples to numerically estimate the expectation of interest. This is where MCMC comes in handy, as the technique allows us to draw samples from almost any form of posterior distribution without having to know its normalizing constant, that is, the denominator in (2).

When one entertains multiple hypotheses and wishes to compare them, this can be achieved using the Bayes factor (BF), which, for two hypotheses H_i and H_j , is defined as the ratio of their marginal likelihoods

$$BF_{ij} = \frac{m(\mathbf{y}|H_i)}{m(\mathbf{y}|H_j)} = \frac{\int f(\mathbf{y}|\boldsymbol{\theta}, H_i)p(\boldsymbol{\theta}|H_i)d\boldsymbol{\theta}}{\int f(\mathbf{y}|\boldsymbol{\theta}, H_j)p(\boldsymbol{\theta}|H_j)d\boldsymbol{\theta}}, \quad (3)$$

where $m(\mathbf{y}|H_i)$ denotes the marginal likelihood under hypothesis H_i . The Bayes factor has several attractive features as a model selection measure. First, the Bayes factor is related to the posterior hypothesis probability—the probability of a hypothesis being true given observed data. That is, from a set of BFs computed for each pair of competing hypotheses, the posterior probability of hypothesis $p(H_i|\mathbf{y})$, $i = 1, \dots, q$, is given as $p(H_i|\mathbf{y}) = BF_{ik} / \sum_{j=1}^q BF_{jk}$, $i = 1, \dots, q$ for any choice of $k = 1, \dots, q$, under the assumption of equal prior probabilities $p(H_i) = 1/q$ for all i 's. Further, Bayes factor-based model selection automatically adjusts for model complexity and avoids overfitting, thereby representing a formal implementation of Occam's razor. What this means is that BF selects the one, among a set of competing hypotheses, that provides the simplest explanation of the data.

Another attractive feature of the Bayes factor, that is particularly fitting for evaluating order constrained hypotheses, is that the model selection measure is applicable for choosing between hypotheses that vary in the number of parameters but also importantly, for comparing multiple informed hypotheses that posit different order constraints but share a common set of parameters. For example, consider the following three hypotheses: $H_1 : \mu_1, \mu_2, \mu_3$; $H_2 : \mu_1, \{\mu_2 < \mu_3\}$; $H_3 : \mu_1 < \mu_2 < \mu_3$. It is worth noting here that commonly used selection criteria like the Akaike Information Criterion (AIC, [1]) and the Bayesian Information Criterion (BIC, [38]) that only consider the number of parameters in their complexity penalty term are inappropriate in this case. This is because the two criteria treat the above three hypotheses equally complex (or flexible), which is obviously not the case.

Accompanying these desirable properties of the Bayes factor are some important caveats. First of all, the Bayes factor can be ill-defined and cannot be used under certain improper priors. An improper prior, by definition, does not integrate finitely so we will have $\int p(\boldsymbol{\theta})_{improper} d\boldsymbol{\theta} = \infty$. For example, the prior $p(\theta) \propto 1/\theta$ is improper over the parameter range $0 < \theta < \infty$, and so is the uniform prior $p(\theta) = c$ for an unspecified constant c over the same range of the parameter θ . To illustrate, suppose that each element of the data vector $\mathbf{y} = (y_1, \dots, y_n)$ is an independent sample from a normal distribution $N(\mu, \sigma^2)$ with unknown mean μ but known variance σ^2 . In this case, the sample mean \bar{y} is a sufficient statistic for parameter μ . The likelihood is then given by

$$f(\bar{y}|\mu) = \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{n})} \exp\left(-\frac{1}{2\sigma^2/n}(\bar{y} - \mu)^2\right) \quad (4)$$

as a function of parameter μ . If we were to use the improper uniform prior $p(\mu) = c$ for $-\infty < \mu < \infty$, the marginal likelihood $m(\bar{y}) = \int_{-\infty}^{+\infty} f(\bar{y}|\mu)p(\mu)d\mu$ would contain the 'unspecified constant' c , and as such, the Bayes factor value in (3) would be undetermined.⁵ Interestingly however, for the present example, it is easy to see that the posterior distribution $p(\mu|\bar{y})$ is proper with

⁵ An exception to this 'undetermined' Bayes factor case is when the marginal likelihood of the other hypothesis being compared against the current one also contains

its finite normalizing constant. This is because the unspecified constant c “conveniently” cancels out in the application of Bayes rule to find the posterior

$$p(\mu|\bar{y}) = \frac{f(\bar{y}|\mu)p(\mu)}{\int f(\bar{y}|\mu)p(\mu) d\mu} = \frac{f(\bar{y}|\mu)}{\int f(\bar{y}|\mu) d\mu} \quad (5)$$

which integrates to one for $-\infty < \mu < \infty$. An important implication is that in a case like this, posterior-based inferences such as Bayesian confidence interval estimation and Deviance Information Criterion (DIC, [41]) based model selection are well-defined and applicable whereas the Bayes factor is not. We will come back to this later in this chapter.

