Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 4012–4025
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Assessing post-data weight of evidence☆

C. Andy Tsao*

Department of Applied Mathematics, National Dong Hwa University, Hualien 97401, Taiwan

Received 11 June 2002; received in revised form 10 March 2005; accepted 10 March 2005
Available online 3 May 2005

Abstract

Under a robust Bayesian framework, we study p-values as post-data weights of evidence in hypothesis testing problems. For one-sided and two-sided hypothesis testing problems for the normal mean, p-values are considered as maximum likelihood estimates for some functions of the mean. We show that in contrast to Bayes estimates for reasonable families of priors, p-values are extreme for one-sided hypothesis testing problems and are moderate for two-sided problems. Implications on the controversies of ir/reconcilability of p-values in Casella and Berger [1987, J. Am. Statist. Assoc. 82, 106–111], Berger and Sellke [1987, J. Am. Statist. Assoc. 82, 112–122] and Berger and Delampady [1987, Statist. Sci. 2, 317–352] will also be addressed.

MSC: Primary 62F03; 62F35; Secondary 62F15; 62C10

Keywords: p-Values; Bayes estimates; Conjugate priors; Robust Bayesian analysis; Bayes-classical conflict

1. Introduction

p-Values have long been considered as post-data evidences against null hypothesis in hypothesis testing problems. The interpretation "the smaller the p-value, the more significant evidence against null hypothesis" is taught in many standard statistics textbooks and widely used in applications. However, the rigorous interpretation of p-values remains controversial. One important line of research was started with Berger and Sellke (1987)

☆ The research is supported by NSC 90-2118-M-259-005.

* Tel.: +886 3863 3520; fax: +886 386 33510.
E-mail address: chtsao@mail.ndhu.edu.tw.

0378-3758/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2005.03.002

and followed by, for example, Berger and Delampady (1987), Casella and Berger (1987), Hwang et al. (1992) and more recently, Oh and DasGupta (1999). Using the estimated accuracy approach, these results suggest that p-values are too liberal for two-sided problems; while p-values are reconcilable but usually remain the most liberal ones compared to their Bayesian contenders for one-sided problems.

Along this line, we investigate the post-data properties of p-values as an estimation problem under the robust Bayesian framework. However, our formulation is slightly different from the estimated accuracy approach. Precisely, let $X_1, \ldots, X_n \sim \mathrm{N}(\theta, \sigma^2)$ with unknown $\theta \in \Theta = \mathscr{R}$ and known $\sigma^2 > 0$. Consider the hypothesis testing problems:

$$\mathrm{H}_0 : \theta \leqslant 0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta > 0, \tag{1}$$

$$\mathrm{H}_0 : \theta = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \neq 0. \tag{2}$$

It is well-known that uniformly most powerful (UMP)/uniformly most powerful unbiased (UMPU) α level test exists for (1) and (2), respectively. The corresponded p-values with respect to (1) and (2) are

$$p_1(x) = P\left(Z > \frac{\hat{\theta}}{\sigma_n}\right) = \Phi\left(\frac{-\hat{\theta}}{\sigma_n}\right), \tag{3}$$

$$p_2(x) = P\left(|Z| > \frac{|\hat{\theta}|}{\sigma_n}\right) = 2\Phi\left(\frac{-|\hat{\theta}|}{\sigma_n}\right), \tag{4}$$

where $\hat{\theta} = \bar{x} = \sum_i x_i/n$, $\sigma_n = \sigma/\sqrt{n}$ and Z is a standard normal random variable with cdf $\Phi(x) = P[Z \leqslant x]$ and $\Phi(z_\alpha) = 1 - \alpha$. Because of the invariance property, p_1 and p_2 can be viewed, respectively, as maximum likelihood estimates of

$$\beta_1(\theta) = \Phi\left(\frac{-\theta}{\sigma_n}\right), \tag{5}$$

$$\beta_2(\theta) = 2\Phi\left(\frac{-|\theta|}{\sigma_n}\right). \tag{6}$$

Note that $\beta_1(\theta)$, $\beta_2(\theta)$ can be considered as smooth versions of $1_{[\theta<0]}$ and $1_{[\theta=0]}$, respectively. The smaller n renders a smoother function. Our investigation is thus an alternative formulation to the estimated accuracy approach used in, for example, Casella and Berger (1987), Berger and Sellke (1987) and Hwang et al. (1992). We will comment on the difference of these two formulations in our conclusion.

In this paper, we study the problem of estimating $\beta_1(\theta)$ and $\beta_2(\theta)$ from a robust Bayesian viewpoint. p-Values are compared with Bayes estimates with respect to some reasonable families of priors. It is found that for the family of continuous, bounded and symmetric priors, p-values are extreme for one-sided hypothesis testing problems when the observed sample mean is positive. For two-sided problem, p-values are moderate for two-sided problems in contrast to Bayes estimates with respect to family of scale mixtures of normal priors with mean zero.