A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Lecture 2: Bayesian Computation

Hedibert Freitas Lopes

The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL 60637
http://faculty.chicagobooth.edu/hedibert.lopes

hlopes@ChicagoBooth.edu

Outline

1. A bit of history
2. MC integration
3. MC via IS
4. Rejection method
5. SIR method
6. MH algorithms
7. Simulated annealing
8. Gibbs sampler
9. SUR model
10. BVAR
11. References

MC in the 40s and 50s

Stan Ulam soon realized that computers could be used in this fashion to answer questions of neutron diffusion and mathematical physics;

He contacted John Von Neumann and they developed many Monte Carlo algorithms (importance sampling, rejection sampling, etc);

In the 1940s Nick Metropolis and Klari Von Neumann designed new controls for the state-of-the-art computer (ENIAC);

Metropolis and Ulam (1949) The Monte Carlo method. *Journal of the American Statistical Association*.
Metropolis *et al.* (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics*.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Monte Carlo methods

We introduce several Monte Carlo (MC) methods for integrating and/or sampling from nontrivial densities.

- MC integration
  - Simple MC integration
  - MC integration via importance sampling (IS)
- MC sampling
  - Rejection method
  - Sampling importance resampling (SIR)
- Iterative MC sampling
  - Metropolis-Hastings algorithms
  - Simulated annealing
  - Gibbs sampler

Based on the book by Gamerman and Lopes (1996).

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

A few references

- MC integration (Geweke, 1989)
- Rejection methods (Gilks and Wild, 1992)
- SIR (Smith and Gelfand, 1992)
- Metropolis-Hastings algorithm (Hastings, 1970)
- Simulated annealing (Metropolis *et al.*, 1953)
- Gibbs sampler (Gelfand and Smith, 1990)

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Two main tasks

1. Compute high dimensional integrals:

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

2. Obtain

$$a \ sample \ \{\theta_1, \ldots, \theta_n\} \ from \ \pi(\theta)$$

when only

$$a \ sample \ \{\tilde{\theta}_1, \ldots, \tilde{\theta}_m\} \ from \ q(\theta)$$

is available.

$q(\theta)$ is known as the *proposal/auxiliary* density.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Bayes via MC

MC methods appear frequently, but not exclusively, in modern Bayesian statistics.

Posterior and predictive densities are hard to sample from:

$$\text{Posterior} \quad : \quad \pi(\theta) = \frac{f(x|\theta)p(\theta)}{f(x)}$$

$$\text{Predictive} \quad : \quad f(x) = \int f(x|\theta)p(\theta)d\theta$$

Other important integrals and/or functionals of the posterior and predictive densities are:

- Posterior modes: $\max_\theta \pi(\theta)$;
- Posterior moments: $E_\pi[g(\theta)]$;
- Density estimation: $\hat{\pi}(g(\theta))$;
- Bayes factors: $f(x|M_0)/f(x|M_1)$;
- Decision: $\max_d \int U(d,\theta)\pi(\theta)d\theta$.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

MC integration

The objective here is to compute moments

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

If $\theta_1, \ldots, \theta_n$ is a random sample from $\pi(\cdot)$ then

$$\bar{h}_{mc} = \frac{1}{n}\sum_{i=1}^{n} h(\theta_i) \rightarrow E_\pi[h(\theta)] \qquad \text{as } n \rightarrow \infty.$$

If, additionally, $E_\pi[h^2(\theta)] < \infty$, then

$$V_\pi[\bar{h}_{mc}] = \frac{1}{n}\int \{h(\theta) - E_\pi[h(\theta)]\}^2\pi(\theta)d\theta$$

and

$$v_{mc} = \frac{1}{n^2}\sum_{i=1}^{n}(h(\theta_i) - \bar{h}_{mc})^2 \rightarrow V_\pi[\bar{h}_{mc}] \qquad \text{as } n \rightarrow \infty.$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example i. MC integration

The objective here is to compute[1]

$$p = \int_0^1 [cos(50\theta) + sin(20\theta)]^2 d\theta$$

by noticing that the above integral can be rewritten as

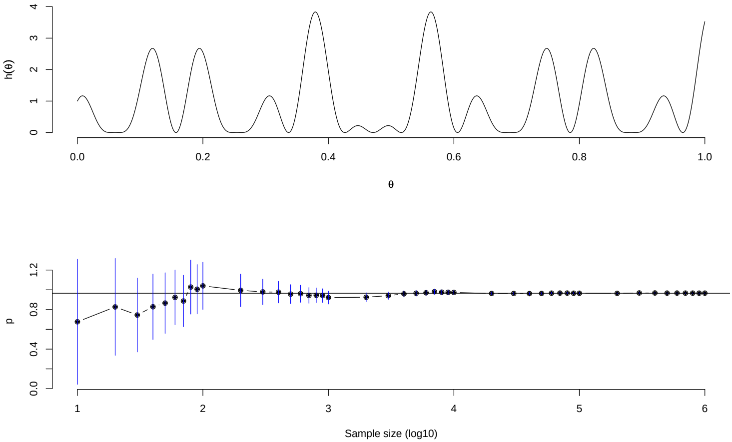
$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

where $h(\theta) = [cos(50\theta) + sin(20\theta)]^2$ and $\pi(\theta) = 1$ is the density of a $U(0,1)$. Therefore

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} h(\theta_i)$$

where $\theta_1, \ldots, \theta_n$ are i.i.d. from $U(0,1)$.

[1] True value is 0.965.


h(θ)
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
θ
p
0.0
0.4
0.8
1.2
1
2
3
4
5
6
Sample size (log10)

MC via IS

The objective is still the same, ie to compute

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

by noticing that

$$E_\pi[h(\theta)] = \int \frac{h(\theta)\pi(\theta)}{q(\theta)} q(\theta)d\theta$$

where $q(\cdot)$ is an *importance function*.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

If $\theta_1, \ldots, \theta_n$ is a random sample from $q(\cdot)$ then

$$\Rightarrow \bar{h}_{is} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(\theta_i)\pi(\theta_i)}{q(\theta_i)} \rightarrow E_\pi[h(\theta)]$$

as $n \longrightarrow \infty$.

Ideally, $q(\cdot)$ should be

- As *close* as possible to $h(\cdot)\pi(\cdot)$, and
- Easy to sample from.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example ii. Cauchy tail

The objective here is to estimate

$$p = Pr(\theta > 2) = \int_2^\infty \frac{1}{\pi(1+\theta^2)} d\theta = 0.1475836$$

where θ is a standard Cauchy random variable.

A natural MC estimator of p is

$$\hat{p}_1 = \frac{1}{n} \sum_{i=1}^{n} I\{\theta_i \in (2, \infty)\}$$

where $\theta_1, \ldots, \theta_n \sim$ Cauchy(0,1).

A more elaborated estimator based on a change of variables from θ to $u = 1/\theta$ is

$$\hat{p}_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{u_i^{-2}}{2\pi[1 + u_i^{-2}]}$$

where $u_1, \ldots, u_n \sim U(0, 1/2)$.

The true value is $p = 0.147584$.

n	$\hat{p}_1$	$\hat{p}_2$	$v_1^{1/2}$	$v_2^{1/2}$
100	0.100000	0.1467304	0.030000	0.001004
1000	0.137000	0.1475540	0.010873	0.000305
10000	0.148500	0.1477151	0.003556	0.000098
100000	0.149100	0.1475591	0.001126	0.000031
1000000	0.147711	0.1475870	0.000355	0.000010

With only $n = 1000$ draws, $\hat{p}_2$ has roughly the same precision that $\hat{p}_1$, which is based on $1000n$ draws, ie. three orders of magnitude.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Rejection method

The objective is to draw from a target density

$$\pi(\theta) = c_\pi \tilde{\pi}(\theta)$$

when only draws from an auxiliary density

$$q(\theta) = c_q \tilde{q}(\theta)$$

is available, for normalizing constants c_π and c_q.

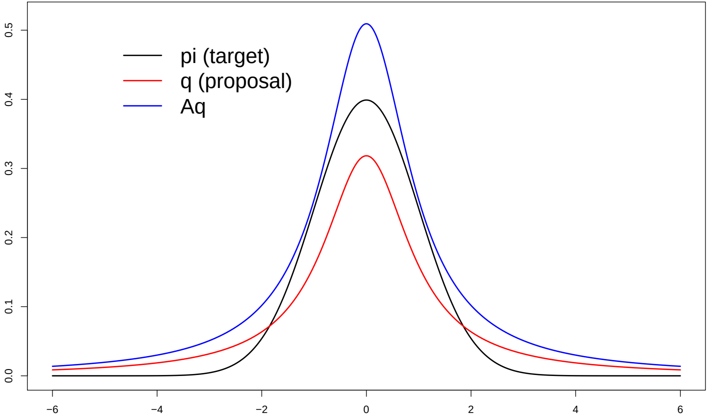
If there exist a constant $A < \infty$ such that

$$0 \leq \frac{\tilde{\pi}(\theta)}{A\tilde{q}(\theta)} \leq 1 \quad \text{for all } \theta$$

then $q(\theta)$ becomes a *blanketing density* or an *envelope* and A the *envelope constant*.

