An Introduction to Estimating Monte Carlo Standard Errors with R Package `mcmcse`

Dootika Vats

August 16, 2021

Contents

1 Introduction

The R package `mcmcse` provides estimates of Monte Carlo standard errors for Markov chain Monte Carlo (MCMC) when estimating means or quantiles of functions of the MCMC output. In addition to MCMC output, the package can be used for time series and other correlated processes.
The package is predominantly useful after MCMC output has been obtained by the user. In addition to estimating the Monte Carlo standard errors, the package also provides univariate and multivariate estimates of effective sample size and tools to determine whether enough Monte Carlo samples have been obtained. There are also some graphical tools to ascertain the behavior of the Monte Carlo estimates.

2 An MCMC Example

To illustrate the use of our package, we consider sampling from a bivariate normal distribution using a Gibbs sampler. For $\omega_1, \omega_2 > 0$ and ρ such that $\rho^2 < \omega_1\omega_2$, the target distribution is

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \omega_1 & \rho \\ \rho & \omega_2 \end{pmatrix}\right)$$

The function BVN_Gibbs in the package draws samples from the above model.

```
#library(mcmcse)
mu = c(2, 50)
sigma = matrix(c(1, 0.5, 0.5, 1), nrow = 2)

# Monte Carlo sample size is N
N <- 5e3
set.seed(100)
chain <- BVN_Gibbs(N, mu, sigma)
```

For using the `mcmcse` package the rows of the MCMC output should store each iteration of the algorithm and so the output should have n rows and p columns. We will denote each row i of the output as $y_i = (y_i^{(1)}, y_i^{(2)})$.

```
#Rows has observations (samples) and each comlumn is a component.
head(chain)

##            Y1       Y2
## [1,] 1.497808 49.86281
```

```
## [2,] 1.863062 50.69951
## [3,] 2.451055 50.50147
## [4,] 1.746889 50.49225
## [5,] 1.531428 49.45406
## [6,] 1.804876 49.98581
```

This vignette will discuss estimating two sets of features of interest of F.

- $\mathrm{E}_F y$: For estimating $\mu = \mathrm{E}_F y$, the estimator is the Monte Carlo sample mean

$$\mu_n = \frac{1}{n} \sum_{t=1}^{n} y_t.$$

  In `R`, μ_n is obtained using the usual `colMeans` function. If $p = 1$, then use `mean` instead of `colMeans`.

```
colMeans(chain)
```

```
##        Y1        Y2
##  2.006982 50.006033
```

- $\mathrm{E}_F \left(y^{(1)2} + y^{(2)2} \right)$: When interested in estimating the sum of the second moments of each component of y, define the function $g : \mathbb{R}^2 \to \mathbb{R}$ as $g((x_1, x_2)) = x_1^2 + x_2^2$. This is defined in `R` by creating a function that implements the function g, row-wise.

```
g <- function(x)
{
  return(sum(x^2))
}
```

  The Monte Carlo estimator for g is

$$\mu_{g,n} = \frac{1}{n} \sum_{t=1}^{n} g(y_t),$$

```r
# Apply the function g to each row
gofy <- apply(chain, 1, g)

# Monte Carlo estimate
mean(gofy)

## [1] 2506.55
```

Thus, to obtain Monte Carlo estimates from MCMC output, the base package is sufficient (generally). However, Monte Carlo estimates must be reported with Monte Carlo standard error. That is, if the following central limit theorems hold

$$\sqrt{n}(\mu_n - \mu) \xrightarrow{d} N_p(0, \Sigma)\,, \tag{1}$$

and

$$\sqrt{n}(\mu_{g,n} - \mu) \xrightarrow{d} N_p(0, \Sigma_g)\,, \tag{2}$$

then estimates of Σ and Σ_g must be reported, directly or indirectly. Since the samples obtained are correlated, these quantities require more sophisticated tools than usual sample estimators. (Note that a Markov chain CLT is not always guaranteed to hold. In fact, it depends on the rate of convergence of the Markov chain. Most of the functions in this package assume that a Markov chain CLT holds. Such an assumption is also made when using many of the convergence diagnostics).

