Practical Markov Chain Monte Carlo[1]

Galin Jones

University of Minnesota

06 September 2023

[1] An homage to Geyer (1992, Statisical Science)

This Document

This document was written in Rmarkdown.

All simulations, calculations, analyses, and plots were produced internally to the document and should be fully reproducible.

The code used to produce the document is available on my GitHub.

I am quite good at making coding errors. If you find an error in my work, I would deeply appreciate it if you would let me know so that it can be corrected.

GOFMC via Toy Example

Suppose the goal is to calculate an integral, say,

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$$

If f is a pdf:

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} \frac{f(x)}{f(x)} dx = \int_{-\infty}^{\infty} \left[\frac{e^{-\frac{1}{2}x^2}}{f(x)} \right] f(x) dx = E_f \left[\frac{e^{-\frac{1}{2}x^2}}{f(x)} \right]$$

GOFMC via Toy Example

Suppose $X_1, \ldots, X_m \overset{iid}{\sim} f$

$$\int_{-\infty}^{\infty} \left[\frac{e^{-\frac{1}{2}x^2}}{f(x)} \right] f(x) dx \approx \frac{1}{m} \sum_{i=1}^{m} \frac{e^{-\frac{1}{2}x_i^2}}{f(x_i)}$$

<u>Question</u> How large should m be for the result to be reliable?

<u>Answer</u> Use the width of a confidence interval for a population mean to decide

GOFMC via Toy Example

Suppose $f(x)$ is the pdf for $N(0, 4)$.

The truth is $\sqrt{2\pi} = 2.5066283$.

After 100 simulated observations the sample mean is 2.2555

A 95% confidence interval is (1.8978, 2.6131)

After 100 samples this is not a good estimate.

Need more samples.

GOFMC via Toy Example

After 10^4 samples the sample mean is 2.4869

A 95% confidence interval is (2.4518, 2.5219)

The truth is $\sqrt{2\pi} = 2.5066283$

Metropolis-Hastings

Practically relevant settings prohibit $X_1, \ldots, X_m \overset{ind}{\sim} f$

Metropolis-Hastings

Given $X_t = x$, draw $Y \sim q_\gamma(\cdot \mid x)$

Draw $U \sim Unif(0,1)$ and set $X_{t+1} = y$ if

$$u \leq \frac{f(y) q_\gamma(x \mid y)}{f(x) q_\gamma(y \mid x)}$$

otherwise set $X_{t+1} = x$.

Choice of γ is crucial.

MH via Toy Example

Recall that f is a $N(0,4)$ pdf. Suppose q_γ is a $N(x,\gamma^2)$ pdf.

Given $X_t = x$, draw $Y \sim N(x,\gamma^2)$

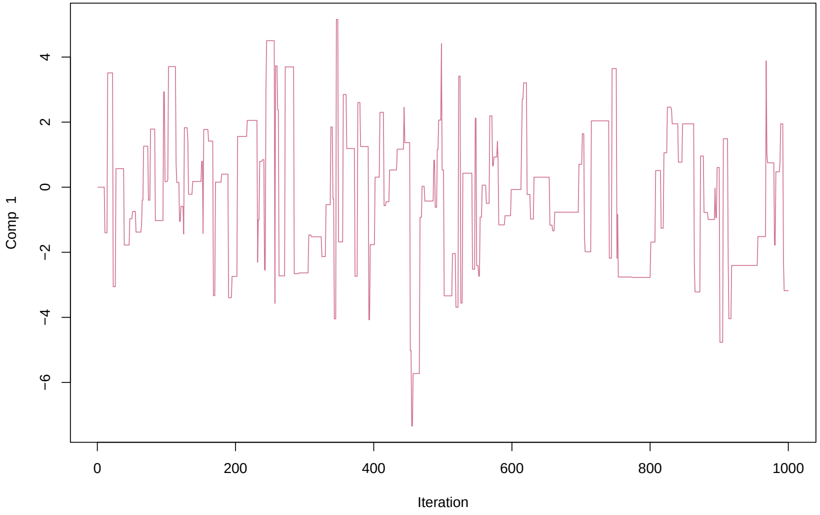
Draw $U \sim Unif(0,1)$ and set $X_{t+1} = y$ if

$$u \leq \frac{f(y)}{f(x)} = \frac{e^{-\frac{1}{8}y^2}}{e^{-\frac{1}{8}x^2}} = e^{-\frac{1}{8}(y^2 - x^2)}$$

otherwise set $X_{t+1} = x$.