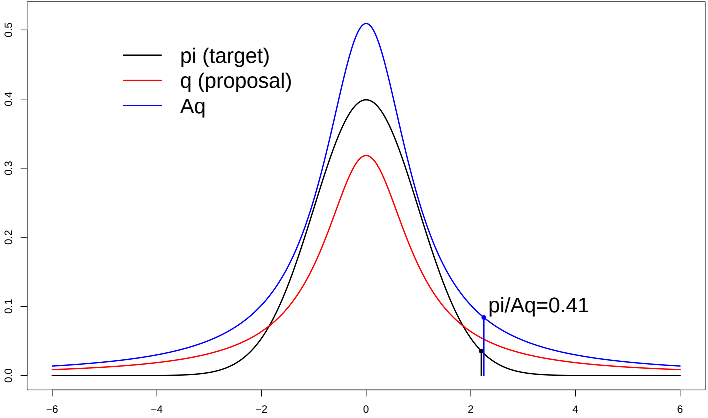
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Blanket distribution



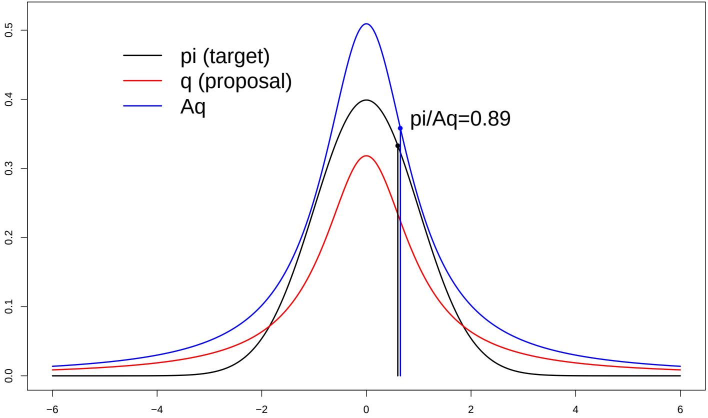
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Bad draw

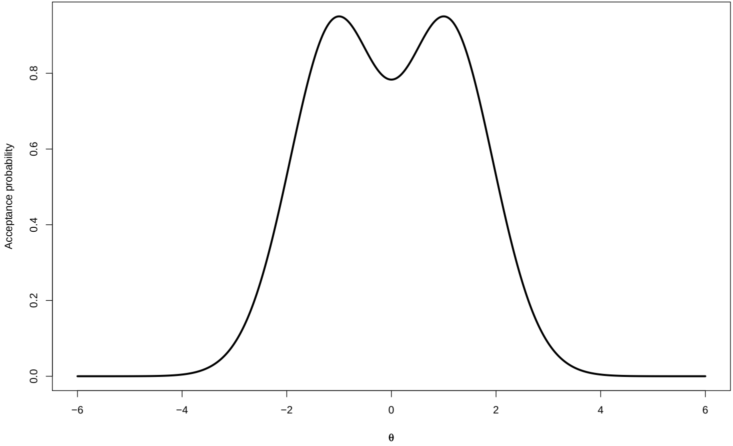


A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Good draw



Acceptance probability



Algorithm

Drawing from $\pi(\theta)$.

1. Draw θ^* from $q(\cdot)$;
2. Draw u from $U(0,1)$;
3. Accept θ^* if $u \leq \frac{\tilde{\pi}(\theta^*)}{A\tilde{q}(\theta^*)}$;
4. Repeat 1, 2 and 3 until n draws are accepted.

Normalizing constants c_π and c_q are not needed.

The theoretical acceptance rate is $\frac{c_q}{Ac_\pi}$.

The smaller the A, the larger the acceptance rate.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example iii. Sampling $N(0,1)$

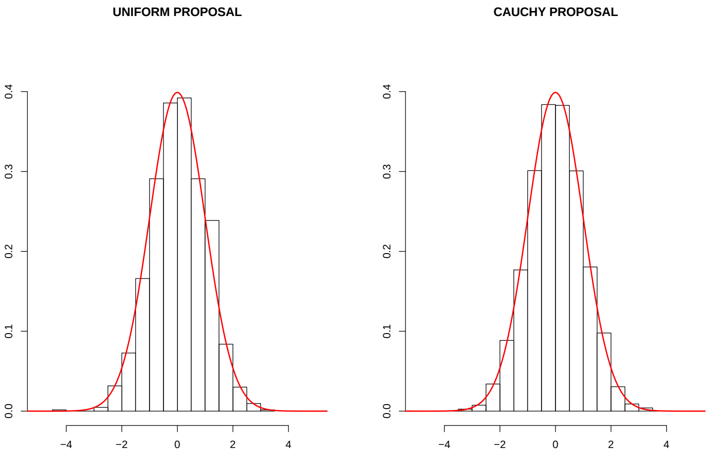
Enveloping the standard normal density

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}} \exp\{-0.5\theta^2\}$$

by a Cauchy density $q_C(\theta) = 1/(\pi(1+\theta^2))$, or a uniform density $q_U(\theta) = 0.05$ for $\theta \in (-10, 10)$.

Bad proposal:The maximum of $\pi(\theta)/q_U(\theta)$ is roughly $A_U = 7.98$ for $\theta \in (-10, 10)$. The theoretical acceptance rate is 12.53%.

Good proposal: The max of $\pi(\theta)/q_C(\theta)$ is equal to $A_C = \sqrt{2\pi/e} \approx 1.53$. The theoretical acceptance rate is 65.35%.



Empirical rates: 0.1265 (Uniform) and 0.6483 (Cauchy)
Theoretical rates: 0.1253 (Uniform) and 0.6535 (Cauchy)

SIR method

No need to rely on the existance of A!

Algorithm

1. Draw $\theta_1^*, \ldots, \theta_n^*$ from $q(\cdot)$
2. Compute (unnormalized) weights

$$\omega_i = \pi(\theta_i^*)/q(\theta_i^*) \qquad i = 1, \ldots, n$$

3. Sample θ from $\{\theta_1^*, \ldots, \theta_n^*\}$ such that

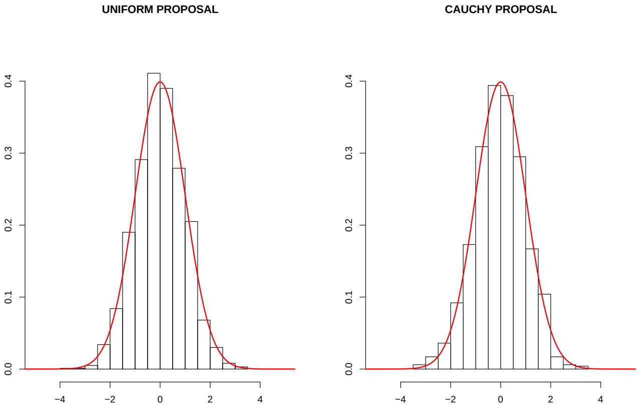
$$Pr(\theta = \theta_i^*) \propto \omega_i \qquad i = 1, \ldots, n.$$

4. Repeat m times step 3.

Rule of thumb: $n/m = 20$.
Ideally, $\omega_i = 1/n$ and $Var(\omega) = 0$.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example iii. revisited



Fraction of redraws: 0.391 (Uniform) and 0.1335 (Cauchy)
Variance of weights: 4.675 (Uniform) and 0.332 (Cauchy)

Example iv. 3-component mixture

Assume that we are interested in sampling from

$$\pi(\theta) = \alpha_1 p_N(\theta; \mu_1, \Sigma_1) + \alpha_2 p_N(\theta; \mu_2, \Sigma_2) + \alpha_3 p_N(\theta; \mu_3, \Sigma_3)$$

where $p_N(\cdot; \mu, \Sigma)$ is the density of a bivariate normal distribution with mean vector μ and covariance matrix Σ. The mean vectors are

$$\mu_1 = (1,4)' \quad \mu_2 = (4,2)' \quad \mu_3 = (6.5,2),$$

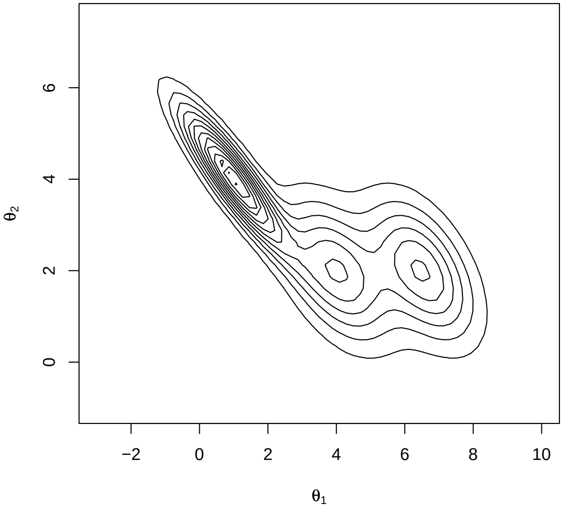
the covariance matrices are

$$\Sigma_1 = \begin{pmatrix} 1.0 & -0.9 \\ -0.9 & 1.0 \end{pmatrix} \text{ and } \Sigma_2 = \Sigma_3 = \begin{pmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{pmatrix},$$

and weights $\alpha_1 = \alpha_2 = \alpha_3 = 1/3$.