3 Estimating Monte Carlo Standard Error

In this package, the functions `mcse`, `mcse.mat`, `mcse.multi`, and `mcse.initseq` estimate the Monte Carlo standard error of μ_n (or $\mu_{g,n}$).

- `mcse`: consistent estimates of $\sqrt{\Sigma/n}$ (standard error) when Σ is 1×1.
- `mcse.mat`: consistent estimates of the square root of the diagonals of Σ/n.
- `mcse.multi`: consistent estimates of Σ.
- `mcse.initseq`: asymptotically conservative estimates of Σ using initial sequence estimators.

Using the `mcmcse` package we can estimate Σ in (1) with the `mcse.multi` and `mcse.initseq` function.

```r
# Batch means estimator
mcerror_bm <- mcse.multi(x = chain, method =  "bm", r = 1,
                         size = NULL, g = NULL, adjust = TRUE,
                         blather = TRUE)

# Overlapping batch means estimator
mcerror_obm <- mcse.multi(x = chain, method =  "obm", r = 1,
                         size = NULL, g = NULL, adjust = TRUE,
                         blather = TRUE)

# Spectral variance estimator with Bartlett window
mcerror_bart <- mcse.multi(x = chain, method =  "bartlett", r = 1,
                           size = NULL, g = NULL, adjust = TRUE,
                           blather = TRUE)

# Spectral variance estimator with Tukey window
mcerror_tuk <- mcse.multi(x = chain, method =  "tukey", r = 1,
                          size = NULL, g = NULL, adjust = TRUE,
                          blather = TRUE)

# Initial sequence estimator, unadjusted
mcerror_is <- mcse.initseq(x = chain, g = NULL,
                           adjust = FALSE, blather = TRUE)

# Initial sequence estimator, adjusted
mcerror_isadj <- mcse.initseq(x = chain, g = NULL,
                              adjust = TRUE, blather = TRUE)
```

- `x` takes the $n \times p$ MCMC data. `x` can take only numeric entries in the form of a matrix or data frame. The rows of `x` are the iterations of the MCMC.
- `method = bm, obm, bartlett, tukey` calculates the estimate using the batch means method and spectral variance methods with the modified-Bartlett and Tukey-Hanning windows.
- `r` is the lugsail parameter that indicates how much to "lift" the lag window (this also applies to `bm` and `obm`). Higher values will increasingly remove underestimation of Σ but may yield more variable estimators. Values more than 5 are not advised and negative values are

not allowed. Reasonable choices are $r = 1, 2, 3$, where $r = 3$ yields the lugsail estimator, $r = 2$ is the flat-top estimator, and $r = 1$ is the vanilla estimator.

- `size` is the batch size for the `bm` method and the truncation point for `tukey` and `bartlett` methods. Default batch size is calulcated using the exported `batchSize` function. Other accepted values are `size = sqroot`, which sets the size as $\lfloor \sqrt{n} \rfloor$ and `size = cuberoot` which sets it at $\lfloor n^{1/3} \rfloor$. An integer value of `size` less than n is also valid as long as $n/$`size` > 1.

  For reference on `bm` (batch means estimators) see Jones et al. (2006) and Vats et al. (2019).

  For reference on `bartlett` and `tukey` (spectral variance estimators)see Flegal et al. (2010) and Vats et al. (2018).

  For reference on lugsail estimation see Liu and Flegal (2018) and Vats and Flegal (2018).

- `g` is a function that is applied to each row of `x` and represents the features of interest of the process. Since here we are interested in only means, `g` is `NULL`. `g` will be explained in later examples.

- `adjust` is a logical argument indicating whether the resulting matrix should be adjusted in order to retain positive-definiteness. By default this is set to be `TRUE`.

- `blather` when TRUE outputs under the hood information about the estimation process. The default is set to `FALSE` since most users should be interested in only `cov` and `est`.

  For reference on `mcse.initseq` (initial sequence estimators) see Dai and Jones (2017).

