Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 128 (2005) 475–487

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Accuracy of posterior approximations via χ^2 and harmonic divergences

Gustavo L. Gilardoni*

Departamento de Estatística, Universidade de Brasília, Brasília DF 70910–900, Brazil

Received 12 July 2002; accepted 26 November 2003

Abstract

We propose to use χ^2 and Harmonic divergences as global measures of accuracy of an approximation $\hat{\pi}$ to a posterior density of interest π. We prove some inequalities which relate these measures to the precision of the corresponding approximations for posterior expectations. In practice these divergences ought to be approximated somehow, and here we propose importance sampling type estimates based on a sample from $\hat{\pi}$.

Unlike the more familiar precision estimates based on Central Limit type theorems for Monte Carlo based $\hat{\pi}$, our proposal (i) can be applied to approximations obtained from virtually every method available; (ii) requires to compute only one measure of accuracy which can then be reused to assess precision of the approximations for many posterior expectations and (iii) since its rationale is external to the method used to obtain $\hat{\pi}$, it avoids the danger of circular reasoning present for instance in Markov chain Monte Carlo algorithms, whereby both the validity of the approximation and of its estimated precision depend on convergence of the simulated chain, which in practice may be difficult to assess.

MSC: 62F15; 65C40; 94A17

Keywords: Approximate Bayesian inference; Chi-squared and harmonic divergences; Hierarchical models; Importance sampling; Markov chain Monte Carlo

1. Introduction

Let π be the posterior density of unobservables θ, and $\hat{\pi}$ an approximation to π. To a large extent, how $\hat{\pi}$ was obtained is irrelevant for the following exposition,

* Correspondence address. Dipartimento di Matematica, Instituto Guido Castelnuovo, Universitad egli Studi di Roma “La Sapienza”, Piazzale Aldo Moro 2, Rome 00185, Italy.

E-mail address: gilardon@unb.br (G.L. Gilardoni).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.11.013

and in fact for the examples in Section 4 we consider approximations based both on asymptotic grounds, such as the familiar Normal approximation, and Monte Carlo simulation. The objective of this paper is to discuss the use of the two divergences $\chi^2 = \chi^2(\hat{\pi}, \pi) = \int \hat{\pi}^2/\pi - 1$ and $H^2 = H^2(\hat{\pi}, \pi) = H^2(\pi, \hat{\pi}) = 1 - 2\int \pi\hat{\pi}/(\pi + \hat{\pi})$ to assess the accuracy of the approximation.

Of course, any distance or divergence between $\hat{\pi}$ and π can conceivable be used for this purpose, but χ^2 and H^2 can be related to the precision of corresponding approximations of posterior expectations $E_\pi g = \int g\pi$, so that knowing χ^2 or H^2 will almost automatically produce a bound for the error $|E_{\hat{\pi}} g - E_\pi g|$. Indeed, we show in Section 2 that

$$|E_{\hat{\pi}} g - E_\pi g|^2 \leqslant [\mathrm{Var}_\pi g]\chi^2 \tag{1}$$

and

$$|E_{\hat{\pi}} g - E_\pi g|^2 \leqslant \frac{\mathrm{Var}_{\hat{\pi}} g + \mathrm{Var}_\pi g}{2} \, \frac{4H^2}{1 - H^2}, \tag{2}$$

where for (1) we assume that the support of $\hat{\pi}$ is contained in the support of π. Inequality (1) was considered before by Kass et al. (1989) and formally proved by Weiss (1996).

The terms involving variances on the right-hand side of (1) and (2) are easy to approximate, and in fact it would usually be sufficient to use $\mathrm{Var}_\pi g \approx \mathrm{Var}_{\hat{\pi}} g$. However, exact computation of χ^2 or H^2 is typically impossible in applications. Here we propose to use a simulated sample from $\hat{\pi}$ to compute importance sampling type approximations, although it is also possible to approximate the divergences simulating from π or even from a third distribution.

Assessing the accuracy of asymptotic-based approximations is quite difficult. For instance, the precision of the normal and Laplace approximations can be estimated using respectively the results in Johnson (1970) and Achcar (1992). However, they require calculating higher order derivatives which may be difficult in practice. On the contrary, one of the main advantages of Monte Carlo based approximations is that, by using versions of the central limit theorem (CLT), they allow for estimation of precision. For instance, precision of Markov chain Monte Carlo (MCMC) approximations can be estimated by studying the asymptotic behavior of the chain. In fact, for uniformly ergodic chains $\{\theta_j\}_{j \geqslant 1}$, and provided that g^2 is π-integrable, $\bar{g} = M^{-1}\sum_{j=1}^{M} g(\theta_j)$ has for large M an approximate $\mathrm{N}(E_\pi g, \delta^2/M)$ distribution, where $\delta^2 = \sigma^2(1 + 2\sum_{k=1}^{\infty} \rho_k)$ and σ^2 and ρ_k are respectively the variance and the lag-k autocorrelation of the sequence $\{g(\theta_j)\}$ (Tierney, 1994). Estimation of δ^2 is not straightforward though, as substituting for the obvious estimates of σ^2 and the ρ_k's and truncating the infinite sum leads to inconsistency. Geyer (1992), Geweke (1992) and more recently, Kosorok (1998) have proposed methods based on times series ideas. For instance, when the spectral density $S_g(\omega)$ of the process $\{g(\theta_j)\}$ is continuous at frequency zero, an estimate of $\delta/\sqrt{M}$ is given by $\sqrt{\hat{S}_g(0)/M}$, called the *numerical standard error* (NSE) by Geweke. Another popular idea is to divide the sequence $\{g(\theta_j)\}$ into batches of successive values and computing the mean of each batch. The size and number of the batches are determined so that the correlation between successive means is very low, say less than 0.05. Then δ^2 is