MCMC Output Analysis with R package mcmcse

University of Warwick

Dootika Vats

22/06/2017

Collaborators

James Flegal - University of California, Riverside

John Hughes - University of Colorado, Denver

Ning Dai - University of Minnesota, Twin-Cities

R package `mcmcse`

Goal: output analysis for $\underbrace{\text{Markov chain Monte Carlo}}_{review}$

Highlights

- Univariate and multivariate standard errors for MCMC
- multivariate effective sample size
- minimum effective sample size required

The MCMC Story

There is a complicated integral (expectation).

$$\theta = \int g(x) \underbrace{f(x)}_{\text{prob. density fn}} dx \in \mathbb{R}^d .$$

Draw samples $X_1, X_2, \ldots, X_n$ from distribution with pdf $f(x)$.

$$\hat{\theta}_n = \frac{1}{n} \sum_{t=1}^{n} g(X_t) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta) \overset{d}{\to} N_p(0, \underbrace{\Lambda}_{p \times p}) .$$

The MCMC Story

There is a complicated integral (expectation).

$$\theta = \int g(x) \underbrace{f(x)}_{\text{prob. density fn}} dx \in \mathbb{R}^d .$$

Draw samples $X_1, X_2, \dots, X_n$ from distribution with pdf $f(x)$.

$$\hat{\theta}_n = \frac{1}{n} \sum_{t=1}^{n} g(X_t) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N_p(0, \underbrace{\Lambda}_{p \times p}) .$$

Individual standard errors for $\hat{\theta}_n^{(i)}$ is $\lambda_{ii}/\sqrt{n}$. sample variance

$$\hat{\lambda}_{ii}^2 = \frac{1}{n-1} \sum_{t=1}^{n} \left(g(X_t^{(i)}) - \hat{\theta}_n^{(i)} \right)^2 .$$

Standard error matrix for $\hat{\theta}_n$. sample covariance

$$\hat{\Lambda} = \frac{1}{n-1} \sum_{t=1}^{n} \left(g(X_t) - \hat{\theta}_n \right) \left(g(X_t) - \hat{\theta}_n \right)^T .$$

The MCMC Story

- Drawing iid samples is often impossible/hard, so $X_1, X_2, \dots, X_n$ samples a Markov chain with stationary distribution having pdf $f(x)$
  - $X_1, X_2, \dots, X_n$ are correlated .
- However, the usual method still works

$$\widehat{\theta}_n = \frac{1}{n}\sum_{t=1}^{n} g(X_t) \quad \text{and} \quad \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N_p(0, \underbrace{\Sigma}_{\neq \Lambda}) .$$

- Standard errors are tough!

The MCMC Story

- Drawing iid samples is often impossible/hard, so $X_1, X_2, \ldots, X_n$ samples a Markov chain with stationary distribution having pdf $f(x)$
  - $X_1, X_2, \ldots, X_n$ are correlated .
- However, the usual method still works

$$\hat{\theta}_n = \frac{1}{n}\sum_{t=1}^{n} g(X_t) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N_p(0, \underbrace{\Sigma}_{\neq \Lambda}) \, .$$

- Standard errors are tough! correlated samples means Σ is difficult to estimate.

The MCMC Story

- Drawing iid samples is often impossible/hard, so $X_1, X_2, \ldots, X_n$ samples a Markov chain with stationary distribution having pdf $f(x)$
  - $X_1, X_2, \ldots, X_n$ are correlated .
- However, the usual method still works