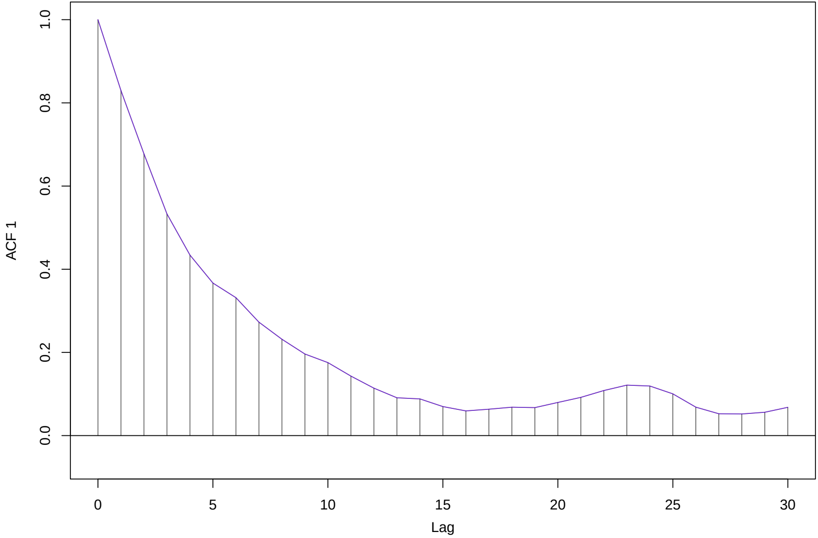
MH via Toy Example

After 1000 iterations using $\gamma = 16$



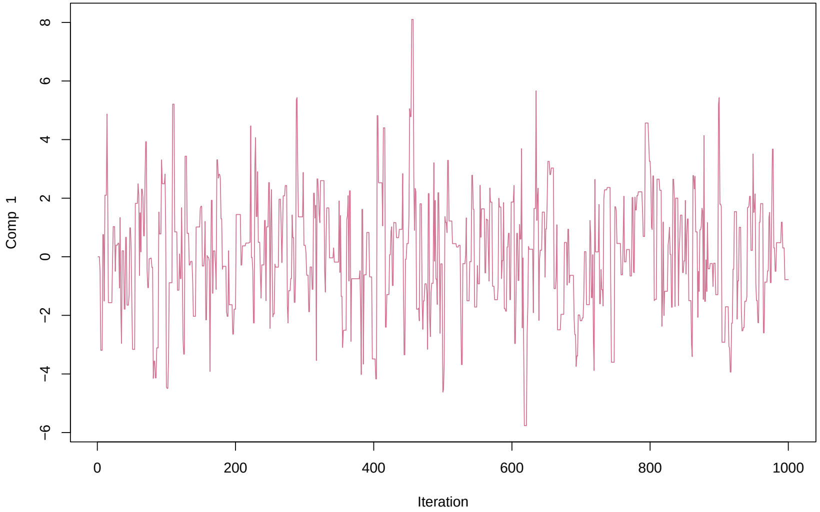
MH via Toy Example

After 1000 iterations using $\gamma = 16$



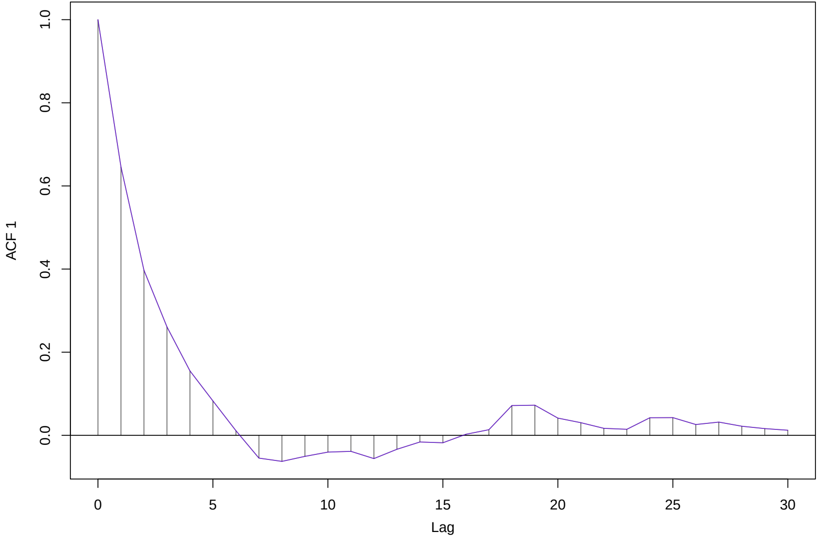
MH via Toy Example

After 1000 iterations using $\gamma = 4$



MH via Toy Example

After 1000 iterations using $\gamma = 4$



MH via Toy Example

Recall that $\sqrt{2\pi} = 2.5066283$

After 1000 iterations of MH using $\gamma = 16$ the estimate is 2.2893 and a 95% confidence interval is (1.6844, 2.8942)

After 1000 iterations of MH using $\gamma = 4$ the estimate is 2.5394 and a 95% confidence interval is (2.2645, 2.8047)

More samples are required to have a reliable estimate.

MH via Toy Example

After 5×10^4 iterations of MH using $\gamma = 4$ the estimate is 2.5104 and a 95% confidence interval is (2.4696, 2.5509)

Markov chain Monte Carlo

A typical goal of an MCMC simulation experiment is to estimate

$$\mu = \int g(x) f(x) dx = E_f[g(X)]$$

by simulating a realization of a Markov chain

$$X_1, X_2, X_3, \ldots$$

which satisfies $X_m \xrightarrow{d} f$ as $m \to \infty$

Eventually a representative (if dependent and non-identically distributed) sample from f will be produced

