Efficient Bernoulli factory MCMC for intractable posteriors[1]

Dootika Vats
Indian Institute of Technology Kanpur

CS Katha Barta Talk Series, NISER

October 4, 2023

[1] Joint with F.B. Gonçalves, K. Łatuszyński, G.O. Roberts

Intractable target distributions

Consider a Bayesian model for parameter θ:

$$\underbrace{\pi(\theta|y)}_{\text{Posterior}} \propto \underbrace{f(y|\theta)}_{\text{Likelihood}} \underbrace{\pi(\theta)}_{\text{Prior}} := \tilde{\pi}(\theta|y)\,.$$

The posterior is often complicated enough that it is only known up to the unnormalized $\tilde{\pi}(\theta|y)$.

Markov chain Monte Carlo (MCMC) algorithms may be used to sample from $\pi(\theta|y)$.

Accept-Reject based MCMC

An accept-reject MCMC algorithm $(k+1)$ update:

1. Generate $\theta^* \sim q(\theta^*|\theta_k)$
2. Set $\theta_{k+1} = \theta^*$ with probability $\alpha(\theta_k, \theta^*)$.
3. Else, $\theta_{k+1} = \theta_k$.

Of course $\alpha(\theta, \theta^*)$ is chosen to satisfy posterior invariance.

Accept-Reject based MCMC

An accept-reject MCMC algorithm $(k+1)$ update:

1. Generate $\theta^* \sim q(\theta^*|\theta_k)$
2. Set $\theta_{k+1} = \theta^*$ with probability $\alpha(\theta_k, \theta^*)$.
3. Else, $\theta_{k+1} = \theta_k$.

Of course $\alpha(\theta, \theta^*)$ is chosen to satisfy posterior invariance.

If $\alpha(\theta, \theta^*)$ can be evaluated, then obtaining an event with prob. $\alpha(\theta, \theta^*)$ is by:

Get $U \sim U(0,1)$ and check is $U \leq \alpha(\theta, \theta^*)$

Metropolis-Hastings (MH)

A popular acceptance probability used is the Metropolis-Hastings acceptance probability:

$$\alpha_{MH}(\theta, \theta^*) = \min\left\{1, \frac{\pi(\theta^*|y)\, q(\theta|\theta^*)}{\pi(\theta|y)\, q(\theta^*|\theta)}\right\} = \min\left\{1, \frac{\tilde{\pi}(\theta^*|y)\, q(\theta|\theta^*)}{\tilde{\pi}(\theta|y)\, q(\theta^*|\theta)}\right\}$$

Metropolis-Hastings (MH)

A popular acceptance probability used is the Metropolis-Hastings acceptance probability:

$$\alpha_{MH}(\theta, \theta^*) = \min\left\{1, \frac{\pi(\theta^*|y)\, q(\theta|\theta^*)}{\pi(\theta|y)\, q(\theta^*|\theta)}\right\} = \min\left\{1, \frac{\tilde{\pi}(\theta^*|y)\, q(\theta|\theta^*)}{\tilde{\pi}(\theta|y)\, q(\theta^*|\theta)}\right\}$$

Of course, if $\tilde{\pi}(\theta|y)$ is known, then MH can be implemented easily.

Intractable posteriors

Consider problems that yield targets that *cannot* be evaluated. This may be for example, because

$$\pi(\theta|y) = \int_\eta g(\theta, \eta|y) d\eta \,.$$

This problem arises in

- Priors on constrained spaces
- Missing data - imputation
- Bayesian inference for diffusions

Intractable posteriors

Consider problems that yield targets that *cannot* be evaluated. This may be for example, because

$$\pi(\theta|y) = \int_{\eta} g(\theta, \eta|y) d\eta \, .$$

This problem arises in

- Priors on constrained spaces
- Missing data - imputation
- Bayesian inference for diffusions

Here,

$$\alpha_{MH}(\theta, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*|y)\, q(\theta_k|\theta^*)}{\pi(\theta|y)\, q(\theta^*|\theta_k)} \right\}$$

cannot be evaluated.

