ADVANCED BAYESIAN MODELING

USING MCMC IN PRACTICE:
MONTE CARLO ERROR & SAMPLE SIZE

Monte Carlo Error

Consider estimating $\mathrm{E}(\psi \mid y)$ for some scalar ψ with finite posterior variance.

The natural estimate based on m chains of length n is the overall average

$$\overline{\psi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\overline{\psi}_{.j} = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}\psi_{ij}$$

Its **Monte Carlo standard error** is

$$\sqrt{\mathrm{var}\left(\overline{\psi}_{..}\right)}$$

or, in practice, a time-series estimate of this.

Sample Size

If the posterior could be sampled independently, how many such draws of ψ would be needed for their average to have the same standard error as $\overline{\psi}_{..}$?

For ψ, this is the **effective sample size** n_{eff}.

Since an average of n_{eff} independent samples has variance $\text{var}(\psi \mid y)/n_{\text{eff}}$,

$$n_{\text{eff}} \;=\; \frac{\text{var}(\psi \mid y)}{\text{var}\left(\overline{\psi}_{..}\right)}$$

Effective sample sizes of at least 400 are recommended for accuracy.

In practice, instead use an asymptotic version

$$n_{\text{eff}} = \frac{mn}{1 + 2\sum_{t=1}^{\infty} \rho_t}$$

where ρ_t is autocorrelation for ψ at lag t.

Estimation procedure: BDA3, Sec. 11.5

Note: Should be based on chains that have converged (after burn-in).

Thinning

Can **thin** a sequence of iterates by using only every kth iterate, for some $k > 1$

When autocorrelations are high, using a thinned sequence is almost as efficient as using the whole sequence.

Use thinning to

- Reduce memory requirements (fewer samples to store)
- Reduce computation time (if storing values is a bottleneck)