Discussion of “Asymptotic variance estimation for Adaptive Markov chain Monte Carlo”

Luke Bornn

Dept. of Statistics, University of British Columbia

Jan 4, 2011

- The asymptotic variance

$$\sigma^2(h) \equiv \lim_{n\to\infty} Var\left(n^{1/2}\sum_{k=1}^{n} h(X_k)\right) \tag{1}$$

  is importance for assessing performance of Monte Carlo methods. Due to serial dependence in Markov chains, generally $\sigma^2(h) \neq Var_\pi h$
- Obtaining a quality estimate $\hat{\sigma}_n^2$ allows one to
  - construct confidence intervals for $E_\pi h$
  - create intelligent stopping rules for your Markov chain
- Several (but not enough!) recent papers have explored this in more detail. See Flegal and Jones (2010, AoS) and references therein.

- Several estimators of the asymptotic variance $\sigma^2(h)$ have been developed:
- Strong consistency of Batch Means and Regenerative Simulation explored in Jones et al. (2006) and Hobert et al. (2002) for (ergodic) Markov chains.
- Further results on Overlapping Batch Means and Spectral Variance in Flegal and Jones (2010).
- Current talk studies the properties of a Spectral Variance estimator under adaptation.

- It can be seen from the form of the SV estimator that increased autocorrelation results in increased variance estimation:

$$\Gamma_n^2 = \sum_{k=-n}^{n} \omega(kb_n)\gamma_n(k). \quad (2)$$

As such, perhaps such an estimate could be used to build a stopping rule for adaptation, providing an automated way to end adaptation and ensure convergence of lesser understood algorithms.

- Consider algorithms which reduce the adaptation step-size only when some criterion is met. For example, the Wang-Landau algorithm divides the support into d regions, only decreasing the adaptation parameter γ_t when the sampler adequately explores all d regions (as measured by a flat-histogram criterion).
- Perhaps similar conditions could be built around $\hat{\sigma}^2(h)$ to adjust adaptation on the fly. In some sense, this amounts to creating micro-stopping rules.

- This brings us two questions to consider:
    1. Will these theoretical results on adaptive MCMC extend to other estimators (BM, OBM, RS, etc.)? If not, what are the barriers to be overcome?
    2. Could this approach be used to construct stopping rules for lesser understood adaptive algorithms?
    3. Aside from constructing stopping rules, might these asymptotic variance estimators be used to help build “adaptive” adaptation rates?