Data-580 Lab 4

Dhun Sheth

2023-10-05

Question 1

Part A

```
m <- 10000; n <- 100; sigma <- 5
M <- matrix(rnorm(m*n, mean = 3, sd = sigma), nrow=n)
xbar <- apply(M, 2,mean)

estimate <- mean(xbar>4.5)
print(estimate)
```

```
## [1] 7e-04
```

```
theory <- 1-pnorm((4.5-3)/sqrt(25/n))
print(theory)
```

```
## [1] 0.001349898
```

Estimate:

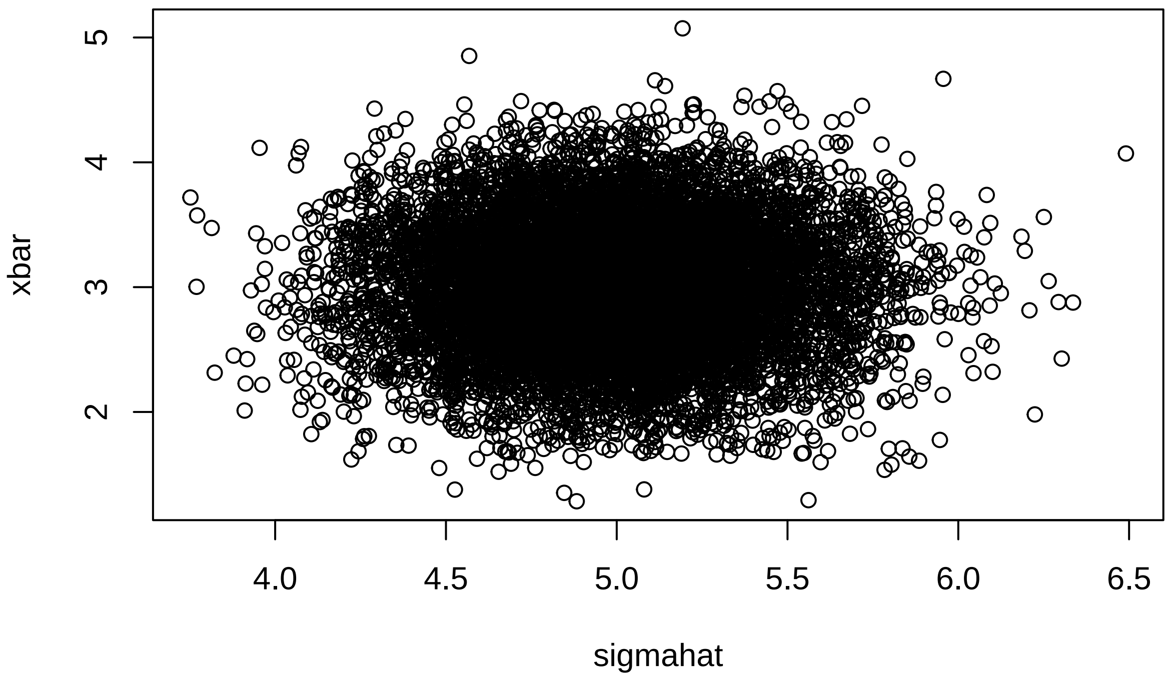
$$P(x_n > 4.5) = 0.0012$$

Theoretical:

$$P(x_n > 4.5) = 0.001349898$$

Part B

```
sigmahat=apply(M, 2,sd)
plot(xbar ~ sigmahat)
```



Based on the plot there is pattern between the sample mean and sample standard deviation, thus they are independent.