Barker's algorithm

Barker (1965) proposed the acceptance function:

$$\alpha_B(\theta, \theta^*) = \frac{\pi(\theta^*|y)\, q(\theta|\theta^*)}{\pi(\theta|y)\, q(\theta^*|\theta) + \pi(\theta^*|y)\, q(\theta|\theta^*)}$$

Barker's algorithm is not very popular due to Peskun's ordering result.

Peskun Ordering (Peskun, 1973)

Let $\bar{X}_h = n^{-1} \sum_t h(X_t)$ be a Monte Carlo estimator for a function h. Let P_B and P_{MH} be Barker's and MH Markov kernels. Then

Peskun Ordering (Peskun, 1973)

Let $\bar{X}_h = n^{-1} \sum_t h(X_t)$ be a Monte Carlo estimator for a function h. Let P_B and P_{MH} be Barker's and MH Markov kernels. Then

$$\mathrm{var}_\pi(P_{MH}, h) \leq \mathrm{var}_\pi(P_B, h) \leq 2\,\mathrm{var}_\pi(P_{MH}, h) + \mathrm{Var}_\pi(h)$$

where $\mathrm{var}_\pi(P, h) = \lim_{n\to\infty} n\mathrm{Var}(\bar{X}_h)$ is the asymptotic variance when X_t is produced from P.

So although Barker's is more inefficient, it is not too much so.

Barker's for intractable posteriors

But Barker's still doesn't solve our problem since $\pi(\theta|y)$ still appears in the function:

$$\alpha_B(\theta, \theta^*) = \frac{\pi(\theta^*|y)\, q(\theta|\theta^*)}{\pi(\theta|y)\, q(\theta^*|\theta) + \pi(\theta^*|y)\, q(\theta|\theta^*)}$$

Barker's for intractable posteriors

But Barker's still doesn't solve our problem since $\pi(\theta|y)$ still appears in the function:

$$\alpha_B(\theta, \theta^*) = \frac{\pi(\theta^*|y)\, q(\theta|\theta^*)}{\pi(\theta|y)\, q(\theta^*|\theta) + \pi(\theta^*|y)\, q(\theta|\theta^*)}$$

To the rescue: Bernoulli factory!

Bernoulli factory

A Bernoulli factory is an algorithm that generates a Bernoulli event of probability $h(p)$ for some function $h(\cdot)$, using Bernoulli(p) events.

Bernoulli factory

A Bernoulli factory is an algorithm that generates a Bernoulli event of probability $h(p)$ for some function $h(\cdot)$, using Bernoulli(p) events.

Want events of prob. $\alpha_{\mathsf{B}}(\theta, \theta^*)$ without evaluating it.

Bernoulli factory

A Bernoulli factory is an algorithm that generates a Bernoulli event of probability $h(p)$ for some function $h(\cdot)$, using Bernoulli(p) events.

Want events of prob. $\alpha_{\mathrm{B}}(\theta, \theta^*)$ without evaluating it. Gonçalves et al. (2017) proposed the following. Suppose we can, find c_θ

$$\pi(\theta|y)q(\theta^*|\theta) \leq c_\theta \,.$$

Then set $\pi(\theta|y)q(\theta^*|\theta) = c_\theta p_\theta$ where

$$p_\theta = \frac{\pi(\theta|y)q(\theta^*|\theta)}{c_\theta} \leq 1,$$

Bernoulli factory

A Bernoulli factory is an algorithm that generates a Bernoulli event of probability $h(p)$ for some function $h(\cdot)$, using Bernoulli(p) events.

Want events of prob. $\alpha_{\text{B}}(\theta, \theta^*)$ without evaluating it. Gonçalves et al. (2017) proposed the following. Suppose we can, find c_θ

$$\pi(\theta|y)q(\theta^*|\theta) \leq c_\theta \,.$$

Then set $\pi(\theta|y)q(\theta^*|\theta) = c_\theta p_\theta$ where

$$p_\theta = \frac{\pi(\theta|y)q(\theta^*|\theta)}{c_\theta} \leq 1,$$

Then to generate events with probability

$$\frac{\pi(\theta^*|y)q(\theta^*|\theta)}{\pi(\theta|y)q(\theta|\theta^*) + \pi(\theta^*|y)q(\theta^*|\theta)} = \frac{c_{\theta^*} p_{\theta^*}}{c_\theta p_\theta + c_{\theta^*} p_{\theta^*}}$$

they propose a *two-coin* algorithm.

