Monte Carlo Solution of Integral Equations

(of the second kind)

Adam M. Johansen

a.m.johansen@warwick.ac.uk
www2.warwick.ac.uk/fac/sci/statistics/staff/academic/johansen/talks/

March 7th, 2011
MIRaW – Monte Carlo Methods Day

Outline

- Background
  - Importance Sampling (IS)
  - Fredholm Equations of the 2nd Kind
  - Sequential IS and 2nd Kind Fredholm Equations
- Methodology
  - A Path-space Interpretation
  - MCMC for 2nd Kind Fredholm Equations
- Results
  - A Toy Example
  - An Asset-Pricing Example

The Monte Carlo Method

- Consider estimating:

$$I_\varphi = \mathbb{E}_f[\varphi(X)] = \int f(x)\varphi(x)dx$$

- Given $X_1, \ldots, X_N \overset{\text{iid}}{\sim} f$:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \varphi(x_i) \to I_\varphi$$

- Alteratively, approximate:

$$\widehat{f}(x) = \frac{1}{N}\sum_{1=1}^{N} \delta(x - x_i)$$

and use

$$\mathbb{E}_{\widehat{f}}[\varphi | X_1, \ldots, X_N] \approx \mathbb{E}_f[\varphi].$$

The Monte Carlo Method

- Consider estimating:

$$I_\varphi = \mathbb{E}_f[\varphi(X)] = \int f(x)\varphi(x)dx$$

- Given $X_1, \ldots, X_N \overset{\text{iid}}{\sim} f$:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \varphi(x_i) \to I_\varphi$$

- Alteratively, approximate:

$$\widehat{f}(x) = \frac{1}{N}\sum_{1=1}^{N} \delta(x - x_i)$$

and use

$$\mathbb{E}_{\widehat{f}}[\varphi|X_1, \ldots, X_N] \approx \mathbb{E}_f[\varphi].$$

Importance Sampling

- If $f \ll g$ (i.e. $g(x) = 0 \Rightarrow f(x) = 0$):

$$I\varphi = \int g(x)\frac{f(x)}{g(x)}\varphi(x)dx$$
$$= \mathbb{E}_g\Bigg[\underbrace{\frac{f(X)}{g(X)}}_{w(X)}\varphi(X)\Bigg]$$

- Given $X_1, \ldots, X_N \overset{\text{iid}}{\sim} g$:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} w(X_i)\varphi(x_i) \to I_\varphi$$

- Alteratively, approximate:

$$\widehat{f}(x) = \frac{1}{N}\sum_{1=1}^{N} w(X_i)\delta(x - X_i) \qquad \mathbb{E}_{\widehat{f}}[\varphi|X_1 \ldots X_N] \approx \mathbb{E}_f[\varphi].$$

Importance Sampling

- If $f \ll g$ (i.e. $g(x) = 0 \Rightarrow f(x) = 0$):

$$I\varphi = \int g(x) \frac{f(x)}{g(x)} \varphi(x) dx$$
$$= \mathbb{E}_g \Bigg[\underbrace{\frac{f(X)}{g(X)}}_{w(X)} \varphi(X) \Bigg]$$

- Given $X_1, \ldots, X_N \overset{\text{iid}}{\sim} g$:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} w(X_i) \varphi(x_i) \to I_\varphi$$

- Alteratively, approximate:

$$\widehat{f}(x) = \frac{1}{N} \sum_{1=1}^{N} w(X_i) \delta(x - X_i) \qquad \mathbb{E}_{\widehat{f}}[\varphi | X_1 \ldots X_N] \approx \mathbb{E}_f[\varphi].$$

Fredholm Equations of the Second Kind

- Consider the equation:

$$f(x) = \int_E f(y)K(x,y)dy + g(x)$$

- in which
  - source $g(x)$ and
  - transition $K(x,y)$

  are known, but
  - $f(x)$ is unknown.
- This is a Fredholm equation of the second kind.
- Finding $f(x)$ is an inverse problem.