Markov chain Monte Carlo

Markov chain SLLN, as $m \to \infty$,

$$\bar{\mu}_m = \frac{1}{m}\sum_{i=1}^{m} g(X_i) \to \int g(x)f(x)dx = \mu$$

Markov chain CLT, as $m \to \infty$,

$$\sqrt{m}(\bar{\mu}_m - \mu) \xrightarrow{d} N(0, \sigma^2)$$

But

$$\sigma^2 = Var_f[g(X)] + 2\sum_{k=1}^{\infty} Cov_f[g(X_1), g(X_k)]$$

$(1-\alpha)100\%$ confidence interval

$$\bar{\mu}_m \pm t_{\alpha,df}\frac{\hat{\sigma}_m}{\sqrt{m}}$$

Estimating σ^2

$$\sigma^2 = Var_f[g(X)] + 2\sum_{k=1}^{\infty} Cov_f[g(X_1), g(X_k)]$$

Initial Sequence Estimators (Geyer (1992, Statistical Science), Dai and Jones (2017, J. Multivariate Analysis))

Spectral Variance Estimators (Flegal and Jones (2010, Annals of Statistics), Vats, Flegal, and Jones (2017, Bernoulli))

Batch means (Jones, Haran, Caffo, Neath (2006, J. American Statistical Association), Vats, Flegal, and Jones (2019, Biometrika))

Implemented in SimTools, available on GitHub

Markov chain CLT

The Markov chain CLT can easily fail.

Suppose f is an Exponential(1) density

$$f(x) = e^{-x} I(x \geq 0)$$

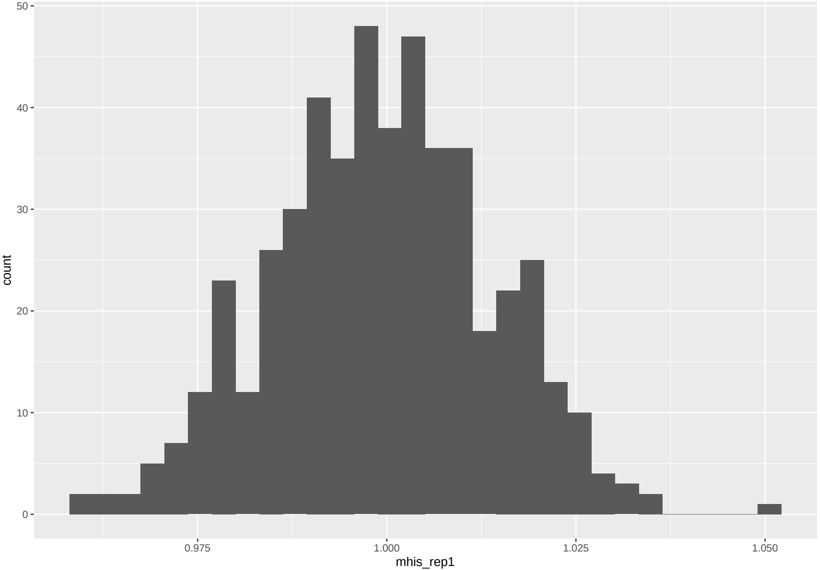
Consider using MH with proposal Exponential(γ) density.