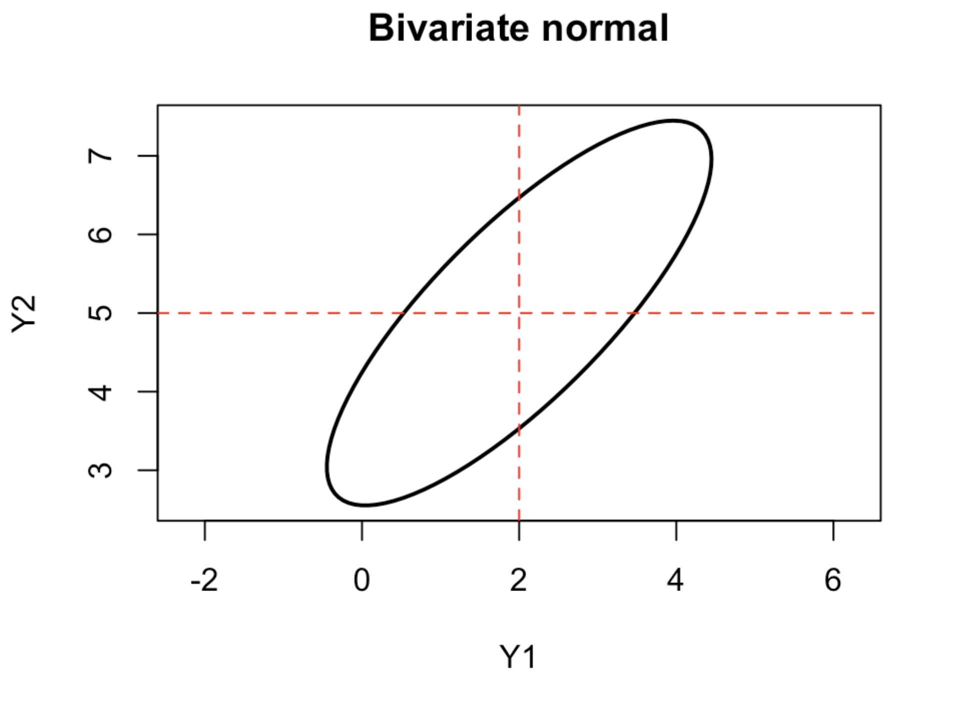
$$\hat{\theta}_n = \frac{1}{n}\sum_{t=1}^{n} g(X_t) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N_p(0, \underbrace{\Sigma}_{\neq \Lambda}) .$$

- Standard errors are tough! correlated samples means Σ is difficult to estimate.
- `mcmcse` estimates Σ and its diagonals σ^2_{ii} for MCMC.

Simple Example

Goal: Estimate mean of $N_2\left(\begin{pmatrix}2\\5\end{pmatrix}, \begin{pmatrix}1 & .8\\.8 & 1\end{pmatrix}\right)$