Two-coin algorithm

1. Draw $C_1 \sim \text{Bern}\left(\frac{c_{\theta^*}}{c_\theta + c_{\theta^*}}\right)$
2. If $C_1 = 1$, then
   2.1 Draw $C_2 \sim \text{Bern}(p_{\theta^*})$
   2.2 If $C_2 = 1$, then output 1
   2.3 If $C_2 = 0$, then goto Step 1
3. If $C_1 = 0$, then
   3.1 Draw $C_2 \sim \text{Bern}(p_\theta)$
   3.2 If $C_2 = 1$, then output 0
   3.3 If $C_2 = 0$, then go to Step 1

The above algorithm outputs 1 w.p. $\alpha_B(\theta, \theta^*)$.

Two-coin algorithm

1. Draw $C_1 \sim \text{Bern}\left(\frac{c_{\theta^*}}{c_\theta + c_{\theta^*}}\right)$
2. If $C_1 = 1$, then
   2.1 Draw $C_2 \sim \text{Bern}(p_{\theta^*})$
   2.2 If $C_2 = 1$, then output 1
   2.3 If $C_2 = 0$, then goto Step 1
3. If $C_1 = 0$, then
   3.1 Draw $C_2 \sim \text{Bern}(p_\theta)$
   3.2 If $C_2 = 1$, then output 0
   3.3 If $C_2 = 0$, then go to Step 1

The above algorithm outputs 1 w.p. $\alpha_B(\theta, \theta^*)$. But how do we sample $\text{Bern}(p_\theta)$?

Two-coin algorithm

To sample Bern(p_θ), note that

$$p_\theta = \frac{\pi(\theta|y)q(\theta^*|\theta)}{c_\theta} = \frac{q(\theta^*|\theta)\int g(\theta,\eta|y)d\eta}{c_\theta}$$

Suppose support of η is bounded.

Two-coin algorithm

To sample Bern(p_θ), note that

$$p_\theta = \frac{\pi(\theta|y)q(\theta^*|\theta)}{c_\theta} = \frac{q(\theta^*|\theta)\int g(\theta,\eta|y)d\eta}{c_\theta}$$

Suppose support of η is bounded. Then draw $N \sim$ Uniform within support of η and set

$$M_\theta = \frac{q(\theta^*|\theta)g(\theta^*,N|y)}{c_\theta} \quad \text{and } \mathrm{E}\left(M_\theta\right) = p_\theta\,.$$

So if $C_2 \sim \text{Bern}(M_\theta)$, then

$$\Pr(C_2 = 1) = \mathrm{E}\left(\mathbb{I}(C_2 = 1)\right) = \mathrm{E}\left(\mathrm{E}\left(\mathbb{I}(C_2 = 1)|M_\theta\right)\right) = p_\theta\,.$$

So $C_2 \sim \text{Bern}(p_\theta)$

Two-coin algorithm

1. Draw $C_1 \sim \text{Bern}\left(\frac{c_{\theta^*}}{c_\theta + c_{\theta^*}}\right)$
2. If $C_1 = 1$, then
   2.1 Draw $C_2 \sim \text{Bern}(p_{\theta^*})$
   2.2 If $C_2 = 1$, then output 1
   2.3 If $C_2 = 0$, then goto Step 1
3. If $C_1 = 0$, then
   3.1 Draw $C_2 \sim \text{Bern}(p_\theta)$
   3.2 If $C_2 = 1$, then output 0
   3.3 If $C_2 = 0$, then go to Step 1

Algorithm restarts often if p_θ or p_{θ^*} are small. That is, if we propose unlikely values in the Barker's algorithm, algorithm gets stuck in a loop.

Portkey Barker's algorithm

Here, the acceptance probability need not be the ratio of the target densities.

