Discussion on the presentation of Radu Craiu
Recent Advances in Regional Adaptation for MCMC

Jean-Michel Marin

Université Montpellier 2

Thank you very much for this nice presentation. It's a pleasure to have the opportunity to discuss around such an interesting work.

I have a few remarks on the practical aspect of your work.

Comment 1: What about the initialization step?

"*Adaptive algorithms are particularly vulnerable to this - quality of initial sample is central to the performance of the sampler.*"

A major difficulty with all adaptive algorithms is that the initialization is paramount to achieve proper performances.

According to the "what-you-get-is-what-you-see" features of such algorithms, it is quite difficult to recover from a poor starting sample, the adaptivity only focussing on the visited parts of the simulation space.

It is the same for MCMC and importance sampling methods: papers on adaptive methods do not focus enough on that problem.

Your proposal: "*Run in parallel a number (5-10) of RWM algorithms with fixed kernels started in different regions of the sample space (if the local modes are known start there) for an initialization period of M steps.*"

How to choose the different regions, i.e. the starting points of the parallel chains?

In the same way: even if the RAPTOR proposal adapts the partition of the sample space, it depends on an initial partition. In practice, it seems difficult to set this partition.

Adaptive Multiple Importance Sampling

We proposed an initialization procedure based on the fact that the ESS can be used to calibrate the initial importance distribution.

If we consider an importance sample of size N having used the importance distribution Q, the Effective Sample Size (ESS) is given by

$$\frac{N}{1 + \mathbb{V}_Q[\pi(\mathbf{y})/q(\mathbf{y})]} .$$

Cornuet, Marin, Mira and Robert (2010)

In practice,

$$\mathbb{V}_Q[\pi(\mathbf{y})/q(\mathbf{y})] = \int \{\pi(\mathbf{y})/q(\mathbf{y}) - \mathbb{E}_Q[\pi(\mathbf{y})/q(\mathbf{y})]\}^2 \, q(\mathbf{y})\nu(\mathrm{d}\mathbf{y})$$

can be estimated using the coefficient of variation of the importance weights.

When $\Pi(\Omega) \subseteq \mathbb{R}^p$, our proposal is:

i. Use an inverse logistic transformation that maps the p components into the unit interval and then to sample from a uniform distribution over this hypercube: independently generate N particles from the product of p standard logistic distributions;

ii. Rescale component-wise this sample by maximizing the ESS.

The Nelder and Mead (1965) algorithm is used to maximize the ESS.

This solution is far from being fool-proof and we favour an informed alternative implementation as soon as pieces of information on the target distribution are a priori available.

For RAPTOR, you could imagine using the same strategy and derive the Voronoï partition associated to a weighted k-means procedure.

Comment 2: How to fix the number of component of the mixture proposal?

There exist different criteria: AIC, BIC, ICL. Which one should be used? There is a compromise between bias and variance to assess.

In high dimension, it is almost impossible to use a large number of components, unless you constraint the covariance structure.