International Journal of Approximate Reasoning 50 (2009) 270–278

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Importance sampling for Bayesian sensitivity analysis

B. O'Neill

Australian National University, School of Finance and Applied Statistics, ACT, 0200, Australia

ARTICLE INFO

Article history:
Available online 6 April 2008

Keywords:
Bayesian statistics
Robustness
Sensitivity analysis
Imprecise probability
Imprecise prevision
Importance sampling

ABSTRACT

We examine the use of sensitivity analysis with a particular focus on calculating the bounds of imprecise previsions in Bayesian statistics. We explain the use of importance sampling in approximating the range of these imprecise previsions and we develop an approximation function for the imprecise posterior prevision based on generating a finite number of random variables. We develop a convergence theorem that shows that this approximation converges almost surely to the posterior prevision as we generate more and more random variables. We also develop a useful accuracy bound for the approximation for a large finite number of generated random variables. We test the efficiency of this approximation using a simple example involving the imprecise Dirichlet model.

1. Introduction

Robustness testing is an important part of Bayesian statistics, particularly in its application to real world decision problems. The kind of invariance conditions that underlie the operational Bayesian paradigm lead to very general model forms. To reach the stage of prescribing a decision in any specific problem, it is often the case that assumptions are required which may go beyond the manifestly reasonable, giving rise to legitimate debate. As with any application of mathematics this is quite natural; after all, mathematics only tells us the logical consequences of our own assumptions. In any case, we are led naturally to enquire as to the sensitivity of our analysis to those assumptions that we believe are open to debate.

2. Sensitivity analysis

This question of sensitivity should not be viewed as an afterthought. In fact, this position can be based on foundational arguments which reject the strict measurability requirements inherent in the basic quantitative coherence axioms of Bayesian statistics. Walley [17] provides such a derivation and Ríos Insua and Ruggeri [15] give a summary of the field. Berger [5] explains the situation as follows:

> There is a common perception that foundational arguments lead to subjective Bayesian analysis as the only coherent method of behaviour. ... Subjective Bayesian analysis is, indeed, the only coherent mode of behaviour, *but only if it is assumed that one can make arbitrarily fine discriminations in judgement about unknowns and utilities*. ... It is less well known that realistic foundational systems exist, based on axiomatics of behaviour which acknowledge that arbitrarily fine discrimination is impossible. ... The conclusion of these foundational systems is that a type of *robust Bayesian analysis* is the coherent behaviour. Roughly, coherent behaviour corresponds to having *classes* of models, priors, and utilities, which yield a range of possible Bayesian answers (corresponding to the answers obtained through combination of all model-prior-utility triples from the classes). If this range of answers is too large, the question of interest may not, of course, be settled, but that is only realistic... (pp. 6–7, emphasis in original)

E-mail address: ben.oneill@anu.edu.au

0888-613X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2008.03.015

3. Global sensitivity analysis

The determination of the range of possible Bayesian answers referred to above is called *global sensitivity analysis*, following Leamer [13]. Suppose that we have the standard Bayesian situation[1] where invariance conditions lead us to a family of plausible sampling measures which is indexed by a parameter $\theta \in \Theta$. We have some prior belief about this parameter with prior density p and we observe data x giving us the likelihood function L_x. The posterior density is then determined by Bayes rule:

$$p(\theta|x) \propto L_x(\theta)p(\theta).$$

Given some fixed action, we suppose that we have utility function m which is a function of the parameter θ. It follows that the posterior prevision (that is, the posterior expected utility) is given by[2]

$$R_m \equiv \mathrm{E}(m(\theta)|x) = \frac{\int m(\theta)L_x(\theta)p(\theta)\,\mathrm{d}\theta}{\int L_x(\theta)p(\theta)\,\mathrm{d}\theta}.$$

For any measurable function k of the parameter θ we define

$$B_k \equiv \int k(\theta)L_x(\theta)p(\theta)\,\mathrm{d}\theta.$$

This allows us to express the posterior prevision as $R_m = B_m/B_1$.

Under a global sensitivity analysis, we vary the utility function, likelihood function and prior over some other reasonable range to obtain a range of possible posterior previsions. To make matters simpler, we can index the plausible utility functions, likelihood functions and prior densities by some parameters υ, λ and π respectively for some reasonable ranges $\upsilon \in \Upsilon$, $\lambda \in \Lambda$ and $\pi \in \Pi$. Conditioning on these parameters we have posterior density

$$p(\theta|x,\lambda,\pi) \propto L_x(\theta|\lambda)p(\theta|\pi)$$

and posterior prevision

$$R_m(\upsilon,\lambda,\pi) \equiv \mathrm{E}(m(\theta|\upsilon)|x,\lambda,\pi) = \frac{\int m(\theta|\upsilon)L_x(\theta|\lambda)p(\theta|\pi)\,\mathrm{d}\theta}{\int L_x(\theta|\lambda)p(\theta|\pi)\,\mathrm{d}\theta}.$$

For any measurable function k of the parameter θ we define

$$B_k(\upsilon,\lambda,\pi) \equiv \int k(\theta|\upsilon)L_x(\theta|\lambda)p(\theta|\pi)\,\mathrm{d}\theta$$

so that $R_m(\upsilon,\lambda,\pi) = B_m(\upsilon,\lambda,\pi)/B_1(\upsilon,\lambda,\pi)$.

Indexation by the parameters υ, λ and π will generally be chosen so that the utility function, likelihood and prior densities are analytical functions of these respective parameters. That is, these indexes will be 'parameters' of the various functions. In any case, we obtain the range of possible posterior previsions:

$$\Omega_m(\Upsilon,\Lambda,\Pi) \equiv \left\{\frac{B_m(\upsilon,\lambda,\pi)}{B_1(\upsilon,\lambda,\pi)} : \upsilon \in \Upsilon, \lambda \in \Lambda, \pi \in \Pi\right\}.$$

In this case we can be satisfied that $R_m \in \Omega_m(\Upsilon,\Lambda,\Pi)$ even though we are not confident of specifying our utility, likelihood or prior with any greater certainty.

In robust Bayesian analysis we refer to such a judgement as an *imprecise prevision*. It is useful to note that by replacing the utility function with an indicator function for an event we obtain the probability of that event. In this case, we refer to our judgement as an *imprecise probability*.

Walley [17] shows that, under basic foundational assumptions of consistency, these imprecise beliefs form a convex hull, so that if we consider two specifications of such a value to be reasonable then we must consider any point between them to also be reasonable. This means that the range of the imprecise prevision $\Omega_m(\Upsilon,\Lambda,\Pi)$ will be a single connected interval having lower bound (called the *lower prevision*):

$$\underline{R}_m = \inf \Omega_m(\Upsilon,\Lambda,\Pi) = \inf_{\upsilon\in\Upsilon} \inf_{\lambda\in\Lambda} \inf_{\pi\in\Pi} \frac{B_m(\upsilon,\lambda,\pi)}{B_1(\upsilon,\lambda,\pi)}$$

and upper bound (called the *upper prevision*):

$$\overline{R}_m = \sup \Omega_m(\Upsilon,\Lambda,\Pi) = \sup_{\upsilon\in\Upsilon} \sup_{\lambda\in\Lambda} \sup_{\pi\in\Pi} \frac{B_m(\upsilon,\lambda,\pi)}{B_1(\upsilon,\lambda,\pi)}.$$

[1] Set out in Chapter 4 of Bernardo and Smith [3].

[2] For convenience we will follow the notation in Evans and Swartz [9].