MCMC Approximation of the Posterior in Linear Models

Galin L. Jones
School of Statistics
University of Minnesota

July 31, 2006

Introduction

Hierarchical Linear Models

Suppose

$$Y \mid \beta, u, R, D \sim \mathrm{N}_n(X\beta + Zu, R^{-1})$$
$$\beta \mid u, R, D \sim \mathrm{N}_p(\beta_0, B^{-1})$$
$$u \mid D, R \sim \mathrm{N}_q(0, D^{-1})$$

The posterior is then characterized by

$$\pi(\beta, u, R, D \mid y) \propto f(y \mid \beta, u, R, D) f(\beta \mid u, R, D) f(u \mid D, R) f(R) f(D)$$

where $f(R)$ and $f(D)$ are unspecified priors.

Hierarchical Linear Models

The posterior

$$\pi(\beta, u, R, D \mid y)$$

is analytically intractable in the sense that even if we choose proper conjugate priors the integrals required for inference are not available in closed form.

MCMC provides a method for approximating these integrals.

Estimation

Let π be a probability distribution. I want the value of

$$E_\pi g := \int g(x)\pi(dx)$$

but this is intractable.

MCMC Approximation:
Simulate a Markov chain $X := \{X_n\}$ with invariant distribution π. Then X "converges" to π and

$$\bar{g}_n := \frac{1}{n}\sum_{i=1}^{n} g(X_i) \overset{a.c.}{\to} E_\pi g \text{ as } n \to \infty$$

How can we produce the sample? Metropolis-Hastings-Green or Gibbs is the "usual" recipe.

Initial value

How should we choose X_1? (Reduce bias in $\bar{g}_n$.)

Standard practice is to set $X_1 = x_1$ and discard the first b iterations in the hope that the distribution of X_{b+1} is close to that of π. How should b be chosen?

Suppose we can construct an approximation to π say $\hat{\pi}$ such that we can make iid draws from $\hat{\pi}$. Moreover, we require that with probability $1 - \alpha$

$$\|\pi - \hat{\pi}\| \leq \epsilon$$

where $\|\cdot\|$ is the total variation norm.

Then taking $X_1 \sim \hat{\pi}$ would be sensible.

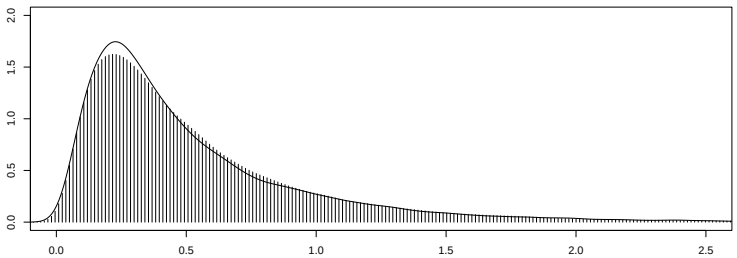
Example

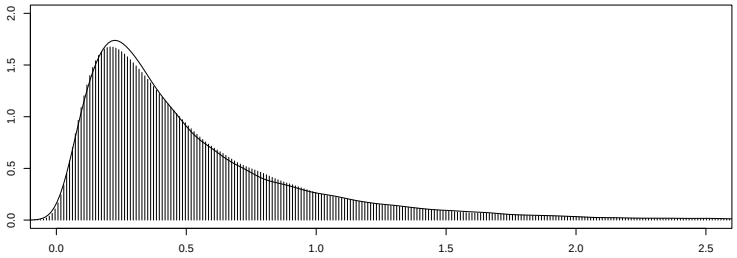
Consider the following special case of the hierarchical linear model

