ST219: Assignment 2

J.P. Meagher

Wednesday, February 17, 2016

Assignment 2: Question 2 - Simulations

We need to verify that $MSE_\theta(\hat{\theta}_3(\mathbf{X})) \leq MSE_\theta(\hat{\theta}_2(\mathbf{X}))$ through a simulation study. We will then construct confidence intervals for our estimators.

To perform our simulation study we:

1. Set θ_0, the underlying parameter.
2. Set n, the number of observations we want in each dataset.
3. Set M, the number of datasets we want to generate.
4. For each dataset $\mathbf{X}_m$ we compute $\hat{\theta}_1(\mathbf{X}_m), \hat{\theta}_2(\mathbf{X}_m)$ and $\hat{\theta}_3(\mathbf{X}_m)$.
5. Approximate $MSE_\theta(\hat{\theta}_i(\mathbf{X}))$ by calculating $\frac{1}{M}\sum_{j=1}^{M}(\hat{\theta}_i(\mathbf{X}_j) - \theta_0)$

This can be repeated across multiple values of θ_0

First we initialise our values

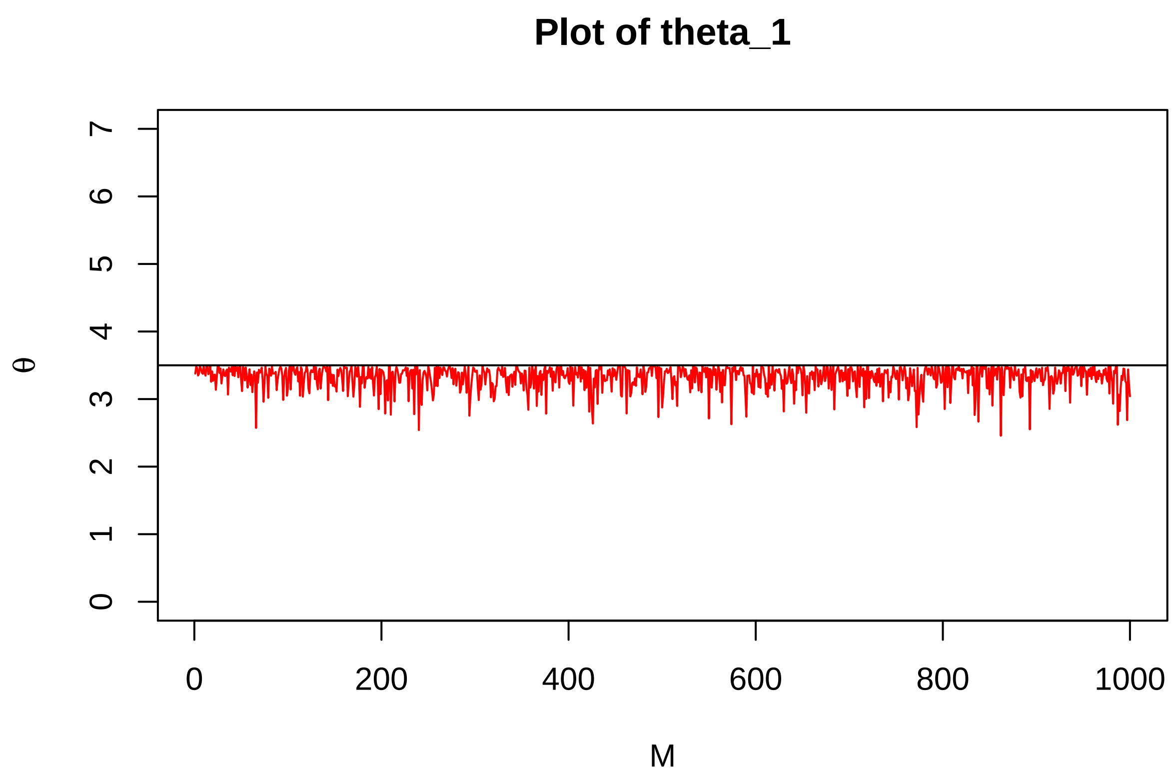
```
n <- 20 # observations
M <- 1000 # Datasets
theta_0 <- 3.5 # parameter
```

