Estimating the integrated likelihood via posterior simulation using the harmonic mean equality: A discussion

Christian P. Robert and Nicolas Chopin
Université Paris Dauphine and CREST-INSEE, University of Bristol

June 16, 2006

Comparison
The issue of approximating marginal densities obviously remains an up-to-date concern for Bayesian Statistics, since two invited talks at this conference are centred around it, namely the present paper and the alternative proposal of Skilling (this volume). While we are not completely convinced of the advantages of nested sampling (see our discussion in this volume), we would welcome the authors’ opinion of the respective worths of both approaches. In particular, Skilling’s (this volume) perspective is completely in line with the second approach of the present paper, that is, based on a (prior or posterior) distribution of the likelihood function, since Skilling’s marginal is expressed as $\pi(y) = \mathbb{E}_\pi[L]$, while Raftery et al.’s marginal is $\pi(y) = \mathbb{E}[L^{-1}|y]$.

Potential dangers
Let us first state that we find the representation of Newton and Raftery [1994] quite interesting in that it allows for an approximation of the marginal density based on the output of the MCMC simulation of the posterior $\pi(\theta|y)$ (rather than from the prior as in nested sampling). Its major drawback is however the disastrous feature of a potential infinite variance, against which the Rao-Blackwellised solution proposed in this version does not always work. We can take for instance the case of the normal variance, $y|\sigma \sim \mathcal{N}(0, \sigma^2)$ and $\pi(\sigma^2|y) \sim \mathcal{I}\mathcal{G}(3/2, 1 + y^2/2)$ where nested sampling provides an approximation of $\pi(y)$ in agreement with a decrease of the error in $\sqrt{n}$ (see Fig. 1 (left), while the harmonic mean approximation has no variance and obviously varies much more for the same computational effort. A further difficulty is that, in complex settings, the infinite variance of the harmonic estimator may remain undetected. For instance, a run up to 5,000,000 iterations produces an apparent decrease of the error in $\sqrt{n}$ in Fig. 1 (right).

Figure 1: Comparison of the error evolution for nested sampling and harmonic approximation: (left) Evolution of the error in Skilling’s nested sampling approximation of the marginal density when $y = 5$ in a $y|\sigma \sim \mathcal{N}(0, \sigma^2)$ and $\pi(\sigma^2|y) \sim \mathcal{I}\mathcal{G}(3/2, 1 + y^2/2)$ model, using $N$ initial simulations from the exponential prior and $j = N/2$ replications. (right) Evolution of the error of the harmonic approximation for the same problem, using 1000 times more simulations from the posterior than in nested sampling.
Bayes factors
A feature also common to both Skilling’s and Raftery et al.’s approaches is that they do not easily adapt to multiple models environments, as those encountered in model choice and the computation of Bayes factors, for which generic approximations methods like path sampling [Gelman and Meng, 1998] are readily available, being based on simulations from the (alternative) posterior distributions.

Asymptotic approximations
To go back to the Gamma approximation, we are a bit concerned with the log(1 − α) in the expression of log π(y) in eqn. (14). Would the Gamma approximation in eqn. (10) be exact (by chance, or because the model has some specific structure, e.g. Gaussian), then log π(y) would be infinite. This does not seem intuitive, and it also casts some doubt on the practicality of this approach. Beyond this specific point, one must always be wary of approximations methods that do not provide a way of evaluating, even roughly, the approximation error (like BIC). In that respect, a sequence of increasingly accurate (and possibly increasingly expensive to compute) approximations, would be preferable, as the authors suggest briefly in the conclusion.

Remark
A final bibliographical remark is about the study of DIC in missing data models: Celeux et al. [2006] give a detailed analysis of the multiple possible interpolations of pD and of DIC in such setups, agreeing with the authors about its instability.

References