Portkey Barker's algorithm

Here, the acceptance probability need not be the ratio of the target densities. We propose the *Portkey Barker's* algorithm for $d(\theta, \theta^*) \geq 0$.

$$\alpha_P(\theta, \theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + d(\theta, \theta^*)}$$

Portkey Barker's algorithm

Here, the acceptance probability need not be the ratio of the target densities. We propose the *Portkey Barker's* algorithm for $d(\theta, \theta^*) \geq 0$.

$$\alpha_P(\theta, \theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + d(\theta, \theta^*)}$$

Theorem

If $d(\theta, \theta^*) = d(\theta^*, \theta)$, then $\alpha_P(\theta, \theta^*)$ yields a π-invariant Markov chain.

Portkey Barker's algorithm

Here, the acceptance probability need not be the ratio of the target densities. We propose the *Portkey Barker's* algorithm for $d(\theta, \theta^*) \geq 0$.

$$\alpha_P(\theta, \theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + d(\theta, \theta^*)}$$

Theorem

If $d(\theta, \theta^*) = d(\theta^*, \theta)$, then $\alpha_P(\theta, \theta^*)$ yields a π-invariant Markov chain.

We consider, for $\beta > 0$,

$$\alpha_\beta(\theta, \theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + \frac{1-\beta}{\beta}(c_{\theta^*} + c_\theta)}$$

$\beta = 1$ is Barker's.

Portkey Barker's algorithm

$$\alpha_\beta(\theta,\theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + \frac{1-\beta}{\beta}(c_{\theta^*} + c_\theta)}$$

Ideally, choose $\beta \approx 1$ so as to remain close to the Barker's algorithm. Because:

Portkey Barker's algorithm

$$\alpha_\beta(\theta,\theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + \frac{1-\beta}{\beta}(c_{\theta^*} + c_\theta)}$$

Ideally, choose $\beta \approx 1$ so as to remain close to the Barker's algorithm. Because:

Theorem

For $\beta > 0$,

$$var_\pi(h, P_B) \leq \beta\, var_\pi(h, P_\beta) + (\beta - 1) Var_\pi(h)\,.$$

Portkey Barker's algorithm

$$\alpha_\beta(\theta, \theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + \frac{1-\beta}{\beta}(c_{\theta^*} + c_\theta)}$$

Ideally, choose $\beta \approx 1$ so as to remain close to the Barker's algorithm. Because:

Theorem

For $\beta > 0$,

$$var_\pi(h, P_B) \leq \beta\, var_\pi(h, P_\beta) + (\beta - 1) Var_\pi(h)\,.$$

and if there exists $\gamma > 0$ such that $p_{\theta^*} > \gamma$ and $p_\theta > \gamma$, then

$$var_\pi(h, P_\beta) \leq \left(1 + \frac{1-\beta}{\gamma\beta}\right) var_\pi(h, P_B) + \frac{1-\beta}{\gamma\beta} Var_\pi(h)\,.$$

Then why use Portkey Barker's?

Portkey Two-coin algorithm

1. Draw $S \sim \text{Bern}(\beta)$ (S is the portkey[2])
2. If $S = 0$, output 0.
3. If $S = 1$,
   - 3.1 Draw $C_1 \sim \text{Bern}\left(\frac{c_{\theta^*}}{c_\theta + c_{\theta^*}}\right)$
   - 3.2 If $C_1 = 1$, then
     - 3.2.1 Draw $C_2 \sim \text{Bern}(p_{\theta^*})$
     - 3.2.2 If $C_2 = 1$, then output 1
     - 3.2.3 If $C_2 = 0$, then goto Step 1
   - 3.3 If $C_1 = 0$, then
     - 3.3.1 Draw $C_2 \sim \text{Bern}(p_\theta)$
     - 3.3.2 If $C_2 = 1$, then output 0
     - 3.3.3 If $C_2 = 0$, then go to Step 1

