Working Paper 01-27
Statistics and Econometrics Series 17
May 2001

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

BAYESIAN ROBUSTNESS OF THE POSTERIOR PREDICTIVE P-VALUE.

J. de la Horra Navarro and M.T. Rodríguez Bernal*

Abstract

In this paper, the Bayesian robustness of the posterior predictive p-value is studied. First of all, it is proved that Lavine´s linearization technique can be extended for analyzing this problem. Then, the result is applied to the ε-contamination class of prior distributions.

Keywords: Bayesian robustness, ε-contamination class, linearization technique, posterior odds, posterior predictive p-value.

*de la Horra, Departamento de Matemáticas, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Tfno: 91-3964945, e-mail: julian.delahorra@uam.es ; Rodríguez, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain, Tfno: 91-6245853, e-mail: mrodrigu@est-econ.uc3m.es.

1 Introduction

In the last years, several papers have been published analyzing possible alternatives to the classical p-value. The posterior predictive p-value is a Bayesian-motivated alternative. The concept was first introduced by Guttman (1967) and Rubin (1984), who used the posterior predictive distribution of a test statistic to calculate the tail-area probability corresponding to the observed value of the statistic. Such a tail-area probability is called posterior predictive p-value by Meng (1994) (who extended the concept by using discrepancy variables), whereas the tail-area probability used by Box (1980) can be called prior predictive p-value. Different aspects of the posterior predictive p- value have been further studied by Gelman et al. (1996) and De la Horra and Rodríguez-Bernal (1997, 1999, 2000, 2001a, 2001b). Recently, Bayarri and Berger (1999, 2000) have introduced the conditional predictive p-value and the partial posterior predictive p-value.

The concept of posterior predictive p-value, $p(x, \Theta_0)$, for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \notin \Theta_0$ is briefly introduced in Section 2.

The posterior predictive p-value depends on the prior distribution. It is important to know whether this p-value is seriously influenced by small modifications on the prior distribution. This is a typical study of Bayesian robustness. If the prior distribution ranges over a class Γ of prior distributions, we have to compute the difference between $\sup_{\pi \in \Gamma} p(x, \Theta_0)$ and $\inf_{\pi \in \Gamma} p(x, \Theta_0)$. In Section 3, it is proved that Lavine's linearization technique can be extended for analyzing this problem.

In Section 4, the result is applied to the particular case of an ε-contamination class of prior distributions, and an example is completely developed.

2 Posterior predictive p-value

Let x be a random sample from the random variable X taking values in $\mathcal{X}$ and having density function $f(x|\theta)$, where $\theta \in \Theta$. We want to test $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$.

The posterior predictive p-value is a Bayesian-motivated alternative to the classical p-value. Let $\pi(\theta)$ be the prior density summarizing the prior information about θ, and let $D(x, \theta)$ be a discrepancy variable, where a discrepancy variable is a function $D : \mathcal{X} \times \Theta_0 \to \Re^+$ measuring (in some reasonable way) the "discrepancy" between the observation x and the parameter θ. The concept of discrepancy variable $D(x, \theta)$ was introduced by Tsui and Weerahandi (1989) and is nothing but a generalization of a test statistic $D(x)$.

We next give the definition of posterior predictive p-value, such as it was introduced by Meng (1994):

Definition

The posterior predictive p-value for testing $H_0 : \theta \in \Theta_0$ *versus* $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$, *when the discrepancy variable* $D(x, \theta)$ *is used, is defined as*

$$\begin{aligned} p(x, \Theta_0) &= Pr\{(y, \theta) \in \mathcal{X} \times \Theta_0 : D(y, \theta) \geq D(x, \theta)|x, \Theta_0\} \\ &= \int_A f(y, \theta|x, \Theta_0) dy d\theta, \end{aligned}$$

where $A = \{(y, \theta) \in \mathcal{X} \times \Theta_0 : D(y, \theta) \geq D(x, \theta)\}$.

We may express $p(x, \Theta_0)$ in an alternative way, by means of the well-known classical p-value for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$:

$$p(x, \theta_0) = Pr\{y \in \mathcal{X} : D(y, \theta_0) \geq D(x, \theta_0)|\theta_0\} = \int_{A_{\theta_0}} f(y|\theta_0) dy,$$

where $A_{\theta_0} = \{y \in \mathcal{X} : D(y, \theta_0) \geq D(x, \theta_0)\}$. We have:

$$\begin{aligned} p(x, \Theta_0) &= \int_A f(y, \theta|x, \Theta_0) dy d\theta \\ &= \int_{\Theta_0} \left[\int_{A_\theta} f(y|\theta) dy \right] d\pi(\theta|x, \Theta_0) \\ &= \int_{\Theta_0} p(x, \theta) d\pi(\theta|x, \Theta_0) \qquad (1) \end{aligned}$$

This expression will be used for analyzing Bayesian robustness in the next section.