Part C

```
left_side = mean(xbar/sigmahat)
right_side = mean(xbar)*mean(1/sigmahat)
print(left_side)
```

```
## [1] 0.604954
```

```
print(right_side)
```

```
## [1] 0.6049844
```

Based on values

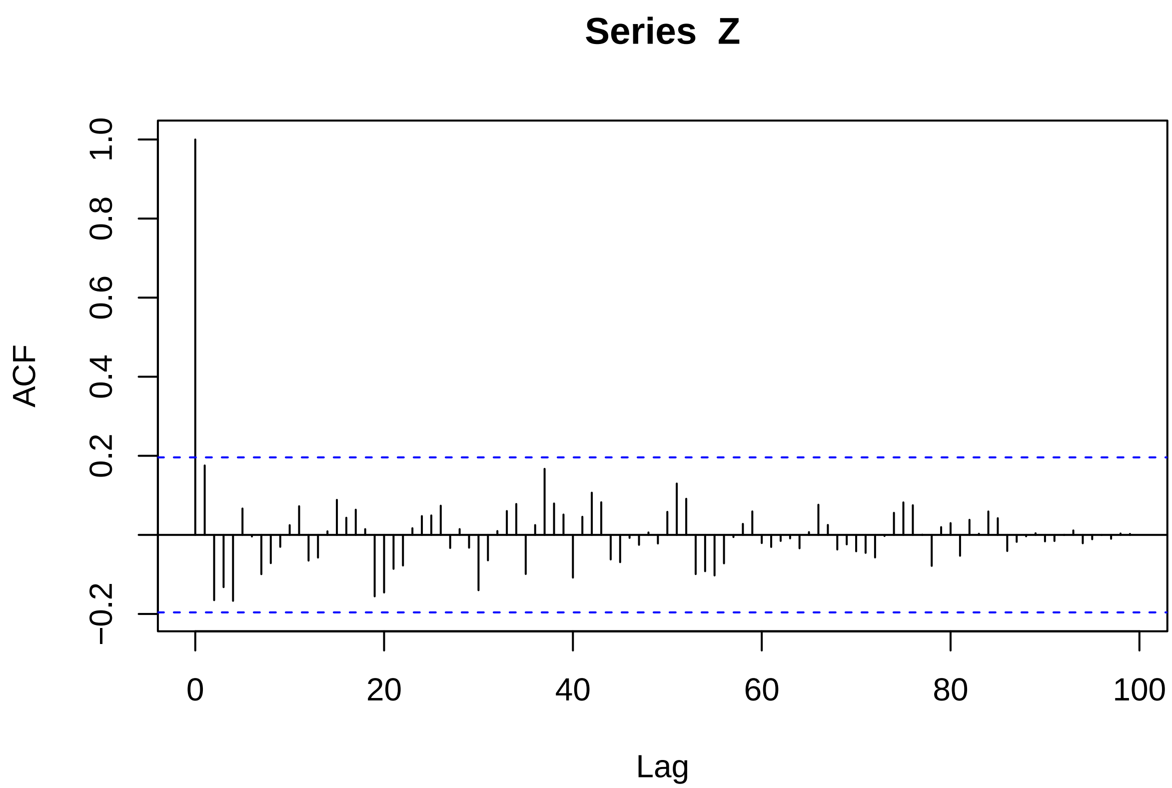
$$E[X_n/S_n] = E[x_n] * E[1/S_n]$$

Question 2

```r
phi1 <- 0.2
Z <- arima.sim(100, model = list(ar = phi1), sd = 1)
```