`mcse.multi` and `mcse.initseq` return an S3 class with multiple components. When `blather = FALSE` `cov` stores the estimate of Σ obtained using the method chosen, `est` stores the estimate of the mean of g applied on the Markov chain. In addition, `nsim` stores the no of Markov chain samples, `eigen-values` stores the eigen values of the estimated Σ and `cov.adj` stores the adjusted covariance matrix if `adjust = TRUE` (see `mcse.multi` for more details). When `blather = TRUE` the following are returned in addition to the above: `size` which indicates the size of batches/truncation, `method` used, `Adjustment-used` indicating whether an adjusted estimator

was used (`adjust`) and `message` containing additional information about the estimation process (like the numerical adjustments possibly made to keep the estimate mathematically consistent).

Note: The Monte Carlo estimates of μ are not affected by the choice of the method.

Note: for consistent estimation, the batch means estimators are significantly faster to calculate than the spectral variance estimators. The user is advised to use the default `method = ''bm''` for large input matrices.

Note: `cov` returns an estimate of Σ and not Σ/n.

If the diagonals of Σ are σ_{ii}^2, the function `mcse` and `mcse.mat` returns $\sigma_{ii}/\sqrt{n}$. `mcse` does it for one component and `mcse.mat` does it for all diagonals.

```
mcse(x = chain[,1], method = "bm", g = NULL)

## $est
## [1] 2.006982
##
## $se
## [1] 0.01792272
##
## $nsim
## [1] 5000

mcse.mat(x = chain, method = "bm", g = NULL)

##          est         se
## Y1  2.006982 0.01792272
## Y2 50.006033 0.01814934
```

In order to estimate $\mu_{n,g}$ and Σ_g as in (2), we use the `R` function `g` we had defined before. Recall that `g` should be a function that takes vector inputs.

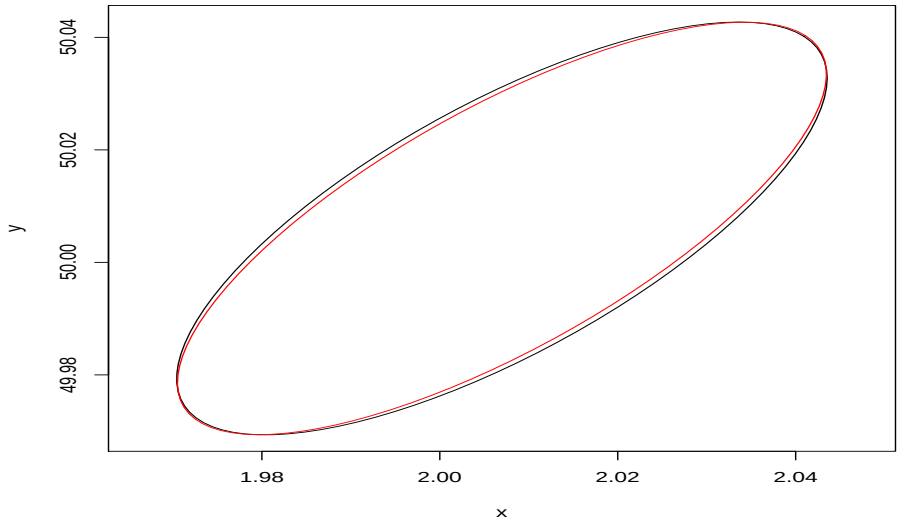
```
g

## function(x)
## {
##   return(sum(x^2))
```

```
## }
## <bytecode: 0x0000021ea5e4eb88>

mcerror_g_bm <- mcse.multi(x = chain, g = g, blather = TRUE)
mcerror_g_is <- mcse.initseq(x = chain, g = g, blather = TRUE)

mcerror_g_bm$cov

##           [,1]
## [1,] 17464.29

# Initial Sequence error is larger than batch means, as expected.
mcerror_g_is$cov

##           [,1]
## [1,] 16091.86

# Returned value is asymptotic variance.
# So we calculate the standard error here.
sqrt(mcerror_g_bm$cov/N)

##           [,1]
## [1,] 1.868919

sqrt(mcerror_g_is$cov/N)

##           [,1]
## [1,] 1.793982
```