3 Bayesian robustness

The posterior predictive p-value depends on the prior distribution π we are taking over the parameter space Θ. It is important to know whether this p-value is seriously affected or not by small variations on the prior distribution. This is a typical study of Bayesian robustness. If the prior distribution ranges over a class Γ of prior distributions, we have to compute the difference between $\sup_{\pi \in \Gamma} p(x, \Theta_0)$ and $\inf_{\pi \in \Gamma} p(x, \Theta_0)$.

Lavine (1991) and Lavine et al.(1991) developed the linearization technique for computing, in an easy way,

$$\rho^* = \sup_{\pi \in \Gamma} E_\pi[\phi(\theta)|x] \qquad \text{and} \qquad \rho_* = \inf_{\pi \in \Gamma} E_\pi[\phi(\theta)|x],$$

where $E_\pi[\phi(\theta)|x] = \int_\Theta \phi(\theta) d\pi(\theta|x)$, being $\phi(\theta)$ an amount of interest.

By means of suitable choices of the function $\phi(\theta)$, $E_\pi[\phi(\theta)|x]$ becomes the "posterior mean", the "posterior probability of a subset B of the parameter space", ... Therefore, first of all, we will study whether the posterior predictive p-value is also a particular case. We have, by equation (1):

$$\begin{aligned} p(x,\Theta_0) &= \int_{\Theta_0} p(x,\theta) d\pi(\theta|x,\Theta_0) \\ &= \int_{\Theta} \left[\frac{I_{\Theta_0}(\theta) p(x,\theta)}{Pr(\Theta_0|x)} \right] d\pi(\theta|x) \\ &= E_\pi \left[\frac{I_{\Theta_0}(\theta) p(x,\theta)}{Pr(\Theta_0|x)} \; |x \right], \end{aligned} \tag{2}$$

where $I_{\Theta_0}(\theta) = 1$ if $\theta \in \Theta_0$ (zero, otherwise).

The null hypothesis, Θ_0, and data, x, remain fixed through all the robustness study, but the prior distribution,π, is not fixed, and it is used for computing $Pr(\Theta_0|x)$. Therefore, we actually have to compute

$$\rho^* = \sup_{\pi\in\Gamma} p(x,\Theta_0) = \sup_{\pi\in\Gamma} E_\pi[\phi(\theta,\pi)|x]$$

and

$$\rho_* = \inf_{\pi\in\Gamma} p(x,\Theta_0) = \inf_{\pi\in\Gamma} E_\pi[\phi(\theta,\pi)|x],$$

where $\phi(\theta,\pi) = I_{\Theta_0}(\theta) p(x,\theta)/Pr(\Theta_0|x)$.

So, we are trying to solve a more general problem. We next prove that it is possible to extend Lavine's linearization technique for analyzing this case.

Lemma

For any $q \in (0,1)$, we have:

a) $\rho^* = \sup_{\pi\in\Gamma} p(x,\Theta_0) \le q$ if and only if

$$\sup_{\pi\in\Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta)-q) d\pi(\theta) \le 0.$$

b) $\rho_* = \inf_{\pi\in\Gamma} p(x,\Theta_0) \ge q$ if and only if

$$\inf_{\pi\in\Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta)-q) d\pi(\theta) \ge 0.$$

Proof.- a) We have:

$$\begin{aligned} \rho^* &= \sup_{\pi\in\Gamma} p(x,\Theta_0) \le q \\ &\Leftrightarrow \text{ For all } \pi \in \Gamma : \int_{\Theta} \left[\frac{I_{\Theta_0}(\theta) p(x,\theta)}{Pr(\Theta_0|x)} \right] d\pi(\theta|x) \le q \qquad \text{by (2)} \\ &\Leftrightarrow \text{ For all } \pi \in \Gamma : \frac{\int_{\Theta_0} p(x,\theta) d\pi(\theta|x)}{\int_{\Theta_0} d\pi(\theta|x)} \le q \end{aligned}$$

$$\Leftrightarrow \quad \text{For all } \pi \in \Gamma : \frac{\int_{\Theta_0} p(x,\theta)\frac{f(x|\theta)}{m(x)} d\pi(\theta)}{\int_{\Theta_0} \frac{f(x|\theta)}{m(x)} d\pi(\theta)} \leq q$$

$$\Leftrightarrow \quad \text{For all } \pi \in \Gamma : \int_{\Theta_0} p(x,\theta) f(x|\theta) d\pi(\theta) \leq \int_{\Theta_0} q f(x|\theta) d\pi(\theta)$$

$$\Leftrightarrow \quad \sup_{\pi \in \Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta) - q) d\pi(\theta) \leq 0,$$

being $m(x) = \int_{\Theta} f(x|\theta) d\pi(\theta)$.