Then generate the datasets and calculate estimators

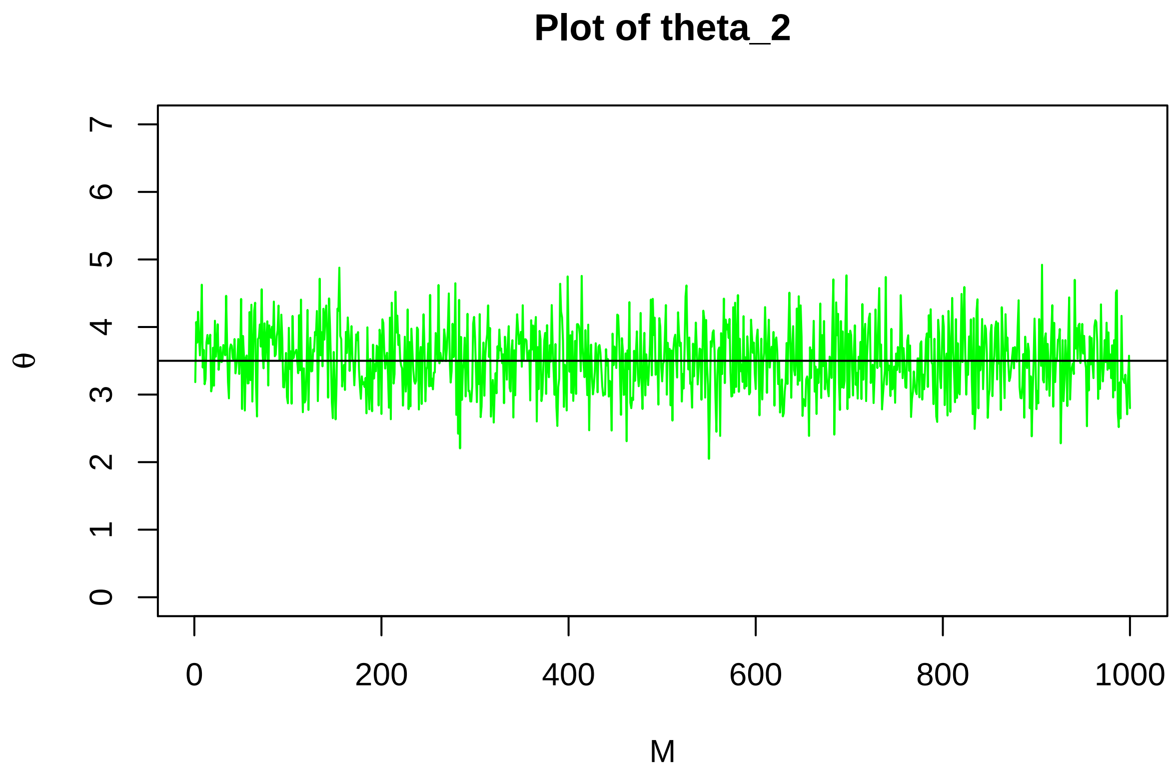
```
X <- matrix(runif(n*M, 0, theta_0), ncol = M, nrow = n)

# Calculate estimators
theta_1 <- apply(X, 2, max)
theta_2 <- 2*(apply(X, 2, mean))
theta_3 <- ((n+1)/n)*theta_1
```