[2] Yes, this is a Harry Potter reference

Flipped Portkey's

Notice that if we divide Portkey Barker's throughout by

$$\pi(\theta^*|y)q(\theta|\theta^*)\ \pi(\theta|y)q(\theta^*|\theta)$$

then,

$$\alpha_P(\theta,\theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + d(\theta,\theta^*)}$$

$$= \frac{(\pi(\theta|y)q(\theta^*|\theta))^{-1}}{(\pi(\theta|y)q(\theta^*|\theta))^{-1} + (\pi(\theta^*|y)q(\theta|\theta^*))^{-1} + d'(\theta,\theta^*)}$$

Flipped Portkey's

Notice that if we divide Portkey Barker's throughout by

$$\pi(\theta^*|y)q(\theta|\theta^*)\ \pi(\theta|y)q(\theta^*|\theta)$$

then,

$$\alpha_P(\theta,\theta^*) = \frac{\pi(\theta^*|y)q(\theta|\theta^*)}{\pi(\theta|y)q(\theta^*|\theta) + \pi(\theta^*|y)q(\theta|\theta^*) + d(\theta,\theta^*)}$$

$$= \frac{(\pi(\theta|y)q(\theta^*|\theta))^{-1}}{(\pi(\theta|y)q(\theta^*|\theta))^{-1} + (\pi(\theta^*|y)q(\theta|\theta^*))^{-1} + d'(\theta,\theta^*)}$$

So if we can *lower bound* $\pi(\theta|y)q(\theta^*|\theta)$, we can implement a similar Portkey two-coin algorithm.

Application: Bayesian Correlation Estimation

Suppose

$$y_1, \ldots, y_n | R \stackrel{iid}{\sim} N(0, R)$$

where R is a $p \times p$ correlation matrix.

Application: Bayesian Correlation Estimation

Suppose

$$y_1, \ldots, y_n | R \overset{iid}{\sim} N(0, R)$$

where R is a $p \times p$ correlation matrix.

Let S_p^+ be the set of $p \times p$ correlation matrices. Liechty et al. (2009) set priors:

$$f(R \mid \mu, \sigma^2) = L(\mu, \sigma^2) \prod_{i<j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(r_{ij} - \mu)^2}{2\sigma^2} \right\} \mathbb{I}\{R \in S_p^+\}, \text{ where}$$

$$L^{-1}(\mu, \sigma^2) = \int_{R \in S_p^+} \prod_{i<j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(r_{ij} - \mu)^2}{2\sigma^2} \right\} dr_{ij}$$

Further, $\mu \sim N(0, \tau^2)$ and $\sigma^2 \sim IG(a_0, b_0)$ are chosen. Interest is in the posterior distribution for (R, μ, σ^2).

MCMC steps

Let $l = p(p-1)/2$. Implement a component-wise algorithm:

$$f(r_{ij} \mid r_{-ij}, \mu, \sigma^2) \propto |R|^{-n/2} \exp\left\{-\frac{\text{tr}(R^{-1}Y^TY)}{2}\right\} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} \mathbb{I}_{\{l_{ij} \le r_{ij} \le u_{ij}\}}$$

We use Metropolis-Hastings update with a Gaussian proposal for each r_{ij}.

MCMC steps

Let $l = p(p-1)/2$. Implement a component-wise algorithm:

$$f(r_{ij} \mid r_{-ij}, \mu, \sigma^2) \propto |R|^{-n/2} \exp\left\{-\frac{\mathrm{tr}(R^{-1}Y^TY)}{2}\right\} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} \mathbb{I}_{\{l_{ij} \le r_{ij} \le u_{ij}\}}$$

We use Metropolis-Hastings update with a Gaussian proposal for each r_{ij}.

$$f(\mu \mid R, \sigma^2) \propto L(\mu, \sigma^2) \prod_{i<j} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} \exp\left\{-\frac{\mu^2}{2\tau^2}\right\},$$

$$f(\sigma^2 \mid R, \mu) \propto L(\mu, \sigma^2) \prod_{i<j} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} \left(\frac{1}{\sigma^2}\right)^{a_0+l/2+1} \exp\left\{-\frac{b_0}{\sigma^2}\right\},$$

Running Metropolis steps for the conditional updates of μ and σ^2 is not possible.

Liechty et al. (2009) use an approximate inference shadow prior approach.

