Convergence results for two Gibbs samplers

Jim Hobert
U. of Florida

Collaborators: Kshitij Khare & Jorge Román

Outline: I. Why is the convergence rate important?

II. One slide on the geometric drift function

III. A Gibbs sampler for Bayesian quantile regression

IV. A Gibbs sampler for Bayesian linear mixed models

V. Yet another warning about improper priors

I. Why is the convergence rate important?

Classical Monte Carlo estimation of $\mathrm{E}_\pi g := \int_{\mathbb{R}^p} g(x)\,\pi(x)\,dx$

Theory: Let $X_1, X_2, X_3, \ldots$ be iid π and form $\overline{g}_n := \frac{1}{n}\sum_{i=1}^{n} g(X_i)$

SLLN: If $\mathrm{E}_\pi|g| < \infty$, then $\overline{g}_n \to \mathrm{E}_\pi g$ a.s. as $n \to \infty$

I. Why is the convergence rate important?

Classical Monte Carlo estimation of $E_\pi g := \int_{\mathbb{R}^p} g(x)\,\pi(x)\,dx$

Theory: Let $X_1, X_2, X_3, \ldots$ be iid π and form $\overline{g}_n := \frac{1}{n}\sum_{i=1}^{n} g(X_i)$

SLLN: If $E_\pi|g| < \infty$, then $\overline{g}_n \to E_\pi g$ a.s. as $n \to \infty$

CLT: If $E_\pi g^2 < \infty$, then $\sqrt{n}\,\big(\overline{g}_n - E_\pi g\big)/\hat{\sigma}_n \xrightarrow{d} \mathrm{N}(0,1)$ where

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(g(X_i) - \overline{g}_n\big)^2$$

So, for large n, $\Pr\left(\overline{g}_n - \frac{2\hat{\sigma}_n}{\sqrt{n}} < E_\pi g < \overline{g}_n + \frac{2\hat{\sigma}_n}{\sqrt{n}}\right) \approx 0.95$

I. Why is the convergence rate important?

Classical Monte Carlo estimation of $E_\pi g := \int_{\mathbb{R}^p} g(x)\,\pi(x)\,dx$

Theory: Let $X_1, X_2, X_3, \ldots$ be iid π and form $\bar{g}_n := \frac{1}{n}\sum_{i=1}^n g(X_i)$

SLLN: If $E_\pi|g| < \infty$, then $\bar{g}_n \to E_\pi g$ a.s. as $n \to \infty$

CLT: If $E_\pi g^2 < \infty$, then $\sqrt{n}\,(\bar{g}_n - E_\pi g)/\hat{\sigma}_n \xrightarrow{d} N(0,1)$ where

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(g(X_i) - \bar{g}_n\right)^2$$

So, for large n, $\Pr\left(\bar{g}_n - \frac{2\hat{\sigma}_n}{\sqrt{n}} < E_\pi g < \bar{g}_n + \frac{2\hat{\sigma}_n}{\sqrt{n}}\right) \approx 0.95$

Application: Fix n, simulate $X_1, \ldots, X_n$ iid π

Asymptotic 95% CI for $E_\pi g$: $\bar{g}_n \pm 2\hat{\sigma}_n/\sqrt{n}$

Can we honestly replace the iid sequence with a MC?

Let $X_0, X_1, X_2, \ldots$ be a well-behaved MC converging to $\pi(x)$

As in the iid case, let $\overline{g}_n := \frac{1}{n} \sum_{i=0}^{n-1} g(X_i)$

Ergodic Theorem: If $\mathrm{E}_\pi |g| < \infty$, then $\overline{g}_n \to \mathrm{E}_\pi g$ a.s. as $n \to \infty$

So the MCMC estimator also converges a.s. to $\mathrm{E}_\pi g$, but...

Can we honestly replace the iid sequence with a MC?

Let $X_0, X_1, X_2, \ldots$ be a well-behaved MC converging to $\pi(x)$

As in the iid case, let $\overline{g}_n := \frac{1}{n}\sum_{i=0}^{n-1} g(X_i)$

Ergodic Theorem: If $E_\pi|g| < \infty$, then $\overline{g}_n \to E_\pi g$ a.s. as $n \to \infty$

So the MCMC estimator also converges a.s. to $E_\pi g$, but...