We can plot each of these estimators

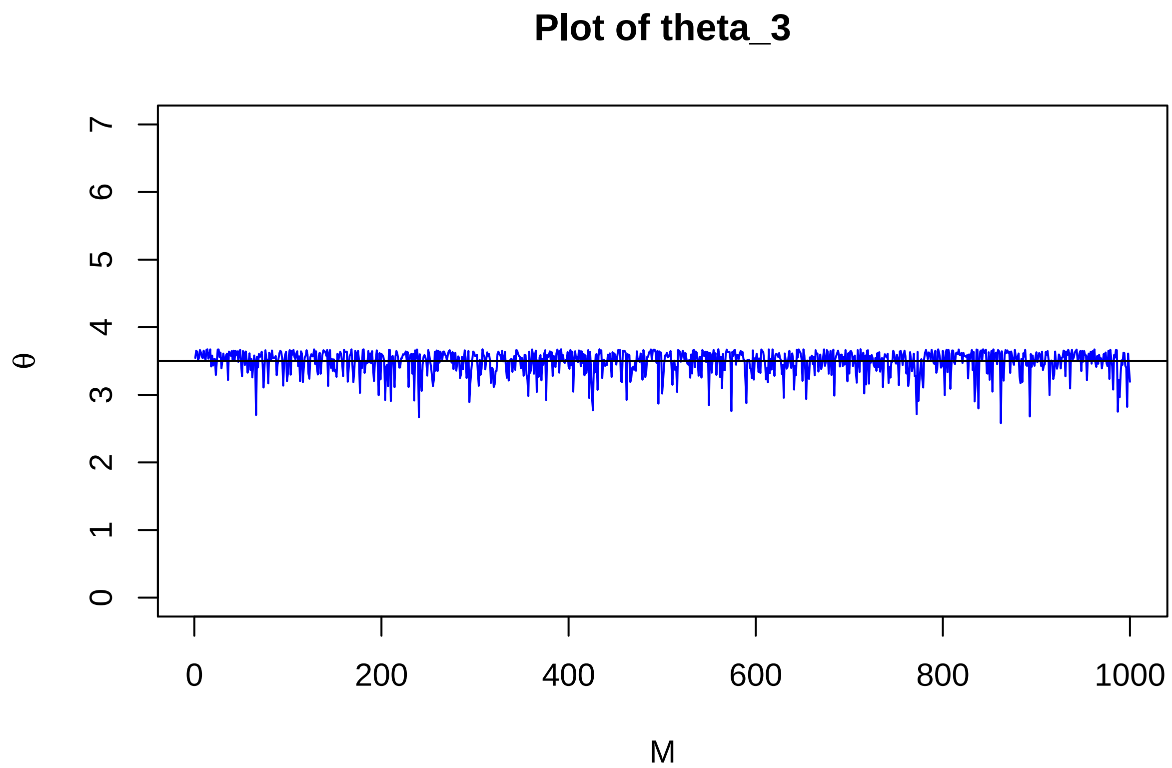
```
plot(1:1000, theta_1, type = 'l', col = 'red', ylim = c(0, 2*theta_0), xlab = 'M',
     ylab = expression(theta), main = 'Plot of theta_1')
abline(h = 3.5)
```




```
plot(1:1000, theta_2, type = 'l', col = 'green', ylim = c(0, 2*theta_0), xlab = 'M',
     ylab = expression(theta), main = 'Plot of theta_2')
abline(h = 3.5)
```

