

Batch means standard errors for MCMC

Batch means is one method used to compute Monte Carlo standard errors. There is a vast literature on sophisticated methods for computing Monte Carlo standard errors for MCMC output. However batch means has the advantage of being easy to implement and it appears to work reasonably well in practice.

Suppose we are interested in estimating an expectation $\mu = E(g(X))$ where X is a draw from a distribution f . The batch means method works as follows: Run the Markov chain $\{X_n\}$ for $N=ab$ iterations (we can assume a and b are integers).

Let

$$Y_k = \frac{\sum_{i=(k-1)b+1}^{kb} g(X_i)}{b}$$

If we think of the Markov chain as having been divided into “ a ” batches of size “ b ” each, Y_k is then the Monte Carlo estimate of μ based on the k th ‘batch’. Let

$$\hat{\sigma}^2 = \frac{b}{a-1} \sum_{k=1}^a (Y_k - \hat{\mu})^2.$$

Then the batch means estimate of Monte Carlo standard error is $\frac{\hat{\sigma}}{\sqrt{N}}$.

How do we pick b (and hence a) ?

b (batch size) should be large enough so Y_k s are approximately independent. This is difficult to determine in general so a reasonable heuristic is to use $b=\sqrt{N}$. This recommendation is related to the work on the “Consistent Batch Means” method due to Jones, Haran, Caffo and Neath (2005) where the standard error based on the batch means method is used to determine a stopping time for an MCMC simulation. The idea is fairly simple: Determine a desired level of accuracy (in terms of Monte Carlo standard errors) for your parameters. Usual frequentist notions of the desired width of 95% confidence intervals for the parameters can help determine the desired Monte Carlo standard error. Periodically compute standard errors for the parameters using consistent batch means as described above. Stop the chain when all standard errors attain desired levels. Note that one should make sure that the chain is run for a minimum initial length (depending on the problem, observed autocorrelations etc.) so that the standard error estimates themselves are not too unreliable.