Efficient Implementation of MCMC When Using An Unbiased Likelihood Estimator

Arnaud Doucet
Department of Statistics, University of Oxford
& Oxford-Man Institute
Joint work with Mike Pitt (Warwick) & Robert Kohn (UNSW)

Assume one is interested in sampling from

\[ \pi(\theta) = \frac{\gamma(\theta)}{\int \gamma(\theta') d\theta'} \]

where \( \gamma : \Theta \rightarrow \mathbb{R}^+ \) is known pointwise but its integral is not.
Assume one is interested in sampling from

$$\pi(\theta) = \frac{\gamma(\theta)}{\int \gamma(\theta') d\theta'}$$

where $\gamma: \Theta \to \mathbb{R}^+$ is known pointwise but its integral is not.

Bayesian inference

$$\pi(\theta) = p(\theta | y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

where $\gamma(\theta) = p(y|\theta)p(\theta)$ is known pointwise but $p(y)$ is not.
Assume one is interested in sampling from

\[ \pi(\theta) = \frac{\gamma(\theta)}{\int \gamma(\theta') d\theta'} \]

where \( \gamma : \Theta \rightarrow \mathbb{R}^+ \) is known pointwise but its integral is not.

Bayesian inference

\[ \pi(\theta) = p(\theta | y) = \frac{p(y | \theta)p(\theta)}{p(y)} \]

where \( \gamma(\theta) = p(y | \theta)p(\theta) \) is known pointwise but \( p(y) \) is not.

MCMC have been used over 60 years to sample approximately from \( \pi(\theta) \).
Consider now the scenario where $p(y|\theta)$ cannot be evaluated pointwise.
Consider now the scenario where \( p(y|\theta) \) cannot be evaluated pointwise.

For latent variable models

\[
p(y|\theta) = \int p(x,y|\theta) \, dx
\]

where \( p(x,y|\theta) \) is often known pointwise but \( p(y|\theta) \) is not.
Consider now the scenario where $p(y|\theta)$ cannot be evaluated pointwise.

For latent variable models

\[ p(y|\theta) = \int p(x, y|\theta) \, dx \]

where $p(x, y|\theta)$ is often known pointwise but $p(y|\theta)$ is not.

Standard MCMC approaches consists of sampling from

\[ p(\theta, x|y) = \frac{p(x, y|\theta) \, p(\theta)}{p(y)} \]

by updating successively $x$ and $\theta$. 

An alternative approach based on the availability of a non-negative unbiased likelihood estimator is possible.
An alternative approach based on the availability of a non-negative unbiased likelihood estimator is possible.

An alternative approach based on the availability of a non-negative unbiased likelihood estimator is possible.


Particle Marginal Metropolis Hastings sampler for state-space models (Andrieu, D. & Holenstein, 2009, 2010).
Denote \( \hat{p}(y|\theta, u) \) the unbiased non-negative likelihood estimator function of the r.v. \( u \) of density \( m(u|\theta) \); i.e.

\[
p(y|\theta) = \int \hat{p}(y|\theta, u) m(u|\theta) \, du
\]
Denote $\hat{p}(y|\theta, u)$ the unbiased non-negative likelihood estimator function of the r.v. $u$ of density $m(u|\theta)$; i.e.

$$ p(y|\theta) = \int \hat{p}(y|\theta, u) m(u|\theta) \, du $$

Given $(\theta, \hat{p}(y|\theta, u))$ then sample $\theta' \sim q(\cdot|\theta)$, $u' \sim m(\cdot|\theta')$ and accept $(\theta', \hat{p}(y|\theta', u'))$ w.p.

$$ 1 \wedge \frac{\hat{p}(y|\theta', u') p(\theta') q(\theta|\theta')}{\hat{p}(y|\theta, u) p(\theta) q(\theta'|\theta)}. $$
MCMC with Intractable Likelihood Function

- Denote $\hat{p}(y|\theta, u)$ the unbiased non-negative likelihood estimator function of the r.v. $u$ of density $m(u|\theta)$; i.e.

$$p(y|\theta) = \int \hat{p}(y|\theta, u) m(u|\theta) \, du$$

- Given $(\theta, \hat{p}(y|\theta, u))$ then sample $\theta' \sim q(\cdot|\theta)$, $u' \sim m(\cdot|\theta')$ and accept $(\theta', \hat{p}(y|\theta', u'))$ w.p.

$$1 \wedge \frac{\hat{p}(y|\theta', u') p(\theta') q(\theta|\theta')}{\hat{p}(y|\theta, u) p(\theta) q(\theta'|\theta)}.$$ 

- The transition kernel of this algorithm admits $p(\theta|y)$ as invariant distribution whatever being the variance of $\hat{p}(y|\theta, u)$. 

This algorithm is a M-H sampler targetting
\[ \tilde{\pi}(\theta, u) \propto p(y|\theta, u) \, m(u|\theta) \, p(\theta) \]
using the proposal
\[ q(\theta'|\theta) \, m(u'|\theta') . \]
**MCMC with Intractable Likelihood Function**

- This algorithm is a M-H sampler targetting
  \[ \hat{\pi}(\theta, u) \propto p(y|\theta, u) \, m(u|\theta) \, p(\theta) \]
  using the proposal
  \[ q(\theta'|\theta) \, m(u'|\theta') \].

- Indeed, we have
  \[
  \frac{\hat{\pi}(\theta', u') \, q(\theta|\theta') \, m(u|\theta)}{\hat{\pi}(\theta, u) \, q(\theta'|\theta) \, m(u'|\theta')}
  = \frac{\hat{p}(y|\theta', u') \, m(u'|\theta') \, p(\theta') \, q(\theta|\theta') \, m(u|\theta)}{\hat{p}(y|\theta, u) \, m(u|\theta) \, p(\theta) \, q(\theta'|\theta) \, m(u'|\theta')}
  = \frac{\hat{p}(y|\theta', u') \, p(\theta') \, q(\theta|\theta')}{\hat{p}(y|\theta, u) \, p(\theta) \, q(\theta'|\theta)}.\]
This algorithm is a M-H sampler targeting
\[ \hat{\pi} (\theta, u) \propto p (y|\theta, u) m (u|\theta) p (\theta) \]
using the proposal
\[ q (\theta'|\theta) m (u'|\theta') . \]