There is no free lunch!

In the MC context: $E_\pi g^2 < \infty \not\Rightarrow \sqrt{n}\left(\overline{g}_n - E_\pi g\right) \xrightarrow{d} N(0, \gamma^2)$

But, if the chain is geometrically ergodic (G.E.), there are CLTs

Target density: $\pi : \mathsf{X} \to (0,\infty)$ where $\mathsf{X} \subseteq \mathbb{R}^d$

Markov chain: $\{X_n\}_{n=0}^{\infty}$ has Mtd $k : \mathsf{X} \times \mathsf{X} \to (0,\infty)$ satisfying

$$\int_{\mathsf{X}} k(x'|x)\,\pi(x)\,dx = \pi(x')$$

Density of X_n given $X_0 = x$: $k^n(x'|x) := \int_{\mathsf{X}} k(x'|z)\,k^{n-1}(z|x)\,dz$

Definition: If there exist $\rho \in [0,1)$ and $M : \mathsf{X} \to [0,\infty)$ st

$$\int_{\mathsf{X}} \left|k^n(x'|x) - \pi(x')\right| dx' \le M(x)\rho^n \quad \text{for all } n \in \mathbb{N}$$

then $\{X_n\}_{n=0}^{\infty}$ is called G.E.

If the chain is G.E. and $\mathrm{E}_\pi |g|^{2+\epsilon} < \infty$ for some $\epsilon > 0$, then

$$\sqrt{n}\left(\overline{g}_n - \mathrm{E}_\pi g\right) \xrightarrow{d} \mathrm{N}(0,\gamma^2)$$

II. One slide on the geometric drift function

Mtd: $k : \mathsf{X} \times \mathsf{X} \to (0, \infty)$ satisfying $\int_{\mathsf{X}} k(x'|x)\, \pi(x)\, dx = \pi(x')$

Assume $k(x'|\cdot)$ is continuous (or just l.s.-c.) for each $x' \in \mathsf{X}$

Definition: $V : \mathsf{X} \to [0, \infty)$ is *unbounded off compact sets* if

$$\left\{x \in \mathsf{X} : V(x) \leq c\right\} \text{ is compact for all } c \geq 0$$

Example: If $\mathsf{X} = \mathbb{R}$, $V(x) = x^2$ is u.o.c.s., but $V(x) = \frac{1}{x^2+1}$ isn't

The chain $\{X_n\}_{n=0}^{\infty}$ is G.E. if, for all $x \in \mathsf{X}$,

$$\mathrm{E}\left[V(X_{n+1}) \,|\, X_n = x\right] = \int_{\mathsf{X}} V(x')\, k(x'|x)\, dx' \leq \lambda V(x) + L$$

where V is u.o.c.s., $\lambda \in [0, 1)$ and $L \in \mathbb{R}$

III. A Gibbs sampler for Bayesian quantile regression

Let $Y_1, \ldots, Y_m$ be indep random variables st

$$Y_i = x_i^T \beta + \sigma \varepsilon_i$$

- x_i is a $p \times 1$ vector of known covariates
- β is a $p \times 1$ vector of unknown parameters
- σ is a univariate scale parameter
- $\{\varepsilon_i\}_{i=1}^n$ are iid with common density

$$r(1-r)\Big[e^{(1-r)\varepsilon} I(\varepsilon \leq 0) + e^{-r\varepsilon} I(\varepsilon > 0)\Big]$$

so that $\Pr(\varepsilon_1 < 0) = r$.

III. A Gibbs sampler for Bayesian quantile regression

Let $Y_1, \dots, Y_m$ be indep random variables st

$$Y_i = x_i^T \beta + \sigma \varepsilon_i$$

- x_i is a $p \times 1$ vector of known covariates
- β is a $p \times 1$ vector of unknown parameters
- σ is a univariate scale parameter
- $\{\varepsilon_i\}_{i=1}^n$ are iid with common density

$$r(1-r)\Big[e^{(1-r)\varepsilon} I(\varepsilon \leq 0) + e^{-r\varepsilon} I(\varepsilon > 0)\Big]$$

  so that $\Pr(\varepsilon_1 < 0) = r$.