4 Confidence Regions

Using the function `confRegion` in the package, the user can create joint confidence regions for two parameters. The input for this function is the output list from the `mcse.multi` or `mcse.initseq` function. The function uses the attributes cov, `est`, and `nsim` from the output list. If the `mcse.initseq` is input and `adjust = TRUE` had been used, then `cov.adj` is used instead of `cov`. mcse.multi also uses the attribute `size`.

```
plot(confRegion(mcerror_bm, which = c(1,2), level = .90), type = 'l', asp = 1)
lines(confRegion(mcerror_bart, which = c(1,2), level = .90), col = "red")
```



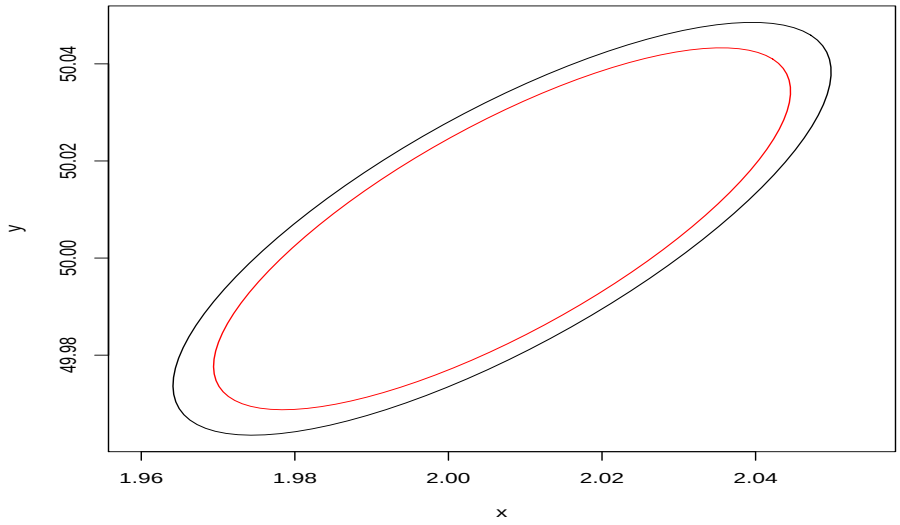
- `which` should be a vector of size 2 that indicates the two components for which the confidence ellipse is to be constructed.
- `level` is the confidence level of the confidence region. The default is .95

NOTE: The argument `confRegion` calls on the function `ellipse` in package `ellipse` to draw the ellipse.

NOTE: Since the confidence region is created for two parameters only, the size of the ellipse is determined by setting $p = 2$ irrespective of the original dimension of the problem.

To determine the effect of the confidence level, we draw two regions with difference confidence levels. We use `mcse.initseq` this time.

```
plot(confRegion(mcerror_is, which = c(1,2), level = .95), type = 'l', asp = 1)
lines(confRegion(mcerror_is, which = c(1,2), level = .90), col = "red")
```



5 Effective Sample Size

Reporting $p \times p$ covariance matrix estimates is impractical and uninterpretable. The motivation of estimating Monte Carlo standard error is to ensure that said error is small. This is essentially the idea behind estimating effective sample size and ensuring that the estimated effective sample size is larger than a prespecified lower bound.
Before sampling the Markov chain, the user is advised to used the function `minESS` to ascertain what is the minimum effective sample size needed for stable analysis. See Vats et al. (2019) for theoretical details.

```
# For mu
minESS(p = 2, alpha = .05, eps = .05)

## minESS
##   7529

#For mu_g
minESS(p = 1, alpha = .05, eps = .05)

## minESS
##   6146
```

- `p` is the dimension of the estimation problem.

- `alpha` is the confidence level
- `eps` is the tolerance level. Default is .05. Reasonable levels are anywhere from .01 to .05. The smaller the tolerance, the larger the minimum effective samples. `eps` represents a tolerance level relative to the variability in the target distribution. It is akin to the idea of margin-of-error.

