Journal of Statistical Planning and Inference 146 (2014) 133–138

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Decision-theoretic justifications for Bayesian hypothesis testing using credible sets

Måns Thulin *

Department of Mathematics, Uppsala University, Box 480, 751 06 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 17 December 2012
Received in revised form
24 September 2013
Accepted 26 September 2013
Available online 8 October 2013

Keywords:
Bayesian inference
Credible set
Decision theory
Directional conclusion
Hypothesis testing
Three-decision problem

ABSTRACT

In Bayesian statistics the precise point-null hypothesis $\theta=\theta_0$ can be tested by checking whether θ_0 is contained in a credible set. This permits testing of $\theta=\theta_0$ without having to put prior probabilities on the hypotheses. While such inversions of credible sets have a long history in Bayesian inference, they have been criticized for lacking decision-theoretic justification.

We argue that these tests have many advantages over the standard Bayesian tests that use point-mass probabilities on the null hypothesis. We present a decision-theoretic justification for the inversion of central credible intervals, and in special case HPD sets, by studying a three-decision problem with directional conclusions. Interpreting the loss function used in the justification, we discuss when tests based on credible sets are applicable.

We then give some justifications for using credible sets when testing composite hypotheses, showing that tests based on credible sets coincide with standard tests in this setting.

1. Introduction

The first step of the standard solution to Bayesian testing of the point-null (or sharp, or precise) hypothesis $\theta\in\Theta_0=\{\theta_0\}$ is to assign a prior probability to the hypothesis. This is necessary if one wishes to use e.g. tests based on Bayes factors, as an absolutely continuous prior for θ would yield $\mathrm{P}(\theta\in\Theta_0|x)=0$ regardless of the data x. We shall refer to this procedure, described in detail e.g. by Robert (2007, Section 5.2.4), as the standard test or the standard solution in the following. A continuous prior can however be utilized in a simpler alternative strategy, in which the evidence against Θ_0 is evaluated indirectly by using inverted credible sets for testing. In this procedure a credible set Θ_c, such that $\mathrm{P}(\theta\in\Theta_c|x)\geq 1-\alpha$, is computed and the null hypothesis is rejected if $\theta_0\notin\Theta_c$. Using credible sets for testing avoids the additional complication of imposing explicit probabilities on the hypotheses, avoids Lindley's paradox and allows for the use of non-informative priors.

While it has been argued that tests of point-null hypotheses are unnatural from a Bayesian point of view, they are undeniably of considerable practical interest (Berger and Delampady, 1987; Ghosh et al., 2006, Section 2.7.2; Robert, 2007, Section 5.2.4). Inverting credible sets is arguably the simplest way to test such hypotheses. For this reason, inverting credible sets as a means to test hypotheses has long been, and continues to be, a part of Bayesian statistics. Tests of this type have been studied by Box and Tiao (1965), Hsu (1982), Kim (1991), Drummond and Rambaut (2007) and Kruschke, 2011, among many others. Zellner (1971, Section 10.2) credited this procedure to Lindley (1965, Section 5.6), who in turn credited it to Jeffreys (1961). Lindley suggested that it may be useful when the prior information is vague or diffuse, with the caveat that

* Tel.: +46 184713389.
E-mail address: thulin@math.uu.se

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.09.014

the distribution of θ must be reasonably smooth around θ_0 for the inversion to be sensible. Koch (2007, Section 3.4) motivated this procedure by its resemblance to frequentist methods. Ghosh et al. (2006, Section 2.7) described the inversion of credible sets as “a very informal way of testing”. Zellner remarked that “no decision theoretic justification appears available for this procedure”.

The purpose of this short note is to discuss such justifications and to put inverted credible sets on a decision-theoretic footing, making them formal Bayesian tests. These justifications shed light on when tests based on credible sets should be used. Throughout the text we assume that the parameter space $\Theta \subseteq \mathbb{R}$ and that the posterior distribution of θ is proper and absolutely continuous.

We will study three types of credible sets. A *highest posterior density* (*HPD*) set contains all θ for which the posterior density is greater than some k_α. Its appeal lies in the fact that among all $1-\alpha$ credible sets, the HPD set has the shortest length.

With $q_\alpha(\theta|x)$ denoting the posterior quantile function of θ, such that $\mathrm{P}(\theta < q_\alpha(\theta|x)) = \alpha$, the interval $(q_{\alpha/2}(\theta|x), q_{1-\alpha/2}(\theta|x))$ is called a $1-\alpha$ *central interval* (also known as an equal-tailed, symmetric or probability centered interval). Central intervals are often used due to the fact that they are computationally simpler than HPD sets, especially when the HPD set is not connected. They are also more in line with frequentist practice.