Secondly, another caveat is about using the Bayes factor for the comparison of two nested models. It is well known that the Bayes factor can be highly sensitive to the choice of priors, especially under diffuse priors with relatively large variances. In other words, the Bayes factor value can fluctuate widely and nonsensically to incidental minor variations of the priors. This is connected to the *Lindley’s paradox* (e.g., [31]). Therefore, for nested models, Bayes factors under diffuse priors must be interpreted with great care.

The last, and by no means least, challenge for the Bayes factor as a model selection measure is a heavy computational burden. The Bayes factor is non-trivial to compute. To date, there exists no general purpose numerical method for routinely computing the required marginal likelihood, especially for non-linear models with many parameters and non-conjugate priors.

Addressing these issues and challenges in Bayes factor calculations, Klugkist, Hoijtink and their colleagues (Chapter 4, [24], [25]) have developed an elegant technique for estimating the Bayes factor from prior and posterior samples for order-restricted hypotheses, without having to directly compute their marginal likelihoods. In following section, we provide a critical review of the essentials of the method, which may be called the *encompassing prior Bayes factor* approach, or the encompassing Bayes approach in short.

2.3 Encompassing Prior Bayes Factors

The encompassing Bayes approach has been developed specifically for model selection with informed hypotheses. Specifically, the approach requires the setting of two nested hypotheses, H_1 and H_2 , that share the same set of parameters but differ from each other in the form of parametric constraints, for example, $H_1 : \mu_1, \mu_2, \mu_3$ and $H_2 : \mu_1, \{\mu_2 < \mu_3\}$. For simplicity, in this section we assume that hypothesis H_2 is nested within hypothesis H_1 . Another condition required for the application of the encompassing Bayes approach is that the prior distribution of the smaller hypothesis H_2 is obtained from the prior distribution of the larger hypothesis H_1 simply by restricting the

the same constant c so both ‘unspecified constants’ do cancel each other out in the calculation of the ratio of the two marginal likelihoods.

parameter space of H_1 in accordance with the order constraints imposed by H_2 . Formally, this condition can be stated as

$$p(\boldsymbol{\theta}|H_2) \propto \begin{cases} p(\boldsymbol{\theta}|H_1) & \text{if } \boldsymbol{\theta} \text{ is in agreement with } H_2, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

With these two conditions met, it has been shown that the Bayes factor can be approximated as a ratio of two proportions (Chapter 4, [25])

$$BF_{21} \approx \frac{r_{post21}}{r_{pre21}}. \quad (7)$$

In the equation, r_{post21} denotes the proportion of samples from the posterior distribution of hypothesis H_1 , $p(\boldsymbol{\theta}|\mathbf{y}, H_1)$, that satisfy the order constraints of hypothesis H_2 . Similarly, r_{pre21} denotes the proportion of samples from the prior distribution $p(\boldsymbol{\theta}|H_1)$ that also satisfy the order constraints of hypothesis H_2 . The beauty of the encompassing Bayes lies in that its implementation requires only the ability to sample from the prior and the posterior of the larger of the two hypotheses, without having to deal with their marginal likelihoods, which can be quite difficult to compute, as mentioned earlier.

The Bayes factor calculated using the computational ‘trick’ in (7) may have large variances especially when the smaller hypothesis is too highly constrained to yield stable estimates of the proportions. r_{post} and r_{pre} . In such cases, one may resort to the following more efficient estimation method. We first note that the Bayes factor for two nested hypotheses H_q and H_1 where $H_q \subset H_1$ can be rewritten in terms of a series of (artificial) Bayes factors corresponding to pairs of nested hypotheses created by recursively constraining the parameter space of H_1 as

$$BF_{q1} = BF_{q(q-1)} \cdot BF_{(q-1)(q-2)} \cdots BF_{21} \quad (8)$$

for $H_q \subset H_{q-1} \subset \dots \subset H_2 \subset H_1$. Using this equality, one can then compute the desired BF_{q1} as a product of BF_{ij} ’s, each of which is in turn estimated from an equation analogous to (7) using any standard MCMC algorithms or the ones that are specifically tailored to order constrained hypotheses (e.g., [10]).

The encompassing Bayes approach is quite an ingenious idea that allows one to routinely compute Bayes factors simply by sampling from prior and posterior distributions, thereby bypassing the potentially steep hurdle of computing the marginal likelihood. As demonstrated in various chapters of this book, the approach has been successfully applied to comparing order constrained hypotheses that arise in a wide range of data analysis problems including analysis of variance, analysis of covariance, multilevel analysis, and analysis of contingency tables.

There is, however, one assumption of the encompassing Bayes approach that may limit its general application. This is the requirement that all hypotheses, constrained or unconstrained, be of the same dimension. To illustrate, consider the following two hypotheses:

$$\begin{aligned} H_1 &: \mu_1, \mu_2, \mu_3, \\ H_2 &: \mu_1 = \mu_2 < \mu_3. \end{aligned} \tag{9}$$

Note that H_1 has three free parameters whereas H_2 has two. In this case, the Bayes factor in (7) is undefined as both prior and posterior proportions is effectively equal to zero. Klugkist, in Chapter 4, outlines a heuristic procedure that may be employed to approximate the Bayes factor for equality-constrained hypotheses. Briefly, according to the procedure, we first construct a series of ‘near equality’ hypotheses of varying degrees,

$$H_2(\delta_i) : |\mu_1 - \mu_2| < \delta_i, \{\mu_1 < \mu_3\}, \{\mu_2 < \mu_3\}, \quad (i = 1, 2, \dots, q) \tag{10}$$

for $\delta_1 > \delta_2 > \dots > \delta_q > 0$. We then estimate the Bayes factor using the formulation in (8) by letting $\delta_q \rightarrow 0$, provided that the estimate converges to a constant. This is quite an elegant trick, though a problem may arise if the estimate does not converge, meaning that the final estimate is highly dependent upon the particular choice of limiting sequences $\{\delta_1, \delta_2, \dots, \delta_q\}$ and/or upon the choice of priors. Further theoretical work showing that this is not generally the case is clearly needed.