`minESS` is independent of the Markov chain or process, and is only a function of the p, α, and ϵ. The user should find `minESS` and then sample their process until the required minimum samples are achieved.
Alternatively, we often don't have to luxury of obtaining a lot of samples, and reaching a minimum effective sample size os not possible. In such a scneario, it is useful to know the ϵ tolerance level the number of estimated effective samples correspond to. So if we can only obtain 1000 effective samples,

```
# For mu
minESS(p = 2, alpha = .05, ess = 1000)

##  Epsilon
## 0.137196

#For mu_g
minESS(p = 1, alpha = .05, ess = 1000)

##  Epsilon
## 0.123959
```

Thus, if you obtained a sample with estimates effective sample size equaling 1000 for estimating μ_g and $\mu_{n,g}$, then the precision level of your estimate is $\epsilon = .137$ and $\epsilon = .124$, respectively. `multiESS` and `ess` are two functions that calculate the effective sample size of a correlated sample. `ess` calculations are based on Gong and Flegal (2016) and is component-wise, and `multiESS` utilizes the multivariate nature of the problem.
Since `ess` produces a different estimate for each component, conservative practice dictates choosing the smallest of the values. `multiESS` returns one estimate of the effective sample size based on the whole sample. The function calls `mcse.multi` function to obtain a batch means estimate of Σ. The user can provide another estimate of Σ using the `covmat` argument.

```r
multiESS(chain)

## [1] 4024.471

# Using spectral variance estimators
multiESS(chain, covmat = mcerror_bart$cov)

## [1] 4377.303

# Using initial sequence estimators
# Since this is a conservative estimator, ess will be smaller
multiESS(chain, covmat = mcerror_is$cov)

## [1] 4263.715
```

Since the effective sample size is less than the minimum effective samples, we should simulate more. Looking at the formula of ESS, we might need around $9,200$ Monte Carlo samples.

```r
set.seed(100)
chain <- BVN_Gibbs(9200, mu, sigma)

# larger than 7529
multiESS(chain)

## [1] 7544.004

# larger than 7529
multiESS(chain, covmat = mcerror_bart$cov)

## [1] 8137.827

# larger than 7529
multiESS(chain, covmat = mcerror_is$cov)

## [1] 7926.657
```

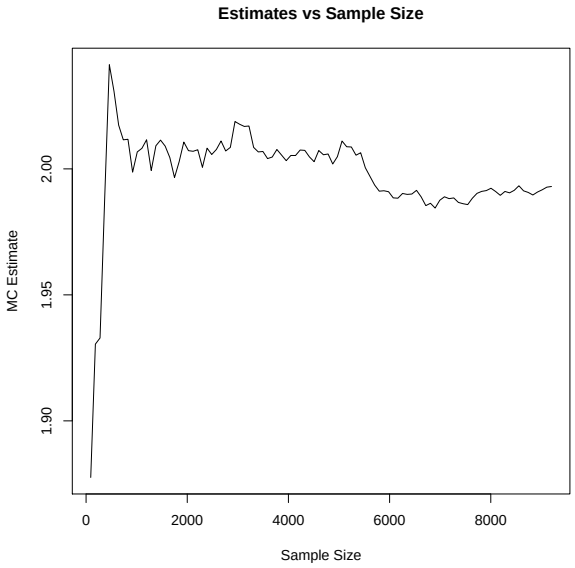
So no matter which estimator we choose for the Monte Carlo standard error, 28000 Monte Carlo samples are sufficient to have $\epsilon = .05$ relative tolerance. **NOTE:** Ideally, we want to get more samples using the last iteration of the previous Markov chain. However, BVN_Gibbs does not allow user specified

starting values and starts from stationarity itself, so to demonstrate the use of `minESS` and `multiESS`, we get a new sample altogether.