A Markov chain CLT holds if $\gamma \leq 1$ but if $\gamma > 2$, then

$$\sigma^2 = \infty$$

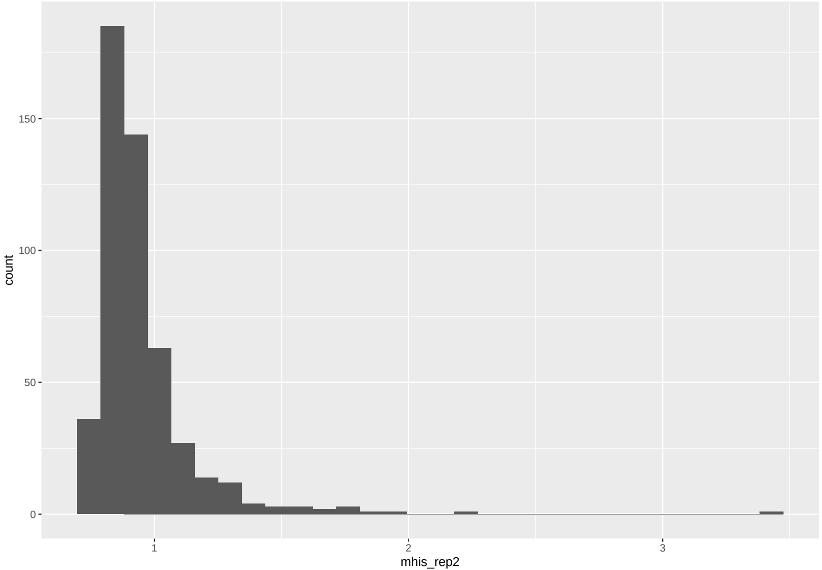
Markov chain CLT

500 sample means, for 10^4 observations using Exp(0.5) proposal



Markov chain CLT

500 sample means, for 10^4 observations using Exp(3) proposal



Ensuring Reliable MCMC

Recall that MH is constructed to ensure that

$$X_m \xrightarrow{d} f \qquad m \to \infty$$

The key is that this convergence has to be fast.

Question Can we identify when it will be slow?

Identifying Slow Convergence of MH

$$A(x) = \int \left(1 \wedge \frac{f(y)q_\gamma(x \mid y)}{f(x)q_\gamma(y \mid x)}\right) q_\gamma(y \mid x)dy$$

Then

$$1 \geq dist(\mathcal{L}(X_t), f) \geq (1 - A(x))^t$$

If $A(x) \approx 0$, then the convergence will be slow.

Identifying Slow Convergence of MH

$$
\begin{aligned}
A(x) &= \int \left(1 \wedge \frac{f(y)q_\gamma(x \mid y)}{f(x)q_\gamma(y \mid x)}\right) q_\gamma(y \mid x)dy \\
&= \int \left(\frac{q_\gamma(y \mid x)}{f(y)} \wedge \frac{q_\gamma(x \mid y)}{f(x)}\right) f(y)dy \\
&\leq \int \left(\frac{q_\gamma(x \mid y)}{f(x)}\right) f(y)dy
\end{aligned}
$$

Then

$$
1 \geq dist(\mathcal{L}(X_t), f) \geq \left(1 - \int \left(\frac{q_\gamma(x \mid y)}{f(x)}\right) f(y)dy\right)^t
$$

Exp(1) Example

f is an Exp(1) and the MH proposal is an Exp(γ) density.