Continuing the discussion on the model selection problem with informed hypotheses involving equality constrained hypotheses, one can think of at least two alternative methods, other than Klugkist’s procedure described above.

The first is the *completing and splitting* method that is introduced in Chapter 7. To illustrate, consider again the two hypotheses: $H_1 : \mu_1, \mu_2, \mu_3; H_2 : \mu_1 = \mu_2 < \mu_3$. The basic idea of the completing and splitting method is to add a third ‘surrogate’ hypothesis H_3 to the original two. The new hypothesis is constructed by removing the order constraint from H_2 but keeping the equality constraint, that is, $H_3 : \{\mu_1 = \mu_2\}, \mu_3$. Note that H_3 is of the same dimension (i.e., 2) as H_2 so one can apply the encompassing Bayes approach to obtain the Bayes factor for these two hypotheses. Now, the desired Bayes factor BF_{21} we wanted to compute is then expressed in terms of the ‘surrogate’ hypothesis H_3 as $BF_{21} = BF_{23} \cdot BF_{31}$. In this expression, the first factor BF_{23} in the right hand side is calculated using the encompassing Bayes approach in (7). As for the second factor BF_{31} for two unconstrained hypotheses that differ in dimensions, this quantity may be computed by using an appropriate prior distribution with the usual Bayesian computational methods, or alternatively, with data-based prior methods such as the intrinsic Bayes factor [3] and the fractional Bayes factor [32]. Incidentally, it would be of interest to examine whether Klugkist’s procedure would yield the same Bayes factor value as the completing and splitting method.

The second approach for dealing with equality hypotheses represents a departure from Bayes factor based model selection. Model selection criteria proposed under this approach may be termed collectively *posterior predictive selection methods* and are discussed in great detail in Chapter 8. In the following section, we provide a critical review of these methods and their relations to Bayes factors.

2.4 Posterior Predictive Selection Criteria

The posterior predictive model selection criteria discussed in Chapter 8 are the L-measure ([4], [14]), the Deviance Information Criterion (DIC, [11], [41]), and the Logarithm of the Pseudomarginal Likelihood (LPML, [8], [15]). All three measures are defined with respect to the posterior predictive distribution (ppd) of future, yet-to-be-observed data \mathbf{z}

$$f_{ppd}(\mathbf{z}|\mathbf{y}_{obs}) = \int f(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{obs})d\boldsymbol{\theta}, \quad (11)$$

where $\mathbf{y}_{obs} = (y_{1,obs}, \dots, y_{n,obs})$ is the currently observed data.⁶ Samples from this predictive distribution represent predictions for future observations from the same process that has generated the observed data.

A posterior predictive criterion is designed to assess a model's or hypothesis's predictive accuracy for future samples. The above three criteria differ from one another in the form of the predictive accuracy measure employed

$$\begin{aligned} L\text{-measure} &= E(\mathbf{z} - \mathbf{y}_{obs})^2, \\ DIC &= E[-2 \ln f(\mathbf{z}|\bar{\boldsymbol{\theta}}(\mathbf{y}_{obs}))], \\ LPML &= \sum_{i=1}^n \ln f_{ppd}(y_{i,obs}|\mathbf{y}_{obs}^{(-i)}), \end{aligned} \quad (12)$$

where $\bar{\boldsymbol{\theta}}$ denotes the posterior mean, $\mathbf{y}_{obs}^{(-i)}$ denotes \mathbf{y}_{obs} with the i -th observation deleted, and finally, all expectations $E(\cdot)$ are taken with respect to the posterior predictive distribution $f_{ppd}(\mathbf{z}|\mathbf{y}_{obs})$. Under suitable assumptions, each of the above 'theoretical' measures is approximately estimated by the following 'computable' expression

$$\begin{aligned} L\text{-measure} &= \sum_{i=1}^n (E_{\boldsymbol{\theta}|\mathbf{y}_{obs}} [E_{\mathbf{y}|\boldsymbol{\theta}}(z_i^2|\boldsymbol{\theta})] - \mu_i^2) + \nu \sum_{i=1}^n (\mu_i - y_{i,obs})^2, \\ DIC &= D(\bar{\boldsymbol{\theta}}) + 2p_D, \\ LPML &= \sum_{i=1}^n \ln E_{\boldsymbol{\theta}|\mathbf{y}_{obs}^{(-i)}} [f(y_{i,obs}|\boldsymbol{\theta})]. \end{aligned} \quad (13)$$