b) The proof is analogous. •

Theorem

$$a) \qquad \rho^* = \sup_{\pi \in \Gamma} p(x, \Theta_0) = \inf \left\{ q \in (0,1) : \sup_{\pi \in \Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta) - q) d\pi(\theta) \leq 0 \right\}.$$

$$b) \qquad \rho_* = \inf_{\pi \in \Gamma} p(x, \Theta_0) = \sup \left\{ q \in (0,1) : \inf_{\pi \in \Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta) - q) d\pi(\theta) \geq 0 \right\}.$$

Proof.- We can now apply results in Lavine et al. (1991). •

Therefore, in practice, ρ^* will be the value of q satisfying the equation

$$\sup_{\pi \in \Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta) - q) d\pi(\theta) = 0, \tag{3}$$

and ρ_* will be the value of q satisfying the equation

$$\inf_{\pi \in \Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta) - q) d\pi(\theta) = 0. \tag{4}$$

4 ε-contamination class

In this section, we will consider the ε-contamination class of prior distributions $\Gamma = \{(1-\varepsilon)\pi_0 + \varepsilon\pi\}$, where π_0 is the basic prior, $\varepsilon \in (0,1)$ is the degree of contamination, and π is any distribution.

The ε-contamination class has been very used in the analysis of Bayesian robustness; see, for instance, Huber (1973), Berger and Berliner (1986), Sivaganesan (1988), Wasserman (1989) and De la Horra and Fernández (1994).

This class is easy to analyze and is a large class, because any distribution π is allowed to contaminate the basic prior π_0. When the result of the robustness analysis with this class is that $\rho^* - \rho_*$ is small, we can conclude that the amount we are analyzing shows a robust behaviour; otherwise, we must analyze narrower classes of priors.

ρ^* will be the value of q satisfying equation (3):

$$\begin{aligned} 0 &= \sup_{\pi \in \Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta) - q)[(1-\varepsilon)d\pi_0(\theta) + \varepsilon d\pi(\theta)] \\ &= (1-\varepsilon)\int_{\Theta_0} f(x|\theta)(p(x,\theta) - q)d\pi_0(\theta) \\ &+ \varepsilon \sup_{\pi \in \Gamma} \int_{\Theta_0} f(x|\theta)(p(x,\theta) - q)d\pi(\theta) \\ &= (1-\varepsilon)\int_{\Theta_0} f(x|\theta)(p(x,\theta) - q)d\pi_0(\theta) \\ &+ \varepsilon \sup_{\theta \in \Theta_0} f(x|\theta)(p(x,\theta) - q). \end{aligned}$$

Therefore, ρ^* will the value of q satisfying the equation:

$$(1-\varepsilon)\int_{\Theta_0} f(x|\theta)(p(x,\theta) - q)d\pi_0(\theta) + \varepsilon \sup_{\theta \in \Theta_0} f(x|\theta)(p(x,\theta) - q) = 0.$$

Analogously, from (4), ρ_* will the value of q satisfying the equation:

$$(1-\varepsilon)\int_{\Theta_0} f(x|\theta)(p(x,\theta) - q)d\pi_0(\theta) + \varepsilon \inf_{\theta \in \Theta_0} f(x|\theta)(p(x,\theta) - q) = 0.$$

We next give an algorithm for computing (in an approximate way) ρ^* and ρ_*:

1. Choose values $\theta_i \in \Theta_0$ $(i = 1, ..., m)$ and values $q_j \in (0,1)$ $(j = 1, ..., n)$, with m and n large enough for providing a dense grid for $\Theta_0 \times (0,1)$.

2. For each θ_i, compute $p(x, \theta_i)$.

3. For each θ_i and q_j, compute $g(\theta_i, q_j) = f(x|\theta_i)(p(x,\theta_i) - q_j)$.

4. For each q_j, compute $\sup_i g(\theta_i, q_j)$ and $\inf_i g(\theta_i, q_j)$.

5. For each q_j, compute $\int_{\Theta_0} f(x|\theta)(p(x,\theta) - q_j)d\pi_0(\theta)$. For doing that (in an approximate way), simulate values of θ from the basic prior π_0, and approximate the integral by means of the sample mean.