A credible set is a *credible bound* (or a one-sided credible interval) if it is of the form $\{\theta : \theta \le q_{1-\alpha}(\theta|x)\}$ or $\{\theta : \theta \ge q_\alpha(\theta|x)\}$. Such sets are of interest when a lower or upper bound for the unknown parameter θ is desired.

First, we consider tests of the point-null hypothesis $\Theta_0 = \{\theta_0\}$. We begin by comparing test based on credible sets compared to the standard solution with point-mass priors in Section 2.1, arguing that the former have many advantages. We then review a decision-theoretic justification of inverted HPD sets proposed by Madruga et al. (2001) in Section 2.2. This justification involves a loss function that is non-standard in that it depends on the sample x. In Section 2.3 we propose a justification of inverted central intervals by recasting the problem of testing Θ_0 in a three-decision setting. This allows us to use a weighted 0–1 loss to justify tests based on central intervals, avoiding losses that depend on x. In Section 2.4 we argue that tests using central intervals are preferable to tests using HPD sets. We then turn our attention to tests of composite hypotheses in Section 3 and show that inverting credible sets lead to standard tests in this setting.

2. Testing point-null hypotheses

2.1. Contrasting the standard solution and credible sets

Consider tests of the point-null hypothesis $\Theta_0 = \{\theta_0\}$ against $\Theta_1 = \{\theta : \theta \ne \theta_0\}$. We will discuss two problems associated with the standard solution to point-null hypothesis testing; the use of mixture priors and its poor frequentist properties that are overcome by using inverted credible sets instead. We then discuss the main drawback of tests based on credible sets, namely the lack of direct measures of evidence. Below, $\pi_0 > 0$ denotes the prior probability of Θ_0 in the standard solution, with $\pi_1 = 1-\pi_0$ being the prior probability of Θ_1.

1. *The use of mixture priors in the standard test*: One could argue, as Berger and Sellke (1987, Section 2), that when using the standard solution it is in general reasonable to have $\pi_0 \ge 1/2$ so that Θ_0 is not rejected because of its low prior probability. This is common in practice. It is however much more sensible to adjust the loss function if stronger evidence is required before Θ_0 is rejected. The prior distribution should reflect the prior beliefs (or lack of them) and not the severity of rejecting Θ_0. Choosing a large π_0 corresponds to using a prior with a sharp spike close to θ_0. Such priors are only reasonable if there is a very strong prior belief in the null hypothesis, and cannot be used in an analysis where at least some degree of objectivity is desired.

When lacking prior information or striving for complete objectivity, it is common practice to use a prior that in some sense is non-informative or objective. The point-null hypothesis is often a simplification of the hypothesis that $|\theta-\theta_0| < \varepsilon$ for some small ε, in which case $\pi_0 = \mathrm{P}(|\theta-\theta_0| < \varepsilon)$. If $\mathrm{P}(|\theta-\theta_0| < \varepsilon)$ is computed under an objective prior on θ, π_0 will invariably be extremely small. The “objective” prior $\pi_0 = 1/2$ is in fact heavily biased towards Θ_0.

Berger and Sellke (1987, Section 2) claim that a point-null test only makes sense to a Bayesian if the prior actually has a sharp spike near θ_0. But it is perfectly reasonable to formulate a hypothesis before constructing an informative prior or to require that an objective prior is used when no prior information is available. It is furthermore reasonable to require that the prior distribution can be used for more than one type of statistical decision, e.g. both estimation and hypothesis testing; the prior should reflect prior beliefs and not the type of decision that one is interested in.

2. *The poor frequentist performance of the standard test*: There are many situations in which it is reasonable to require that a statistical procedure works well from both a Bayesian and a frequentist perspective, particularly when an objective analysis is desirable. A well-known complication with the standard solution to testing point-null hypotheses is the asymptotic discrepancy between Bayesian and frequentist analysis that is known as Lindley's paradox (Lindley, 1957), in which, for any fixed prior, $\mathrm{P}(\theta \in \Theta_0|x)$ goes to 1 as the sample size increases for values of a test statistic that correspond to a fixed (small) p-value. Thus a frequentist and a Bayesian will reach different conclusions as more and more data is collected. This is another example of how the standard Bayesian test favours the null hypothesis.

There is no such discrepancy for tests based on credible sets. On the contrary, when based on non-informative priors, credible sets tend to have favourable properties when treated as frequentist confidence intervals; see e.g. Brown et al. (2001); and hypothesis testing using a credible set is often a valid frequentist procedure.