Indeed, we have
\[
\frac{\hat{\pi} (\theta', u')}{\hat{\pi} (\theta, u)} = \frac{q (\theta|\theta') m (u|\theta)}{q (\theta'|\theta) m (u'|\theta')}
\]
\[
= \frac{\hat{p} (y|\theta', u') m (u'|\theta') p (\theta')}{\hat{p} (y|\theta, u) m (u|\theta) p (\theta)} \frac{q (\theta|\theta') m (u|\theta)}{q (\theta'|\theta) m (u'|\theta')}
\]
\[
= \frac{\hat{p} (y|\theta', u') p (\theta')}{\hat{p} (y|\theta, u) p (\theta)} \frac{q (\theta|\theta')}{q (\theta'|\theta)} .
\]

Crucially unbiasedness provides
\[ \hat{\pi} (\theta) = p (\theta|y) . \]
Assume one has

\[ p(y|\theta) = \int p(x,y|\theta) \, dx. \]
Importance Sampling Estimator

- Assume one has
  \[ p(y|\theta) = \int p(x, y|\theta) \, dx. \]

- Let \( g(x|y, \theta) \) be an Importance Sampling (IS) density then
  \[
  \hat{p}(y|\theta, u) = \frac{1}{N} \sum_{k=1}^{N} \omega(x^k, \theta),
  \]
  where the \( x^k \) are iid samples from \( g(x|y; \theta) \), \( u \) is the vector of r.v. used to generate the \( x^k \) and
  \[
  \omega(x, \theta) = \frac{p(x, y|\theta)}{g(x|y; \theta)}.
  \]
**Importance Sampling Estimator**

- Assume one has

\[ p(y|\theta) = \int p(x, y|\theta) \, dx. \]

- Let \( g(x|y, \theta) \) be an Importance Sampling (IS) density then

\[ \hat{p}(y|\theta, u) = \frac{1}{N} \sum_{k=1}^{N} \omega(x^k, \theta), \]

where the \( x^k \) are iid samples from \( g(x|y; \theta) \), \( u \) is the vector of r.v. used to generate the \( x^k \) and

\[ \omega(x, \theta) = \frac{p(x, y|\theta)}{g(x|y; \theta)}. \]

- \( \hat{p}(y|\theta, u) \) is unbiased of variance inversely proportional to \( N \).
Sequential Monte Carlo Estimator

- Suppose that $\{x_t\}_{t \geq 0}$ is a latent Markov process of initial density $\mu(x_0 | \theta)$ and transition density $f(x_{t+1} | x_t, \theta)$. We have access to conditionally independent observations $\{y_t\}_{t \geq 1}$ arising from $g(y_t | x_t, \theta)$ so that the likelihood of the obs. $y$ is given by

$$p(y | \theta) = \int p(y | x, \theta) p(x | \theta) dx$$

where denoting $x = x_0: T = (x'_0, ..., x'_T)'$

$$p(x | \theta) = \mu(x_0 | \theta) \prod_{t=1}^{T} f(x_t | x_{t-1}, \theta),$$

$$p(y | x, \theta) = \prod_{t=1}^{T} g(y_t | x_t, \theta).$$

SMC can provide an unbiased estimator which has a relative variance $O(T/N)$ (Del Moral et al., 2010, Whiteley 2012).
Suppose that \( \{x_t\}_{t \geq 0} \) is a latent Markov process of initial density \( \mu(x_0|\theta) \) and transition density \( f(x_{t+1}|x_t,\theta) \). We have access to conditionally independent observations \( \{y_t\}_{t \geq 1} \) arising from \( g(y_t|x_t,\theta) \) so that the likelihood of the obs. \( y \) is given by

\[
p(y|\theta) = \int p(y|x,\theta) p(x|\theta) dx
\]

where denoting \( x = x_0:T = (x'_0, ..., x'_T) \)'

\[
p(x|\theta) = \mu(x_0|\theta) \prod_{t=1}^{T} f(x_t|x_{t-1},\theta),
\]

\[
p(y|x,\theta) = \prod_{t=1}^{T} g(y_t|x_t,\theta).
\]

SMC can provide an unbiased estimator which has a relative variance in \( O(T/N) \) (Del Moral et al., 2010, Whiteley 2012).
Two species $X^1_t$ (prey) and $X^2_t$ (predator)

$$
\Pr \left( X^1_{t+dt} = x^1_{t} + 1, X^2_{t+dt} = x^2_{t} \mid x^1_{t}, x^2_{t} \right) = \alpha x^1_{t} dt + o(dt),
$$
$$
\Pr \left( X^1_{t+dt} = x^1_{t} - 1, X^2_{t+dt} = x^2_{t} + 1 \mid x^1_{t}, x^2_{t} \right) = \beta x^1_{t} x^2_{t} dt + o(dt),
$$
$$
\Pr \left( X^1_{t+dt} = x^1_{t}, X^2_{t+dt} = x^2_{t} - 1 \mid x^1_{t}, x^2_{t} \right) = \gamma x^2_{t} dt + o(dt),
$$

observed at discrete times

$$
Y_n = X^1_{n\Delta} + W_n \text{ with } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2).
$$
Two species $X_t^1$ (prey) and $X_t^2$ (predator)

$$\Pr \left( X_{t+dt}^1 = x_t^1 + 1, X_{t+dt}^2 = x_t^2 \bigg| x_t^1, x_t^2 \right) = \alpha x_t^1 dt + o \left( dt \right),$$

$$\Pr \left( X_{t+dt}^1 = x_t^1 - 1, X_{t+dt}^2 = x_t^2 + 1 \bigg| x_t^1, x_t^2 \right) = \beta x_t^1 x_t^2 dt + o \left( dt \right),$$

$$\Pr \left( X_{t+dt}^1 = x_t^1, X_{t+dt}^2 = x_t^2 - 1 \bigg| x_t^1, x_t^2 \right) = \gamma x_t^2 dt + o \left( dt \right),$$

observed at discrete times

$$Y_n = X_{n\Delta}^1 + W_n \text{ with } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N} \left( 0, \sigma^2 \right).$$

We are interested in the kinetic rate constants $\theta = (\alpha, \beta, \gamma)$ a priori distributed as (Boys et al., 2008)

$$\alpha \sim \mathcal{G} \left( 1, 10 \right), \quad \beta \sim \mathcal{G} \left( 1, 0.25 \right), \quad \gamma \sim \mathcal{G} \left( 1, 7.5 \right).$$
Two species $X^1_t$ (prey) and $X^2_t$ (predator)

\[
\begin{align*}
\Pr \left( X^1_{t+dt} = & x^1_t + 1, X^2_{t+dt} = x^2_t \mid x^1_t, x^2_t \right) = \alpha x^1_t dt + o(dt), \\
\Pr \left( X^1_{t+dt} = & x^1_t - 1, X^2_{t+dt} = x^2_t + 1 \mid x^1_t, x^2_t \right) = \beta x^1_t x^2_t dt + o(dt), \\
\Pr \left( X^1_{t+dt} = & x^1_t, X^2_{t+dt} = x^2_t - 1 \mid x^1_t, x^2_t \right) = \gamma x^2_t dt + o(dt),
\end{align*}
\]

observed at discrete times

\[
Y_n = X^1_{n\Delta} + W_n \text{ with } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2).
\]