Then

$$1 \geq dist(\mathcal{L}(X_t), f) \geq \left(1 - \int \left(\frac{q_\gamma(x)}{f(x)}\right) f(y)dy\right)^t = \left(1 - \frac{q_\gamma(x)}{f(x)}\right)^t$$

Suppose $x = 1$, then

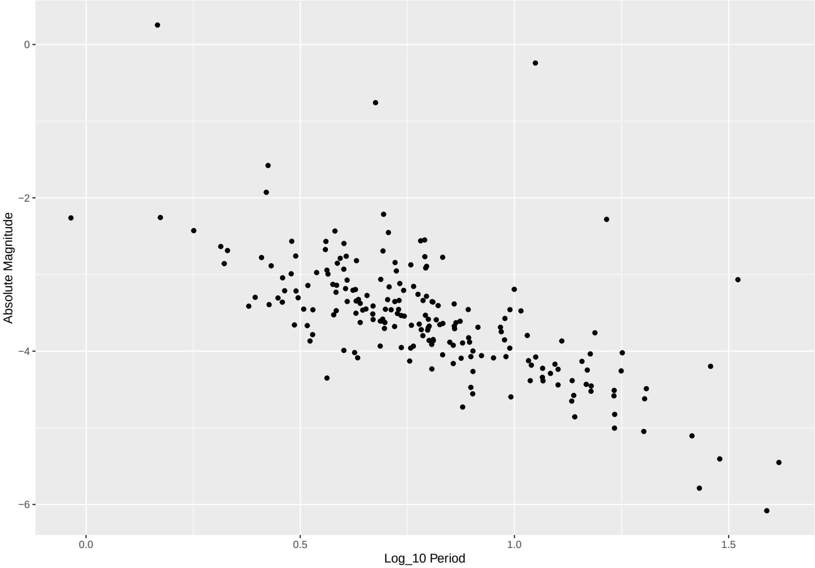
$$1 \geq dist(\mathcal{L}(X_t), f) \geq \left(1 - \gamma e^{(1-\gamma)}\right)^t$$

so if γ is "large" the convergence will be slow.

Take-Home Summary

- The lessons from GOFMC carry over to MCMC
- Terminate the simulation based on the estimated Monte Carlo error
- Theoretical study is required to ensure a CLT
- Theoretical study of MCMC algorithms can be challenging, but there are some easy ways to avoid guaranteed poor behavior
- Starting values should be any point you don't mind having in the sample–use optimization
- Burn-in if you must, but you should take little solace in it
- Convergence diagnostics should also provide little solace

Cepheid Period-Luminosity



Cepheid Period-Luminosity

Suppose

$$Y_i | X_i, \beta_0, \beta_1, \lambda \sim N(\beta_0 + \beta_1 x_i, \lambda^{-1})$$

with priors

$$\beta_0 \sim N(-1.43, 4) \qquad \beta_1 \sim N(-2.8, 4) \qquad \lambda \sim Gamma(2, 2)$$

