BAYESIAN STATISTICS 9,
J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,
D. Heckerman, A. F. M. Smith and M. West (Eds.)

Using TPA for Bayesian inference: a discussion

Nicolas Chopin and Christian P. Robert
CREST, Paris and Université Paris-Dauphine, CEREMADE
nicolas.chopin@ensae.fr, xian@ceremade.dauphine.fr

Summary

In this discussion, we reflect on the similarities and differences between TPA and NS (nested sampling).

1. IMPLEMENTATION OF TPA

Even though the use of an unknown accronym in the title of a paper is a risky marketing strategy (especially when it also means *Televisão Pública de Angola!*), we congratulate the author on the derivation of the precise bounds on the approximation of $\mu(B)/\mu(B')$. The argument in favour of TPA relating to the computing time, when compared with an accept-reject algorithm, is not completely fair: TPA is presented as an algorithm used to approximate evidence and alternative algorithms (see, e.g., Marin and Robert, 2010) do not include the simple-minded algorithm described by the author in his talk. A more detailed and precise description would have facilitated the lecture of the paper. When considering the implementation of TPA, notations used in the paper are slightly confusing, since "X is a draw from $\mu(A(\beta))$" first gives the impression of a uniform draw on $(0, \mu\{A(\beta)\})$. The distribution π_β in Section 8 is not defined. (We assume this is the Ising distribution although simulating from the Ising model requires perfect simulation.) The fact that the evidence is represented as an artificial integral ratio—rather than as a single integral—opens a Pandora box in that the calibration of the smaller set B' is open to mishandling and prone to errors in realistic problems.

2. NESTED SAMPLING

The paper acknowledges only briefly the connection with the NS method, presented at the last Valencia meeting and published in Skilling (2006, 2007). We were wondering whether or not there is a fundamental difference between both methods since the exploration principle, going from one level set to the next one, is most similar, as shown by the decomposition (3) in Mark Huber's paper.

N. Chopin and C.P. Robert are supported by the 2007–2010 grant ANR-07-BLAN-0237-01 "SP Bayes".

To investigate this further, we distinguish two aspects of the TPA method. First, we discuss the basic algorithm for computing $p = \mu(B)/\mu(B')$, which is the evidence of an artificial model, with prior $\mu/\mu(B)$, and likelihood equal to the indicator function of B'. Simple calculations (omitted for the sake of space) show that TPA can be interpreted as a NS algorithm for this artificial model, up to a tiny modification which we now describe briefly. The NS algorithm stops when the contribution $(t_i - t_{i+1})L_i$, $t_i = e^{-i}$ to the Riemman sum that defines the NS estimate is negligible. But in this artificial model, all these contributions are zero, until one reaches iteration k such that $L_i = 1$ for the first time, and then one may stop and directly add the sum all the remaining contributions (up to $j = +\infty$), i.e. $(t_k - 0) \times 1 = e^{-k}$; this is exactly the TPA estimate.

This means that the results of Chopin and Robert (2010) apply more or less directly to TPA; e.g. the cost with respect to the dimension d of the problem is likely to be $O(d^3)$. (The asymptotic result seems to directly apply to TPA.)

Second, we discuss how the author uses TPA to compute the evidence or marginal likelihood in a Bayesian framework. The parameter truncation scheme bears some resemblance with the nested ellipsoid strategy of Chopin and Robert (2010), but is much less applicable, if only because the first draw requires a starting sample from the posterior. The alternative likelihood truncation scheme is intriguing. Contrary to NS, exploration climbs down the likelihood contours, not up. This may be more efficient a strategy when the likelihood is unbounded (because nested sampling may fail to detect where it should stop increasing). Otherwise, intuition suggests going up or down in likelihood values should take roughly the same time, but we would welcome any comment from the author on this point.

But the most important point, as mentioned by Chopin and Robert (2007, 2010) in the context of nested sampling, and by Professor Roberts in his discussion of TPA, is that simulating from the dominating measure μ within a level set $A(\beta)$ is a difficult (not to say hopeless) problem in most realistic models. The proposal of Skilling (2006) to use an MCMC device like slice sampling—similar to the augmentation used in Section 8—is unsatisfactory in that it creates both a further approximation level and a dependence in the Poisson process that bias the TPA estimate.

REFERENCES

Chopin, N. and C. Robert. 2007. Comments on 'Nested Sampling' by John Skilling. In *Bayesian Statistics 8*, ed. J. M. et al. (eds) Oxford University Press. Bernardo, 491–524.

—. 2010. Properties of nested sampling. *Biometrika* To appear, doi:10.1093/biomet/asq021.

Marin, J. and C. Robert. 2010. Importance sampling methods for Bayesian discrimination between embedded models. In *Frontiers of Statistical Decision Making and Bayesian Analysis*, eds. M.-H. Chen, D. Dey, P. Müller, D. Sun, and K. Ye. Springer-Verlag, New York. To appear.

Skilling, J. 2006. Nested sampling for general Bayesian computation. *Bayesian Analysis* 1(4): 833–860.

—. 2007. Nested sampling for general Bayesian computation. In *Bayesian Statistics 8*, eds. J. Bernardo, M. Bayarri, J. Berger, A. David, D. Heckerman, A. Smith, and M. West.