Joint density of the data, $(Y_1, \dots, Y_m)$, is

$$f(y; \beta, \sigma) = r^m (1-r)^m \sigma^{-m} \exp\left\{ -\frac{1}{\sigma} \sum_{i=1}^{m} \psi_r(y_i - x_i^T \beta) \right\}$$

where $\psi_r(u) = u\big[r - I(u < 0)\big]$

The model: $f(y;\beta,\sigma) = r^m(1-r)^m\sigma^{-m}e^{-\frac{1}{\sigma}\sum_{i=1}^m \psi_r(y_i - x_i^T\beta)}$

Let $\pi(\beta,\sigma)$ be a prior. The intractable posterior density is

$$\pi(\beta,\sigma|y) = \frac{f(y;\beta,\sigma)\,\pi(\beta,\sigma)}{m(y)}$$

Kozumi & Kobayashi (2011): Let $\{(Y_i, Z_i)\}_{i=1}^m$ be indep pairs st

$$Y_i|Z_i = z \sim \mathrm{N}\big(x_i^T\beta + \theta z, \sigma z \tau^2\big) \quad \& \quad Z_i \sim \mathrm{Exp}(\sigma)$$

The model: $f(y;\beta,\sigma) = r^m(1-r)^m\sigma^{-m}e^{-\frac{1}{\sigma}\sum_{i=1}^m \psi_r(y_i - x_i^T\beta)}$

Let $\pi(\beta,\sigma)$ be a prior. The intractable posterior density is

$$\pi(\beta,\sigma|y) = \frac{f(y;\beta,\sigma)\,\pi(\beta,\sigma)}{m(y)}$$

Kozumi & Kobayashi (2011): Let $\{(Y_i, Z_i)\}_{i=1}^m$ be indep pairs st

$$Y_i|Z_i = z \sim \mathrm{N}\big(x_i^T\beta + \theta z, \sigma z \tau^2\big) \quad \& \quad Z_i \sim \mathrm{Exp}(\sigma)$$

$$\int_{\mathbb{R}_+^n} f(y, z;\beta,\sigma)\, dz = f(y;\beta,\sigma) \qquad (*)$$

Now define

$$\pi(\beta,\sigma,z|y) = \frac{f(y,z;\beta,\sigma)\,\pi(\beta,\sigma)}{m(y)}$$

It follows from $(*)$ that $\int_{\mathbb{R}_+^n} \pi(\beta,\sigma,z|y)\, dz = \pi(\beta,\sigma|y)$

$$Y_i|Z_i = z \sim \mathrm{N}\left(x_i^T\beta + \theta z, \sigma z \tau^2\right) \quad \& \quad Z_i \sim \mathrm{Exp}(\sigma)$$

$$\pi(\beta, \sigma|y) = \int_{\mathbb{R}_+^n} \pi(\beta, \sigma, z|y)\, dz = \int_{\mathbb{R}_+^n} \frac{f(y, z; \beta, \sigma)\, \pi(\beta, \sigma)}{m(y)}$$

Normal $\times$ Inverse Gamma prior for (β, σ) yields π st:

- $\beta|\sigma, z, y \sim \mathrm{N}_p(\cdot, \cdot)$
- $\sigma|\beta, z, y \sim$ Inverse Gamma$(\cdot, \cdot)$
- $\left(\frac{1}{z_1}, \ldots, \frac{1}{z_m}\right)\Big|\beta, \sigma, y \sim \prod_{i=1}^m$ Inverse Gaussian$(\cdot, \cdot)$

Let $\Phi = \left\{(\beta_n, \sigma_n)\right\}_{n=0}^\infty$ be a Markov chain on $\mathbb{R}^p \times \mathbb{R}_+$ with Mtd

$$k\left(\beta', \sigma' \mid \beta, \sigma\right) = \int_{\mathbb{R}_+^n} \pi\left(\beta'|\sigma', z, y\right) \pi\left(\sigma'|\beta, z, y\right) \pi\left(z|\beta, \sigma, y\right) dz$$

Invariance: $\pi(\beta', \sigma'|y) = \int_{\mathbb{R}^p} \int_{\mathbb{R}_+} k\left(\beta', \sigma' \mid \beta, \sigma\right) \pi(\beta, \sigma|y)\, d\sigma\, d\beta$

