Journal of Statistical Planning and Inference 141 (2011) 3131–3141

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

The Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean

Brendan Kline

Northwestern University, Department of Economics, 302 Arthur Andersen Hall, 2001 Sheridan Road, Evanston, IL 60208-2600, United States

ARTICLE INFO

Article history:
Received 18 June 2010
Received in revised form
30 March 2011
Accepted 31 March 2011
Available online 8 April 2011

Keywords:
Testing
Posterior probability
p-value
One-sided hypothesis
Moment inequality conditions

ABSTRACT

This paper compares the Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean. First, this paper proposes a simple way to assign a Bayesian posterior probability to one-sided hypotheses about a multivariate mean. The approach is to use (almost) the exact posterior probability under the assumption that the data has multivariate normal distribution, under either a conjugate prior in large samples or under a vague Jeffreys prior. This is also approximately the Bayesian posterior probability of the hypothesis based on a suitably flat Dirichlet process prior over an unknown distribution generating the data. Then, the Bayesian approach and a frequentist approach to testing the one-sided hypothesis are compared, with results that show a major difference between Bayesian reasoning and frequentist reasoning. The Bayesian posterior probability can be substantially smaller than the frequentist *p*-value. A class of example is given where the Bayesian posterior probability is basically 0, while the frequentist *p*-value is basically 1. The Bayesian posterior probability in these examples seems to be more reasonable. Other drawbacks of the frequentist *p*-value as a measure of whether the one-sided hypothesis is true are also discussed.

1. Introduction

Suppose that n i.i.d. realizations $\{X_i\}_{i=1}^n$ of a d-dimensional random variable X are observed. The distribution of X is assumed to have first and second moments μ_0 and Σ_0, respectively. It is often important to test the null hypothesis that $\mu_0 \geq 0$, either from the Bayesian perspective (e.g., a posterior probability) or the frequentist perspective (e.g., a p-value). Under affine transformation of X, this describes also statements that various components of μ_0 are weakly greater than (or weakly less than) known constants. Inference on $\mu_0 \geq 0$ has been previously studied by Perlman (1969), who takes a frequentist perspective and assumes that X is known to be multivariate normal. Related literature from a frequentist perspective includes Kodde and Palm (1986), Kudô (1963), Nüesch (1966), and Wolak (1989).

The hypothesis that $\mu_0 \geq 0$ is of interest in many situations, just a few of which are described here. Consider an experimental study where there are $d+1$ "treatments," each of which result in a scalar outcome. Suppose that a total of n experimental subjects experience each of the treatments and the resulting outcomes $\{Y_{i,t}\}_{i=1,t=1}^{n,d+1}$ are observed.[1] Define $X_{it} = Y_{i1} - Y_{i,t+1}$, the difference in outcome between treatments 1 and $t+1$. The experimental question may be whether

E-mail address: b-kline@northwestern.edu

[1] It may be reasonable also to use this framework when fewer treatments, or even only one treatment, can be logically assigned to each subject, under additional assumptions about the assignment of treatments to subjects and assumptions like that different subjects, each of which experience one treatment, may be considered to be the "same" subject.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2011.03.034

treatment 1 dominates all of the remaining treatments on average, corresponding to the statement that $\mu_0 \geq 0$ with respect to X.

Alternatively, the experimental question may concern a treatment with many scalar outcomes. It could be the case that a treatment is intended to improve one outcome, but with possible side effects on other outcomes. An experimental question may be whether the treatment has non-negative average effect on all of these side effect outcomes.