The Von Neumann Expansion

- We have the expansion:

$$
\begin{aligned}
f(x) &= \int_E f(y)K(x,y)dy + g(x) \\
&= \int_E \left[\int_E f(z)K(y,z)dx + g(y)\right] K(x,y)dy + g(x) \\
&\;\;\vdots \\
&= g(x) + \sum_{n=1}^{\infty} \int K^n(x,y)f(y)dy
\end{aligned}
$$

  where:

$$
K^1(x,y) = K(x,y) \quad K^n(x,y) = \int_E K^{n-1}(x,z)K(z,y)dz
$$

- For convergence, it is sufficient that:

$$
|g(x)| + \sum_{n=1}^{\infty} \int |K^n(x,y)g(y)dy| < \infty
$$

A Sampling Approach

- Choose μ a pdf on E.
- Define

$$E' := E \cup \{\dagger\}$$

- Let $M(x, y)$ be a Markov transition kernel on E' such that:
  - $\dagger$ is absorbing
  - $M(x, \dagger) = P_d$ and
  - $M(x, y) > 0$ whenever $K(x, y) > 0$.
- The pair (μ, M) defines a Markov chain on E'.

Sequential Importance Sampling (SIS) Algorithm

- Simulate N paths $\left\{X^{(i)}_{0:k^{(i)}+1}\right\}_{i=1}^{N}$ until $X^{(i)}_{k^{(i)}+1} = \dagger$.
- Calculate weights:

$$W\left(X^{(i)}_{0:k^{(i)}}\right) = \begin{cases} \frac{1}{\mu\left(X^{(i)}_0\right)} \left(\prod_{k=1}^{k^{(i)}} \frac{K\left(X^{(i)}_{k-1}, X^{(i)}_k\right)}{M\left(X^{(i)}_{k-1}, X^{(i)}_k\right)}\right) \frac{g\left(X^{(i)}_{k^{(i)}}\right)}{P_d} & \text{if } k^{(i)} \geq 1, \\ \frac{g\left(X^{(i)}_0\right)}{\mu\left(X^{(i)}_0\right) P_d} & \text{if } k^{(i)} = 0. \end{cases}$$

- Approximate $f(x_0)$ with:

$$\widehat{f}(x_0) = \frac{1}{N} \sum_{i=1}^{N} W\left(X^{(i)}_{0:k^{(i)}}\right) \delta\left(x_0 - X^{(i)}_0\right)$$

Von Neumann Representation Revisited

- Starting from

$$f(x_0) = g(x_0) + \sum_{n=1}^{\infty} \int_E K^n(x_0, x_n) f(x_n) dx_n$$

- We have directly:

$$\begin{aligned} f(x_0) =& g(x_0) + \sum_{n=1}^{\infty} \int_{E^n} \left[\prod_{p=1}^{n} K(x_{p-1}, x_p) \right] g(x_n) dx_1 \dots dx_n \\ =& f_0(x_0) + \sum_{n=1}^{\infty} \int_{E^n} f_n(x_{0:n}) dx_1 \dots dx_n \end{aligned}$$

with $f_n(x_{0:n}) := g(x_n) \prod_{p=1}^{n} K(x_{p-1}, x_p)$.

A Path-Space Interpretation of SIS

- Define:

$$\pi(n, x_{0:n}) = p_n \pi_n(x_0 : n) \qquad \text{on } F = \cup_{k=0}^{\infty} \{k\} \times E'^{k+1}$$

  where

$$p_n = \mathbb{P}(X_{0:n} \in E^{n+1}, X_{n+1} = \dagger) = (1 - P_d)^n P_d$$

$$\pi_n(x_{0:n}) = \frac{\mu(x_0) \prod_{k=1}^{n} M(x_{k-1}, x_k)}{(1 - P_d)^n}$$

- Define also:

$$f(k, x_{0:k}) = f_0(x_0)\delta_{0,k} + \sum_{n=1}^{\infty} f_n(x_{0:n})\delta_{n,k}$$

- Approximate with:

$$\hat{f}(x_0) = \frac{1}{N} \sum_{i=1}^{N} \frac{f(k^{(i)}, X^{(i)}_{0:k^{(i)}})}{\pi(k^{(i)}, X^{(i)}_{0:k^{(i)}})} \delta(x_0 - X^{(i)}_0)$$

Limitations of SIS

- Arbitary geometric distribution for k.
- Arbitary initial distribution.
- Importance weights

$$\frac{1}{\mu\left(X_0^{(i)}\right)}\left(\prod_{k=1}^{k^{(i)}}\frac{K\left(X_{k-1}^{(i)},X_k^{(i)}\right)}{M\left(X_{k-1}^{(i)},X_k^{(i)}\right)}\right)\frac{g\left(X_{k^{(i)}}^{(i)}\right)}{P_d}$$

- Resampling *will not* help.