Quantile regression model: $Y_i = x_i^T \beta + \sigma \varepsilon_i$

Prior for (β, σ): Normal $\times$ Inverse Gamma

Markov chain: $\Phi = \{(\beta_n, \sigma_n)\}_{n=0}^{\infty}$ has Mtd

$$k(\beta', \sigma' \mid \beta, \sigma) = \int_{\mathbb{R}_+^n} \pi(\beta' | \sigma', z, y) \pi(\sigma' | \beta, z, y) \pi(z | \beta, \sigma, y) \, dz$$

and invariant density $\pi(\beta, \sigma | y)$

Proposition (Khare & H, 2011): Φ is G.E.

Proof uses the drift function $V : \mathbb{R}^p \times \mathbb{R}_+ \to (0, \infty)$ given by

$$V(\beta, \sigma) = \sigma + \frac{1}{\sigma} + \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \beta^T \Sigma^{-1} \beta$$

IV. A Gibbs sampler for Bayesian linear mixed models

Let $Y_{N\times 1}$ follow the general linear mixed model

$$Y = X\beta + \sum_{i=1}^{r} Z_i u_i + e$$

- $X, Z_1, \ldots, Z_r$ are known matrices
- $u_i \sim \mathrm{N}_{q_i}\big(0, I\sigma^2_{u_i}\big)$ and $e \sim \mathrm{N}_N\big(0, I\sigma^2_e\big)$
- β and $\sigma^2 = \big(\sigma^2_e \;\; \sigma^2_{u_1} \cdots \sigma^2_{u_r}\big)$ are unknown parameters

Improper prior density for (β, σ^2):

$$\pi(\beta, \sigma^2) = \big(\sigma^2_e\big)^{-(a_e+1)} \prod_{i=1}^{r} \big(\sigma^2_{u_i}\big)^{-(a_i+1)}$$

$Y = X\beta + \sum_{i=1}^{r} Z_i u_i + e\,,\ \ u_i \sim \mathrm{N}_{q_i}\left(0, I\sigma^2_{u_i}\right)\ \ \&\ \ e \sim \mathrm{N}_N\left(0, I\sigma^2_e\right)$

Prior density: $\pi(\beta, \sigma^2) = \left(\sigma^2_e\right)^{-(a_e+1)} \prod_{i=1}^{r} \left(\sigma^2_{u_i}\right)^{-(a_i+1)}$

Set $\theta = (\beta, u) = (\beta, u_1, u_2, \ldots, u_r)$. Define

$$\pi^*(\theta, \sigma^2 | y) = f(y | u; \beta, \sigma^2)\, f(u; \sigma^2)\, \pi(\beta, \sigma^2)$$

Necessary conditions for propriety:

$$\int_{\mathbb{R}^{p+q}} \pi^*(\theta, \sigma^2 | y)\, d\theta < \infty \quad \& \quad \int_{\mathbb{R}^r_+} \pi^*(\theta, \sigma^2 | y)\, d\sigma^2 < \infty$$

When these hold, $\pi^*(\theta, \sigma^2 | y)$ has “conditionals” given by:

- $\theta | \sigma^2, y \sim \mathrm{N}_{p+q}(\cdot, \cdot)$
- $\sigma^2 | \theta, y \sim \prod_{i=1}^{r+1}$ Inverse Gamma$(\cdot, \cdot)$

$Y = X\beta + \sum_{i=1}^{r} Z_i u_i + e,\;\; u_i \sim \mathrm{N}_{q_i}\left(0, I\sigma_{u_i}^2\right)$ & $e \sim \mathrm{N}_N\left(0, I\sigma_e^2\right)$

Prior density: $\pi(\beta, \sigma^2) = \left(\sigma_e^2\right)^{-(a_e+1)} \prod_{i=1}^{r} \left(\sigma_{u_i}^2\right)^{-(a_i+1)}$