We are interested in the kinetic rate constants $\theta = (\alpha, \beta, \gamma)$ a priori distributed as (Boys et al., 2008)

\[
\alpha \sim \mathcal{G}(1, 10), \quad \beta \sim \mathcal{G}(1, 0.25), \quad \gamma \sim \mathcal{G}(1, 7.5).
\]

We sample from $p(\theta \mid y_{1:T})$ using a Metropolis-Hastings sampler where $p(y_{1:T} \mid \theta)$ is evaluated numerically using a particle filter.
Autocorrelation of $\alpha$ (left) and $\beta$ (right) of the MH sampler for various $N$. 
A key issue from a practical point of view is how one to select $N$?
A key issue from a practical point of view is how one to select $N$?
- If $N$ is too small, then the algorithm mixes poorly and will require many MCMC iterations.
A key issue from a practical point of view is how one to select $N$?

- If $N$ is too small, then the algorithm mixes poorly and will require many MCMC iterations.
- If $N$ is too large, then each MCMC iteration is expensive.
A key issue from a practical point of view is how one to select $N$?

- If $N$ is too small, then the algorithm mixes poorly and will require many MCMC iterations.
- If $N$ is too large, then each MCMC iteration is expensive.

**Aim:** We would like to provide guidelines on how to select $N$. 
Let $z = \log \hat{p}_N (y|\theta, u) - \log p(y|\theta)$ be the error in log-likelihood.
Let $z = \log \hat{p}_N (y|\theta, u) - \log p(y|\theta)$ be the error in log-likelihood.

The proposal from which $z$ arises is denoted $g_N (z|\theta)$. 
MCMC with Intractable Likelihood Function

- Let $z = \log \hat{p}_N (y|\theta, u) - \log p(y|\theta)$ be the error in log-likelihood.
- The proposal from which $z$ arises is denoted $g_N (z|\theta)$.
- We can rewrite the extended target

$$\hat{\pi}_N (\theta, z) = \pi(\theta) \exp(z) g_N (z|\theta)$$

which is directly related to $\hat{\pi}_N (\theta, u)$ through the many-to-one transformation from $u$ to $z$. 
MCMC with Intractable Likelihood Function

- Let $z = \log \hat{p}_N(y|\theta, u) - \log p(y|\theta)$ be the error in log-likelihood.
- The proposal from which $z$ arises is denoted $g_N(z|\theta)$.
- We can rewrite the extended target

$$\hat{\pi}_N(\theta, z) = \pi(\theta) \exp(z) g_N(z|\theta)$$

which is directly related to $\hat{\pi}_N(\theta, u)$ through the many-to-one transformation from $u$ to $z$.
- The previous algorithm proposes $\theta' \sim q(\cdot|\theta)$ and $z' \sim g_N(z|\theta')$, accepting $(\theta', z')$ w.p.

$$\alpha_Q(\theta, z; \theta', z') = \min \left\{ 1, \exp(z' - z) \omega(\theta'; \theta)/\omega(\theta; \theta') \right\},$$

where $\omega(\theta'; \theta) = \pi(\theta')/q(\theta'|\theta)$. 

Inefficiency Measure

Consider a stationary Markov chain \( \{\theta_j\} \) with invariant density \( \pi(\theta) \) and \( h : \Theta \rightarrow \mathbb{R} \) with \( \mathbb{V}_\pi[h(\theta)] < \infty \). Define

\[
\mu_h = \mathbb{E}_\pi[h(\theta)] \quad \text{and} \quad \hat{\mu}_{h,n} = n^{-1} \sum_{j=1}^{n} h(\theta_j).
\]

Then the IACT of the Markov chain \( IF_h \) is given by

\[
\lim_{n \rightarrow \infty} n \mathbb{V}_\pi(\hat{\mu}_{h,n}) = \mathbb{V}_\pi[h(\theta)] \quad IF_h \quad \text{with} \quad IF_h = 1 + 2 \sum_{\tau=1}^{\infty} \rho_h(\tau),
\]

where \( \rho_h(\tau) \) is the autocorrelation at lag \( \tau \) of the stationary sequence \( \{h(\theta_j)\} \).
Consider a stationary Markov chain \( \{\theta_j\} \) with invariant density \( \pi(\theta) \) and \( h : \Theta \to \mathbb{R} \) with \( \mathbb{V}_\pi [h(\theta)] < \infty \). Define

\[
\mu_h = \mathbb{E}_\pi [h(\theta)] \quad \text{and} \quad \hat{\mu}_{h,n} = n^{-1} \sum_{j=1}^{n} h(\theta_j).
\]

Then the IACT of the Markov chain \( IF_h \) is given by

\[
\lim_{n \to \infty} n \mathbb{V}_\pi (\hat{\mu}_{h,n}) = \mathbb{V}_\pi [h(\theta)] \quad IF_h \quad \text{with} \quad IF_h = 1 + 2 \sum_{\tau=1}^{\infty} \rho_h(\tau),
\]

where \( \rho_h(\tau) \) is the autocorrelation at lag \( \tau \) of the stationary sequence \( \{h(\theta_j)\} \).