In the first equation defining the *L-measure* criterion, z_i is a future response with the sampling distribution as $f(\mathbf{y}|\boldsymbol{\theta})$, ν is a tuning parameter to be fixed between 0 and 1, and $\mu_i = E_{\boldsymbol{\theta}|\mathbf{y}_{obs}} [E_{\mathbf{y}|\boldsymbol{\theta}}(z_i|\boldsymbol{\theta})]$ with the first expectation defined with respect to the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{obs})$ and the second expectation defined with respect to the sampling distribution $f(\mathbf{y}|\boldsymbol{\theta})$. In the

⁶ The subscript *obs* in \mathbf{y}_{obs} is inserted to explicitly indicate the observed data so as to avoid confusions with another symbol \mathbf{y} , which is used in an equation below to denote a random vector.

second expression defining DIC, $D(\boldsymbol{\theta})$ is the deviance function given data vector \mathbf{y}_{obs} defined as $D(\boldsymbol{\theta}) = -2 \ln f(\mathbf{y}_{obs}|\boldsymbol{\theta})$ (see, e.g., [29]), $\bar{\boldsymbol{\theta}}$ denotes the mean of $\boldsymbol{\theta}$ with respect to the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{obs})$, and finally, p_D is the effective number of model parameters, or a model complexity (flexibility) measure, defined as $p_D = \bar{D}(\boldsymbol{\theta}) - D(\bar{\boldsymbol{\theta}})$. In the third expression regarding LPML, the expectation is with regard to the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{obs}^{(-i)})$. For L-measure and DIC, the smaller their value, the better the model. The opposite is true for LPML.

The three model selection criteria in (13) differ in at least two important ways from the Bayes factor. First, they are predictive measures the goal of which is to pick a model or hypothesis that achieves best predictions for future data. In contrast, the goal of Bayes factor model selection is to find the model with highest posterior model probability. Second, all three criteria are defined based on samples from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{obs})$. As such, it is straightforward to compute the criteria with any standard MCMC methods for order constrained hypotheses and even for equality constrained hypotheses, which can be particularly thorny for Bayes factor computation.

Notwithstanding these attractive features of the predictive model selection criteria, one may object them on the grounds that they may be intuitive but are based on arbitrary measures of predictive accuracy. That is, one may ask questions such as: Why the squared error loss function in L-measure, or for that matter, the deviance function in DIC?; which of the three is the “best”; what should we do if their model choices disagree with one another? Further, regarding DIC, it is known to violate the reparameterization invariance rule [41]. Reparameterization invariance means that a model’s data fitting capability does not change, as it should, when the model’s equation is rewritten under a reparameterization. For instance, the model equation $y = \exp(-\theta x)$ can be re-expressed as $y = \eta^{-x}$ through the reparameterization $\eta = \exp(\theta)$. DIC is generally not reparameterization-invariant as the posterior mean $\bar{\boldsymbol{\theta}}$ in the DIC equation (13) does change its value under reparameterization. In short, the reader should be aware of these issues and interprets the results from the application of the posterior predictive criteria with a grain of salt.

3 Hierarchical Bayes Order-constrained Analysis of Variance

In this section, we present and discuss an exemplary application of the Bayesian approach for analyzing ANOVA-like data. In particular, we implement and demonstrate a hierarchical Bayes framework. Also discussed in the example application is a comparison between the results from Bayes factor model selection and those from posterior predictive model selection using DIC.

3.1 Blood Pressure Data and Informed Hypotheses

We consider blood pressure data that is discussed in Maxwell and Delaney's book [28] on experimental designs. These are hypothetical data created to illustrate certain statistical ideas in their book. The data are imagined to be from an experiment in which a researcher wants to study the effectiveness of diet, drugs, and biofeedback for treating hypertension. The researcher designs a $2 \times 3 \times 2$ between-subjects factorial experiment in which the diet factor varies over two levels (absent and present), the drug factor over three levels (drug X, Y and Z) and the biofeedback factor over two levels (absent and present). Blood pressure is measured for six individuals in each of twelve cells.

The full data are reported and summarized in Tables 8.12 and 8.13 of Maxwell and DeLaney [28]. Some general trends can be noticed from these tables. Both diet and biofeedback seem to be effective in lowering blood pressure. Also, among the three drugs, it appears that drug X is the most effective, and that drug Z seems better than drug Y, though the latter differences may be due to sampling error. Results from an analysis-of-variance applied to these data and reported in Table 8.14 of the book indicate that all three main effects are statistically significant, with each p-value being less than 0.001, and that one of the two-way interactions and the three-way interaction are marginally significant, i.e., $p = 0.06$ and $p = 0.04$, respectively.