Part B

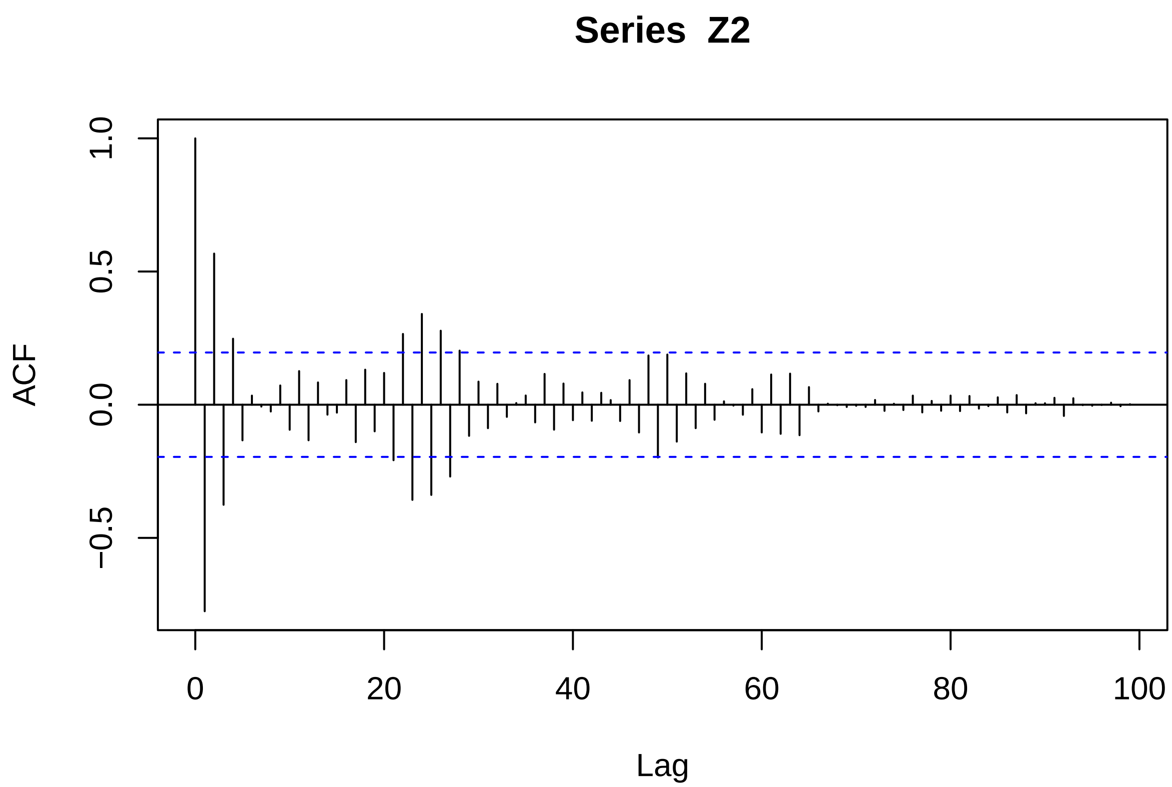
```r
acf(Z, lag=100)
```



Based on the ACF plot, there is an instance around lag 10 and lag 38 where the bars break the tolerance limits. It can be seen that the autocorrelation is oscillating and decreasing as we go towards lag 100.

Part C

```r
phi2 <- -0.8
Z2 <- arima.sim(100, model = list(ar = phi2), sd = 1)
acf(Z2, lag=100)
```



Comparing the ACF plots with parameter 0.2 vs. -0.8, the -0.8 parameter starts with much higher correlation during the initial lags but both ACF plots oscillate and decrease by lag 100. Parameter 0.2 oscillates slower than parameter -0.8.

Question 3

Part A

```
set.seed(1333)

x <- seq(0, 1, length=100)
eps_i <- rnorm(100, sd = 1)
y0 <- 2 + 3*x + eps_i

set.seed(1333)
eps_ii <- arima.sim(100, model = list(ar = 0.8), sd = 1)
y1 <- 2 + 3*x + eps_ii
```

Part B

```
y0.lm <- lm(y0 ~ x)
y1.lm <- lm(y1 ~ x)
```

```r
print(coef(y0.lm))
```

```
## (Intercept)           x
##    2.078381    2.877188
```

```r
print(coef(y1.lm))
```

```
## (Intercept)           x
##    2.052800    3.417216
```

Case 1:

$$Y = 2.877x + 2.078$$

Case 2:

$$Y = 3.417x + 2.053$$

Part C

```r
x=0.8
Y0 <- 2.877*x + 2.078
Y1 <- 3.417*x + 2.053

true_exp <- 2+3*x

print(Y0)
```

```
## [1] 4.3796
```

```r
print(Y1)
```

```
## [1] 4.7866
```

```r
print(true_exp)
```

```
## [1] 4.4
```

Case 1 is closer to the true expected value.

Part D

The noise term has dependency in case 2, as seen from the ACF plot in question 2 for AR(1), and based on the regression results, having a noise term with dependency leads to a less accurate regression model, as we compare the regression estimates from case 1 (no dependency in noise term) and case 2 (dependency in noise term) to the true expected value.