The IACT, \( IF_h \), quantifies how many times more samples are required from the Markov chain relative to using iid samples from \( \pi(\theta) \) to achieve a given precision.
Proposition (Kipnis & Varadhan): Consider a stationary, ergodic, reversible Markov chain \( \{\theta_j\} \) with invariant density \( \pi(\theta) \). Suppose that \( h(\theta) \) is a square integrable function of \( \theta \) with respect to \( \pi(\theta) \), with \( \nabla_\pi [h(\theta)] < \infty \) then

\[
\rho_h(\tau) = \mathbb{E}_\pi \left[ e^{\mathbb{E}_\theta [h(\theta)]} \right],
\]

where \( e^{\mathbb{E}_\theta [h(\theta)]} \) is a positive measure on \( (1, 1) \).
Proposition (Kipnis & Varadhan): Consider a stationary, ergodic, reversible Markov chain \( \{\theta_j\} \) with invariant density \( \pi(\theta) \). Suppose that \( h(\theta) \) is a square integrable function of \( \theta \) with respect to \( \pi(\theta) \), with \( \mathbb{V}_\pi [h(\theta)] < \infty \) then

The autocorrelation and IACT are given by

\[
\rho_h(\tau) = \int_{-1}^{1} |\tau| \tilde{E}_h(d\lambda), \quad IF_h = 1 + 2 \sum_{\tau=1}^{\infty} \rho_h(\tau) = \int_{-1}^{1} \frac{1 + \lambda}{1 - \lambda} \tilde{E}_h(d\lambda),
\]

where \( \tilde{E}_h \) is a positive measure on \((-1, 1)\).
**Proposition** (Kipnis & Varadhan): Consider a stationary, ergodic, reversible Markov chain \( \{ \theta_j \} \) with invariant density \( \pi(\theta) \). Suppose that \( h(\theta) \) is a square integrable function of \( \theta \) with respect to \( \pi(\theta) \), with \( \mathbb{V}_\pi [h(\theta)] < \infty \) then

1. The autocorrelation and IACT are given by

\[
\rho_h(\tau) = \int_{-1}^{1} \lambda^{\mid\tau\mid} \tilde{E}_h(d\lambda), \quad IF_h = 1 + 2 \sum_{\tau=1}^{\infty} \rho_h(\tau) = \int_{-1}^{1} \frac{1 + \lambda}{1 - \lambda} \tilde{E}_h(d\lambda),
\]

where \( \tilde{E}_h \) is a positive measure on \((-1, 1)\).

2. If \( IF_h < \infty \), then \( \sqrt{n}(\hat{\mu}_{h,n} - \mu_h) \longrightarrow \mathcal{N}(0; \mathbb{V}_\pi [h(\theta)] IF_h) \).
For the sake of analysis, we introduce an alternative $Q^*$ chain.
A Bounding Chain

- For the sake of analysis, we introduce an alternative $Q^*$ chain.
- Let $(\theta, z)$ be the current state of the Markov chain.
For the sake of analysis, we introduce an alternative $Q^*$ chain.

Let $(\theta, z)$ be the current state of the Markov chain.

1. Propose $\theta' \sim q(\cdot | \theta)$ and $z' \sim g_N(\cdot | \theta')$. 
A Bounding Chain

- For the sake of analysis, we introduce an alternative $Q^*$ chain.
- Let $(\theta, z)$ be the current state of the Markov chain.

1. Propose $\theta' \sim q(\cdot | \theta)$ and $z' \sim g_N(\cdot | \theta')$.
2. Accept $\theta'$ w.p. $\alpha_{Q^{\text{EX}}}(\theta; \theta') = \min \{1, \omega(\theta'; \theta) / \omega(\theta; \theta')\}$. 
For the sake of analysis, we introduce an alternative $Q^*$ chain.

Let $(\theta, z)$ be the current state of the Markov chain.

1. Propose $\theta' \sim q(\cdot|\theta)$ and $z' \sim g_N(\cdot|\theta')$.

2. Accept $\theta'$ w.p. $\alpha_{Q^{EX}}(\theta; \theta') = \min \{1, \omega(\theta'; \theta) / \omega(\theta; \theta')\}$.

3. Accept $z'$ w.p. $\alpha_{Q^z}(z; z') = \min \{1, \exp(z' - z)\}$.
For the sake of analysis, we introduce an alternative $Q^*$ chain.

Let $(\theta, z)$ be the current state of the Markov chain.

1. Propose $\theta' \sim q(\cdot | \theta)$ and $z' \sim g_N(\cdot | \theta')$.
2. Accept $\theta'$ w.p. $\alpha_{Q^{EX}}(\theta; \theta') = \min \{1, \omega(\theta'; \theta) / \omega(\theta; \theta')\}$.
3. Accept $z'$ w.p. $\alpha_{Q^z}(z; z') = \min \{1, \exp(z' - z)\}$.
4. $(\theta', z')$ is accepted if and only if there is acceptance in both criteria.
Lemma: The Markov chain $Q^*$ has the following properties:

1. We have $\alpha_{Q}(\theta, z; \theta_0, z_0) = \alpha_{Q}(\theta, z; \theta_0, z_0) = \alpha_{Q}(\theta; \theta_0)$.

2. $Q$ is a reversible Markov chain with invariant density $\pi N(\theta, z)$.

3. For any function $h(\theta)$ the IACT is higher for the $Q$ chain than for the $Q$ chain; i.e. $IF_{Q} h > IF_{Q} h$ (Peskun, 1973; Tierney 1998).

Remark: We have $\alpha_{Q}(\theta, z; \theta_0, z_0) = \alpha_{Q}(\theta, z; \theta_0, z_0) = \alpha_{Q}(\theta; \theta_0)$ when the likelihood is known exactly and $\alpha_{Q}(\theta, z; \theta_0, z_0) = \alpha_{Q}(\theta, z; \theta, z)$ when the proposal is perfect, i.e. $q(\theta_0|\theta) = \pi(\theta_0)$. 

Lemma: The Markov chain $Q^*$ has the following properties:

1. We have

$$\alpha_Q(\theta, z; \theta', z') \geq \alpha_{Q^*}(\theta, z; \theta', z') = \alpha_{Q^{\text{EX}}}(\theta; \theta') \times \alpha_{Q^z}(z; z').$$
Lemma: The Markov chain $Q^*$ has the following properties:

1. We have

$$\alpha_Q(\theta, z; \theta', z') \geq \alpha_{Q^*}(\theta, z; \theta', z') = \alpha_{Q^{\text{EX}}}(\theta; \theta') \times \alpha_{Q^z}(z; z').$$

2. $Q^*$ is a reversible Markov chain with invariant density $\hat{\pi}_N(\theta, z)$. 


Lemma: The Markov chain $Q^*$ has the following properties:

1. We have

$$\alpha_Q (\theta, z; \theta', z') \geq \alpha_{Q^*} (\theta, z; \theta', z') = \alpha_{Q^{EX}} (\theta; \theta') \times \alpha_{Q^z} (z; z').$$

2. $Q^*$ is a reversible Markov chain with invariant density $\hat{\pi}_N (\theta, z)$.

3. For any function $h(\theta)$ the IACT is higher for the $Q^*$ chain than for the $Q$ chain; i.e. $IF_{h}^{Q^*} \geq IF_{h}^{Q}$ (Peskun, 1973; Tierney 1998).
**Lemma**: The Markov chain $Q^*$ has the following properties:

1. We have
   \[
   \alpha_Q(\theta, z; \theta', z') \geq \alpha_{Q^*}(\theta, z; \theta', z') = \alpha_{Q^*}^{\text{EX}}(\theta; \theta') \times \alpha_{Q^*}(z; z').
   \]

2. $Q^*$ is a reversible Markov chain with invariant density $\hat{\pi}_N(\theta, z)$.