Based these analysis-of-variance results, to illustrate a hierarchical Bayes order-restricted inference framework, we consider five hypotheses. They include the null hypothesis, H_0 , with no order constraints, and four informed hypotheses, $H_1 - H_4$, with varying degrees of order constraints on the population cell means:

$$\begin{aligned}
 H_0 : & \text{Unconstrained } \mu_{ijk}'s \text{ for all } i, j, k, \\
 H_1 : & \mu_{DB\bullet} < \{\mu_{D\bar{D}\bullet}, \mu_{D\bar{D}\bullet}\}; \{\mu_{D\bar{D}\bullet}, \mu_{D\bar{D}\bullet}\} < \mu_{D\bar{D}\bullet}, \\
 H_2 : & \mu_{DB\bullet} < \mu_{D\bar{D}\bullet} < \mu_{D\bar{D}\bullet} < \mu_{D\bar{D}\bullet}, \\
 H_3 : & \mu_{DBk} < \{\mu_{D\bar{D}k}, \mu_{D\bar{D}k}\}; \{\mu_{D\bar{D}k}, \mu_{D\bar{D}k}\} < \mu_{D\bar{D}k} \text{ for all } k, \\
 & \mu_{ijX} < \mu_{ijZ} < \mu_{ijY} \text{ for all } i, j, \\
 H_4 : & \mu_{DBk} < \mu_{D\bar{D}k} < \mu_{D\bar{D}k} < \mu_{D\bar{D}k} \text{ for all } k, \\
 & \mu_{ijX} < \mu_{ijZ} < \mu_{ijY} \text{ for all } i, j.
 \end{aligned} \tag{14}$$

In the above equation the subscript i denotes the level of the diet factor (D : present; \bar{D} : absent), the subscript j denotes the level of the biofeedback factor (B : present; \bar{B} : absent), and finally, the subscript k denotes the drug type (X, Y or Z). The subscript \bullet indicates that the result is averaged across all levels of the corresponding factor.

Shown in Figure 1 are the four informed hypotheses in graphical form. The data are found to violate none of the order constraints specified by hypothesis H_1 or by hypothesis H_2 . In contrast, as marked by the asterisk symbol $*$ in the figure, three violations of the order constraints under H_3 and four violations of the order constraints under H_4 are observed in the data. A question one

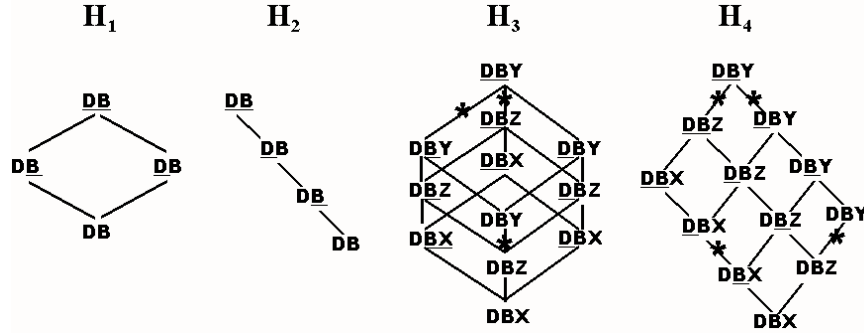


Fig. 1. The four informed hypotheses defined in (14). In each connected graph, for two treatment conditions that are connected to each other, the one that is positioned above the other has a higher population mean value and as such is *less effective* in treating high blood pressure than the other condition. The asterisk mark * indicates a violation of the corresponding ordinal prediction in the data.

might ask, then, are these violations “real” or just sampling errors? In the following section, we present a hierarchical Bayesian analysis that attempts to answer questions such as this.

3.2 Hierarchical Bayesian Analysis

Given the five hypotheses in (14) the model selection problem is to identify the hypothesis that best describes the blood pressure data. To this end, we present a hierarchical Bayesian framework and discuss results from its application to the data.

A defining feature of hierarchical Bayesian modelling is the set-up of multi-level dependency relationships between model parameters such that lower-level parameters are specified probabilistically in terms of higher-level parameters, known as *hyper-parameters*, which themselves may in turn be given another probabilistic specification in terms of even higher-level parameters, and so on [11]. The hierarchical modelling generally improves the robustness of the resulting Bayesian inferences with respect to prior specification [33]. Importantly, the hierarchical set-up of parameters is particularly suitable for modelling various kinds of dependence structure that the data might exhibit, such as individual differences in response variables and trial-by-trial dependency of reaction times. Recently, the hierarchical Bayesian modelling has become increasingly popular in cognitive modelling, and its utility and success have been well demonstrated (see, e.g., [26], [27], [36], [37]).

Using standard distributional notation, we now specify the hierarchical Bayesian framework for modelling the blood pressure data as

$$Likelihood : y_{ijkl} \sim N(\mu_{ijk}, \sigma^2),$$

(15)

$$\begin{aligned}
& \mu_{ijk} | \eta, \tau^2 \sim N(\eta, \tau^2) \\
& \eta | \psi^2 \sim N(0, \psi^2) \\
\text{Priors : } & \tau^2 | a, b \sim IG(a, b) \\
& \sigma^2 | c, d \sim IG(c, d),
\end{aligned}$$

where $i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K, l = 1, \dots, n$; N denotes a normal distribution, IG denotes an inverse Gamma distribution⁷, and ψ^2, a, b, c, d are fixed constants. Note in the above equation that η and τ^2 represent two hyperparameters assumed in the model. For the blood pressure data, there were six persons in each of the 12 cells created by the 2 x 2 x 3 factorial design, and as such, we have $I = 2, J = 2, K = 3$ and $n = 6$.