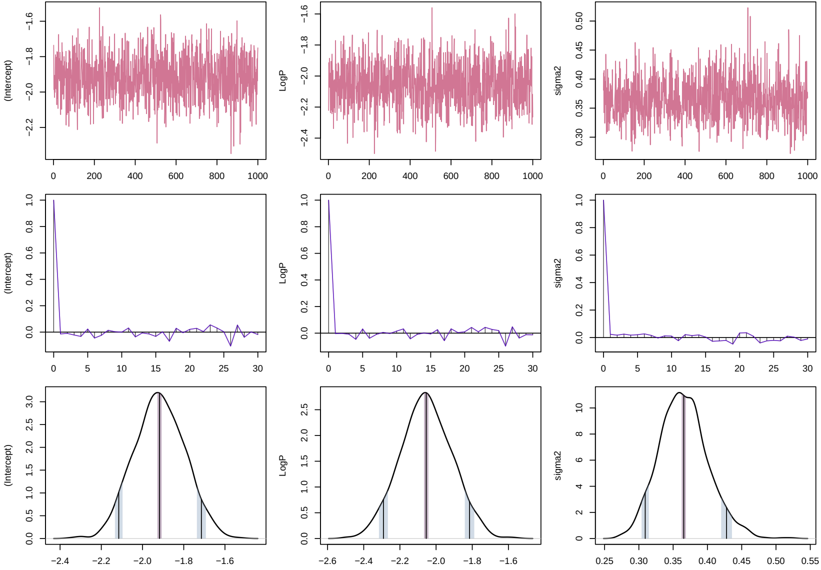
Then the posterior exists

$$q(\beta_0, \beta_1, \lambda \mid y_1, x_1, \ldots, y_n, x_n)$$

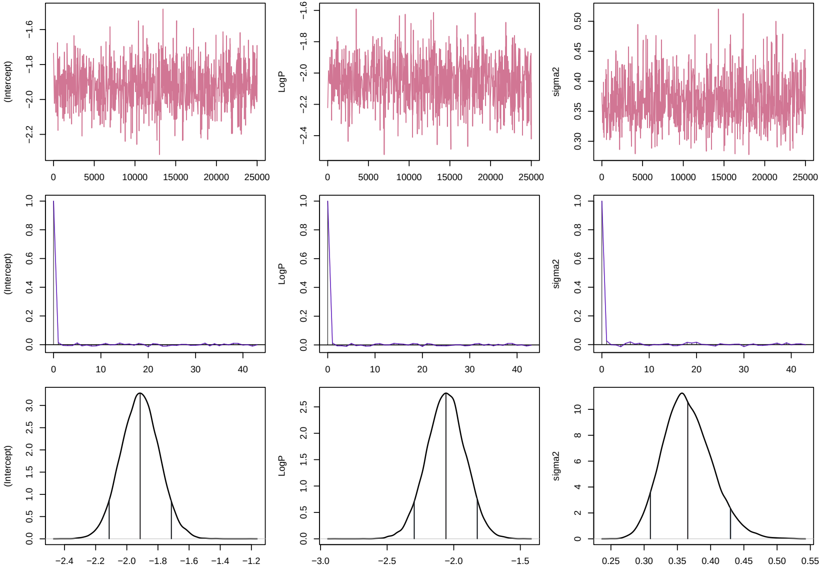
The MCMC algorithm has been proved to converge quickly so a CLT exists.

Plot the marginal posterior densities $q(\beta_0 \mid y_1, x_1, \ldots, y_n, x_n)$, $q(\beta_1 \mid y_1, x_1, \ldots, y_n, x_n)$, $q(\lambda \mid y_1, x_1, \ldots, y_n, x_n)$

Cepheid Period-Luminosity



Cepheid Period-Luminosity



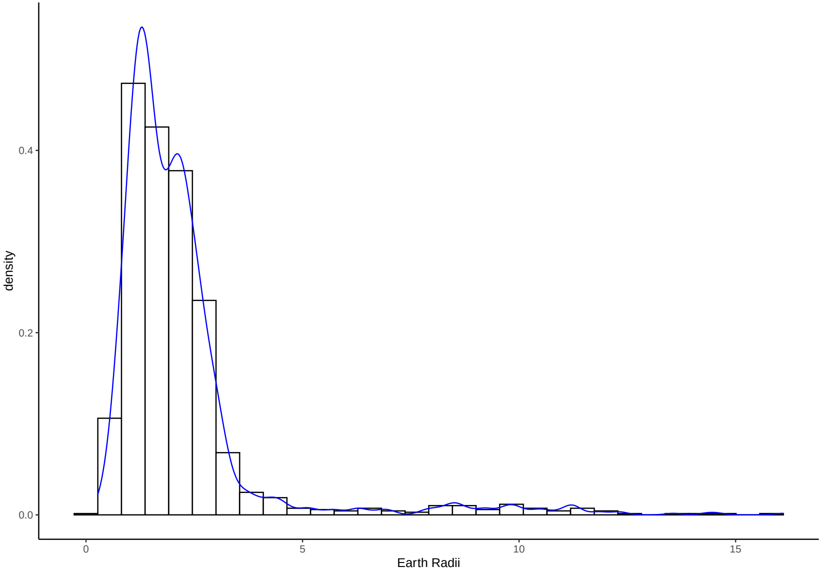
Cepheid Period-Luminosity

The posterior mean for the intercept is -1.914 and the 0.90-credible interval is (-2.113, -1.714).