Limitations of SIS

- Arbitary geometric distribution for k.
- Arbitary initial distribution.
- Importance weights

$$\frac{1}{\mu\left(X_0^{(i)}\right)}\left(\prod_{k=1}^{k^{(i)}}\frac{K\left(X_{k-1}^{(i)},X_k^{(i)}\right)}{M\left(X_{k-1}^{(i)},X_k^{(i)}\right)}\right)\frac{g\left(X_{k^{(i)}}^{(i)}\right)}{P_d}$$

- Resampling *will not* help.

Optimal Importance Distribution

- Minimize variance of absolute value of weights.
- Consider choosing

$$\pi(n, x_{0:n}) = \sum_{i=k}^{\infty} p_k \delta_{k,n} \pi_k(x_{0:k})$$

$$\pi_n(x_{0:n}) = c_n^{-1} |f_n(x_{0:n})| \qquad c_n = \int_{E^{n+1}} |f_n(x_{0:n})| \, dx_{0:n}$$

$$p_n = c_n / c \qquad c = \sum_{n=0}^{\infty} c_n.$$

- We have:

$$\begin{aligned} f(x_0) &= c \operatorname{sgn}(f_0(x_0)) \pi(0, x_0) + \\ & c \sum_{n=1}^{\infty} \int_{E^n} \operatorname{sgn}(f_n(x_{0:n})) \pi(n, x_{0:n}) \, dx_{1:n}. \end{aligned}$$

That's where the MCMC comes in

- Obtain a sample-approximation of π ...
  - using your favourite sampler,
  - Reversible Jump MCMC (2), for instance
- Use the samples to approximate the optimal proposal.
- Combine with importance sampling.
- Estimate 'normalising constant', c.

A Simple RJMCMC Algorithm

Initialization.

- Set $\left(k^{(1)}, X^{(1)}_{0:k^{(1)}}\right)$ randomly or deterministically.

Iteration $i \geq 2$.

- Sample $U \sim \mathrm{U}[0,1]$
- If $U \leq u_{k^{(i-1)}}$
  - $(k^{(i)}, x^{(i)}_{0:k^{(i)}}) \leftarrow$Update Move.
- Else If $U \leq u_{k^{(i-1)}} + b_{k^{(i-1)}}$:
  - $(k^{(i)}, x^{(i)}_{0:k^{(i)}}) \leftarrow$Birth Move.
- Else
  - $(k^{(i)}, x^{(i)}_{0:k^{(i)}}) \leftarrow$Death Move.

Very Simple Update Move

- Set $k^{(i)} = k^{(i-1)}$, sample $J \sim \mathcal{U}\left(\left\{0, 1, \ldots, k^{(i)}\right\}\right)$ and $X_J^* \sim q_u\left(X_J^{(i-1)}, \cdot\right)$.
- With probability

$$\min\left\{1, \frac{\pi\left(k^{(i)}, \left(X_{0:J-1}^{(i-1)}, X_J^*, X_{J+1:k^{(i)}}^{(i-1)}\right)\right) q_u\left(X_J^*, X_J^{(i-1)}\right)}{\pi\left(k^{(i)}, X_{0:k^{(i)}}^{(i-1)}\right) q_u\left(X_J^{(i-1)}, X_J^*\right)}\right\}$$

set $X_{0:k^{(i)}}^{(i)} = \left(X_{0:J-1}^{(i-1)}, X_J^*, X_{J+1:k^{(i)}}^{(i-1)}\right)$,
- otherwise set $X_{0:k^{(i)}}^{(i)} = X_{0:k^{(i-1)}}^{(i-1)}$.

Acceptance probabilities involve:

$$\frac{\pi\left(l, x_{0:l}\right)}{\pi\left(k, x_{0:k}\right)} = \frac{c_l \pi_l\left(x_{0:l}\right)}{c_k \pi_k\left(x_{0:k}\right)} = \left|\frac{f_l\left(x_{0:l}\right)}{f_k\left(x_{0:k}\right)}\right|.$$