Let us define the data vector as $\mathbf{y} = (y_{1111}, \dots, y_{IJKn})$ and the parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\mu}, \eta, \tau^2, \sigma^2)$ where $\boldsymbol{\mu} = (\mu_{111}, \dots, \mu_{IJK})$. The posterior density under the unconstrained hypothesis H_0 in (14) is then given by

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\mu}, \sigma^2) p(\boldsymbol{\mu} | \eta, \tau^2) p(\eta | \psi^2) p(\tau^2 | a, b) p(\sigma^2 | c, d) \quad (16)$$

with the likelihood function of the following form

$$f(\mathbf{y} | \boldsymbol{\mu}, \sigma^2) = \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{l=1}^n \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{1}{2\sigma^2} (y_{ijkl} - \mu_{ijk})^2\right). \quad (17)$$

From these expressions, one can easily derive the full conditional posterior distributions of various parameters as

$$\begin{aligned}
p(\mu_{ijk} | \mathbf{y}, \boldsymbol{\mu}^{(-ijk)}, \eta, \tau^2, \sigma^2) & \sim N\left(\frac{\frac{\sigma^2}{n} \eta + \tau^2 \frac{\sum_{l=1}^n y_{ijkl}}{n}}{\frac{\sigma^2}{n} + \tau^2}, \frac{\frac{\sigma^2}{n} \tau^2}{\frac{\sigma^2}{n} + \tau^2}\right) \\
p(\eta | \mathbf{y}, \boldsymbol{\mu}, \tau^2, \sigma^2) & \sim N\left(\frac{\psi^2}{IJK\psi^2 + \tau^2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \mu_{ijk}, \frac{\psi^2 \tau^2}{IJK\sigma^2 + \tau^2}\right) \\
p(\tau^2 | \mathbf{y}, \boldsymbol{\mu}, \eta, \sigma^2) & \sim IG\left(a + \frac{IJK}{2}, \left[\frac{1}{b} + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\mu_{ijk} - \eta)^2\right]^{-1}\right) \\
p(\sigma^2 | \mathbf{y}, \boldsymbol{\mu}, \eta, \tau^2) & \sim IG\left(c + \frac{IJKn}{2}, \left[\frac{1}{d} + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^n (y_{ijkl} - \mu_{ijk})^2\right]^{-1}\right).
\end{aligned} \quad (18)$$

From these full conditionals for the unconstrained hypothesis, a Gibbs sampler can be devised to draw posterior samples from an informed hypothesis

⁷ The probability density function of the Gamma and Inverse-Gamma distributions are defined as $G(a, b) : f(x|a, b) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b} (a, b > 0; 0 < x < \infty)$; $IG(a, b) : f(x|a, b) = \frac{1}{\Gamma(a)b^a} x^{-a-1} e^{-1/bx} (a, b > 0; 0 < x < \infty)$. Note that $X \sim G(a, b) \iff 1/X \sim IG(a, b)$.

with order constraints of the form $\alpha \leq \theta_i \leq \beta$, specifically, the following *inverse probability sampling* procedure [10]

$$\theta_i = F_i^{-1} [F_i(\alpha) + U \cdot (F_i(\beta) - F_i(\alpha))], \quad (19)$$

where F_i is the cumulative full conditional distribution for θ_i of the unconstrained hypothesis, F_i^{-1} is its inverse, and U is a uniform random number on $[0, 1]$. It should be noted that special care needs to be taken in applying this procedure for hierarchical models with constrained parameters. This is because the normalizing constants for lower-level parameters generally depend upon the values of higher-level parameters so the constants do not cancel one another out, thereby making the implementation of Gibbs sampling difficult, if not impossible. Chen and Shao [5] developed efficient Monte Carlo methods that address the problem. We implemented their methods in our application of the inverse probability sampling procedure.

From posterior samples, one can then compute the DIC criterion in (13) with the deviance function $D(\boldsymbol{\theta})$ for the data model in (15) expressed as

$$\begin{aligned} D(\boldsymbol{\theta}) &= -2 \ln f(\mathbf{y} | \boldsymbol{\mu}, \sigma^2) \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \frac{(\mu_{ijk} - \bar{y}_{ijk})^2}{\sigma^2/n} + IJK \cdot \ln(2\pi\sigma^2/n), \end{aligned} \quad (20)$$

where \bar{y}_{ijk} represents the sample mean for cell ijk . The Bayes factors and the posterior model probabilities for the five hypotheses in (14) are estimated using the encompassing Bayes approach discussed earlier.

The model comparison results are presented in Table 1. Shown in the second column of the table are the p_D values, which measure the effective number of parameters. All five hypotheses assume the same number of parameters (i.e., 15) including the two hyper-parameters of η and τ^2 , and yet, obviously they differ in model complexity (flexibility) as each imposes different degrees of order constraints upon the parameters. Note that the unconstrained hypothesis H_0 has the largest p_D value of 9.61 and then the complexity value decreases from top to bottom of the column. This pattern of result agrees with the intuitive notion that the more order constraints an informed hypothesis assumes, the less complexity the hypothesis presents. The DIC results shown on the third column indicate that among the five informed hypotheses, the simplest one H_4 is the best predicting model from the posterior predictive standpoint.