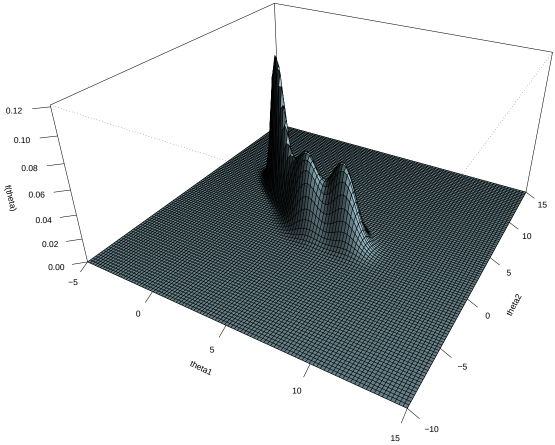
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Target $\pi(\theta)$



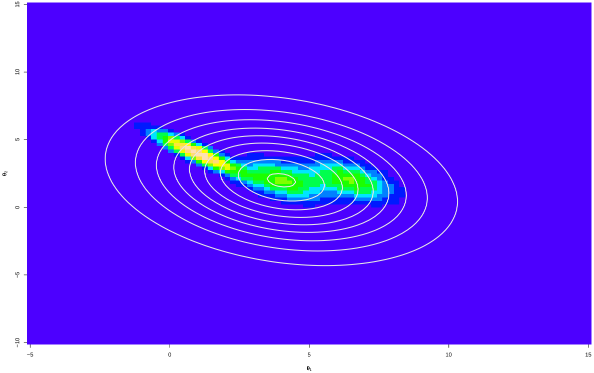
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Target $\pi(\theta)$



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

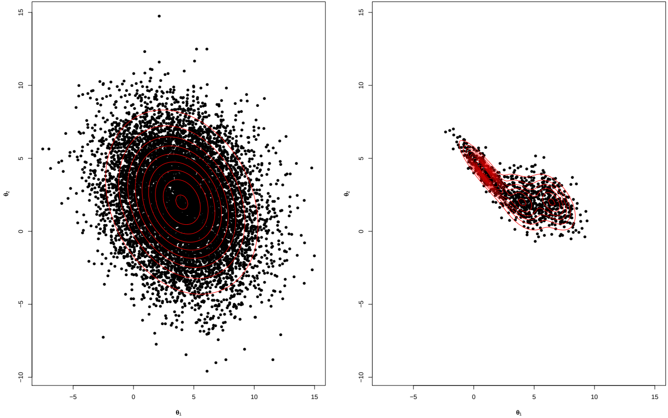
Proposal $q(\theta)$



$q(\theta) \sim N(\mu, \Sigma)$ where

$$\mu_2 = (4,2)' \quad \text{and} \quad \Sigma = 9\begin{pmatrix} 1.0 & -0.25 \\ -0.25 & 1.0 \end{pmatrix}$$

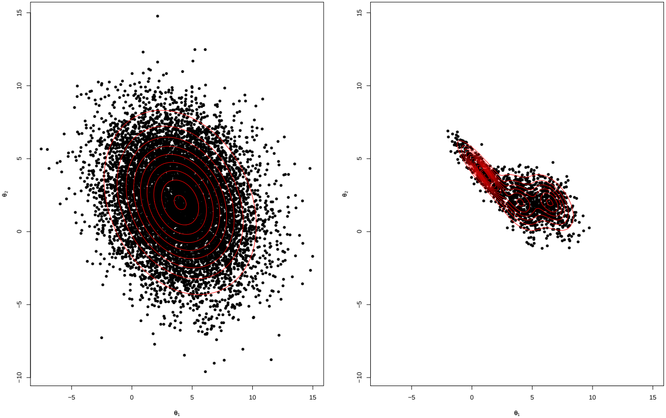
Rejection method



Acceptance rate: 9.91% of $n = 10,000$ draws.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

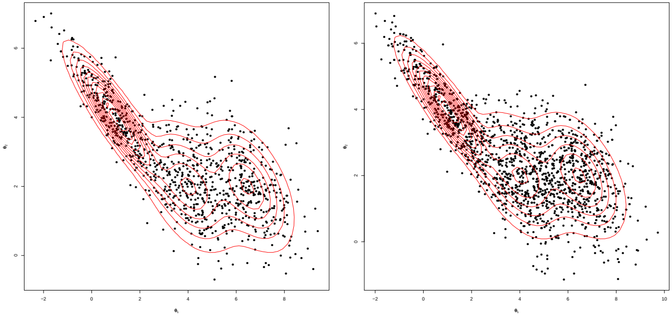
SIR method



Fraction of redraws: 29.45% of ($n = 10,000$, $m = 2,000$).

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Rejection & SIR



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example v. 2-component mixture

Let us now assume that

$$\pi(\theta) = \alpha_1 p_N(\theta; \mu_1, \Sigma_1) + \alpha_3 p_N(\theta; \mu_3, \Sigma_3)$$

where mean vectors are

$$\mu_1 = (1, 4)' \quad \mu_3 = (6.5, 2),$$

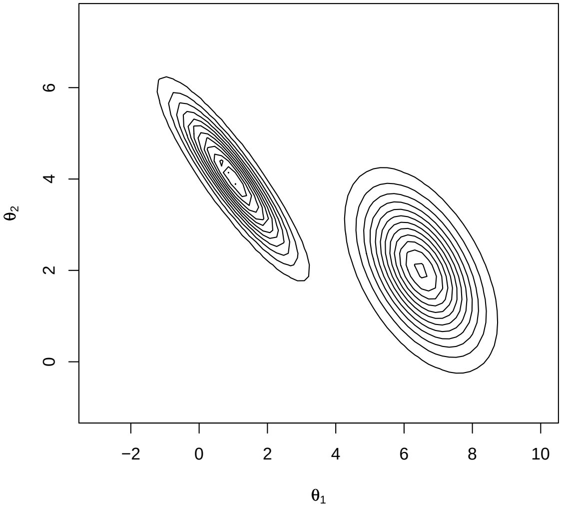
the covariance matrices are

$$\Sigma_1 = \begin{pmatrix} 1.0 & -0.9 \\ -0.9 & 1.0 \end{pmatrix} \text{ and } \Sigma_3 = \begin{pmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{pmatrix},$$

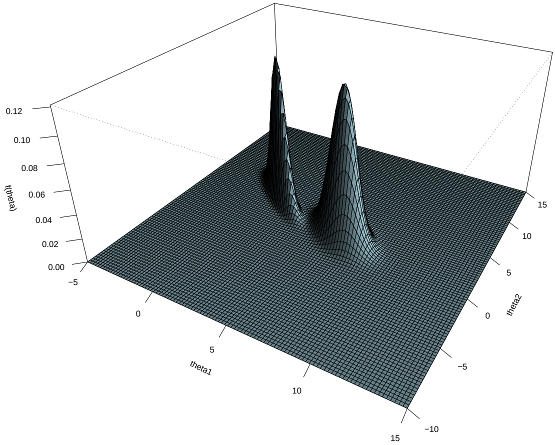
and weights $\alpha_1 = 1/3$ and $\alpha_3 = 2/3$.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Target $\pi(\theta)$

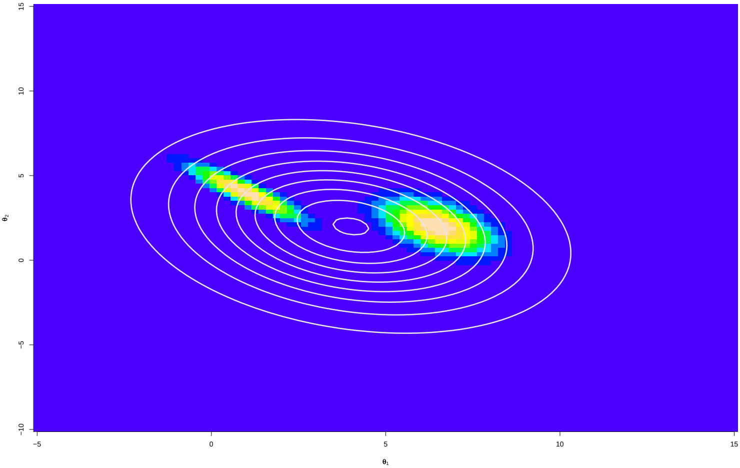


Target $\pi(\theta)$



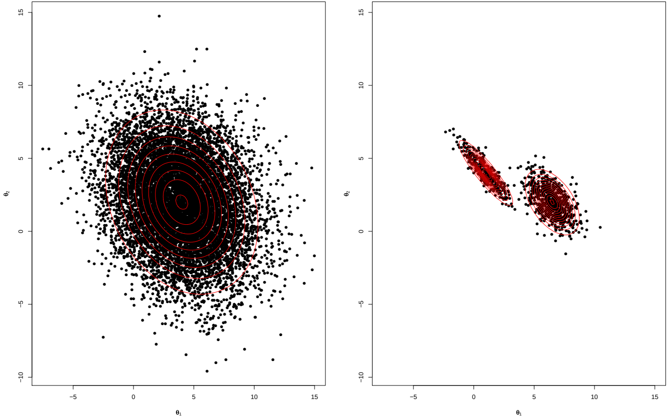
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Proposal $q(\theta)$



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

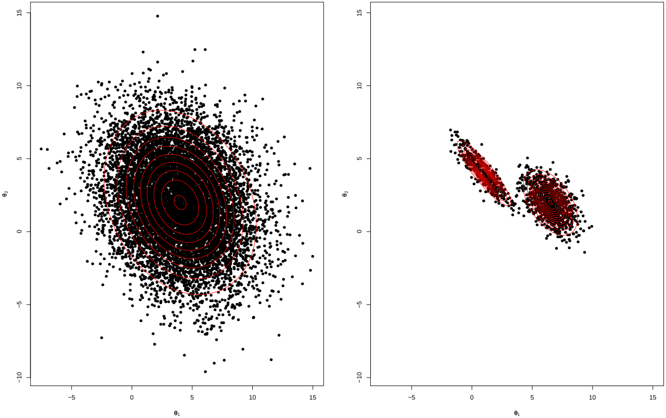
Rejection method



Acceptance rate: 10.1% of $n = 10,000$ draws.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

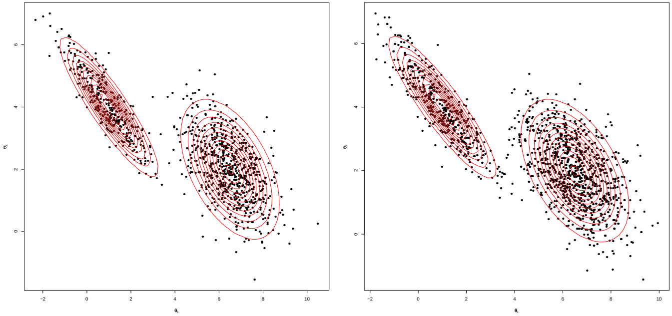
SIR method



Fraction of redraws: 37.15% of $(n = 10,000, m = 2,000)$.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Rejection & SIR



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

MCMC history

Dongarra and Sullivan (2000) list the top algorithms with the greatest influence on the development and practice of science and engineering in the 20th century (in chronological order):

- Metropolis Algorithm for Monte Carlo
- Simplex Method for Linear Programming
- Krylov Subspace Iteration Methods
- The Decompositional Approach to Matrix Computations
- The Fortran Optimizing Compiler
- QR Algorithm for Computing Eigenvalues
- Quicksort Algorithm for Sorting
- Fast Fourier Transform

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

70s and 80s

Metropolis-Hastings:
Hastings (1970) and his student Peskun (1973) showed that Metropolis and the more general Metropolis-Hastings algorithm are particular instances of a larger family of algorithms.