3. For any function $h(\theta)$ the IACT is higher for the $Q^*$ chain than for the $Q$ chain; i.e. $IF_{h^*}^{Q^*} \geq IF_{h}^{Q}$ (Peskun, 1973; Tierney 1998).

**Remark**: We have $\alpha_{Q^*}(\theta, z; \theta', z') = \alpha_Q(\theta, z; \theta', z') = \alpha_{Q^*}^{\text{EX}}(\theta; \theta')$ when the likelihood is known exactly and $\alpha_{Q^*}(\theta, z; \theta', z') = \alpha_Q(\theta, z; \theta^*, z^*) = \alpha_{Q^*}(z; z^*)$ when the proposal is perfect, i.e. $q(\theta'|\theta) = \pi(\theta')$. 

Assumption. Let \( z = \log \hat{p}_N(y|\theta, u) - \log p(y|\theta) \) be the error in the estimator of the log likelihood.
Assumption. Let $z = \log \hat{p}_N(y|\theta, u) - \log p(y|\theta)$ be the error in the estimator of the log likelihood.

We have

$$g_N(z|\theta) = \phi \left( z; -\gamma^2(\theta)/2N, \gamma^2(\theta)/N \right),$$

$$\pi_N(z|\theta) = \exp(z)g_N(z|\theta) = \phi \left( z; \gamma^2(\theta)/2N, \gamma^2(\theta)/N \right)$$

where $\phi(z; a, b^2)$ is a univariate normal of mean $a$, variance $b^2$. 

Making Assumptions to Move Forward

**Assumption.** Let \( z = \log \hat{p}_N(y|\theta, u) - \log p(y|\theta) \) be the error in the estimator of the log likelihood.

1. We have

\[
\begin{align*}
g_N(z|\theta) &= \phi\left(z; -\gamma^2(\theta)/2N, \gamma^2(\theta)/N\right), \\
\pi_N(z|\theta) &= \exp(z)g_N(z|\theta) = \phi\left(z; \gamma^2(\theta)/2N, \gamma^2(\theta)/N\right)
\end{align*}
\]

where \( \phi(z; a, b^2) \) is a univariate normal of mean \( a \), variance \( b^2 \).

2. For a given value of \( \sigma^2 \) we set \( N = N_{\sigma^2}(\theta) = \gamma(\theta)^2 / \sigma^2 \).
Making Assumptions to Move Forward

**Assumption.** Let \( z = \log \hat{p}_N(y|\theta, u) - \log p(y|\theta) \) be the error in the estimator of the log likelihood.

1. We have

\[
g_N(z|\theta) = \phi\left(z; -\gamma^2(\theta)/2N, \gamma^2(\theta)/N\right), \\
\pi_N(z|\theta) = \exp(z)g_N(z|\theta) = \phi\left(z; \gamma^2(\theta)/2N, \gamma^2(\theta)/N\right)
\]

where \( \phi(z; a, b^2) \) is a univariate normal of mean \( a \), variance \( b^2 \).

2. For a given value of \( \sigma^2 \) we set \( N = N_{\sigma^2}(\theta) = \gamma(\theta)^2 / \sigma^2 \).

Under this assumption, both \( g_N(z|\theta) \) and \( \pi_N(z|\theta) \) are functions of \( \sigma^2 \) only and we write \( g_N(z|\theta) \) and \( \pi_N(z|\theta) \) as

\[
g(z|\sigma^2) = \phi(z; -\sigma^2/2, \sigma^2), \\
\pi(z|\sigma^2) = \phi(z; \sigma^2/2, \sigma^2)
\]

and \( \theta \) and \( z \) are independent under \( \hat{\pi}_N(\theta, z) \).
Lemma (Pitt, Giordani et al., 2012). Under the previous assumption, the following results hold for the chain \( \{\theta_j, z_j\} \) arising from \( Q^* \) that uses the perfect proposal \( q(\theta'|\theta) = \pi(\theta') \) denoted \( Q^Z \).
Lemma (Pitt, Giordani et al., 2012). Under the previous assumption, the following results hold for the chain \( \{\theta_j, z_j\} \) arising from \( Q^* \) that uses the perfect proposal \( q(\theta'|\theta) = \pi(\theta') \) denoted \( Q^Z \).

Let \( p(z; \sigma^2) \) be the probability of rejection given the current value \( z \). Then

\[
p(z; \sigma^2) = 1 - \int a_{Q^Z} (z; z') g(z'|\sigma^2) dz' \\
= \Phi(z/\sigma + \sigma/2) - \exp(-z)\Phi(z/\sigma - \sigma/2).
\]
Lemma (Pitt, Giordani et al., 2012). Under the previous assumption, the following results hold for the chain \( \{\theta_j, z_j\} \) arising from \( Q^* \) that uses the perfect proposal \( q(\theta'|\theta) = \pi(\theta') \) denoted \( Q^Z \).

1. Let \( p(z; \sigma^2) \) be the probability of rejection given the current value \( z \). Then

\[
p(z; \sigma^2) = 1 - \int a_{Q^Z}(z; z') g(z'|\sigma^2) dz' \\
= \Phi(z/\sigma + \sigma/2) - \exp(-z)\Phi(z/\sigma - \sigma/2).
\]

2. \( IF^Z(\sigma^2) = \mathbb{E}_{\pi(\cdot|\sigma^2)} \left( \frac{1 + p(z; \sigma^2)}{1 - p(z; \sigma^2)} \right) = \int \frac{1 + p^*(w, \sigma)}{1 - p^*(w, \sigma)} \phi(w) dw, \)

where \( p^*(w, \sigma) = \Phi(w + \sigma) - \exp(-w\sigma - \sigma^2/2)\Phi(w), \Phi(\cdot) \) standard normal cdf. In particular, \( IF^Z(\sigma^2) \) is independent of \( h \).
Main Result

- Under the previous assumption and additional regularity conditions,

\[
IF^Q_h(\sigma^2) \leq IF^{Q*}_h(\sigma^2) \leq IF^U_h(\sigma^2),
\]

\[
IF^U_h(\sigma^2) = \frac{1}{2}(1 + IF^E_h)(1 + IF^Z_h(\sigma^2)) - 1.
\]
Main Result