The remaining columns of the table present the encompassing Bayes results. First of all, recall that the r_{preq0} and r_{postq0} values estimate the proportions of prior and posterior samples, respectively, drawn from the unconstrained hypothesis H_0 that satisfy the order constraints of an informed hypothesis. We note that both of these proportion values exhibit the same decreasing trend as the p_D values, though it is a much steeper for the r_{preq0} and r_{postq0} values. Next, the Bayes factor results, shown in the sixth column, clearly point to H_3 and H_4 as two “winners” in the model selection competition. Between these two, H_4 has a Bayes factor that is about double the

Table 1. Model comparison results for the five hypotheses in (14) and the blood pressure data in Maxwell and Delaney’s book [28]. The DIC results are based on the following parameter values for the hyper priors: $a = 10, b = 0.01, c = 10, d = 0.01; \psi = 4000$. For each hypothesis, the mean DIC value and the 95% confidence interval based on ten independent runs of the inverse probability sampling procedure are shown. The encompassing prior Bayes factors are based on 30 million samples drawn from each of the prior and posterior distributions under the unconstrained hypothesis H_0

Hypothesis	p_D	DIC	r_{preq0}	r_{postq0}	BF_{q0}	$p(H_q \mathbf{y})$
H_0	9.61	37.06 ± 0.11	1.000	1.000	1.00	0.0004
H_1	7.50	34.10 ± 0.52	0.080	0.570	7.15	0.003
H_2	7.03	33.52 ± 1.42	0.041	0.49	12.0	0.005
H_3	5.70	32.14 ± 1.57	5.0e-06	0.0038	711	0.31
H_4	5.03	30.96 ± 1.09	7.7e-07	0.0012	1533	0.67

corresponding factor for H_3 . This result, taking into account the other Bayes factor values in the same column, translates into the posterior hypothesis probabilities of 0.67 and 0.31 for H_4 and H_3 , respectively. So if we were to choose between these two informed hypotheses, it would then be H_4 as the one that is most likely to have generated the data. An implication of this conclusion is that the four violations in the data of the order constraints specified by H_4 (see Figure 1) are judged to be no more than sampling variations, and not due to systematic deviations of the underlying data-generating process from the said hypothesis.

To summarize, both DIC and Bayes factor based selection criteria pick the hypothesis H_4 as the best model among the five competing hypotheses. Therefore, as far as the present data are concerned, the best predicting model turns out to be also the most likely model, which we find is often the case in practice.

4 Concluding Remarks

In this chapter we provided an overview of the recent developments in Bayesian order-restricted inference that are well suited to theory testing in the psychological sciences. We also discussed an application of the Bayesian framework for hierarchical modelling. Fuelled by a series of the computational breakthroughs in the early 1990s, Bayesian statistics has become increasingly popular in various scientific disciplines, in particular, in the biomedical and engineering sciences. We believe that it is the time for psychological researchers to take notice and reap the benefits of applying these powerful and versatile

inference tools to advance our understanding of the mental and behavioral phenomena we are studying. We hope this chapter will serve as another example that demonstrates the power of the Bayesian approach.

We conclude the chapter by reiterating what we said earlier: the Bayesian methods developed over the past decade for testing informed hypotheses are quite impressive in their applicability and success across a wide array of data modelling problems, as illustrated in Chapters 2–5 and 10–13 of this volume. The work is likely to be recognized in the years to come as a major contribution to the field of quantitative data modelling.

Acknowledgement

The authors are supported by United States National Science Foundation Grants SES-0241862 to JIM and SES-0242030 to GK.

References

- [1] Akaike, H.: information theory and an extension of the maximum likelihood principle. In: petrox, B. N., Caski, F. (eds.), Second International Symposium on Information Theory (pp. 267-281). Academia Kiado, Budapest (1973)
- [2] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., Bunk, H. D.: Statistical Inference under Order Restrictions. Wiley, New York (1972)
- [3] Berger, J. O., Pericchi, L. R.: The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, **91**, 109–122 (1996)
- [4] Chen, M.-H., Dey, D. K., Ibrahim, J. G.: Bayesian criterion based model assessment for categorical data. *Biometrika*, **91**, 45–63 (2004)
- [5] Chen, M.-H., Shao, Q.-M.: Monte carlo methods on Bayesian analysis of constrained parameter problems. *Biometrika*, **85**, 73–87 (1998)
- [6] Dunson, D. B., Neelon, B.: Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics*, **59**, 286–295 (2003)
- [7] Efron, B., Tibshirani, R. J.: An Introduction to the Bootstrap. Chapman & Hall, New York (1993)
- [8] Gelfand, A. E., Dey, D. K.: Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B*, **56**, 501–514 (1994)
- [9] Gelfand, A. E., Smith, A. F. M.: Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409 (1990)
- [10] Gelfand, A. E., Smith, A. F. M., Lee, R.-M.: Bayesian analysis of constrained parameter and truncated data problems. *Journal of the American Statistical Association*, **87**, 523–532 (1992)
- [11] Gelman, A., Carlin, J. B., Stern, H.S., Rubin, D. B.: Bayesian Data Analysis (second edition). Chapman & Hall/CRC, Boca Raton, Florida (2004)