Gibbs sampler:

Besag (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems.
Geman and Geman (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images.
Pearl (1987) Evidential reasoning using stochastic simulation.
Tanner and Wong (1987). The calculation of posterior distributions by data augmentation.
Gelfand and Smith (1990) Sampling-based approaches to calculating marginal densities.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

MH algorithms

A sequence $\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots\}$ is drawn from a Markov chain whose *limiting equilibrium distribution* is the posterior distribution, $\pi(\theta)$.

Algorithm

1. Initial value: $\theta^{(0)}$
2. Proposed move: $\theta^* \sim q(\theta^*|\theta^{(i-1)})$
3. Acceptance scheme:

$$\theta^{(i)} = \begin{cases} \theta^* & \text{com prob.} \quad \alpha \\ \theta^{(i-1)} & \text{com prob.} \quad 1-\alpha \end{cases}$$

where

$$\alpha = \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^*|\theta^{(i-1)})}\right\}$$

Special cases

1. Symmetric chains: $q(\theta|\theta^*) = q(\theta^*|\theta)$

$$\alpha = \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta)}\right\}$$

2. Independence chains: $q(\theta|\theta^*) = q(\theta)$

$$\alpha = \min\left\{1, \frac{\omega(\theta^*)}{\omega(\theta)}\right\}$$

where $\omega(\theta^*) = \pi(\theta^*)/q(\theta^*)$.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Random walk Metropolis

The most famous symmetric chain is the random walk Metropolis:

$$q(\theta|\theta^*) = q(|\theta - \theta^*|)$$

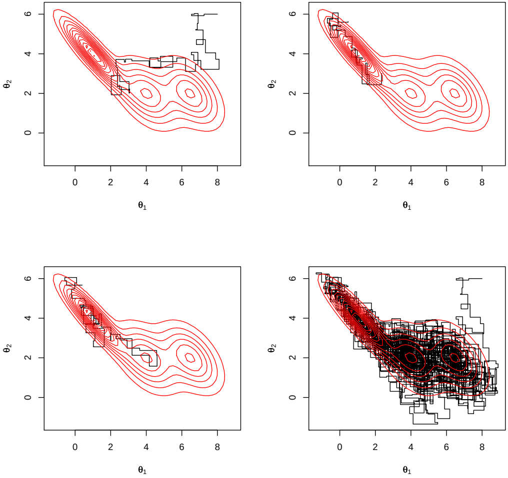
Hill climbing: when

$$\alpha = \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta)}\right\}$$

a value θ^* with higher density $\pi(\theta^*)$ greater than $\pi(\theta)$ is automatically accepted.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

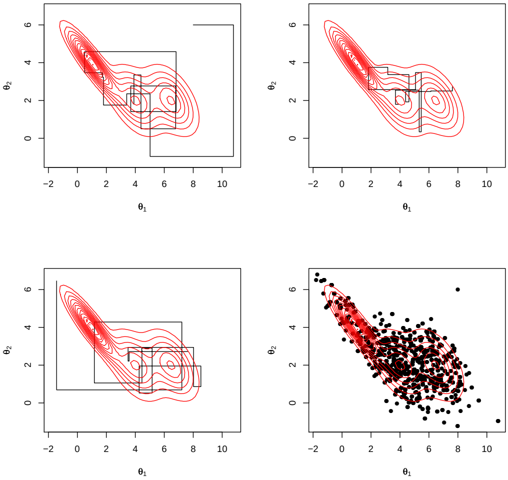
Example iv. RW Metropolis



$q(\theta|\theta_i) \sim N(\theta_i, 0.25\Sigma_2)$.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

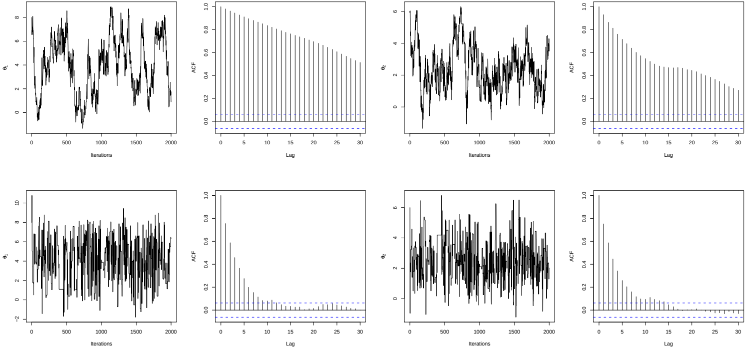
Example iv. Ind. Metropolis



$$q(\theta) \equiv q_{SIR}(\theta) \sim N(\mu, \Sigma).$$

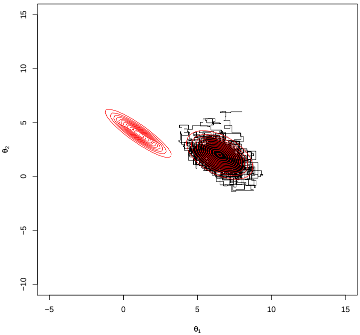
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

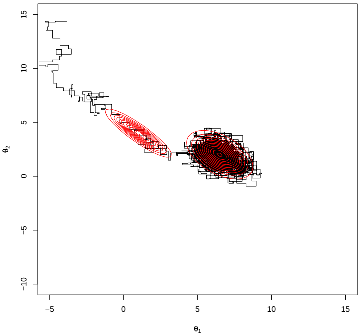
Example iv. Autocorrelations



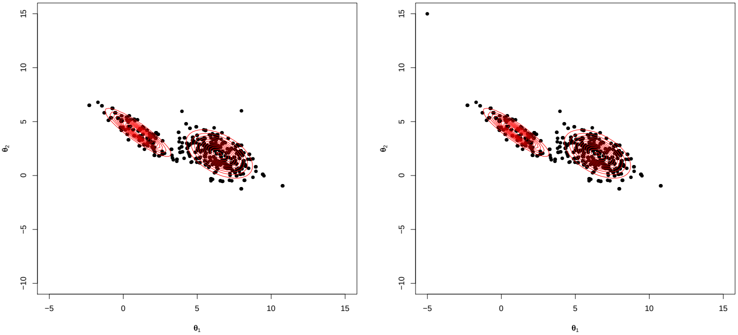
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example v. RW Metropolis

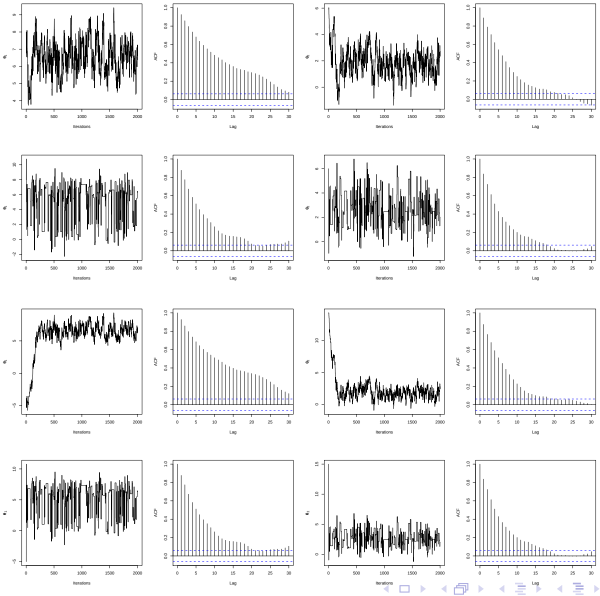




Example v. Ind. Metropolis



Example v. Autocorrelations



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example vi. tuning selection

Tthe target distribution is a two-component mixture of bivariate normal densities, ie:

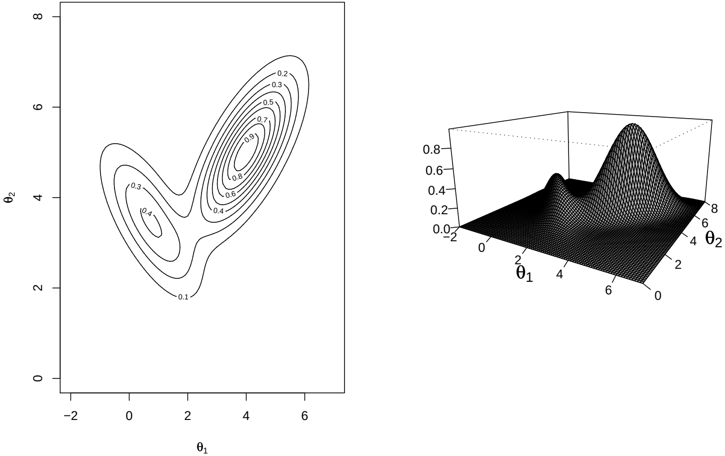
$$\pi(\theta) = 0.7 f_N(\theta; \mu_1, \Sigma_1) + 0.3 f_N(\theta; \mu_2, \Sigma_2).$$

where

$$\begin{aligned}
\mu_1' &= (4.0, 5.0) \\
\mu_2' &= (0.7, 3.5) \\
\Sigma_1 &= \begin{pmatrix} 1.0 & 0.7 \\ 0.7 & 1.0 \end{pmatrix} \\
\Sigma_2 &= \begin{pmatrix} 1.0 & -0.7 \\ -0.7 & 1.0 \end{pmatrix}.
\end{aligned}$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

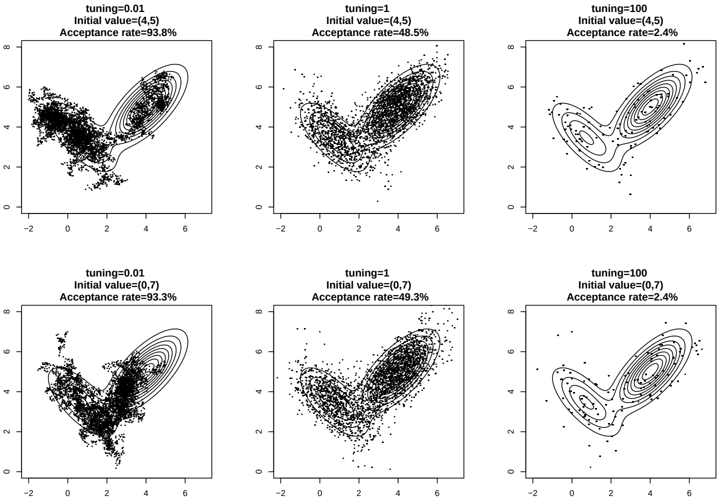
Target distribution



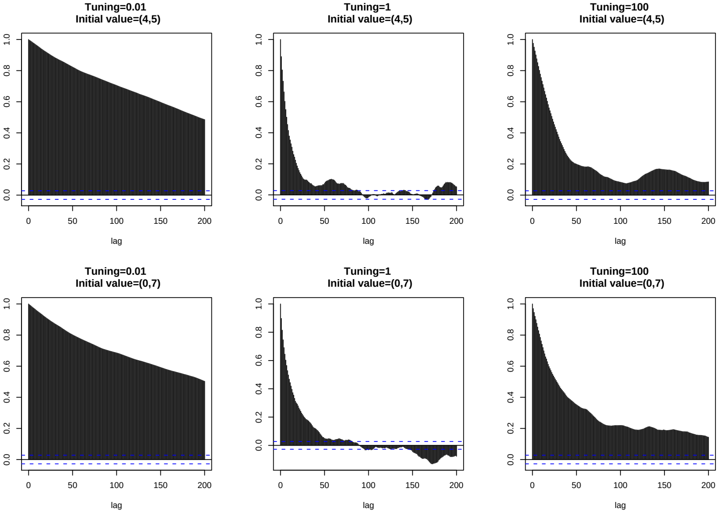
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

RW Metropolis

$q(\theta,\phi) = f_N(\phi;\theta,\nu I_2)$ and ν =tuning.

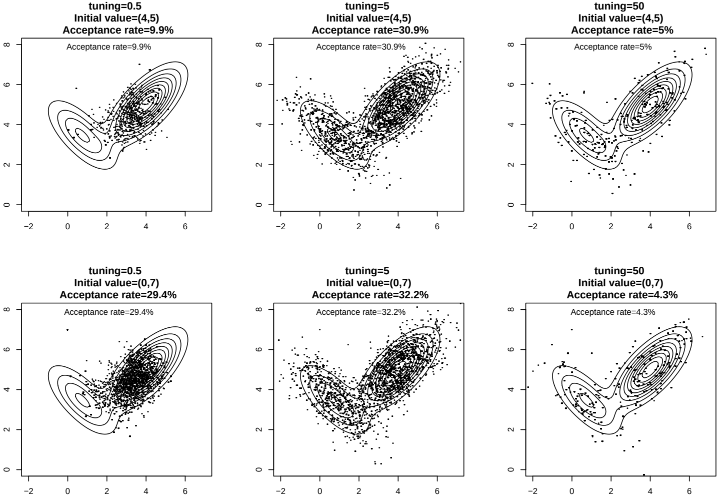


Autocorrelations



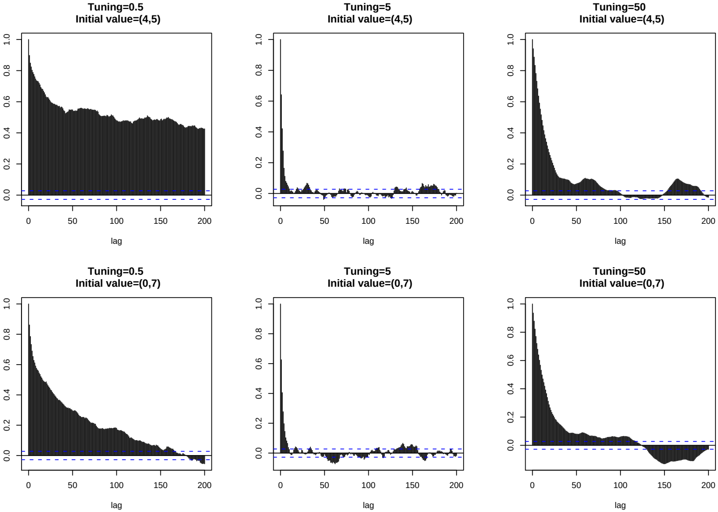
Independent Metropolis

$q(\theta, \phi) = f_N(\phi; \mu_3, \nu I_2)$ and $\mu_3 = (3.01, 4.55)'$.



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Autocorrelations



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Simulated annealing

Simulated annealing[2] is an optimization technique designed to find maxima of functions.

It can be seen as a M-H algorithm that *tempers* with the target distribution:

$$q(\theta) \propto \pi(\theta)^{1/T}$$

where the constant $T > 1$ receives the physical interpretation of system temperature, hence the nomenclature used (Jennison, 1993).

The *heated* distribution q is flattened with respect to π and its density gets closer to the uniform distribution, which is particularly relevant for the case of a distribution with distant modes.

By flattening the modes, the moves required to cover adequately the parameter space become more likely.

[2] Kirkpatrick, Gelatt and Vecchi (1983)

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example vii: Nonlinear surface

Assume that the goal is to find the mode/maximum of

$$\pi(\beta_1, \beta_2) \propto \prod_{i=1}^{4} \frac{e^{(\beta_1+\beta_2 x_i) y_i}}{(1 + e^{\beta_1+\beta_2 x_i})^5},$$

with $x = (-0.863, -0.296, -0.053, 0.727)$ and $y = (0, 1, 3, 5)$.

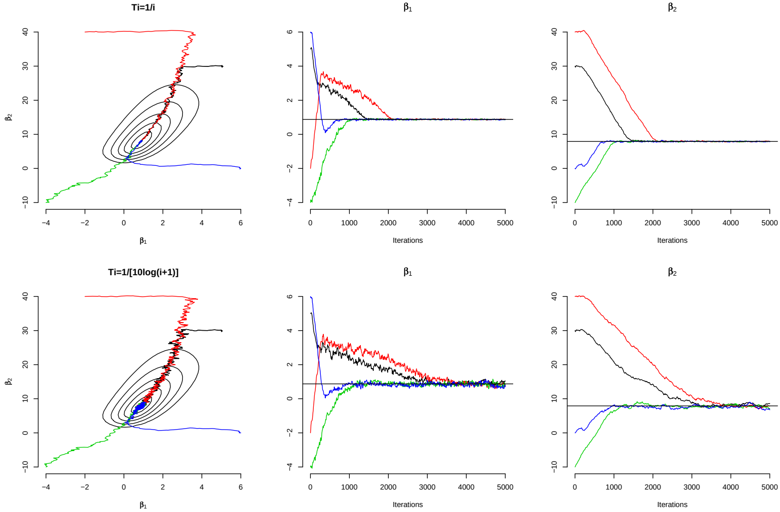
The simulated annealing algorithm is implemented for four initial values:

$$(5, 30) \quad (-2, 40) \quad (-4, -10) \quad (6, 0)$$

and two cooling schedules:

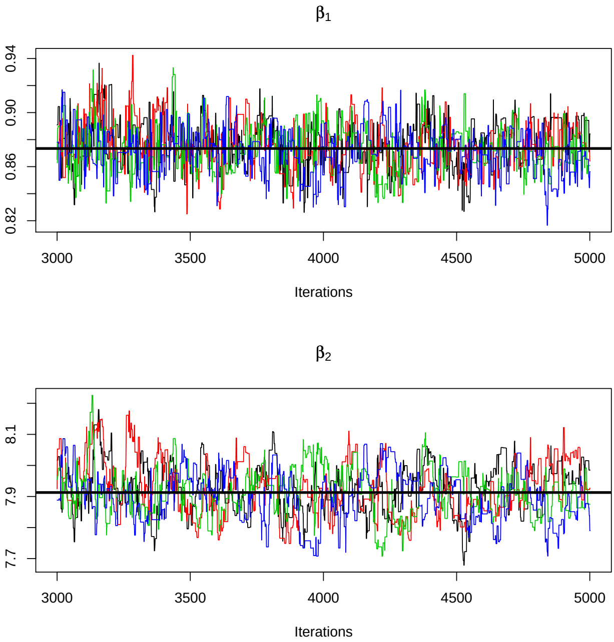
$$T_i = 1/i \quad \text{and} \quad T_i = 1/[10 \log(1 + i)].$$

The proposal distribution is $q(\beta|\beta^{(i)}) = f_N(\beta; \beta^{(i)}, 0.05^2 I_2)$.