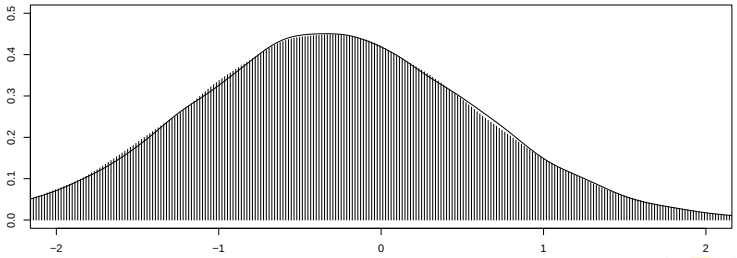
$$\begin{aligned} Y \mid \beta, u, \lambda_R &\sim \mathrm{N}_6(x\beta + u, \lambda_R^{-1} I_6) \\ \beta \mid u &\sim \mathrm{N}(0, 1) \\ u \mid \lambda_D &\sim \mathrm{N}_6(0, \lambda_D^{-1} I_6) \\ \lambda_R &\sim \mathrm{Gamma}(1, 2) \\ \lambda_D &\sim \mathrm{Gamma}(1, 2) \end{aligned}$$

where $x \sim N_6(0, I_6)$.

Example


0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
1.5
2.0
2.5


0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
1.5
2.0
2.5


0.0
0.1
0.2
0.3
0.4
0.5
−2
−1
0
1
2

Minorization

Let $P(x, dy)$ be the Markov transition kernel associated with our sampler.
Minorization Condition: $P(x, B) \geq s(x)Q(B)$
Mixture Representation:

$$\begin{aligned} P(x, dy) &= s(x)Q(dy) + [1 - s(x)]\frac{P(x, dy) - s(x)Q(dy)}{1 - s(x)} \\ &= s(x)Q(dy) + [1 - s(x)]R(x, dy) \end{aligned}$$

Split Chain

$$P(x, dy) = s(x)Q(dy) + [1 - s(x)]R(x, dy)$$

Simulating the split chain: Given $X_i = x$

- generate $\delta_i \sim \text{Ber}(s(x))$
- if $\delta_i = 0$ draw $X_{i+1} \sim R$ but
- if $\delta_i = 1$ draw $X_{i+1} \sim Q$ (Regeneration!)

The result

$$X' = \{(X_0, \delta_0), (X_1, \delta_1), (X_2, \delta_2), \ldots\}$$

Definitions

Split chain:

$$X' = \{(X_0, \delta_0), (X_1, \delta_1), (X_2, \delta_2), \ldots\}$$

Define $\tau = \min\{n \geq 1 : \delta_n = 1\}$,

$$Q_t(\cdot) = \Pr(X_t \in \cdot \mid \tau \geq t)$$

and $p_t = \frac{\Pr(\tau \geq t)}{E(\tau)}$

The Approximation

Theorem

If X is Harris recurrent then for any $B \in \mathcal{B}(\mathcal{X})$
$\pi(B) = \sum_{t=1}^{\infty} Q_t(B) p_t$

This suggests the possibility of drawing from π by randomly drawing an element from the set

$$\{Q_1, Q_2, Q_3, \ldots\}$$

according to the probabilities $p_1, p_2, p_3, \ldots$ Then making a random draw from the chosen Q_t

The Approximation

Drawing from $Q_t(\cdot)$ is easy.

Let T be a random variable with mass function $\Pr(T = t) = p_t$.

Drawing from this distribution is challenging and we don't know how to do it.

But we can simulate the split chain and estimate p_t with $\hat{p}_t$ which allows us to estimate π with

$$\hat{\pi}(B) = \sum_{t=1}^{\infty} Q_t(B)\hat{p}_t$$

so that

$$\|\pi - \hat{\pi}\| \to 0 \quad \text{as} \quad n \to \infty \quad \text{w.p. } 1$$

Concluding Remarks

The approximation $\hat{\pi}$ can be useful for

1. obtaining a preliminary estimate of the target distribution,
2. controlling the bias in $\bar{g}_n$ and
3. finding an overdispersed starting distribution for use in some diagnostics

but it is important to remember that is *not* a way to "fix" a poorly mixing Markov chain.