Another application concerns moment inequality conditions, an important generalization of moment equality conditions. Moment inequality conditions are often found to be a convenient way to characterize a model in which the parameter of interest is only partially identified; see, for example, Manski (2007). A moment inequality condition specifies that a parameter θ satisfies the moment inequality conditions $E(m(Z,\theta)) \geq 0$, where m is a possibly vector-valued function of the observed data Z and the parameter θ. The statistical question concerns determining which parameter values satisfy this condition. The issue of partial identification arises because typically many different values of θ satisfy this condition. If $X = m(Z,\theta)$ then the statement that $\mu_0 \geq 0$ is a statement that θ satisfies the moment inequality conditions. For the moment inequality condition problem in particular there is a large recent literature in econometrics including Andrews and Guggenberger (2009), Andrews and Jia (unpublished), Andrews and Soares (2010), Beresteanu and Molinari (2008), Bugni (2010), Canay (2010), Chernozhukov et al. (2007), Imbens and Manski (2004), Pakes et al. (2006), Romano and Shaikh (2008, 2010), Rosen (2008), and Stoye (2009). None of these papers address the problem from a Bayesian perspective, and some use the extra parametric structure of the moment inequality conditions. Liao and Jiang (2010) address the moment inequality condition problem from a Bayesian perspective, based on a limited information likelihood. The focus of their paper is on the parameter θ, as their procedure admits a prior over the parameter θ that is not admitted in the approach considered here, and results in a posterior distribution for θ. The difference is because the focus of this paper is on the general problem of testing $\mu_0 \geq 0$, including comparing the Bayesian approach to a frequentist approach, and not specifically on the extra parametric structure of the moment inequality conditions. Thus, Liao and Jiang (2010) and this paper provide complementary results to the literature on moment inequality conditions. Another related paper is Moon and Schorfheide (2009) who investigate the relationship between Bayesian and frequentist inference in partially identified models, of which moment inequality conditions are examples. Again, the focus is on the parameter θ. In their model they find a result similar to the result of this paper, that a frequentist confidence set for θ may be larger than the Bayesian credible set for θ. The results of this paper, applied to moment inequality conditions, complement and help explain this result.

The case $d=1$ has been considered previously, including comparing the Bayesian approach with a frequentist approach, for example by Casella and Berger (1987). This paper complements Casella and Berger (1987) by considering similar questions in the case of more than one dimension. Casella and Berger (1987) are able to conclude, in contrast to the case of a point null hypothesis as in Lindley (1957), Berger and Delampady (1987), or Berger and Sellke (1987), that the Bayesian posterior probability can equal the frequentist p-value when considering a one-sided null hypothesis about a one-dimensional parameter, based on an improper flat prior. This paper comes to an opposite conclusion about the case $d > 1$.[2] In the case that $d > 1$, the Bayesian posterior probability of $\mu_0 \geq 0$ can be substantially smaller than the frequentist p-value for testing $\mu_0 \geq 0$.

This paper proposes a Bayesian approach to the problem of testing $\mu_0 \geq 0$ by providing a posterior probability of $\mu_0 \geq 0$, and justifies the approach on Bayesian grounds in two different ways. The asymptotic properties of the Bayesian posterior probability are derived, and a frequentist procedure is also described. Next, the Bayesian approach and the frequentist approach are compared. It is shown that for some data the Bayesian posterior probability and frequentist p-value are very different, especially when the dimension d is large. It is argued that the Bayesian posterior probability seems more reasonable than the frequentist p-value. Also, comparing two different hypotheses, the frequentist p-value may be higher for hypotheses that are logically stronger. The Bayesian posterior probability does not exhibit this incoherent behavior. Moreover, comparing two different possible datasets, the frequentist p-value may be lower after observing data that is apparently more supportive of $\mu_0 \geq 0$. The Bayesian posterior probability exhibits a monotonicity. It is argued by proving an impossibility theorem that at least some of these properties are necessarily generic for any frequentist approach. All proofs are in the appendix.

Part of this paper presumes that it is interesting to compare the Bayesian posterior probability and the frequentist p-value.[3] Consequently this paper is in the tradition of a variety of results that do exactly this. Papers that compare the Bayesian posterior probability (or other Bayesian measure of evidence) and frequentist p-value include some of those already discussed, including Lindley (1957), Berger and Delampady (1987), Berger and Sellke (1987), and Casella and

[2] The Casella and Berger (1987) sense of the Bayesian posterior probability equaling the frequentist p-value involves studying the infimum of the Bayesian posterior probability over a class of reasonable priors. They focus on data that is ''supportive'' of the alternative hypothesis. For some classes of priors, after observing data supportive of the alternative hypothesis the infimum of the Bayesian posterior probability equals the frequentist p-value, while for other classes the infimum may be smaller than the frequentist p-value. The prior achieving the infimum of the posterior probabilities, when it equals the frequentist p-value, is the ''uninformative'' improper flat prior on the real line. Similarly, the approach taken in this paper is to study the Bayesian posterior probability under an ''uninformative'' prior, or when the data dominates the prior. In contrast to the Casella and Berger (1987) results finding equality, the posterior probability proposed here can be substantially smaller than the frequentist p-value when d is sufficiently large, a difference which could only be increased by taking an infimum over a larger class of priors.

[3] I thank a referee for comments that led to a revision and improvement of this discussion.