Let $\Phi = \left\{(\theta_n, \sigma_n^2)\right\}_{n=0}^{\infty}$ be a chain on $\mathbb{R}^{p+q} \times \mathbb{R}_+^r$ with Mtd

$$k\left(\tilde{\theta}, \tilde{\sigma}^2 \mid \theta, \sigma^2\right) = \pi^*\left(\tilde{\sigma}^2 | \tilde{\theta}, y\right) \pi^*\left(\tilde{\theta} | \sigma^2, y\right)$$

Invariance: $\pi^*(\tilde{\theta}, \tilde{\sigma}^2 | y) = \int \int k\left(\tilde{\theta}, \tilde{\sigma}^2 \mid \theta, \sigma^2\right) \pi^*(\theta, \sigma^2 | y)\, d\sigma^2\, d\theta$

$Y = X\beta + \sum_{i=1}^{r} Z_i u_i + e,\ \ u_i \sim \mathrm{N}_{q_i}\left(0, I\sigma_{u_i}^2\right)$ & $e \sim \mathrm{N}_N\left(0, I\sigma_e^2\right)$

Prior density: $\pi(\beta, \sigma^2) = \left(\sigma_e^2\right)^{-(a_e+1)} \prod_{i=1}^{r} \left(\sigma_{u_i}^2\right)^{-(a_i+1)}$

Let $\Phi = \left\{(\theta_n, \sigma_n^2)\right\}_{n=0}^{\infty}$ be a chain on $\mathbb{R}^{p+q} \times \mathbb{R}_+^r$ with Mtd

$$k\left(\tilde{\theta}, \tilde{\sigma}^2 \mid \theta, \sigma^2\right) = \pi^*\left(\tilde{\sigma}^2 | \tilde{\theta}, y\right) \pi^*\left(\tilde{\theta} | \sigma^2, y\right)$$

Invariance: $\pi^*(\tilde{\theta}, \tilde{\sigma}^2 | y) = \int\int k\left(\tilde{\theta}, \tilde{\sigma}^2 \mid \theta, \sigma^2\right) \pi^*(\theta, \sigma^2 | y)\, d\sigma^2\, d\theta$

Proposition (Román & H, 2011): Φ is G.E. if

- $a_i < 0$
- $q_i + 2a_i > q - t + 2$
- $N + 2a_e > p + t + 2$

Notation: $X_{N\times p}, Z_{N\times q} = (Z_1 \cdots Z_r)$ and $t = \mathrm{rank}\left(Z^T (I - P_X) Z\right)$

$$Y = X\beta + \sum_{i=1}^{r} Z_i u_i + e, \;\; u_i \sim \mathrm{N}_{q_i}\left(0, I\sigma^2_{u_i}\right) \;\&\; e \sim \mathrm{N}_N\left(0, I\sigma^2_e\right)$$

Prior density: $\pi(\beta, \sigma^2) = \left(\sigma^2_e\right)^{-(a_e+1)} \prod_{i=1}^{r} \left(\sigma^2_{u_i}\right)^{-(a_i+1)}$

Let $\Phi = \left\{(\theta_n, \sigma^2_n)\right\}_{n=0}^{\infty}$ be a chain on $\mathbb{R}^{p+q} \times \mathbb{R}^r_+$ with Mtd

$$k\left(\tilde{\theta}, \tilde{\sigma}^2 \,|\, \theta, \sigma^2\right) = \pi^*\left(\tilde{\sigma}^2|\tilde{\theta}, y\right) \pi^*\left(\tilde{\theta}|\sigma^2, y\right)$$

Invariance: $\pi^*(\tilde{\theta}, \tilde{\sigma}^2|y) = \int\int k\left(\tilde{\theta}, \tilde{\sigma}^2 \,|\, \theta, \sigma^2\right) \pi^*(\theta, \sigma^2|y)\, d\sigma^2\, d\theta$

Proposition (Román & H, 2011): Φ is G.E. if

- $a_i < 0$ $\quad\quad$ $\left(a_i < 0\right)$
- $q_i + 2a_i > q - t + 2$ $\quad\quad$ $\left(q_i + 2a_i > q - t\right)$
- $N + 2a_e > p + t + 2$ $\quad\quad$ $\left(N + 2a_e > p - 2\sum_{i=1}^{r} a_i\right)$

Notation: $X_{N\times p}, Z_{N\times q} = (Z_1 \cdots Z_r)$ and $t = \mathrm{rank}\left(Z^T(I - P_X)Z\right)$

Markov chain: $\Phi = \left\{(\theta_n, \sigma_n^2)\right\}_{n=0}^{\infty}$ has Mtd

$$k\left(\tilde{\theta}, \tilde{\sigma}^2 | \theta, \sigma^2\right) = \pi^*\left(\tilde{\sigma}^2 | \tilde{\theta}, y\right) \pi^*\left(\tilde{\theta} | \sigma^2, y\right)$$

and invariant density $\pi^*(\theta, \sigma^2 | y)$.