Newton-Raphson mode: $(0.87, 7.91)$.

$T_i = 1/i$: mode is $(0.88, 7.99)$ when $(\beta_1^{(0)}, \beta_2^{(0)}) = (5, 30)$.


β1
0.94
0.90
0.86
0.82
3000
3500
4000
4500
5000
Iterations
β2
8.1
7.9
7.7
3000
3500
4000
4500
5000
Iterations

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Gibbs sampler

Technically, the Gibbs sampler is an MCMC scheme whose transition kernel is the product of the full conditional distributions.

Algorithm

1. Start at $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots)$
2. Sample the components of $\theta^{(j)}$ iteratively:

$$\begin{array}{rcl}
\theta_1^{(j)} & \sim & \pi(\theta_1 | \theta_2^{(j-1)}, \theta_3^{(j-1)}, \ldots) \\
\theta_2^{(j)} & \sim & \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \ldots) \\
\theta_3^{(j)} & \sim & \pi(\theta_3 | \theta_1^{(j)}, \theta_2^{(j)}, \ldots) \\
 & \vdots &
\end{array}$$

The Gibbs sampler opened up a new way of approaching statistical modeling by combining simpler structures (the full conditional models) to address the more general structure (the full model).

Example viii: Bivariate normal

Assume that the target distribution is the bivariate normal with mean vector and covariance matrix given by

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},$$

respectively.

In this case, the two full conditionals are given by

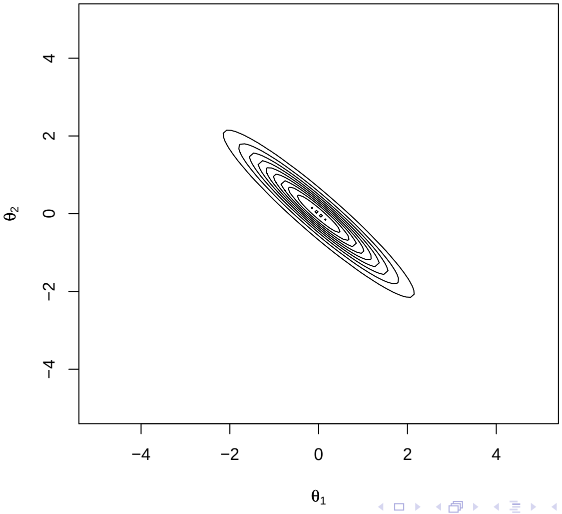
$$\theta_1|\theta_2 \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(\theta_2 - \mu_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right)$$

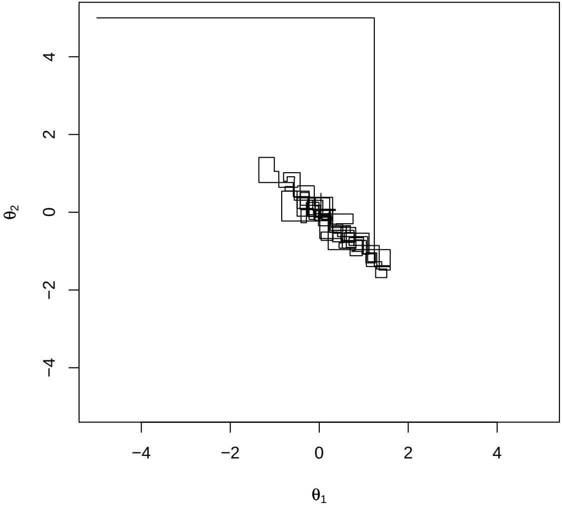
and

$$\theta_2|\theta_1 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(\theta_1 - \mu_1), \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}\right)$$

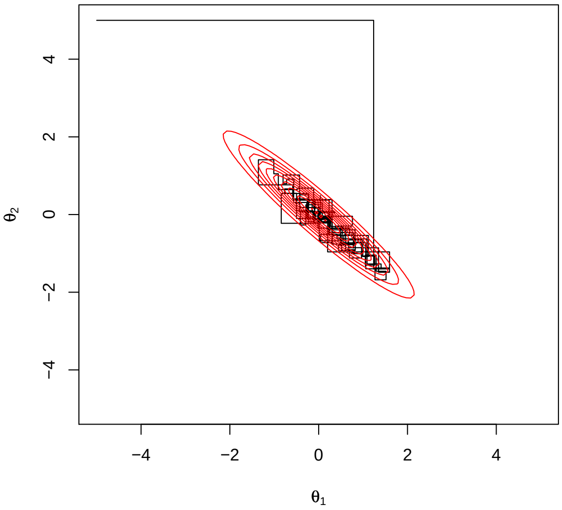
A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

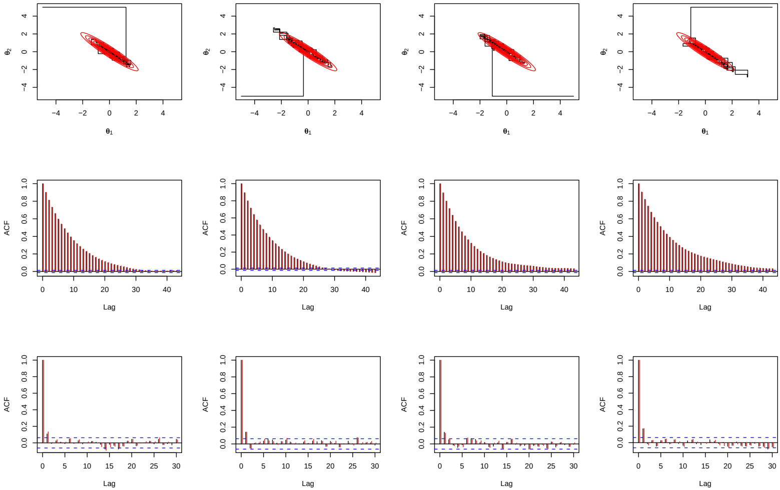
$$\mu_1 = \mu_2 = 0$$
$$\sigma_1^2 = \sigma_2^2 = 1$$
$$\sigma_{12} = -0.95$$



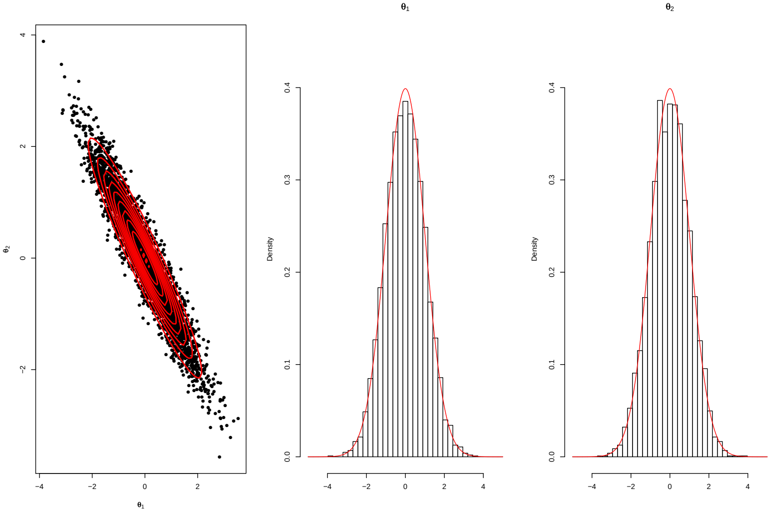

4
2
0
−2
−4
θ_2
−4
−2
0
2
4
θ_1

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References





Middle frame: Based on $M = 21,000$ consecutive draws.
Bottom frame: Based on $M = 1000$ draws, after initial $M_0 = 1000$ draws and saving every 20th draws.


θ₁
θ₂
θ₂
θ₁
Density
Density
-4
-2
0
2
4
0.0
0.1
0.2
0.3
0.4

Example: Seemingly Unrelated Regressions

Investments

Grunfeld (1958), Boot and White (1960) and Zellner (1962,1963) study

$$I_{mt} = \beta_{m1} + \beta_{m2} F_{mt} + \beta_{m3} C_{mt} + \varepsilon_{mt}$$

- Should the parameters be the same across firms?
- Do the ε_{mt} share unobserved common factors?
- Staking the observation for firm m:

$$y_m = X_m \beta_m + \varepsilon_m$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Capital asset pricing

For a given security, the CAPM specifies that

$$r_{mt} - r_{ft} = \alpha_m + \beta_m(r_{mt} - r_{ft}) + \varepsilon_{mt}$$

for return on security m, r_{mt}, return on a risk-free security, r_{ft}, and market return, r_{mt}.