Here we have the luxury of knowing the truth. $\theta = (2\ \ 5)^T$

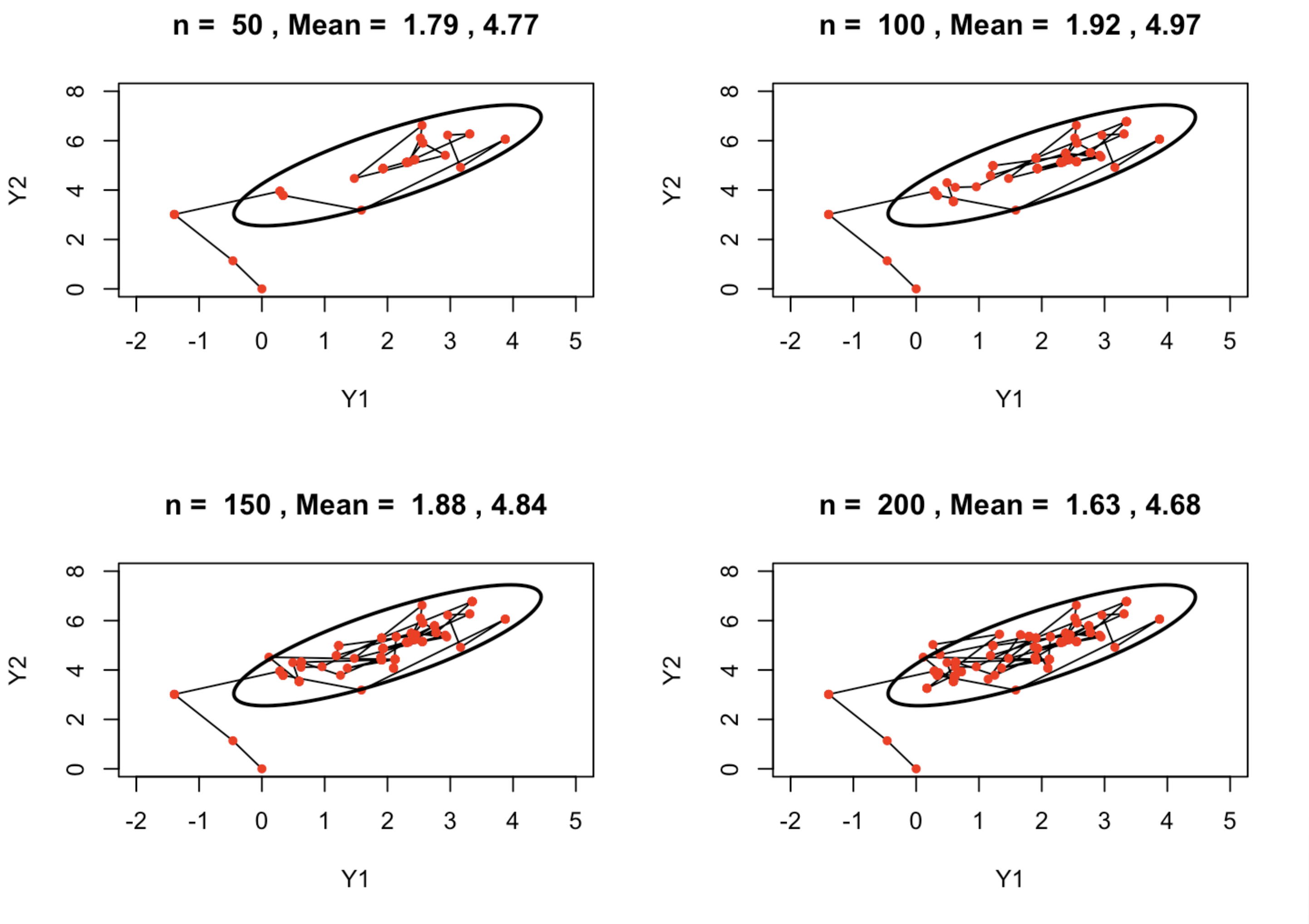


Example: MCMC

We will use a random walk Metropolis sampler to draw correlated, non iid samples.

```r
# Runs mcmc for 1e4 steps
N <- 1e4
out.rwm <- rwm(sigma = 1.5, N = N)$chain
```

Example: MCMC



Example: MCMC

```
# Rows represent samples, columns are components of the Markov chain
head(out.rwm)
```

```
##             [,1]     [,2]
## [1,]  0.0000000 0.000000
## [2,] -0.4617975 1.137128
## [3,] -1.3943455 3.014124
## [4,] -1.3943455 3.014124
## [5,] -1.3943455 3.014124
## [6,]  0.2874315 3.954838
```

```
dim(out.rwm)
```

```
## [1] 10000     2
```

Are 10000 samples enough to estimate the mean here?