- Under the previous assumption and additional regularity conditions,

\[
IF_Q^h(\sigma^2) \leq IF_Q^*(\sigma^2) \leq IF_U^h(\sigma^2),
\]

\[
IF_U^h(\sigma^2) = \frac{1}{2} (1 + IF_{EX}^h)(1 + IF_Z^*(\sigma^2)) - 1.
\]

- The inequality becomes exact when \(\sigma \to 0\) (as \(IF_Z^*(\sigma^2) \to 1\) for \(\sigma \to 0\)) and when the proposal is perfect (as \(IF_{EX}^h = 1\) when \(q(\theta' | \theta) = \pi(\theta')\)).
Main Result

- Under the previous assumption and additional regularity conditions,

\[ IF_h^Q(\sigma^2) \leq IF_h^{Q^*}(\sigma^2) \leq IF_h^U(\sigma^2), \]
\[ IF_h^U(\sigma^2) = \frac{1}{2} (1 + IF_h^{EX})(1 + IF_h^{Z}(\sigma^2)) - 1. \]

- The inequality becomes exact when \( \sigma \to 0 \) (as \( IF_h^{Z}(\sigma^2) \to 1 \) for \( \sigma \to 0 \)) and when the proposal is perfect (as \( IF_h^{EX} = 1 \) when \( q(\theta'|\theta) = \pi(\theta') \)).

- For the unconditional acceptance probability, we have

\[ P_Q^Q(A|\sigma^2) \geq P_Q^{Q^*}(A|\sigma^2) = 2\Phi(-\sigma/\sqrt{2}) \cdot P^{EX}(A). \]

with the bound getting tighter if either \( P^{EX}(A) \to 1 \) or \( P^{Z}(A|\sigma^2) \to 1 \).
We have

\[
\frac{IF_h^Q(\sigma^2)}{IF_h^{EX}} \leq \frac{IF_h^U(\sigma^2)}{IF_h^{EX}} = RIF_h^U(\sigma^2),
\]

\[
RIF_h^U(\sigma^2) = \frac{1}{2} \left( \frac{IF_h^Z(\sigma^2) - 1}{IF_h^{EX}} \right) + \frac{1}{2} (1 + IF_h^Z(\sigma^2)).
\]
Relative Inefficiency

- We have

\[
\frac{IF^Q_h(\sigma^2)}{IF^E_h} \leq \frac{IF^U_h(\sigma^2)}{IF^E_h} = RIF^U_h(\sigma^2),
\]

\[
RIF^U_h(\sigma^2) = \frac{1}{2} \frac{IF^Z_h(\sigma^2) - 1}{IF^E_h} + \frac{1}{2} (1 + IF^Z_h(\sigma^2)).
\]

- For fixed \( \sigma \), \( RIF^U_h(\sigma^2) \) decreases as \( IF^E_h \) increases from a value of \( IF^Z_h(\sigma^2) \) for \( IF^E_h = 1 \) to

\[
RIF^U_h(\sigma^2) \rightarrow \frac{1}{2} (1 + IF^Z_h(\sigma^2)) \leq IF^Z_h(\sigma^2) \quad \text{as} \quad IF^E_h \rightarrow \infty.
\]
Relative Inefficiency

- We have

\[
\frac{IF^Q_h(\sigma^2)}{IF^E_h} \leq \frac{IF^U_h(\sigma^2)}{IF^E_h} = RIF^U_h(\sigma^2),
\]

\[
RIF^U_h(\sigma^2) = \frac{1}{2} \left( \frac{IF^Z_h(\sigma^2) - 1}{IF^E_h} \right) + \frac{1}{2} \left( 1 + IF^Z_h(\sigma^2) \right).
\]

- For fixed \( \sigma \), \( RIF^U_h(\sigma^2) \) decreases as \( IF^E_h \) increases from a value of \( IF^Z_h(\sigma^2) \) for \( IF^E_h = 1 \) to

\[
RIF^U_h(\sigma^2) \to \frac{1}{2} (1 + IF^Z(\sigma^2)) \leq IF^Z(\sigma^2) \text{ as } IF^E_h \to \infty.
\]

- The loss in efficiency from using the estimated likelihood goes down as the proposal deteriorates.
The Computing Time (CT) for $Q$ is defined as

$$CT_h^Q(\sigma^2) = IF_h^Q(\sigma^2) / \sigma^2;$$

i.e. we take into account the computational efforts associated to $\sigma^2 \propto 1 / N$. 

Deﬁne the relative computing time for the ine¢ ciency bound $IF_U^h(\sigma^2)$ as

$$RCT_U^h(\sigma^2) = RIF_U^h(\sigma^2) / \sigma^2.$$ 

Both $RIF_U^h(\sigma^2)$ and $RCT_U^h(\sigma^2)$ are decreasing functions of $IF_EX^h$.
The Computing Time (CT) for Q is defined as

$$CT_h^Q(\sigma^2) = IF_h^Q(\sigma^2) / \sigma^2;$$

i.e. we take into account the computational efforts associated to $\sigma^2 \propto 1/N$.

$$CT_h^Q(\sigma^2) \leq CT_h^U(\sigma^2) \text{ where } CT_h^U(\sigma^2) := IF_h^U(\sigma^2) / \sigma^2.$$
The Computing Time (CT) for $Q$ is defined as

$$CT^Q_h(\sigma^2) = IF^Q_h(\sigma^2)/\sigma^2;$$

i.e. we take into account the computational efforts associated to $\sigma^2 \propto 1/N$.

1. $CT^Q_h(\sigma^2) \leq CT^U_h(\sigma^2)$ where $CT^U_h(\sigma^2) := IF^U_h(\sigma^2)/\sigma^2$.

2. If $IF^\text{EX}_h = 1$, then $CT^U_h(\sigma^2)$ is minimized at $\sigma^U_{opt} = 0.92$ and $IF^Z(\sigma^U_{opt}) = 4.54$, $P^Z(A|\sigma^U_{opt}) = 0.5153$. 
The Computing Time (CT) for $Q$ is defined as

$$CT^Q_h(\sigma^2) = IF^Q_h(\sigma^2) / \sigma^2;$$

i.e. we take into account the computational efforts associated to $\sigma^2 \propto 1/N$.

1. $CT^Q_h(\sigma^2) \leq CT^U_h(\sigma^2)$ where $CT^U_h(\sigma^2) := IF^U_h(\sigma^2) / \sigma^2$.

2. If $IF^E_h = 1$, then $CT^U_h(\sigma^2)$ is minimized at $\sigma^U_{opt} = 0.92$ and $IF^Z(\sigma^U_{opt}) = 4.54$, $P^Z(A|\sigma^U_{opt}) = 0.5153$.