- Are the disturbances correlated across securities?
- Are the α_ms and/or β_ms related in any way?
- Staking the observation for security m:

$$y_m = X_m\beta_m + \varepsilon_m$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Gross State Product

Greene (2008) examines (his examples 9.9 and 9.12) Munnell's (1990) model for output by the 48 continental US states:

$$\begin{aligned} \log GSP_{mt} &= \beta_{m1} + \beta_{m2} \log pcap_{mt} + \beta_{m3} \log hwy_{mt} \\ &+ \beta_{m4} \log water_{mt} + \beta_{m5} \log util_{mt} \\ &+ \beta_{m6} \log emp_{mt} + \beta_{m7} unemp_{mt} + \varepsilon_{mt} \end{aligned}$$

- Should the coefficient vector be the same across states?
- Should the disturbances correlated across states?
- Should the disturbances correlated across time?
- Staking the observation for state m:

$$y_m = X_m \beta_m + \varepsilon_m$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

SUR

For $m = 1, \ldots, M$ and $t = 1, \ldots, T$

$$y_{mt} = x'_{mt}\beta_m + \epsilon_{mt},$$

with x_{mt} a k_m-dimensional vector of regressors.

Let us stack all equations:

$$\begin{array}{rcll} y_t & = & (y_{1t}, \ldots, y_{Mt})' & (M \times 1) \\ \varepsilon_t & = & (\varepsilon_{1t}, \ldots, \varepsilon_{Mt})' & (M \times 1) \\ \beta & = & (\beta'_1, \ldots, \beta'_M)' & (k \times 1) \\ X_t & = & \text{diag}(x'_{1t}, \ldots, x'_{Mt}) & (M \times k) \end{array}$$

where $k = \sum_{m=1}^{M} k_m$. Therefore,

$$y_t = X_t\beta + \varepsilon_t$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

We can now stack all observations $t = 1, \ldots, T$ together:

$$\begin{aligned} y &= (y_1, \ldots, y_T)' \\ \varepsilon &= (\varepsilon_t, \ldots, \varepsilon_T)' \\ X &= (X_1', \ldots, X_T')', \end{aligned}$$

such that

$$y = X\beta + \varepsilon.$$

NLRM: ε_{mt} are i.i.d. $N(0, \sigma^2)$ for all m and t.

SUR: ε_t are i.i.d. $N(0, \Sigma)$ for all t.

This leads to $\varepsilon \sim N(0, \Omega)$, where

$$\Omega = \text{diag}(\Sigma, \ldots, \Sigma) = I_T \otimes \Sigma$$

is an $MT \times MT$ block-diagonal covariance matrix.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Prior distribution

Conditionally conjugate prior for β and $\Phi = \Sigma^{-1}$:

$$p(\beta, \Phi) = p(\beta)p(\Phi),$$

where

$$\beta \sim N(\beta_0, V_0)$$

and

$$\Phi \sim Wishart(\nu_0, \Phi_0).$$

See Dreze and Richard (1983) and Richard and Steel (1988) for further discussion regarding alternative prior specifications.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Full conditionals

The full conditional distributions are

$$\begin{aligned} \beta|y, X, \Sigma &\sim N(\beta_1, V_1) \\ \Phi|y, X, \beta &\sim Wishart(\nu_1, \Phi_1) \end{aligned}$$

where $\nu_1 = \nu_0 + T$,

$$\begin{aligned} V_1^{-1} &= V_0^{-1} + \sum_{t=1}^{T} X_t' \Phi X_t \\ V_1^{-1}\beta_1 &= V_0^{-1}\beta_0 + \sum_{t=1}^{n} X_t' \Phi y_t \\ \Phi_1^{-1} &= \Phi_0^{-1} + \sum_{t=1}^{T} (y_t - X_t\beta)(y_t - X_t\beta)' \end{aligned}$$

Grunfeld's (1958) data

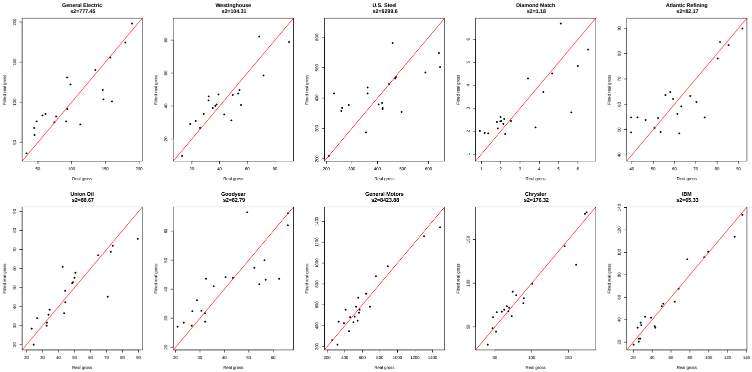
M=10 U.S. firms over T=20 years, 1935-1954.[3]

Variables:
FN = Firm Number; YR = Year;
I = Annual real gross investment;
F = Real value of the firm (shares outstanding); and
C = Real value of the capital stock.

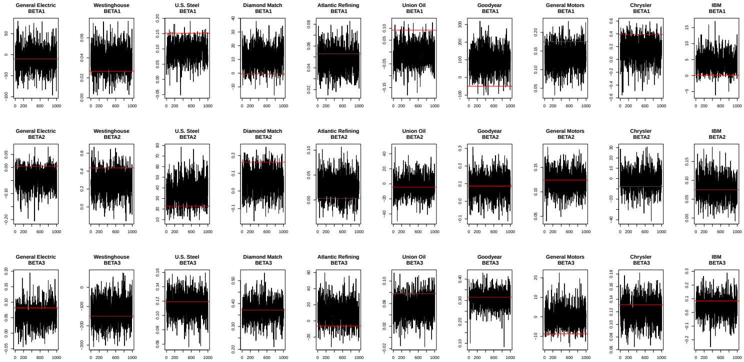
Firms:
General Electric, Westinghouse,
U.S. Steel, Diamond Match,
Atlantic Refining, Union Oil,
Goodyear, General Motors,
Chrysler and IBM

[3] Zellner (1971), pages 240-246.

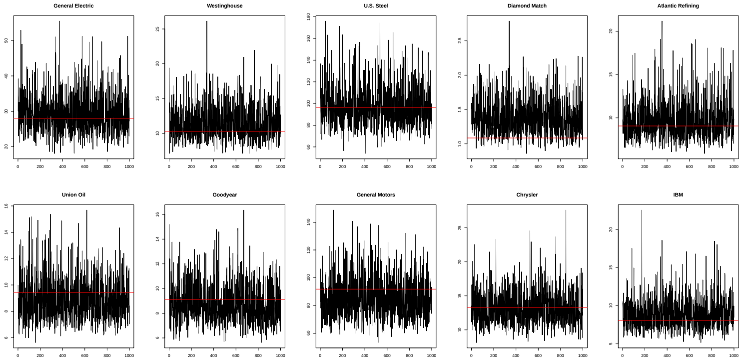
Individual regressions



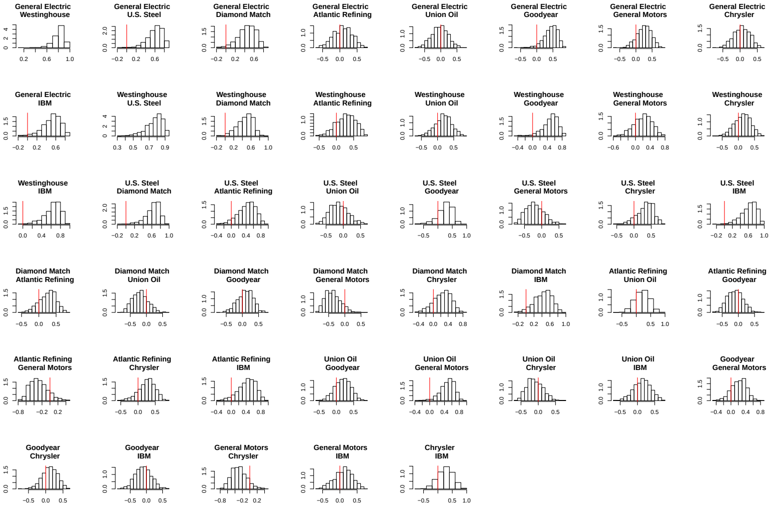
Regression coefficients



Standard deviations



Correlations



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

SUR references

Boot and De Wit (1960) Investment Demand: An Empirical Contribution to the Aggregation. Problem, *International Economic Review*, 1, 3-30.

Zellner (1971) *An Introduction to Bayesian Inference and Econometrics*, Wiley.

Koop (2003) *Bayesian Econometrics*, Wiley.

Kleiber and Zeileis (2010) The Grunfeld Data at 50, *German Economic Review*, 11(4), 404-417.

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Example: Bivariate BVAR(1)

Let $y_t = (y_{t1}, y_{t2})'$ contain 2 time series observed at time t.

The (basic) VAR(1) can be written as

$$(y_t | y_{t-1}, B, \Sigma) \sim N(By_{t-1}, \Sigma)$$

where

$$B = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

VAR(1) as a SUR model

The above VAR(1) model can be rewritten as a SUR model as

$$(y_t|z_t, \beta, \Sigma) \sim N(z_t\beta, \Sigma)$$

where

$$z_t = \begin{pmatrix} y_{t-1,1} & y_{t-1,2} & 0 & 0 \\ 0 & 0 & y_{t-1,1} & y_{t-1,2} \end{pmatrix}$$

and

$$\beta = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{pmatrix}.$$

Therefore,

$$(y|\beta, \Sigma) \sim N(Z\beta, \Sigma)$$

where $y = (y_1', \ldots, y_T')'$ and $Z = (z_1', \ldots, z_T')'$.

Prior of (β, Σ)

We assume that β and Σ are independent *a priori*.

Prior of β:

$$\beta \sim N(b_0, B_0).$$

Prior of Σ:

$$\Sigma \sim IW(v_0, V_0).$$

This conditionally conjugate prior DOES NOT lead to closed form posterior inference, but the implementation of the Gibbs sampler is straightforward.