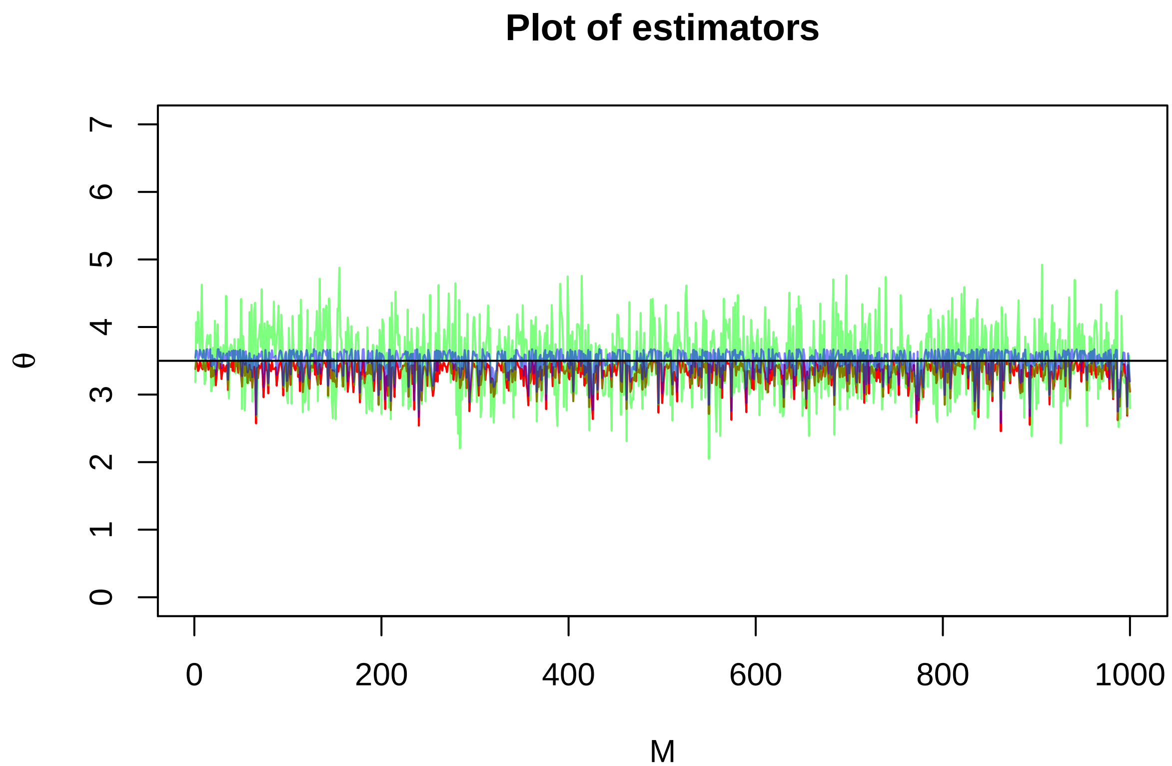



```
plot(1:1000, theta_3, type = 'l', col = 'blue', ylim = c(0, 2*theta_0), xlab = 'M',
     ylab = expression(theta), main = 'Plot of theta_3')
abline(h = 3.5)
```



Or all on a single chart

```
plot(1:1000, theta_1, type = 'l', col = rgb(1,0,0,1), ylim = c(0, 2*theta_0), xlab = 'M',
     ylab = expression(theta), main = 'Plot of estimators')
lines(theta_2, col = rgb(0,1,0,0.5))
lines(theta_3, col = rgb(0,0,1,0.5))
abline(h = 3.5)
```



We can now calculate our approximations for Mean Squared Error. We know that the actual values for each mean squared error are

```
Actual_MSE_2 <- (theta_0^2)/ (3*n)
Actual_MSE_2
```

```
## [1] 0.2041667
```

```
Actual_MSE_3 <- (theta_0^2)/ (n*(n+2))
Actual_MSE_3
```

```
## [1] 0.02784091
```

our estimates are

```
# Approximate MSE
est_MSE_2 <- mean((theta_2 - theta_0)^2)
est_MSE_2
```

```
## [1] 0.234647
```

```
est_MSE_3 <- mean((theta_3 - theta_0)^2)
est_MSE_3
```

```
## [1] 0.03388993
```