6. ρ^* will be the value of q_j satisfying (approximately) the equation:

$$(1-\varepsilon)\int_{\Theta_0} f(x|\theta)(p(x,\theta) - q_j)d\pi_0(\theta) + \varepsilon \sup_i f(x|\theta_i)(p(x,\theta_i) - q_j) = 0.$$

7. ρ_* will be the value of q_j satisfying (approximately) the equation:

$$(1-\varepsilon)\int_{\Theta_0} f(x|\theta)(p(x,\theta)-q_j)d\pi_0(\theta)+\varepsilon\inf_i f(x|\theta_i)(p(x,\theta_i)-q_j)=0.$$

Example

Let x be an observation from the density $f(x|\theta) \sim N(\theta;\sigma^2=1)$. We will compute ρ^* and ρ_* for an ε-contamination class where the basic prior is $\pi_0 \sim N(0;1)$.

First of all, we have to choose the discrepancy variable. The discrepancy variable (in fact, a test statistic)

$$D^*(x)=\frac{Pr(\Theta_1|x)}{Pr(\Theta_0|x)},$$

where the posterior probabilities are obtained from the prior $\pi_0 \sim N(0;1)$, has good properties; see De la Horra and Rodríguez-Bernal (2000) for asymptotic optimality properties and De la Horra and Rodríguez-Bernal (2001b) for coherence properties. Moreover, the computation of $p(x,\theta_i)$ (second step of the algorithm) will be very easy with this test statistic.

We will analyze two cases:

a) We take $\Theta_0=[0,\infty)$ as null hypothesis. In this case, it was proved in De la Horra and Rodríguez-Bernal (2001b) that $p(x,\theta)$ is a monotonically decreasing function in θ, of the form:

$$\begin{aligned} p(x,\theta) &= Pr\{y\in\mathcal{X}: D^*(y)\geq D^*(x)|\theta\} \\ &= Pr\{y\in\mathcal{X}: y\leq x|\theta\} \end{aligned}$$

Table I shows ρ^* and ρ_* for different values of x and different degrees of contamination.

TABLE I

	x=-2		x=-1		x=0	
	ρ_*	ρ^*	ρ_*	ρ^*	ρ_*	ρ^*
$\epsilon=0.01$	0.01	0.01	0.09	0.09	0.30	0.30
$\epsilon=0.05$	0.01	0.01	0.09	0.10	0.29	0.31
$\epsilon=0.10$	0.01	0.02	0.09	0.11	0.29	0.33
$\epsilon=0.15$	0.01	0.02	0.09	0.11	0.28	0.34
$\epsilon=0.20$	0.01	0.02	0.08	0.12	0.27	0.35

	x =1		x =2	
	ρ_*	ρ^*	ρ_*	ρ^*
$\epsilon = 0.01$	0.56	0.57	0.76	0.76
$\epsilon = 0.05$	0.55	0.57	0.73	0.77
$\epsilon = 0.10$	0.53	0.58	0.69	0.78
$\epsilon = 0.15$	0.52	0.60	0.66	0.78
$\epsilon = 0.20$	0.50	0.61	0.62	0.79

In general, differences between ρ^* and ρ_* are very small. Moreover, the biggest differences are obtained for large posterior predictive p-values and, in these cases, there is no influence on the final decision (to accept H_0). So, we can conclude that, in this case, the posterior predictive p-value shows a robust behaviour.

b) We take $\Theta_0 = [-1, 1]$ as null hypothesis. In this case, it was proved in De la Horra and Rodríguez-Bernal (2001b) that $p(x, \theta)$ is a monotonically increasing function in $|\theta|$, of the form:

$$\begin{aligned} p(x,\theta) &= Pr\{y \in \mathcal{X} : D^*(y) \geq D^*(x)|\theta\} \\ &= Pr\{y \in \mathcal{X} : |y| \geq |x| \ |\theta\} \end{aligned}$$

Table II shows ρ^* and ρ_* for different values of x and different degrees of contamination.

