Problem Set 4

Tate Mason

Question 1 - Hansen 7.17

Part A

Part B

Part C

Question 2 - Hansen 7.28

Part B

Part C

Part D

Part E

EQ 1

EQ 2

```r
b0 <- 0
b1 <- 1
n <- 100
sim <- function() {
  X1 <- rexp(n)
  e <- mixtools::rnormmix(n,lambda=c(0.5,0.5),mu=c(-1,2),sigma=c(1,1))
```

```
  Y <- b0 + b1*X1 + e
  x <- cbind(1,X1)
  xx <- t(x)%*%x
  xy <- t(x)%*%Y
  bhat <- solve(xx,xy)
  return(c(bhat[2]))
}
```

These results show that the average of $\hat{\beta}_1 \to \beta_1$ as n grows. The variance also approaches 0. This is consistent with what was derived in class, that as $n \to \infty$, we see the predicted approach the actual, and variance should be 0 as with sufficiently large n, there will be no variance in observations.

```
run_mc <- function(n_sims = 1000) {
  mc_res <- sapply(1:n_sims, function(s) {
    sim()
  })
  cat("Mean b1:", mean(mc_res), "\n")
  cat("Variance of b1:", var(mc_res), "\n")
}
```

```
run_mc()
```

```
Mean b1: 0.993102
Variance of b1: 0.03446937
```

```
n <- 2
run_mc()
```

```
Mean b1: 25.64785
Variance of b1: 660092.2
```

```
n <- 10
run_mc()
```

```
Mean b1: 0.9566694
Variance of b1: 0.6509778
```

```r
n <- 50
run_mc()
```

```
Mean b1: 1.008
Variance of b1: 0.07218543
```

```r
n <- 500
run_mc()
```

```
Mean b1: 1.003312
Variance of b1: 0.006857525
```

As $n \to \infty$, the mean and variance get closer to the true values. This is a showcase of the WLLN.

EQ 3

Part A

```r
b0 <- 0
b1 <- 1
num_sims <- 1000
alpha <- 0.05

sim_test <- function(n,b1_true, b1_null) {
  X <- rexp(n)
  e <- mixtools::rnormmix(n,lambda=c(0.5,0.5),mu=c(-2,2),sigma=c(1,1))
  Y <- b0+b1*X + e
  x <- cbind(1,X)
  xx <- t(x)%*%x
  xy <- t(x)%*%Y
  bhat <- solve(xx,xy)
  b0_hat <- bhat[1]
  b1_hat <- bhat[2]

  yhat <- x %*% bhat
  ehat <- Y - yhat
```

```r
  sigma_sq_hat <- sum(ehat^2)/(n-2)
  var_cov_matrix <- as.numeric(sigma_sq_hat)*solve(xx)
  se_b1_hat <- sqrt(var_cov_matrix[2,2])
  t_stat <- (b1_hat - b1_null)/se_b1_hat

  df <- n-2
  t_crit <- qt(1-alpha/2,df)
  reject <- abs(t_stat) > t_crit
  p_val <- 2*pt(abs(t_stat), df=df, lower.tail=FALSE)

  return(list(
    b1_hat = b1_hat,
    se_b1_hat = se_b1_hat,
    t_stat = t_stat,
    t_crit = t_crit,
    p_val = p_val,
    reject = reject
  ))
}
```

```r
run_hypothesis_test <- function(n, b1_true, b1_null) {
  results <- data.frame(
    b1_hat = numeric(num_sims),
    se_b1_hat = numeric(num_sims),
    t_stat = numeric(num_sims),
    p_val = numeric(num_sims),
    reject = logical(num_sims)
  )

  for (i in 1:num_sims) {
    sim_result <- sim_test(n,b1_true,b1_null)
    results$b1_hat[i] <- sim_result$b1_hat
    results$se_b1_hat[i] <- sim_result$se_b1_hat
    results$t_stat[i] <- sim_result$t_stat
    results$p_val[i] <- sim_result$p_val
    results$reject[i] <- sim_result$reject
  }

  reject_rate <- mean(results$reject)
  mean_b1_hat <- mean(results$b1_hat)
  var_b1_hat <- var(results$b1_hat)
  mean_se_b1_hat <- mean(results$se_b1_hat)
```

```r
  theoretical_var <- mean(results$se_b1_hat^2)

  return(list(
    results = results,
    reject_rate = reject_rate,
    mean_b1_hat = mean_b1_hat,
    var_b1_hat = var_b1_hat,
    mean_se_b1_hat = mean_se_b1_hat,
    theoretical_var = theoretical_var
  ))
}
```

```r
results_100_true <- run_hypothesis_test(n=100,b1_true=1,b1_null=1)
cat("Part a & b: Results for n = 100, H :   = 1 (true value)\n")
```

```
Part a & b: Results for n = 100, H :   = 1 (true value)
```

```r
cat("Theoretical rejection rate at   = 0.05 should be: 0.05\n")
```

```
Theoretical rejection rate at   = 0.05 should be: 0.05
```

```r
cat("Observed rejection rate:", results_100_true$rejection_rate, "\n")
```

```
Observed rejection rate:
```

```r
cat("Mean ^ :", results_100_true$mean_b1_hat, "\n")
```

```
Mean ^ : 0.9991033
```

```r
cat("Variance of ^ :", results_100_true$var_b1_hat, "\n")
```

```
Variance of ^ : 0.0570121
```

```r
cat("Mean standard error of ^ :", results_100_true$mean_se_b1_hat, "\n")
```

```
Mean standard error of ^ : 0.2297288
```

```r
cat("Theoretical variance (from SE):", results_100_true$theoretical_var, "\n\n")
```

```
Theoretical variance (from SE): 0.05392328
```