Proposition (Román & H, 2011): Φ is G.E. if

- $a_i < 0$ $\quad\quad (a_i < 0)$
- $q_i + 2a_i > q - t + 2$ $\quad\quad (q_i + 2a_i > q - t)$
- $N + 2a_e > p + t + 2$ $\quad\quad (N + 2a_e > p - 2\sum_{i=1}^{r} a_i)$

Φ, $\Phi_\theta = \left\{\theta_n\right\}_{n=0}^{\infty}$ and $\Phi_{\sigma^2} = \left\{\sigma_n^2\right\}_{n=0}^{\infty}$ converge at the same rate

Proof uses Φ_{σ^2} and drift function $V : \mathbb{R}_+^{r+1} \to (0, \infty)$ given by

$$V(\sigma^2) = c_1\left[\sigma_e^2 + \left(\sigma_e^2\right)^{-c_2}\right] + \sum_{i=1}^{r}\left[\sigma_{u_i}^2 + \left(\sigma_{u_i}^2\right)^{-c_2}\right]$$

V. Yet another warning about improper priors

$\Phi_\theta = \{\theta_n\}_{n=0}^{\infty}$ has state space $\mathbb{R}^{p+q}$ and Mtd

$$k(\tilde{\theta} \mid \theta) = \int_{\mathbb{R}_+^{r+1}} \pi^*(\tilde{\theta}|\sigma^2, y)\pi^*(\sigma^2|\theta, y)\, d\sigma^2$$

Recall that $\theta = (\beta, u_1, u_2, \ldots, u_r)$, and that

- $\theta|\sigma^2, y \sim \text{N}_{p+q}(\cdot, \cdot)$
- $\sigma^2|\theta, y \sim \prod_{i=1}^{r+1}$ Inverse Gamma$(\cdot, \cdot)$

In particular, $\sigma_{u_i}^2|\theta, y \sim$ Inverse Gamma$\left(\frac{q_i}{2} + a_i, \frac{\|u_i\|^2}{2}\right)$

Technical Problem: It's undefined on $\left\{\theta \in \mathbb{R}^{p+q} : \|u_i\| = 0\right\}$

V. Yet another warning about improper priors

$\Phi_\theta = \{\theta_n\}_{n=0}^{\infty}$ has state space $\mathbb{R}^{p+q}$ and Mtd

$$k(\tilde{\theta} \mid \theta) = \int_{\mathbb{R}_+^{r+1}} \pi^*(\tilde{\theta}|\sigma^2, y)\pi^*(\sigma^2|\theta, y)\, d\sigma^2$$

Recall that $\theta = (\beta, u_1, u_2, \ldots, u_r)$, and that

- $\theta|\sigma^2, y \sim \mathrm{N}_{p+q}(\cdot, \cdot)$
- $\sigma^2|\theta, y \sim \prod_{i=1}^{r+1}$ Inverse Gamma$(\cdot, \cdot)$

In particular, $\sigma^2_{u_i}|\theta, y \sim$ Inverse Gamma$\left(\frac{q_i}{2} + a_i, \frac{\|u_i\|^2}{2}\right)$

Technical Problem: It's undefined on $\left\{\theta \in \mathbb{R}^{p+q} : \|u_i\| = 0\right\}$

Can't we just change state space from $\mathbb{R}^{p+q}$ to $\mathbb{R}^{p+q} \setminus \mathcal{N}$?

Example: If the drift function is $V(x) = x^2$

$\left\{x \in \mathbb{R} : x^2 \leq c\right\}$ is compact, but $\left\{x \in \mathbb{R} \setminus \{0\} : x^2 \leq c\right\}$ isn't