6 Graphical Diagnostics

The function `estvssamp` plots the Monte Carlo estimates versus the sample size for a component of the MCMC output. This plot indicates whether the Monte Carlo estimate has stabilized.

```
estvssamp(chain[,1])
```

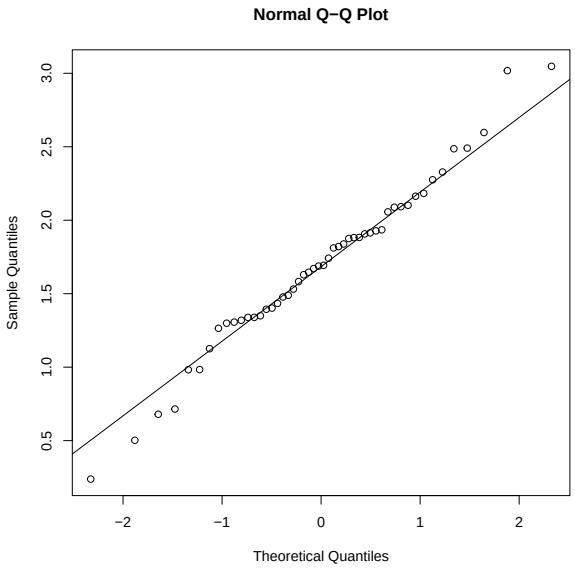


Additionally, if p is not too small, due to the central limit theorem in (1) and an estimate of Σ using the `mcse.multi` function, a QQ plot of the standardized estimates gives an idea of whether asymptopia has been achieved. We generate a new Markov chain with $p = 50$.

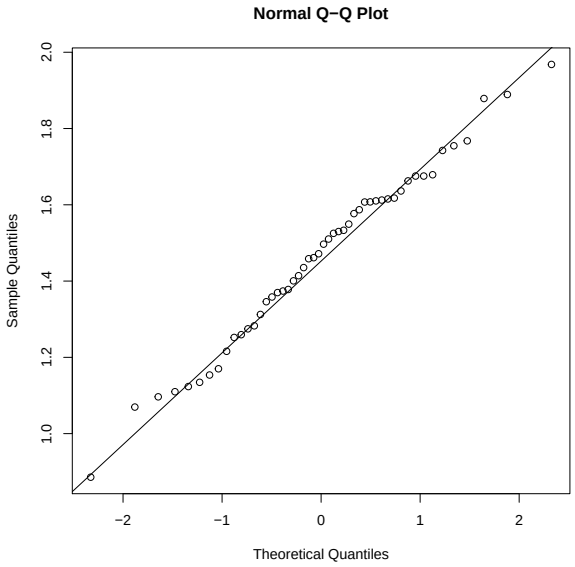
```
library(mAr)
p <- 50
A <- diag(seq(.1, .9, length = p))
C <- diag(rep(2, p))

set.seed(100)
chain <- mAr.sim(w = rep(2,p), A = A, C = C, N = 10000)
```

For this new Markov chain, we find an estimate of Σ to use for the `qqTest` function.

```
mcerror_bm <- mcse.multi(chain, method = "bm", blather = TRUE)
mcerror_isadj <- mcse.initseq(chain, adjust = TRUE, blather = TRUE)
qqTest(mcerror_bm)
```




```
qqTest(mcerror_isadj)
```



Thus, we see here that the chain has reached not reached asymptopia.

References

Dai, N. and Jones, G. L. (2017). Multivariate initial sequence estimators in Markov chain Monte Carlo. *Journal of Multivariate Analysis (to appear)*.

Flegal, J. M., Jones, G. L., et al. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38:1034–1070.

Gong, L. and Flegal, J. M. (2016). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 25(3):684–700.

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101:1537–1547.

Liu, Y. and Flegal, J. (2018). Weighted batch means estimators in Markov chain Monte Carlo. *Electronic Journal of Statistics*, 12:3397–3442.

Vats, D. and Flegal, J. M. (2018). Lugsail lag windows and their application to MCMC. *arXiv preprint arXiv:1809.04541*.

Vats, D., Flegal, J. M., and Jones, G. L. (2018). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli*, 24:1860–1909.

Vats, D., Flegal, J. M., and Jones, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*, 106:321–337.