Very Simple Update Move

- Set $k^{(i)} = k^{(i-1)}$, sample $J \sim \mathcal{U}\left(\left\{0, 1, \ldots, k^{(i)}\right\}\right)$ and $X_J^* \sim q_u\left(X_J^{(i-1)}, \cdot\right)$.
- With probability

$$\min\left\{1, \frac{\pi\left(k^{(i)}, \left(X_{0:J-1}^{(i-1)}, X_J^*, X_{J+1:k^{(i)}}^{(i-1)}\right)\right) q_u\left(X_J^*, X_J^{(i-1)}\right)}{\pi\left(k^{(i)}, X_{0:k^{(i)}}^{(i-1)}\right) q_u\left(X_J^{(i-1)}, X_J^*\right)}\right\}$$

set $X_{0:k^{(i)}}^{(i)} = \left(X_{0:J-1}^{(i-1)}, X_J^*, X_{J+1:k^{(i)}}^{(i-1)}\right)$,

- otherwise set $X_{0:k^{(i)}}^{(i)} = X_{0:k^{(i-1)}}^{(i-1)}$.

Acceptance probabilities involve:

$$\frac{\pi\left(l, x_{0:l}\right)}{\pi\left(k, x_{0:k}\right)} = \frac{c_l \pi_l\left(x_{0:l}\right)}{c_k \pi_k\left(x_{0:k}\right)} = \left|\frac{f_l\left(x_{0:l}\right)}{f_k\left(x_{0:k}\right)}\right|.$$

Very Simple Birth Move

- Sample $J \sim \mathcal{U}\left\{0, 1, \ldots, k^{(i-1)}\right\}$, sample $X_J^* \sim q_b(\cdot)$.
- With probability

$$\min\left\{1, \frac{\pi\left(k^{(i-1)}+1, \left(X_{0:J-1}^{(i-1)}, X_J^*, X_{J:k^{(i-1)}}^{(i-1)}\right)\right) d_{k^{(i-1)}+1}}{\pi\left(k^{(i-1)}, X_{0:k^{(i-1)}}^{(i-1)}\right) q_b\left(X_J^*\right) b_{k^{(i-1)}}}\right\}$$

set $k^{(i)} = k^{(i-1)} + 1$, $X_{0:k}^{(i)} = \left(X_{0:J-1}^{(i-1)}, X_J^*, X_{J:k^{(i-1)}}^{(i-1)}\right)$,

- otherwise set $k^{(i)} = k^{(i-1)}$, $X_{0:k^{(i)}}^{(i)} = X_{0:k^{(i-1)}}^{(i-1)}$.

Very Simple Death Move

- Sample $J \sim \mathcal{U}\left\{0, 1, \ldots, k^{(i-1)}\right\}$.
- With probability

$$\min\left\{1, \frac{\pi\left(k^{(i-1)}-1, \left(X_{0:J-1}^{(i-1)}, X_{J+1:k^{(i-1)}}^{(i-1)}\right)\right) q_b\left(X_J^{(i-1)}\right) b_{k^{(i-1)}-1}}{\pi\left(k^{(i-1)}, X_{0:k^{(i-1)}}^{(i-1)}\right) d_{k^{(i-1)}}}\right\}$$

set $k^{(i)} = k^{(i-1)} - 1$, $X_{0:k^{(i)}}^{(i)} = \left(X_{0:J-1}^{(i-1)}, X_{J+1:k^{(i-1)}}^{(i-1)}\right)$,
otherwise set $k^{(i)} = k^{(i-1)}$, $X_{0:k^{(i)}}^{(i)} = X_{0:k^{(i-1)}}^{(i-1)}$.

Estimating the Normalising Constant

- Estimate p_n as:

$$\widehat{p_n} = \frac{1}{N} \sum_{i=1}^{n} \delta_{n,k^{(i)}}$$

- Estimate directly, also,

$$\widetilde{c_0} \approx \int g(x)dx$$

and hence

$$\widehat{c} = \widetilde{c_0} / \widehat{p_0} \qquad \widehat{c_n} = \widehat{c}\widehat{p_n}$$