Example: MCMC

```
# Monte Carlo estimate for (2, 5)
colMeans(out.rwm)
```

```
## [1] 2.025037 5.010399
```

```
# Standard error?
```

We need the `mcmcse` R package to estimate the standard error!

mcmcse

```r
library(mcmcse)
# Function calculates univariate standard errors for first comp
mcse(out.rwm[,1])
```

```
## $est
## [1] 2.025037
##
## $se
## [1] 0.03808689
```

$$\text{Markov chain CLT: } \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma) \quad \text{mcse returns} \frac{\widehat{\sigma_{ii}}}{\sqrt{n}}$$

```r
# sigma^2
mcse(out.rwm[,1])$se^2*N
```

```
## [1] 14.50611
```

mcmcse

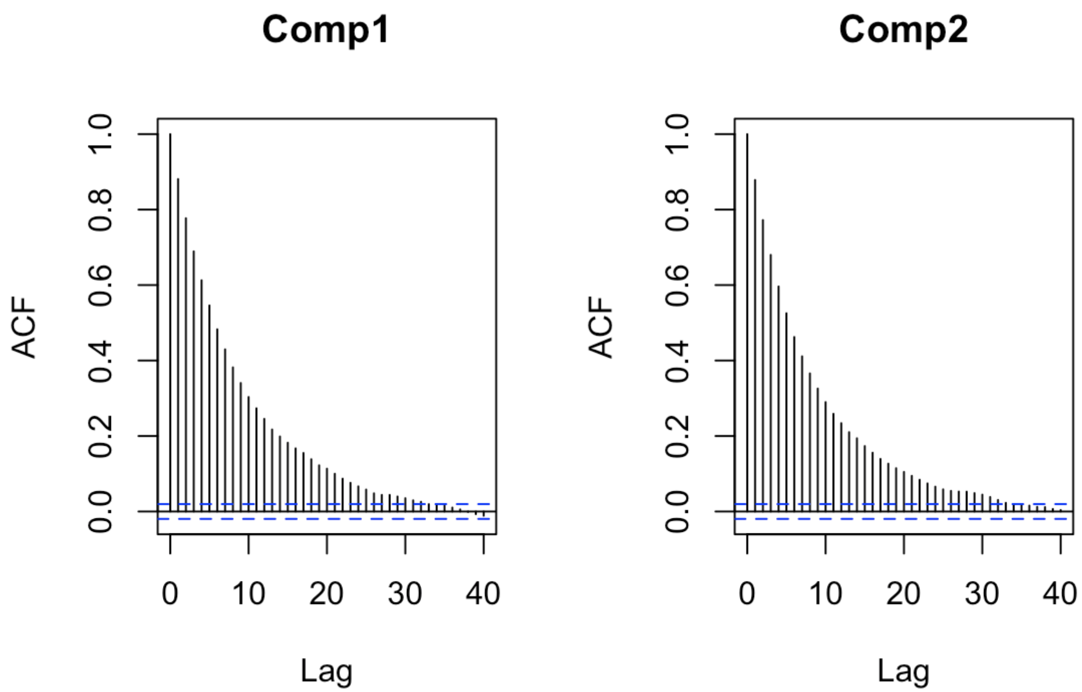
```
# For both components
mcse.mat(out.rwm)
```

```
##            est         se
## [1,] 2.025037 0.03808689
## [2,] 5.010399 0.03857074
```

```
# If IID sampling, variance should have been 1
mcse.mat(out.rwm)[,2]^2*N
```

```
## [1] 14.50611 14.87702
```

Autocorrelation



The autocorrelations inflate the variance. `mcse` accounts for these lag correlations

Standard Errors estimators

To estimate Σ or σ_{ii} consistently, options are

- `bm` Fast
- `tukey` Slow
- `bartlett` Slow

To estimate Σ or σ_{ii} coservatively, options are

- `multi.initseq`

All references Dai and Jones (2016), Vats, Flegal, and Jones (2015a), Vats, Flegal, and Jones (2015b)

ess

One common way of assessing MCMC performance is to know its effective sample size .