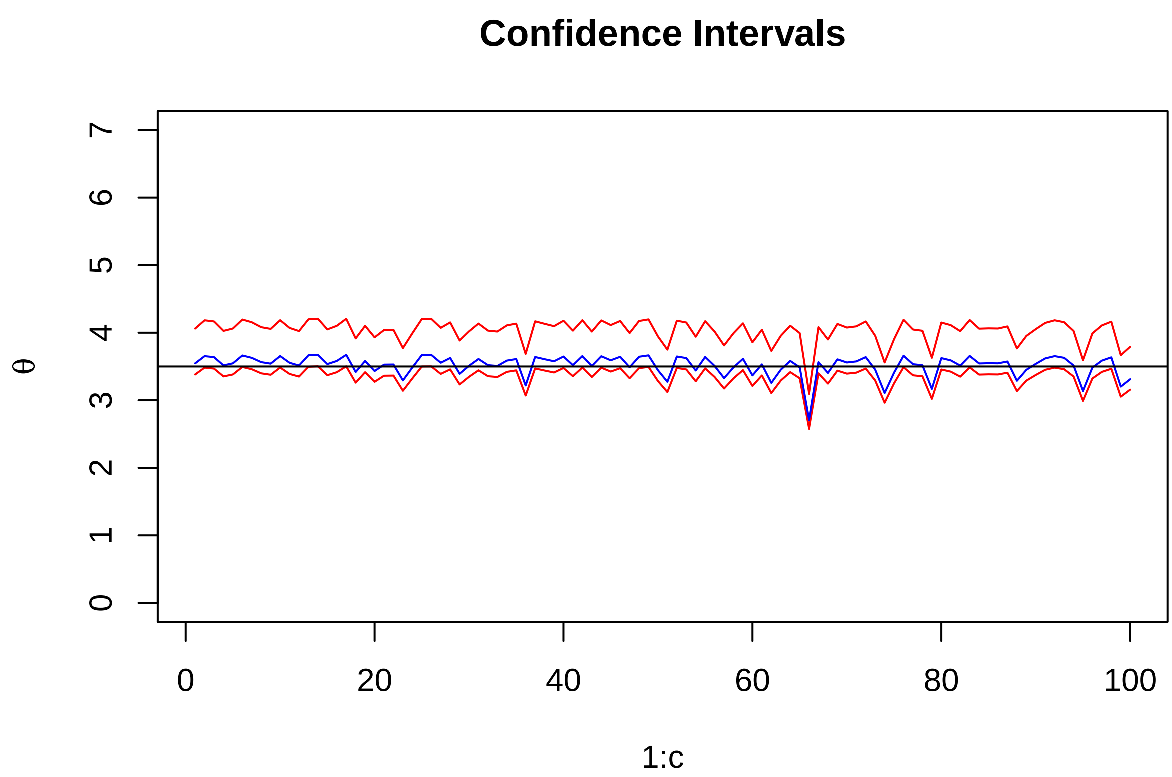
The approximations should be close to the actual values. This can be repeated for various values of θ_0

Now find confidence intervals for each $\hat{\theta}_3(\mathbf{X}_m)$. Note that

$$P(\frac{\hat{\theta}_1(\mathbf{X}_m)}{a} \leq \theta \leq \frac{\hat{\theta}_1(\mathbf{X}_m)}{b}) = b^n - a^n$$

For simplicity chose a, b to produce a symmetric Confidence Interval

```
a <- (0.025^(1/n))
b <- (0.975^(1/n))

int <- rbind(theta_1/b, theta_1/a)

c <- 100

plot(1:c, int[1,1:c], type = 'l', col = 'red', ylim = c(0, 2*theta_0), ylab = expression(theta), main =
lines(int[2,1:c], col = 'red')
lines(theta_3[1:c], col = 'blue')
#lines(theta_1[1:c], col = 'green')
abline(h = 3.5)
```



The proportion of the M datasets for which the true θ_0 is inside the confidence interval for $\hat{\theta}_3(\mathbf{X}_m)$ is

```
1 - ((sum(theta_0 < int[1,]) + sum(theta_0 > int[2,]))/M)
```

```
## [1] 0.936
```