Application: Flipped Portkey Barker's

Let's focus on the μ update:

$$f(\mu \mid R, \sigma^2) \propto L(\mu, \sigma^2) \prod_{i<j} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} \exp\left\{-\frac{\mu^2}{2\tau^2}\right\}$$

Recall,

$$L^{-1}(\mu, \sigma^2) = \int_{R \in S_p^+} \prod_{i<j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} dr_{ij}$$

Application: Flipped Portkey Barker's

Let's focus on the μ update:

$$f(\mu \mid R, \sigma^2) \propto L(\mu, \sigma^2) \prod_{i<j} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} \exp\left\{-\frac{\mu^2}{2\tau^2}\right\}$$

Recall,

$$L^{-1}(\mu, \sigma^2) = \int_{R \in S_p^+} \prod_{i<j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} dr_{ij}$$

Obtaining an unbiased estimate of L^{-1} is simple since

Application: Flipped Portkey Barker's

Let's focus on the μ update:

$$f(\mu \mid R, \sigma^2) \propto L(\mu, \sigma^2) \prod_{i<j} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} \exp\left\{-\frac{\mu^2}{2\tau^2}\right\}$$

Recall,

$$L^{-1}(\mu, \sigma^2) = \int_{R \in S_p^+} \prod_{i<j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} dr_{ij}$$

Obtaining an unbiased estimate of L^{-1} is simple since

$$L^{-1}(\mu, \sigma^2) \leq \left[\Phi\left(\sigma^{-1}(1-\mu)\right) - \Phi\left(\sigma^{-1}(-1-\mu)\right)\right]^{I} := \tilde{c}_\mu$$

which gives us a lower bound for $f(\mu \mid R, \sigma^2)$. So we can use flipped portkey Barker's algorithm.

Application: Flipped Portkey Barker's

Let's focus on the μ update:

$$f(\mu \mid R, \sigma^2) \propto L(\mu, \sigma^2) \prod_{i<j} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} \exp\left\{-\frac{\mu^2}{2\tau^2}\right\}$$

Recall,

$$L^{-1}(\mu, \sigma^2) = \int_{R \in S_p^+} \prod_{i<j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(r_{ij}-\mu)^2}{2\sigma^2}\right\} dr_{ij}$$

Obtaining an unbiased estimate of L^{-1} is simple since

$$L^{-1}(\mu, \sigma^2) \leq \left[\Phi\left(\sigma^{-1}(1-\mu)\right) - \Phi\left(\sigma^{-1}(-1-\mu)\right)\right]^{I} := \tilde{c}_\mu$$

which gives us a lower bound for $f(\mu \mid R, \sigma^2)$. So we can use flipped portkey Barker's algorithm.

We study the correlation of the closing prices of the four major European stocks from 1991-1998.

Number of loops

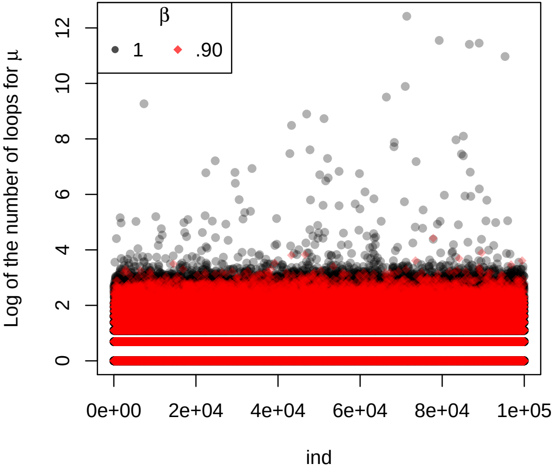

Figure: Log of the ratio of the Bernoulli factory loops for one run of length $1e5$.