Part C

```r
sample_size <- c(10,50,500,1000)
results_varying_n <- list()

for (n in sample_size) {
  results_varying_n[[paste0("n", n)]] <- run_hypothesis_test(n=n, b1_true=1, b1_null=1)
  cat("Results for n =", n, ", H :   = 1 (true value)\n")
  cat("Rejection rate:", results_varying_n[[paste0("n", n)]]$rejection_rate, "\n")
  cat("Mean ^ :", results_varying_n[[paste0("n", n)]]$mean_b1_hat, "\n")
  cat("Variance of ^ :", results_varying_n[[paste0("n", n)]]$var_b1_hat, "\n")
  cat("Mean standard error of ^ :", results_varying_n[[paste0("n", n)]]$mean_se_b1_hat, "\n")
  cat("Theoretical variance (from SE):", results_varying_n[[paste0("n", n)]]$theortical_var,
}
```

```
Results for n = 10 , H :   = 1 (true value)
Rejection rate:
Mean ^ : 1.034403
Variance of ^ : 1.131923
Mean standard error of ^ : 0.9316865
Theoretical variance (from SE):

Results for n = 50 , H :   = 1 (true value)
Rejection rate:
Mean ^ : 0.9990011
Variance of ^ : 0.1208746
Mean standard error of ^ : 0.3382535
Theoretical variance (from SE):

Results for n = 500 , H :   = 1 (true value)
Rejection rate:
Mean ^ : 0.991619
Variance of ^ : 0.009268145
Mean standard error of ^ : 0.1008843
Theoretical variance (from SE):
```

```
Results for n = 1000 , H :    = 1 (true value)
Rejection rate:
Mean ^ : 0.9988795
Variance of ^ : 0.005354894
Mean standard error of ^ : 0.07096727
Theoretical variance (from SE):
```

Part D

```r
results_100_false <- run_hypothesis_test(n=100,b1_true=1,b1_null=0)
cat("Part d: Results for n = 100, H :    = 0 (false null)\n")
```

```
Part d: Results for n = 100, H :    = 0 (false null)
```

```r
cat("Rejection rate (power):", results_100_false$rejection_rate, "\n")
```

```
Rejection rate (power):
```

```r
cat("Mean ^ :", results_100_false$mean_b1_hat, "\n")
```

```
Mean ^ : 0.9941803
```

```r
cat("Variance of ^ :", results_100_false$var_b1_hat, "\n\n")
```

```
Variance of ^ : 0.05532426
```

```r
results_varying_n_false <- list()

for (n in sample_size) {
  set.seed(123)
  results_varying_n_false[[paste0("n", n)]] <- run_hypothesis_test(n = n, b1_true = 1, b1_nul

  cat("Results for n =", n, ", H :    = 0 (false null)\n")
  cat("Rejection rate (power):", results_varying_n_false[[paste0("n", n)]]$rejection_rate, "\
  cat("Mean ^ :", results_varying_n_false[[paste0("n", n)]]$mean_b1_hat, "\n")
  cat("Variance of ^ :", results_varying_n_false[[paste0("n", n)]]$var_b1_hat, "\n")
}
```

```
Results for n = 10 , H :   = 0 (false null)
Rejection rate (power):
Mean ^ : 1.017303
Variance of ^ : 1.029224
Results for n = 50 , H :   = 0 (false null)
Rejection rate (power):
Mean ^ : 1.009133
Variance of ^ : 0.1264878
Results for n = 500 , H :   = 0 (false null)
Rejection rate (power):
Mean ^ : 1.004012
Variance of ^ : 0.009162103
Results for n = 1000 , H :   = 0 (false null)
Rejection rate (power):
Mean ^ : 0.997987
Variance of ^ : 0.005327341
```