- [12] Geyer, C. J.: Constrained maximum likelihood exemplified by isotonic convex logistic regression. *Journal of the American Statistical Association*, **86**, 717–724 (1991)
- [13] Gilks, W. R., Richardson, S., Spiegelhalter, D. J.: *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York (1996)
- [14] Ibrahim, J. G., Chen, M.-H., Sinha, D.: Criterion based methods for Bayesian model assessment. *Statistical Sinica*, **11**, 419–443 (2001)
- [15] Ibrahim, J. G., Chen, M.-H., Sinha, D.: *Bayesian Survival Analysis*. Springer-Verlag, New York (2001)
- [16] Iliopoulos, G., Kateri, M., Ntzoufras, I.: Bayesian estimation of unrestricted and order-restricted association models for a two-way contingency table. *Computational Statistics & Data Analysis*, **51**, 4643–4655 (2007)
- [17] Iverson, G. J., Harp, S.A.: A conditional likelihood ratio test for order restrictions in exponential families. *Mathematical Social Sciences*, **14**, 141–159 (1987)
- [18] Iverson, G. J.: Testing order in pair comparison data. Doctoral dissertation, Department of Psychology, New York University (1983)
- [19] Johnson, V. E., Albert, J. H.: *Ordinal Data Modeling*. Springer, New York (1999)
- [20] Karabatsos, G.: The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, **2**, 389–423 (2001)
- [21] Karabatsos, G., Sheu, C.-F.: Bayesian order-constrained inference for dichotomous models of unidimensional non-parametric item response theory. *Applied Psychological Measurement*, **28**, 110–125 (2004)
- [22] Kass, R. E., Raftery, A. E.: Bayes factor. *Journal of the American Statistical Association*, **90**, 773–795 (1995)
- [23] Kato, B. S., Hoijtink, H.: A Bayesian approach to inequality constrained linear mixed models: estimation and model selection. *Statistical Modelling*, **6**, 1–19 (2006)
- [24] Klugkist, I., Laudy, O., Hoijtink, H.: Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, **10**, 477–493 (2005)
- [25] Klugkist, I., Kato, B., Hoijtink, H.: Bayesian model selection using encompassing priors. *Statistic Neerlandica*, **59**, 57–69 (2005)
- [26] Lee, M. D.: A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cognitive Science*, **30**, 1–26 (2006)
- [27] Lee, M. D.: Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, **15**, 1–15 (2008)
- [28] Maxwell, S. E., Delaney, H. D.: *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (2nd edition). Lawrence Erlbaum Associates, Mahwah, New Jersey (2004)
- [29] McCullagh, P., Nelder, J. A.: *Generalized Linear Models* (2nd edition). Chapman & Hall/CRC, Boca Raton, Florida (1989)
- [30] Myung, J. I., Karabatsos, G., Iverson, G. J.: A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, **49**, 205–225 (2005)
- [31] O’Hagan, A., Forster, J.: *Kendall’s Advanced Theory of Statistics* (2nd ed.), Vol. 2B: Bayesian Inference (pp. 77–78). Arnold, London (2004)

- [32] O'Hagan, A.: Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B*, **57**, 99–138 (1995)
- [33] Robert, C. P.: *The Bayesian Choice* (second edition). Springer, New York (2001)
- [34] Robert, C. P., Casella, G.: *Monte Carlo Statistical Methods* (second edition). Springer, New York (2004)
- [35] Robertson, T., Wright, F. T., Dykstra, R. L.: *Order Restricted Statistical Inference*. Wiley, New York (1988)
- [36] Rouder, J. N., Lu, J.: An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, **12**, 573–604 (2005)
- [37] Rouder, J. N., Lu, J., Speckman, P. L., Sun, D., Jiang, Y.: A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, **12**, 195–223 (2005)
- [38] Schwartz, G: Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464 (1978)
- [39] Sedransk, J., Monahan, J., Chiu, H. Y.: Bayesian estimation of finite population parameters in categorizal data models incorporating order restrictions. *Journal of the Royal Statistical Society, Series B*, **47**, 519–527 (1985)
- [40] Silvapulle, M. J., Sen, P. K.: *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley, Hoboken, New Jersey (2005)
- [41] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A: Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**, 583–639 (2002)
- [42] Wagenmakers, E.-J.: A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, **14**, 779–804 (2007)

Index

- analysis of variance, 12
- Bayes factor, 5
- Bayesian order restricted inference, 2
- completing and splitting method, 9
- deviance information criterion (DIC), 10
- effective number of parameters, 10
- encompassing prior Bayes factor, 7
- full conditional distributions, 14
- Gibbs sampling, 14
- hierarchical Bayes, 11
- hyper-parameters, 13
- hyper-priors, 13
- improper priors, 6
- inverse probability sampling, 14
- isotonic regression, 3
- likelihood ratio test, 3
- Lindley's paradox, 7
- logarithm of pseudomarginal likelihood (LPML), 10
- L-measure, 10
- Markov chain Monte Carlo (MCMC), 3
- model complexity, 11
- Occam's razor, 6
- posterior model probability, 6
- posterior predictive distribution, 10
- posterior predictive model selection, 9