3. The minimizing $\sigma^U_{opt}$ rises with $IF^E_h$ to $\sigma^U_{opt} = 1.0206$ as $IF^E_h \to \infty$. 

Computing Time

- The Computing Time (CT) for $Q$ is defined as
  \[ CT^Q_h(\sigma^2) = IF^Q_h(\sigma^2) / \sigma^2; \]
  i.e. we take into account the computational efforts associated to \( \sigma^2 \propto 1/N \).

1. \( CT^Q_h(\sigma^2) \leq CT^U_h(\sigma^2) \) where \( CT^U_h(\sigma^2) := IF^U_h(\sigma^2) / \sigma^2 \).
2. If \( IF^E_h = 1 \), then \( CT^U_h(\sigma^2) \) is minimized at \( \sigma^U_{\text{opt}} = 0.92 \) and \( IF^Z(\sigma^U_{\text{opt}}^2) = 4.54 \), \( P^Z(A | \sigma^U_{\text{opt}}) = 0.5153 \).
3. The minimizing \( \sigma^U_{\text{opt}} \) rises with \( IF^E_h \) to \( \sigma^U_{\text{opt}} = 1.0206 \) as \( IF^E_h \to \infty \).

- Define the relative computing time for the inefficiency bound \( IF^U_h(\sigma^2) \) as
  \[ RCT^U_h(\sigma^2) = \frac{RIF^U_h(\sigma^2)}{\sigma^2}. \]
  Both \( RIF^U_h(\sigma^2) \) and \( RCT^U_h(\sigma^2) \) are decreasing functions of \( IF^E_h \).
Figure: $RCT^U_h$ (top) and $RIF^U_h$ (bottom) against $1/\sigma^2$ (left) and $\sigma$ (right). Different values of $IF^EX_h$ are shown on each plot.
Example 1: Probit Model

- We use a simple Bernoulli Probit model, where for $t = 1, \ldots, T$

$$y_t = \mathbb{I}(x_t > 0), \ x_t \sim iid \ N(\theta; 1).$$
Example 1: Probit Model

- We use a simple Bernoulli Probit model, where for $t = 1, \ldots, T$

$$y_t = \mathbb{I}(x_t > 0), \ x_t \overset{iid}{\sim} \mathcal{N}(\theta; 1).$$

- The likelihood is known explicitly as $\Pr(y_t = 1) = \Phi(\theta)$ but is estimated through

$$\hat{\Pr}(y_t = 1) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{I}(x_t^{(k)} > 0), \ x_t^{(k)} \overset{iid}{\sim} \mathcal{N}(\theta; 1)$$

and set $\theta \sim \mathcal{N}(0, \sigma^2_{\theta})$ with $\sigma^2_{\theta} \gg 1$. 
Example 1: Probit Model

- We use a simple Bernoulli Probit model, where for \( t = 1, ..., T \)
  \[ y_t = \mathbb{I}(x_t > 0), \quad x_t \sim \mathcal{N}(\theta; 1). \]

- The likelihood is known explicitly as \( \Pr(y_t = 1) = \Phi(\theta) \) but is estimated through
  \[ \hat{\Pr}(y_t = 1) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{I}(x^{(k)}_t > 0), \quad x^{(k)}_t \sim \mathcal{N}(\theta; 1) \]

  and set \( \theta \sim \mathcal{N}(0, \sigma^2_{\theta}) \) with \( \sigma^2_{\theta} \gg 1. \)

- Autoregressive Metropolis proposal
  \[ \theta' = \hat{\theta} + \rho(\theta - \hat{\theta}) + \sqrt{\frac{\sigma^2(1 - \rho^2)}{\nu - 2}} t_5, \]

  where \( \hat{\theta} \) is the posterior mode and \( \hat{\sigma}^2 \) is chosen as the negative inverse of the second derivative of the log posterior.
Example 1: Probit Model

- We use a simple Bernoulli Probit model, where for $t = 1, \ldots, T$
  \[
y_t = \mathbb{I}(x_t > 0), \quad x_t \sim \mathcal{N}(\theta; 1).
\]

- The likelihood is known explicitly as $\Pr(y_t = 1) = \Phi(\theta)$ but is estimated through
  \[
  \hat{\Pr}(y_t = 1) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{I}(x_t^{(k)} > 0), \quad x_t^{(k)} \sim \mathcal{N}(\theta; 1)
  \]
  and set $\theta \sim \mathcal{N}(0, \sigma^2_\theta)$ with $\sigma^2_\theta \gg 1$.

- Autoregressive Metropolis proposal
  \[
  \theta' = \hat{\theta} + \rho(\theta - \hat{\theta}) + \sqrt{\frac{\hat{\sigma}^2(1 - \rho^2)}{\nu - 2}} t_5,
  \]
  where $\hat{\theta}$ is the posterior mode and $\hat{\sigma}^2$ is chosen as the negative inverse of the second derivative of the log posterior.

- We will consider $\rho \in \{0, 0.4, 0.6, 0.9, 0.97\}$. 
Setting the Number of Samples

- We want to assess experimentally whether our upper/lower bounds are sharp.

1. Choose a large initial value for the number of samples, $N_S$.
2. Run the MCMC scheme for a fixed number of iterates recording $\theta$.
3. Record the estimated variance of the log of the likelihood estimator, $V(\theta, N_S) = b \frac{V}{\sigma^2}$.
4. Set $N_\theta = \frac{V(\theta, N_S)}{\sigma^2}$. 

We want to assess experimentally whether our upper/lower bounds are sharp.
We select $N$ so as $\sigma$ to be roughly equal to a pre-specified value.
Setting the Number of Samples

- We want to assess experimentally whether our upper/lower bounds are sharp.
- We select $N$ so as $\sigma$ to be roughly equal to a pre-specified value.

1. Choose a large initial value for the number of samples, $N_S$. 
We want to assess experimentally whether our upper/lower bounds are sharp.

We select $N$ so as $\sigma$ to be roughly equal to a pre-specified value.

1. Choose a large initial value for the number of samples, $N_S$.
2. Run the MCMC scheme for a fixed number of iterates recording $\overline{\theta}$.
We want to assess experimentally whether our upper/lower bounds are sharp.

We select $N$ so as $\sigma$ to be roughly equal to a pre-specified value.