The posterior mean for the slope is -2.059 and the 0.90-credible interval is (-2.297, -1.823).

The posterior mean for the precision is 0.3656 and the 0.90-credible interval is (0.3094, 0.4299).

Radii of Extrasolar Planets



Radii of Extrasolar Planets

For $i = 1, \ldots, n$ assume $X_i \stackrel{iid}{\sim} LogNormal(\mu, \lambda)$

Improper prior $\nu(\mu, \lambda) = 1/\lambda$

The posterior

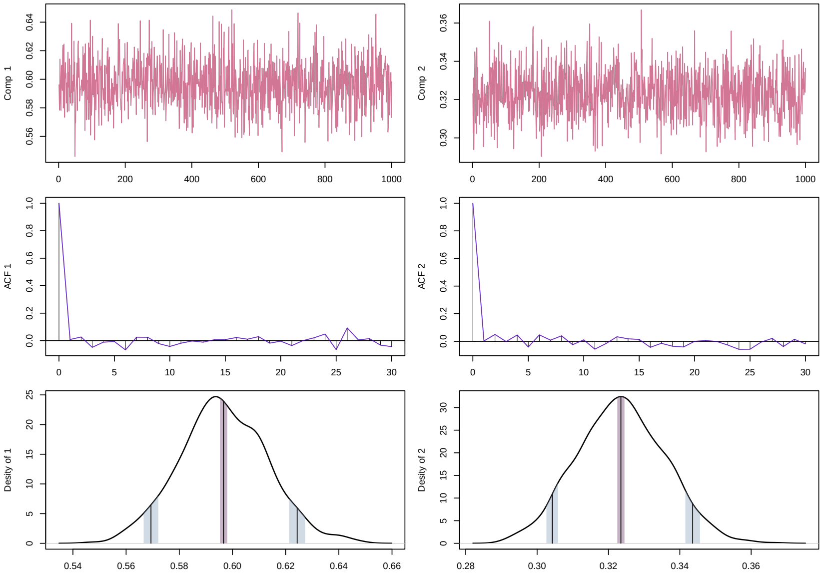
$$q(\mu, \lambda \mid x_1, \ldots, x_n)$$

exists

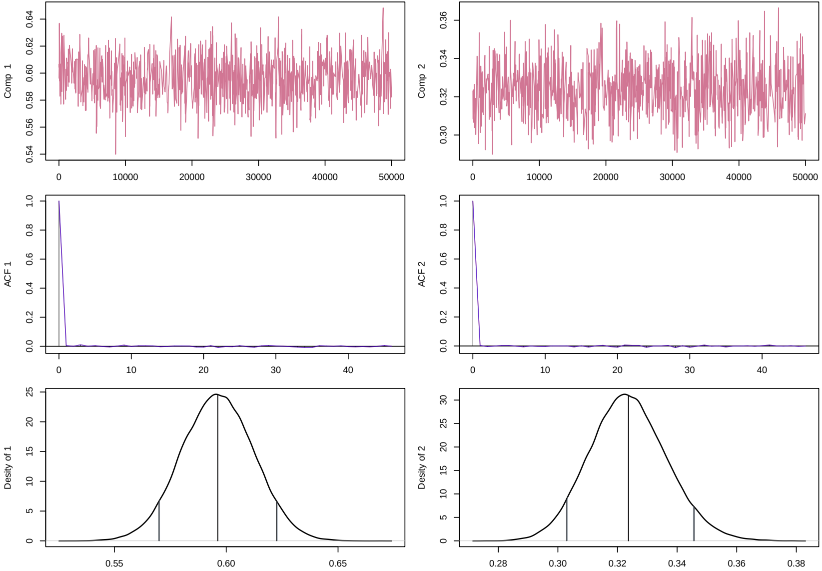
The MCMC algorithm has been proved to converge quickly so a CLT exists.

Plot the marginal posterior densities $q(\mu \mid x_1, \ldots, x_n)$ and $q(\lambda \mid x_1, \ldots, x_n)$

Radii of Extrasolar Planets



Radii of Extrasolar Planets



Radii of Extrasolar Planets

Posterior mean radii is 2.135 and 0.95-credible interval (2.064, 2.209).