If we had taken iid samples

$$\text{CLT: } \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \Lambda)$$

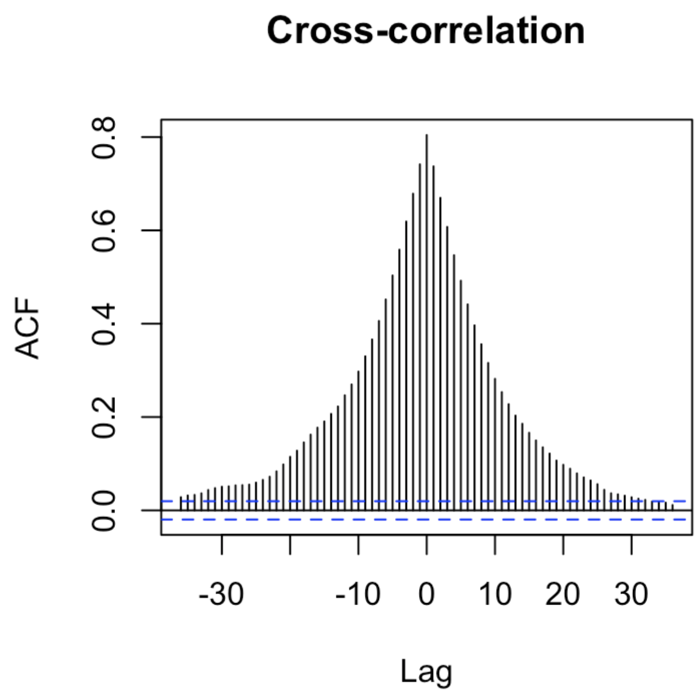
$$\text{ESS}_i = n \frac{\lambda_{ii}^2}{\sigma_{ii}^2}$$

```
# Positive correlation means smaller ess. n = 1e4
ess(out.rwm)
```

```
## [1] 724.9664 701.4952
```

Multivariate ess

But we have two multivariate CLTs. Why a univariate ESS? With univariate ESS we are ignoring cross-correlation



Multivariate ess

If estimating p components

$$\text{ESS} = n \frac{|\Lambda|^{1/p}}{|\Sigma|^{1/p}}$$

```
# Effective sample size for estimating the mean vector
multiESS(out.rwm)
```

```
## [1] 1056.513
```

- Calls function `mcse.multi` which estimates Σ
- Estimates Λ using the usual `cov` function
- Estimates Σ using batch means method by default. Other methods may be used.
- Coded using `Rcpp`

Minimum ess required

But how do we know if we have enough samples?

Minimum ess required

But how do we know if we have enough samples?

To get relative tolerance of $\epsilon = .05$, and in order to make 95% confidence regions we need minimum — effective sample size

```
# Compare to estimated 1056 from 1e4 Monte Carlo samples
minESS(p = 2, eps = .05, alpha = .05)
```

```
## minESS
##   7529
```

- Similar to sample size calculations for one sample t-tests
- This does not depend on the Markov chain. Should be done a priori.

Minimum ess required

So I need some 6500 more **effective samples**.

```
N <- 8e4
out.rwm <- rwm(sigma = 1.5, N = N)$chain
multiESS(out.rwm)
```

```
## [1] 8713.306
```

I overshot a little bit, but I'd rather overshoot than undershoot.

So now I know that with **8e4** Monte Carlo samples, my effective sample size for estimating the mean of this bivariate normal distribution is **8713** for a relative tolerance of $\epsilon = .05$ in order to be **95%** confident in my estimate. Phew!

Conclusions

- Determine `minESS` before starting simulation
- Recommend using `multiESS` over univariate `ess`
- R package `coda` produces only biased univariate estimates
- Example was only for mean. Package can be used for E$[g(x)]$
- Package can also be used for finding standard errors for quantiles

Thank you!

References

Dai, Ning, and Galin Jones. 2016. "Multivariate Initial Sequence Estimators in Markov Chain Monte Carlo." *ArXiv*.

Vats, Dootika, James M Flegal, and Galin L Jones. 2015a. "Multivariate Output Analysis for Markov Chain Monte Carlo." *Preprint*.

———. 2015b. "Strong Consistency of Multivariate Spectral Variance Estimators in Markov Chain Monte Carlo." *Bernoulli to appear*.