1. Choose a large initial value for the number of samples, $N_S$.
2. Run the MCMC scheme for a fixed number of iterates recording $\bar{\theta}$.
3. Record the estimated variance of the log of the likelihood estimator, $V(\bar{\theta}, N_S) = \hat{V} \left[ \log \hat{p}_{N_S}(y|\bar{\theta}, u) \right]$. 
We want to assess experimentally whether our upper/lower bounds are sharp.

We select $N$ so as $\sigma$ to be roughly equal to a pre-specified value.

1. Choose a large initial value for the number of samples, $N_S$.
2. Run the MCMC scheme for a fixed number of iterates recording $\bar{\theta}$.
3. Record the estimated variance of the log of the likelihood estimator, $V(\bar{\theta}, N_S) = \hat{\mathbb{V}} \left[ \log \hat{p}_{N_S}(y|\bar{\theta}, u) \right]$.
4. Set $N_{\bar{\theta}} = V(\bar{\theta}, N_S) / \sigma^2$. 

Acceptance Probabilities for Probit Example

Figure: Probit example $T = 100, \theta = 0.5$. Accept. Proba vs $\sigma(\bar{\theta})$. Estim. proba for the exact MCMC scheme is shown (constant), estim. proba from the simulated likelihood scheme (red) and lower bound given as proba exact scheme times $2\Phi(-\sigma/\sqrt{2})$ (blue).
Figure: $RCT_h^Q$ (top) and $RIF_h^Q$ (bottom) against $N$ (left) and $\sigma(\bar{\theta})$ (right) for various values of $\rho$. 
Empirical vs Asymptotic Distribution of Log-Likelihood Estimator

Figure: Histograms of proposed (red) and accepted (pink) values of $z$ in PMCMC scheme. Overlayed are Gaussian pdfs from our simplifying Assumption for a target of $\sigma = 0.92$. (Oxford, 27th Sep. 2012)
Example 2: Noisy Autoregressive Example

We have

\[ y_t = x_t + \sigma_\varepsilon \varepsilon_t \quad \text{and} \quad x_{t+1} = \mu(1 - \phi) + \phi x_t + \sigma_\eta \eta_t \]

where \( \varepsilon_t \) and \( \eta_t \) are independent standard normal and \( \theta = \left( \phi, \mu, \sigma_\eta^2 \right) \).
Example 2: Noisy Autoregressive Example

We have

\[ y_t = x_t + \sigma \varepsilon_t \quad \text{and} \quad x_{t+1} = \mu(1 - \phi) + \phi x_t + \sigma \eta \eta_t \]

where \( \varepsilon_t \) and \( \eta_t \) are independent standard normal and \( \theta = (\phi, \mu, \sigma^2_\eta) \).

The likelihood can be computed exactly using the Kalman filter.
Example 2: Noisy Autoregressive Example

We have

\[ y_t = x_t + \sigma \varepsilon_t \quad \text{and} \quad x_{t+1} = \mu(1 - \phi) + \phi x_t + \sigma \eta \eta_t \]

where \( \varepsilon_t \) and \( \eta_t \) are independent standard normal and \( \theta = (\phi, \mu, \sigma^2 \eta) \).

The likelihood can be computed exactly using the Kalman filter.

Autoregressive Metropolis proposal for \( \theta \) based on multivariate t-distribution.
Example 2: Noisy Autoregressive Example

- We have

\[ y_t = x_t + \sigma_{\varepsilon} \varepsilon_t \quad \text{and} \quad x_{t+1} = \mu (1 - \phi) + \phi x_t + \sigma_{\eta} \eta_t \]

where \( \varepsilon_t \) and \( \eta_t \) are independent standard normal and \( \theta = \left( \phi, \mu, \sigma_{\eta}^2 \right) \).

- The likelihood can be computed exactly using the Kalman filter.

- Autoregressive Metropolis proposal for \( \theta \) based on multivariate t-distribution.

- \( N \) is selected in the same manner so as to obtain an approximately constant \( \sigma \left( \bar{\theta} \right) \).
Acceptance Probabilities

Figure: AR1 plus noise example with $T = 300$, $\phi = 0.8$, $\mu = 0.5$, $\sigma_\eta^2 = 1$, $\sigma_\varepsilon^2 = 0.5$. Probabilities of acceptance displayed against $\sigma(\bar{\theta})$. 

Relative Inefficiency and Computing Time

Figure: From left to right: $RCT_h^Q$ vs $N$, $RCT_h^Q$ vs $\sigma(\bar{\theta})$, $RIF_h^Q$ against $N$ and $RIF_h^Q$ against $\sigma(\bar{\theta})$ for various values of $\rho$ and different parameters.
Assume a consistent estimator $\hat{\gamma}^2(\theta; N)$ of $\gamma(\theta)$ is available.
Assume a consistent estimator $\hat{\gamma}^2(\theta; N)$ of $\gamma(\theta)$ is available.

We select $N_{\bar{\theta}}$ so that $V(\bar{\theta}, N_{\bar{\theta}}) = \hat{\mathbb{V}} \left[ \log \hat{p}_{N_{\bar{\theta}}}(y|\bar{\theta}, u) \right] = \sigma^2.$
Assume a consistent estimator $\hat{\gamma}^2(\theta; N)$ of $\gamma(\theta)$ is available.

We select $N_{\theta}$ so that $V(\bar{\theta}, N_{\theta}) = \hat{V} \left[ \log \hat{p}_{N_{\theta}}(y | \bar{\theta}, u) \right] = \sigma^2$.

Select $N_{\theta}$ dynamically such that

$$\frac{\hat{\gamma}^2(\theta; N_{\theta})}{N_{\theta}} = \sigma^2.$$
Distribution of the Standard Deviation of the Log-likelihood Estimator

**Figure:** Probit model. Histograms for the standard deviation of the log-likelihood (over post of $\theta$) when $N$ is chosen statically (blue) for the proposal as $N_\theta$ and when it is chosen dynamically (pink) as $N_\theta$ for $\sigma^2 = 0.92$. 
We have provided an approximate analysis of MCMC using unbiased likelihood estimator.
We have provided an approximate analysis of MCMC using unbiased likelihood estimator.

For a general proposal and under simplifying assumptions on the likelihood estimator, we can get guidelines on how to select $\sigma$: *as long as $\sigma$ is around 1 then you are fine.*
Discussion

- We have provided an approximate analysis of MCMC using unbiased likelihood estimator.
- For a general proposal and under simplifying assumptions on the likelihood estimator, we can get guidelines on how to select $\sigma$: \textit{as long as $\sigma$ is around 1 then you are fine.}
- Coming up with a nicer way to adapt $N$ would be useful; e.g. Lee, 2011.