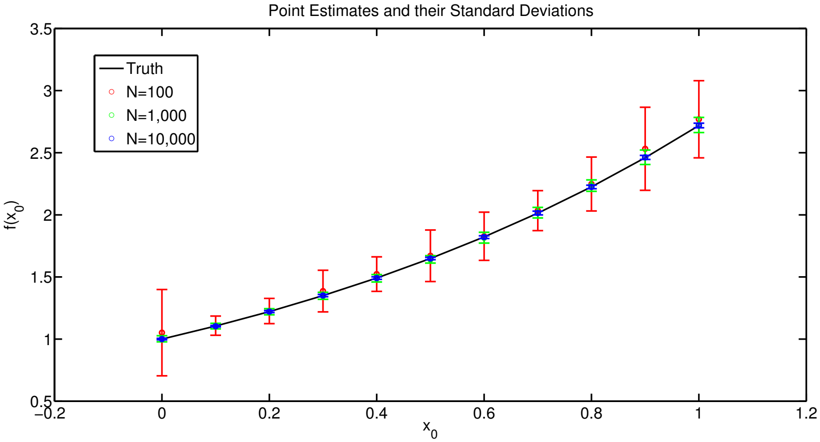
Toy Example

- The simple algorithm was applied to

$$f(x) = \int_0^1 \underbrace{\frac{1}{3}\exp(x-y)}_{K(x,y)} f(y)dy + \underbrace{\frac{2}{3}\exp(x)}_{g(x)}.$$

- The solution $f(x) = \exp(x)$.
- Birth, death and update probabilities were set to $1/3$.
- A uniform distribution over the unit interval was used for all proposals.

Point Estimation for the Toy Example



(Fix x_0 and simulate everything else).

Estimating $f(x)$ from Samples

- Given $\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} W^{(i)}\delta(x - X_0^{(i)})$,
- such that

$$\mathbb{E}\left[\int \hat{f}(x)\varphi(x)\right] = \int f(x)\varphi(x)dx$$

- Simple option: histogram ($\varphi(x) = \mathbb{I}_A(x)$).
- More subtle option:

$$\widetilde{f}(x) = \int K(x,y)\hat{(}f)(y)dy + g(x)$$

$$= g(x) + \frac{1}{N}\sum_{i=1}^{N} W^{(i)}K(x, X_0^{(i)}).$$

Estimating $f(x)$ from Samples

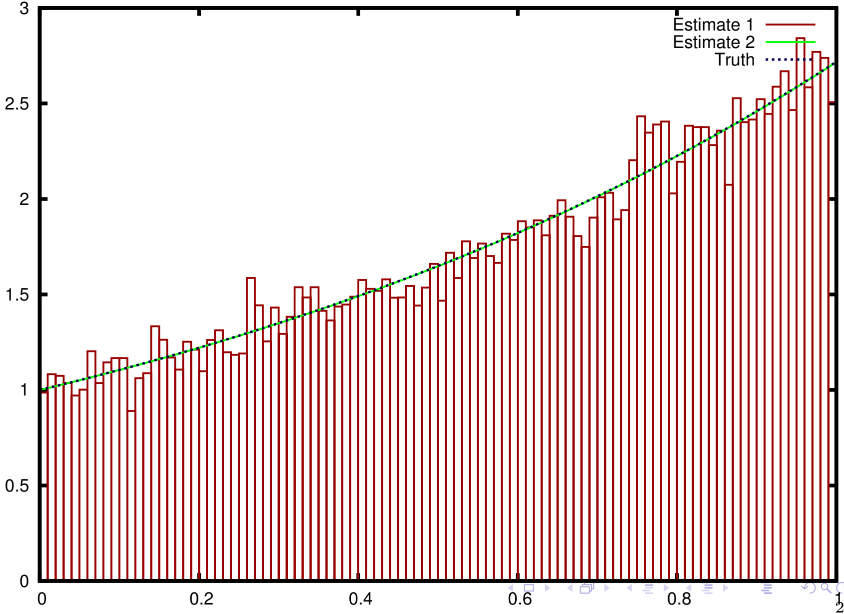
- Given $\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} W^{(i)} \delta(x - X_0^{(i)})$,
- such that

$$\mathbb{E}\left[\int \hat{f}(x)\varphi(x)\right] = \int f(x)\varphi(x)dx$$

- Simple option: histogram ($\varphi(x) = \mathbb{I}_A(x)$).
- More subtle option:

$$\widetilde{f}(x) = \int K(x,y)\hat{(f)}(y)dy + g(x)$$
$$= g(x) + \frac{1}{N}\sum_{i=1}^{N} W^{(i)} K(x, X_0^{(i)}).$$