TABLE II

	x=0		x=0.5		x=1	
	ρ_*	ρ^*	ρ_*	ρ^*	ρ_*	ρ^*
$\epsilon = 0.01$	1.00	1.00	0.66	0.66	0.37	0.38
$\epsilon = 0.05$	1.00	1.00	0.66	0.66	0.37	0.39
$\epsilon = 0.10$	1.00	1.00	0.65	0.67	0.37	0.40
$\epsilon = 0.15$	1.00	1.00	0.65	0.68	0.37	0.41
$\epsilon = 0.20$	1.00	1.00	0.65	0.68	0.36	0.42

	x =1.5		x =2		x=2.5	
	ρ_*	ρ^*	ρ_*	ρ^*	ρ_*	ρ^*
$\epsilon = 0.01$	0.20	0.20	0.09	0.09	0.03	0.04
$\epsilon = 0.05$	0.19	0.21	0.09	0.10	0.03	0.04
$\epsilon = 0.10$	0.19	0.22	0.09	0.11	0.03	0.04
$\epsilon = 0.15$	0.19	0.23	0.08	0.11	0.03	0.05
$\epsilon = 0.20$	0.19	0.24	0.08	0.12	0.03	0.05

	x=3	
	ρ_*	ρ^*
$\epsilon = 0.01$	0.01	0.01
$\epsilon = 0.05$	0.01	0.01
$\epsilon = 0.10$	0.01	0.01
$\epsilon = 0.15$	0.01	0.02
$\epsilon = 0.20$	0.01	0.02

In general, differences between ρ^* and ρ_* are very small. So, we can conclude that, also in this case, the posterior predictive p-value shows a robust behaviour.

ACKNOWLEDGEMENTS

This research has been partially supported by the DGESIC (Spain) under grant number PB97-0021.

BIBLIOGRAPHY

Bayarri, M.J. and Berger, J.O, (1999). "Quantifying surprise in the data and model verification," *Bayesian Statistics 6* (Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. Eds.), Oxford University Press, Oxford, 53-82 (with discussion).

Bayarri, M.J. and Berger, J.O, (2000). "P-values for composite null models," *J. Amer. Statist. Assoc.* 95, 1127-1142.

Berger, J.O. and Berliner, L.M. (1986). "Robust Bayesian and empirical Bayes analysis with ε-contaminated priors," *Ann. Statist.* 14, 461-486.

Box, G. E. P. (1980). "Sampling and Bayes' inference in scientific modelling and robustness," *J. Roy. Statist. Soc. A* 143, 383-430.

De la Horra, J. and Fernández, C. (1994). "Bayesian analysis under ε-contaminated priors: A trade-off between robustness and precision," *J. Statist. Plann. Inf.* 38, 13-30.

De la Horra, J. and Rodríguez-Bernal, M. T. (1997). "Asymptotic behaviour of the posterior predictive p-value," *Commun. Statist.-Theory Meth.* 26, 2689-2699.

De la Horra, J. and Rodríguez-Bernal, M. T. (1999). "The posterior predictive p-value for the problem of goodness of fit," *Test* 8, 117-128.

De la Horra, J. and Rodríguez-Bernal, M.T. (2000). "Optimality of the posterior predictive p-value based on the posterior odds," *Commun. Statis.-Theory Meth.* 29, 181-192.

De la Horra, J. and Rodríguez-Bernal, M. T. (2001a). "Posterior predictive p-values: what they are and what they are not," *Test* (to appear).

De la Horra, J. and Rodríguez-Bernal, M. T. (2001b). "Coherence of the posterior predictive p-value based on the posterior odds," *Stat. Econom. Ser. 16 (Univ. Carlos III)*, Working paper 01-25.

Gelman, A., Meng, X. L. and Stern, H. (1996). "Posterior predictive assessment of model fitness via realized discrepancies," *Statist. Sinica.* 6, 733-807 (with discussion).

Guttman, I. (1967). "The use of the concept of a future observation in goodness-of-fit problems," *J. Roy. Statist. Soc. B* 29, 83-100.

Huber, P.J. (1973). "The use of Choquet capacities in statistics," *Bull. Internat. Statist. Inst.* 45, 181-191.

Lavine, M. (1991). "Sensitivity in Bayesian statistics: the prior and the likelihood," *J. Amer. Statist. Assoc.* 86, 396-399.

Lavine, M., Wasserman, L. and Wolpert, R.L. (1991). "Bayesian inference with specified prior marginals," *J. Amer. Statist. Assoc.* 86, 964-971.

Meng, X. L. (1994). "Posterior predictive p-values," *Ann. Statist.* 22, 1142-1160.

Rubin, D. B. (1984). "Bayesianly justifiable and relevant frequency calculations for the applied statistician," *Ann. Statist.* 12, 1151-1172.

Sivaganesan, S. (1988). "Range of posterior measures for priors with arbitrary contaminations," *Commun. Statist.-Theory. Meth.* 17, 1591-1612.

Tsui, K.W. and Weerahandi, S. (1989). "Generalized p-values in significance testing of hypothesis in the presence of nuisance parameters," *J. Amer.*

Statist. Assoc. 84, 602-607.

Wasserman, L. (1989). "A robust Bayesian interpretation of likelihood regions," *Ann. Statist.* 17, 1387-1393.