Example: ACF plots

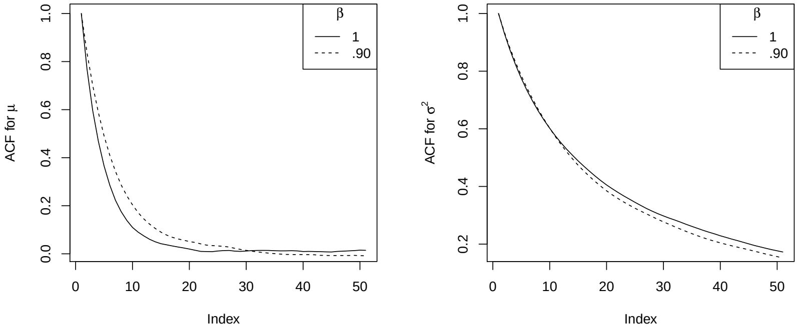

Figure: Autocorrelation plot for one run of length 1e5.

Example: Efficiency

Table: Averaged results from 10 replications of length $1e4$

β	1	.90
ESS	542 (13.50)	496 (9.00)
ESS/s	9.63 (1.992)	14.83 (0.279)
Mean loops μ	218.43 (148.89)	2.99 (0.010)
Mean loops σ^2	3.21 (0.02)	2.49 (0.010)
Max loops μ	2084195 (1491777)	34 (2.94)
Max loops σ^2	38 (1.13)	27 (1.51)

Example: Efficiency

Table: Averaged results from 10 replications of length 1e4

β	1	.90
ESS	542 (13.50)	496 (9.00)
ESS/s	9.63 (1.992)	14.83 (0.279)
Mean loops μ	218.43 (148.89)	2.99 (0.010)
Mean loops σ^2	3.21 (0.02)	2.49 (0.010)
Max loops μ	2084195 (1491777)	34 (2.94)
Max loops σ^2	38 (1.13)	27 (1.51)

Could only do 10 replications as $\beta = 1$ original would get stuck in large loops!

Attempting properly tuned proposal

Agrawal et al. (2023) showed that the optimal acceptance rate for Barker's acceptance function is 15.8%. If the proposal distribution is tuned to this:

Attempting properly tuned proposal

Agrawal et al. (2023) showed that the optimal acceptance rate for Barker's acceptance function is 15.8%. If the proposal distribution is tuned to this:

$\beta = .90$: 10^4 samples in 40s with estimated ESS $= 514$

Attempting properly tuned proposal

Agrawal et al. (2023) showed that the optimal acceptance rate for Barker's acceptance function is 15.8%. If the proposal distribution is tuned to this:

$\beta = .90$: 10^4 samples in 40s with estimated ESS $= 514$

$\beta = 1$: not even 10^3 samples in 24hrs and simulation was forcibly stopped.

Conclusion

Vats, D., Gonçalves, F., Łatusyński, K., Roberts, G. O.,
Efficient Bernoulli Factory MCMC for intractable posteriors, Biometrika, 2022

Advantages

- Markovian dynamics are mildly altered for $\beta \approx 1$
- Exact MCMC
- Significantly more robust

Disadvantages

- Loss of statistical efficiency from MH algorithms
- Finding the bounds c_θ may be challenging.

Conclusion

Vats, D., Gonçalves, F., Łatusyński, K., Roberts, G. O.,
Efficient Bernoulli Factory MCMC for intractable posteriors, Biometrika, 2022

Advantages

- Markovian dynamics are mildly altered for $\beta \approx 1$
- Exact MCMC
- Significantly more robust

Disadvantages

- Loss of statistical efficiency from MH algorithms
- Finding the bounds c_θ may be challenging.

Thank You!

Reference I

Agrawal, S., Vats, D., Łatuszyński, K., and Roberts, G. O. (2023). Optimal scaling of MCMC beyond Metropolis. *Advances in Applied Probability*, 55(2):492–509.

Barker, A. A. (1965). Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. *Australian Journal of Physics*, 18:119–134.

Gonçalves, F. B., Łatuszyński, K., Roberts, G. O., et al. (2017). Barker's algorithm for Bayesian inference with intractable likelihoods. *Brazilian Journal of Probability and Statistics*, 31:732–745.

Liechty, M. W., Liechty, J. C., and Müller, P. (2009). The shadow prior. *Journal of Computational and Graphical Statistics*, 18:368–383.

Peskun, P. (1973). Optimum Monte-Carlo sampling using Markov Chains. *Biometrika*, 89:745–754.