Estimated $f(x)$ for the Toy Problem



An Asset-Pricing Problem

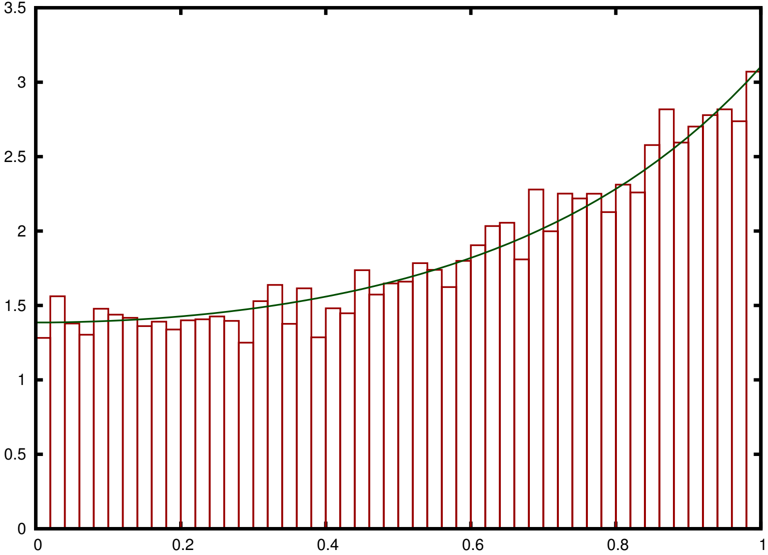
- The rational expectation pricing model (cf. (4)) requires:
  - the price, $V(s)$ of an asset
  - in some state $s \in E$

  satisfies
  $$V(s) = \pi(s) + \beta \int_E V(t)p(t|s)dt.$$
- Where:
  - $\pi(s)$ denotes the return on investment,
  - β is a discount factor and
  - $p(t|s)$ is a Markov kernel which models the state evolution.
- Simple example:

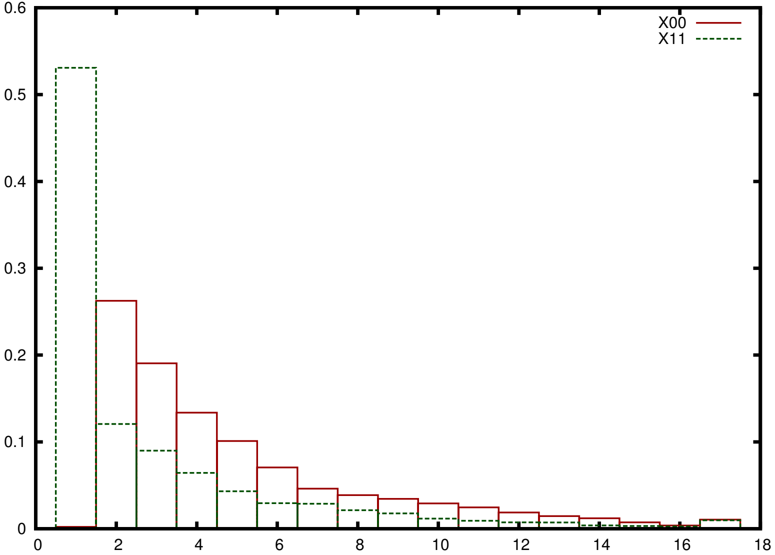
$$E = [0, 1] \quad \beta = 0.85 \quad p(t|s) = \frac{\phi\left(\frac{t - |as+b|}{\sqrt{\lambda}}\right)}{\Phi(\frac{1-[as+b]}{\sqrt{\lambda}}) - \Phi(-\frac{as+b}{\sqrt{\lambda}})}$$

with $a = 0.05$, $b = 0.85$ and $\lambda = 100$.

Estimated $f(x)$ for asset pricing case.



Simulated Chain Lengths for Different Starting Points



Conclusions

- Many Monte Carlo methods can be viewed as importance sampling.
- Many problems can be recast as calculation of expectations.

- Approximate solution of Fredholm equations.
- Also applicable to Volterra Equations (3).
- Any Monte Carlo method could be used.

- Some room for further investigation of this method.

References

[1] A. Doucet, A. M. Johansen, and V. B. Tadić. On solving integral equations using Markov Chain Monte Carlo. *Applied Mathematics and Computation*, 216:2869–2889, 2010.

[2] P. J. Green. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

[3] G. W. Peters, A. M. Johansen, and A. Doucet. Simulation of the annual loss distribution in operational risk via Panjer recursions and Volterra integral equations for value at risk and expected shortfall estimation. *Journal of Operational Risk*, 2(3):29–58, Fall 2007.

[4] J. Rust, J. F. Traub, and H. Wózniakowski. Is there a curse of dimensionality for contraction fixed points in the worst case? *Econometrica*, 70(1):285–329, 2002.