Full conditional of β

It is easy to see that

$$\begin{aligned} p(\beta|\Sigma, y) &\propto \exp\left\{-0.5\left[\beta' B_0^{-1}\beta - 2\beta' B_0^{-1}\beta_0\right]\right\} \\ &\times \exp\left\{-0.5\left[\beta' Z'\Sigma^{-1} Z\beta - 2\beta' Z'\Sigma^{-1} y\right]\right\}. \end{aligned}$$

Therefore,

$$\beta|\Sigma, \gamma, y \sim N(\beta_1, V_1)$$

where

$$\beta_1 = B_1(B_0^{-1}\beta_0 + Z'\Sigma^{-1}y) \quad \text{and} \quad B_1^{-1} = B_0^{-1} + Z'\Sigma^{-1}Z.$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Full conditional of Σ

It is easy to see that

$$\begin{aligned} p(\Sigma|\beta, y) &\propto |\Sigma|^{\frac{q+v_0+1}{2}} \exp\left\{-0.5\text{tr}(\Sigma^{-1}V_0)\right\} \\ &\times |\Sigma|^{\frac{q+T+1}{2}} \exp\left\{-0.5\sum_{t=1}^{T}(y_t - z_t\beta)'\Sigma^{-1}(y_t - z_t\beta)\right\}. \end{aligned}$$

Therefore,

$$\Sigma|\beta, y \sim IW(v_1, V_1)$$

where $v_1 = v_0 + T$ and

$$S = V_0 + \sum_{t=1}^{T}(y_t - z_t\beta)(y_t - z_t\beta)'.$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Simulated data

We simulated $n = 360$ observations (30 years of monthly data) from the above bivariate VAR(1) with

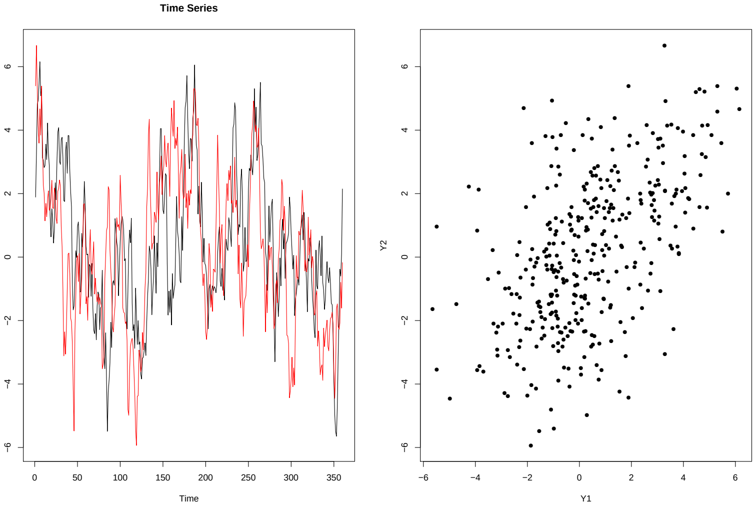
$$B = \begin{pmatrix} 0.85 & 0.10 \\ 0.00 & 0.95 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{pmatrix}$$

A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Data



Posterior inference

The prior hyperparameters are

$$b_0 = 0_4 \quad \text{and} \quad B_0 = 1000 I_4$$

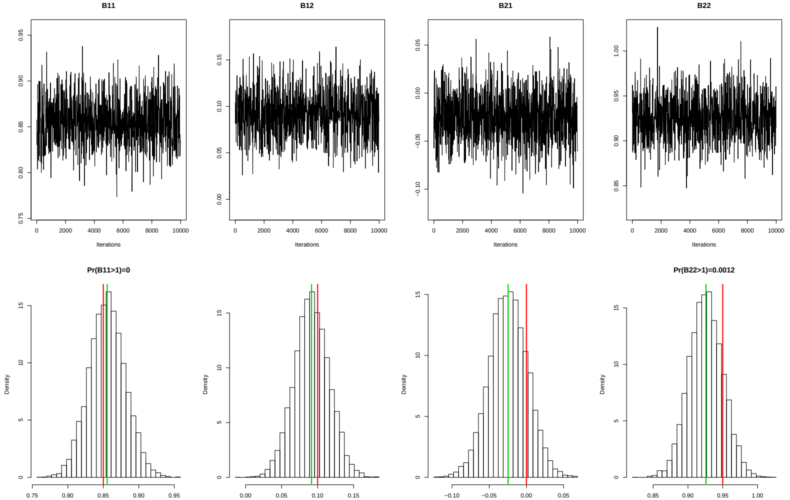
and

$$v_0 = 5 \quad \text{and} \quad V_0 = 0.001$$

We started the Gibbs sampler with $B^{(0)} = B$ (true value).

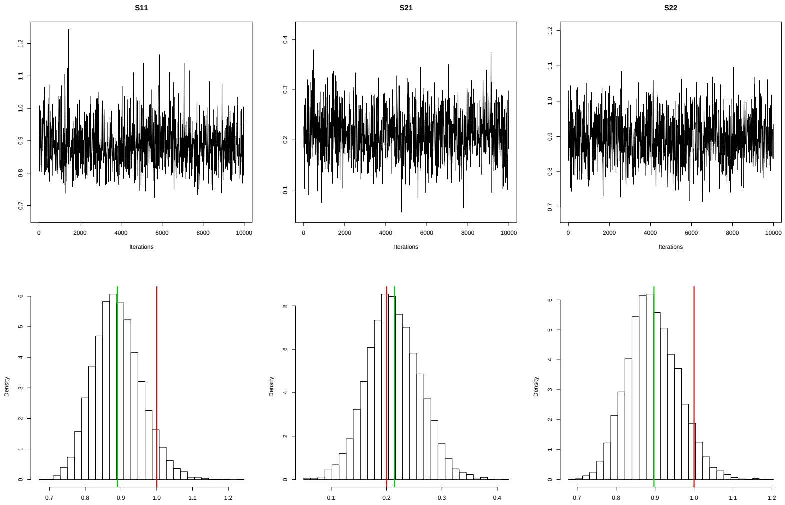
We runt he Gibbs sampler for $M = 10,000$ iterations.

$p(B|\text{data})$



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

$p(\Sigma|\text{data})$



A bit of history
MC integration
MC via IS
Rejection method
SIR method
MH algorithms
Simulated annealing
Gibbs sampler
SUR model
BVAR
References

Book references

1. Ripley (1987) Stochastic Simulation. Wiley. (Chapters 5 and 7).
2. Thisted (1988) Elements of Statistical Computing. Chapman & Hall/CRC. (Chapter 5).
3. Gentle (1998) Random Number Generation and Monte Carlo Methods. Springer. (Chapter 5).
4. Liu (2001) Monte Carlo Strategies in Scientific Computing. Springer. (Chapters 2, 5 and 6).
5. Robert and Casella (2004) Monte Carlo Statistical Methods. Springer. (Chapters 3, 4 and 7).
6. Givens and Hoeting (2005) Computational Statistics. Wiley. (Chapters 6, 7 and 8).
7. **Gamerman and Lopes (2006) MCMC: Stochastic Simulation for Bayesian Inference. Chapman & Hall/CRC. (Chapters 3, 5 and 6).**
8. Rizzo (2008) Statistical Computing with R. Chapman & Hall/CRC. (Chapters 5 and 9).
9. Rubinstein and Kroese (2008) Simulation and the Monte Carlo Method. Wiley. (Chapters 5 and 6).
10. Robert and Casella (2010) Introducing Monte Carlo Methods with R. Springer. (Chapters 3, 6 and 7).

A bit of history

MC integration

MC via IS

Rejection method

SIR method

MH algorithms

Simulated annealing

Gibbs sampler

SUR model

BVAR

References

Seminal papers

1. Metropolis and Ulam (1949) The Monte Carlo method. JASA, 44, 335-341.
2. Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) Equation of state calculations by fast computing machines. Journal of Chemical Physics, Number 21, 1087-1092.
3. Hastings (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57, 97-109.
4. Peskun (1973) Optimum Monte-Carlo sampling using Markov chains. Biometrika, 60, 607-612.
5. Besag (1974) Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, Series B, 36, 192-236.
6. Geman and Geman (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.
7. Jennison (1993) Discussion of the meeting on Gibbs sampling and other Markov chain Monte Carlo methods. JRSS-B, 55, 54-6.
8. Kirkpatrick, Gelatt and Vecchi (1983) Optimization by simulated annealing. Science, 220, 671-80.
9. Pearl (1987) Evidential reasoning using stochastic simulation. Articial Intelligence, 32, 245-257.
10. Tanner and Wong (1987) The calculation of posterior distributions by data augmentation. JASA, 82, 528-550.
11. Geweke (1989) Bayesian inference in econometric models using Monte Carlo integration. Econometrica, 57, 1317-1339.
12. Gelfand and Smith (1990) Sampling-based approaches to calculating marginal densities, JASA, 85, 398-409.
13. Casella and George (1992) Explaining the Gibbs sampler. The American Statistician,46,167-174.
14. Smith and Gelfand (1992) Bayesian statistics without tears: a sampling-resampling perspective. American Statistician, 46, 84-88.
15. Gilks and Wild (1992) Adaptive rejection sampling for Gibbs sampling. Applied Statistics, 41, 337-48.
16. Chib and Greenberg (1995) Understanding the Metropolis-Hastings algorithm. The American Statistician, 49, 327-335.
17. Dongarra and Sullivan (2000) Guest editors' introduction: The top 10 algorithms. Computing in Science and Engineering, 2, 22-23.