More on Metropolis-Hastings

Proposal distributions / jumping rules

It is allowed to propose a value θ^* outside the support of the prior distribution, but the value will be rejected because $p(\theta^*) = 0$, hence

$$r = \frac{p(\theta^*|y) / J(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y) / J(\theta^{t-1}|\theta^*)} = 0$$

Two main types of proposal distributions

- Random walk type: that the proposal depends on the previous value θ^{t-1}, most commonly centered at θ^{t-1}, for example a Normal distr. with mean θ^{t-1}
  - Acceptance rates should not be too high or too low
    - ≈ 0.5 for 1 or 2 dimensional θ
    - ≈ 0.25 for dimensions higher than 5

Does not depend on θ_i^{t-1}, but on $\theta_{(-i)}^{t-1}$

- Gibbs type proposals, which are approximations of the target distribution
  - Acceptance rates should be as high as possible

nov 24-12:30

Acceptance probabilities and acceptance rates

Acceptance probabilities

$$\alpha(\theta^{t-1}, \theta^*) = \min\{1, r\}$$

Acceptance rates: The average of acceptance probabilities over iterations

Can be calculated in different ways:

- Take the average of the acceptance probabilities calculated simulation
- Count the number of accepted proposals, divide by the number of simulations

$\hookrightarrow$ Give comparable results

nov 24-12:44

Effective number of independent simulation draws

Remember: Consider a scalar estimand ψ

c chains, each of length T. Remove half as burn-in, split each chain into two parts, resulting in $m = 2 \cdot c$ of length $n = \frac{T}{2 \cdot 2}$, and in total $m \cdot n$ samples.

- We have
  - the between sequence variance B
  - the within sequence variance W
  - $\widehat{\text{var}}^+(\psi|y) = \frac{n-1}{n} W + \frac{1}{n} B$

- Remember that MCMC by construction gives dependent samples. Hence estimating $\underbrace{E[\psi|y]}_{\text{could be } h(\theta)}$ by $\bar{\psi} = \frac{1}{m \cdot n} \sum_{t=1}^{m \cdot n} \psi^{(t)}$

is not as efficient as if the m·n samples were independent

If the m·n samples were independent, then the estimator of the variance of $\bar{\psi}$ would be

nov 24-12:54

$$\frac{1}{m \cdot n} \cdot \widehat{\text{Var}[\psi|y]} = \frac{1}{m \cdot n} \cdot \frac{1}{(m \cdot n - 1)} \sum_{t=1}^{m \cdot n} (\psi^t - \bar{\psi})^2 \quad (*)$$

Since the samples in fact are dependent, (*) underestimates the variance of $\bar{\psi}$

The effective sample size n_{eff} is such that the variance of $\bar{\psi}$ is

$$\frac{1}{n_{eff}} \cdot \widehat{\text{Var}[\psi|y]} = \frac{1}{n_{eff}} \cdot \frac{1}{(m \cdot n - 1)} \cdot \sum_{t=1}^{m \cdot n} (\psi^t - \bar{\psi})^2$$

In fact

$$\lim_{n \to \infty} m \cdot n \cdot \text{var}(\bar{\psi}) = \left(1 + 2 \sum_{t=1}^{\infty} \rho_t\right) \text{var}(\psi|y)$$

where ρ_t is the autocorrelation of the sequence ψ at lag t.

Hence $$\text{var}(\bar{\psi}) \approx \frac{1}{\frac{m \cdot n}{(1 + 2 \cdot \sum_{t=1}^{\infty} \rho_t)}} \cdot \widehat{\text{var}(\psi|y)}$$

and hence $$n_{eff} = \frac{m \cdot n}{(1 + 2 \cdot \underbrace{\sum_{t=1}^{\infty} \rho_t}_{\text{Need to be estimated}})}$$

nov 24-13:21

Rule of thumb $n_{eff} \geqslant 5 \cdot m = 10 \cdot c$

nov 24-13:33

The BUGS language

- WinBUGS
- OpenBUGS

R interfaces: BRugs
+ 1 more

- An R package called coda: Diagnostics tools

STAN

INLA: Integrated nested Laplace approximation. Software: R-INLA

R-NIMBLE

ABC: Approximative Bayesian computation

nov 24-13:33

nov 24-13:40