Accuracy

A stationary Markov process is still dependent.

A M-H process is typically positively correlated, because X_{n+1} tends to be close to X_n (or even equal to X_n).

If $X_1, \ldots, X_n$ are i.i.d. then the Central Limit Theorem gives us an approximate 95% confidence interval for $\mathbb{E}_\pi h(X) = \int h(x)\pi(x)dx$, namely

$$\overline{h(X)} \pm 2\frac{S}{\sqrt{n}}$$

where $S^2 = \sum_{i=1}^n \left(h(X_i) - \overline{h(X)}\right)^2 / (n-1)$ is the sample variance.

For a positively correlated sequence S will generally underestimate the standard deviation, and our confidence interval will be too small.

Blocking

We can use blocking to get a better confidence interval.

Split our sample up into Q blocks of size a, and let $\overline{h(X)}(i)$ be the average from the i-th block.

If a is large enough then the block means will be close to independent, whence

$$\overline{h(X)} = \frac{1}{n}\sum_{i=1}^{n} h(X_i) = \frac{1}{Q}\sum_{i=1}^{Q} \overline{h(X)}(i)$$

will have the approximate variance

$$S^2 = \frac{1}{Q-1}\sum_{i=1}^{Q}\left(\overline{h(X)}(i) - \overline{h(X)}\right)^2$$

and we can use this to get a 95% CI in the usual way.

WinBUGS uses $a = \lfloor \sqrt{n} \rfloor$ and reports $S/\sqrt{Q}$ as the MC error.

If you have run M chains and n is the number of samples taken from each chain (after discarding the burn-in), then WinBUGS still uses $a = \lfloor \sqrt{n} \rfloor$ but the total number of blocks is now MQ.

We can check for (linear) dependence in the sequence of block means $\left\{ \overline{h(X)}(i) \right\}_{i=1}^{Q}$ using the lag-1 autocorrelation, estimated using

$$\hat{\rho}(1) = \frac{\sum \left(\overline{h(X)}(i+1) - \overline{h(X)} \right) \left(\overline{h(X)}(i) - \overline{h(X)} \right)}{\sum \left(\overline{h(X)}(i) - \overline{h(X)} \right)^2}$$

If the sequence is uncorrelated then $\hat{\rho}(1) \approx N\left(-1/Q, 1/Q\right)$.

Choosing the sample size

The larger the lag-1 autocorrelation is, the larger our blocks have to be before their means look independent.

It's always the case that the more samples/blocks the better, however a good rule of thumb is that the sample standard deviation (from the block means, that is the MC error) should be no more that 5% of the size of the standard deviation of the posterior you are interested in.

A visual rule of thumb is that if you can still see positive correlation in a trace of $\{X_i\}$ then you need more samples. That is, relative to the overall length of the trace, the oscillations need to be short enough that the trace just looks like white noise. If the trace looks like a wooly snake rather than a hairy caterpillar (lindys blewog), then you need more samples.

Effective sample size

The Effective Sample Size gives an estimate of the equivalent number of independent iterations that a stationary ergodic sequence represents.

If $\mathbb{E}\,h(X_i) = \mu$ and $\operatorname{Var} h(X_i) = \sigma^2$ then for large n

$$\operatorname{Var} \frac{1}{n} \sum_{i=1}^{n} h(X_i) \approx \frac{\sigma^2}{n} \left(1 + 2 \sum_{i=1}^{\infty} \rho_i \right)$$

where ρ_i is the lag i autocorrelation:

$$\rho_i = \mathbb{E}(h(X_{n+i}) - \mu)(h(X_n) - \mu)/\sigma^2.$$

The effective sample size is thus

$$\frac{n}{1 + 2 \sum_{i=1}^{\infty} \rho_i}.$$

A crude estimate of ρ_i is given by $\hat{\rho}(1)^i$.
Using this gives us the following estimate of the effective sample size

$$n\frac{1-\hat{\rho}(1)}{1+\hat{\rho}(1)}.$$