QuanGenomicsHW3

Noelle Wheeler

2023-03-03

### Question 1a

If you reject a null hypothesis this does not mean your null hypothesis is definitely wrong. This is because we are only looking at one sample. According to the definition of our alpha cutoff (let α = .05). Then we have a 5% chance of incorrectly rejecting the null hypothesis. This means if we reject the null hypothesis under then we are probably correct but there is a .05 probability that we are wrong.\ If you reject the null hypothesis you cannot interpret this result as evidence the null hy- pothesis is correct because, again, our sample does not give us all the information about our distribution. It is easier to disprove something than prove it, so this is all we can do with out limited information from our sample

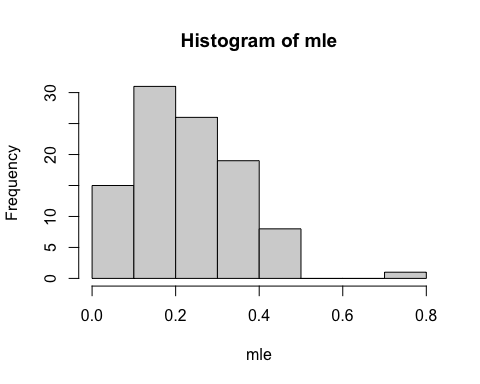
### Question 1b

We can set the Type I error because this is our alpha threshold value that we choose. We cannot control power because it depends on our actual parameter values such as our sample size and how we define our statistic.

### Question a

Assume that the true value of the parameter p = 0.2 (i.e., an unfair coin). Write code that produces 100 distinct i.i.d samples of the random variable X each with n = 10 experimental trials. Then for each sample, calculate the Maximum Likelihood Estimator, and using the function ‘hist()’ plot a histogram of these 100 estimator values.

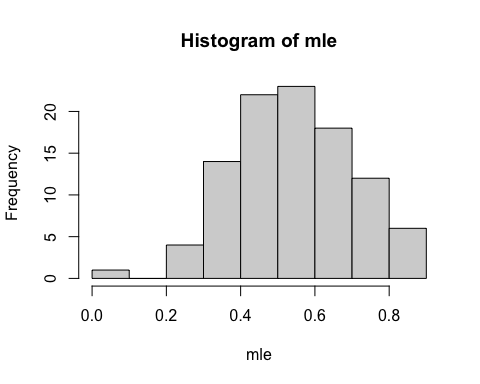
# produces sample with n = 10 experimental trails and p = .2  
y <- rbinom(10, 1, .2)  
# calculate mle  
mle <- c(sum(y / 10))  
# repeat 99 more times to get 100 samples   
for (i in 1:99) {  
 x <- rbinom(10, 1, .2)  
 y <- rbind(y, x)  
 mle <- append( mle, (sum(y[i + 1,]) / 10))  
}  
# plot 100 estimator values  
# set x-tick marks  
interval <- seq(min(mle), max(mle) + .1, by=.1)  
hist(mle, interval, right=FALSE)



### Question b

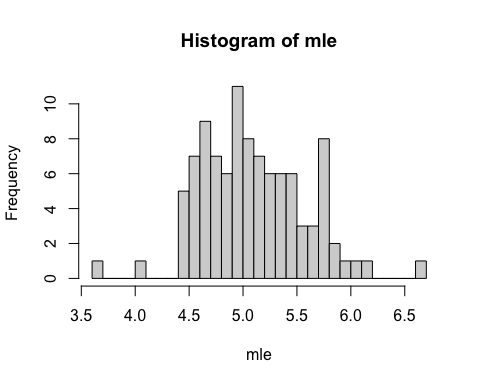
Repeat the exercise in part [a] but assume that p = 0.5

# produces sample with n = 10 experimental trails and p = .5  
y <- rbinom(10, 1, .5)  
# calculate mle  
mle <- c(sum(y / 10))  
# repeat 99 more times to get 100 samples   
for (i in 1:99) {  
 x <- rbinom(10, 1, .5)  
 y <- rbind(y, x)  
 mle <- append( mle, (sum(y[i + 1,]) / 10))  
}  
# plot 100 estimator values  
# set x-tick marks  
interval <- seq(min(mle), max(mle) + .1, by=.1)  
hist(mle, interval, right=FALSE)



Repeat the exercise in part [a] but assume that p = 0.5 and n = 100

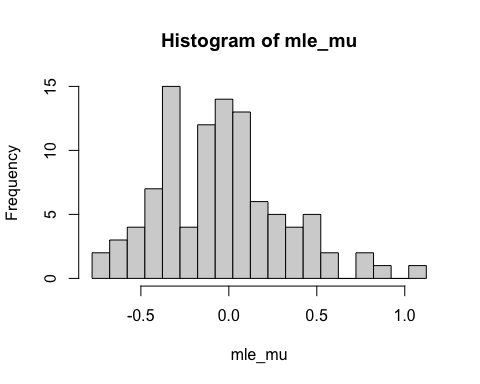
# produces sample with n = 100 experimental trails and p = .5  
y <- rbinom(100, 1, .5)  
# calculate mle  
mle <- c(sum(y / 10))  
# repeat 99 more times to get 100 samples   
for (i in 1:99) {  
 x <- rbinom(100, 1, .5)  
 y <- rbind(y, x)  
 mle <- append( mle, (sum(y[i + 1,]) / 10))  
}  
# plot 100 estimator values  
# set x-tick marks  
interval <- seq(min(mle), max(mle) + .1, by=.1)  
hist(mle, interval, right=FALSE)



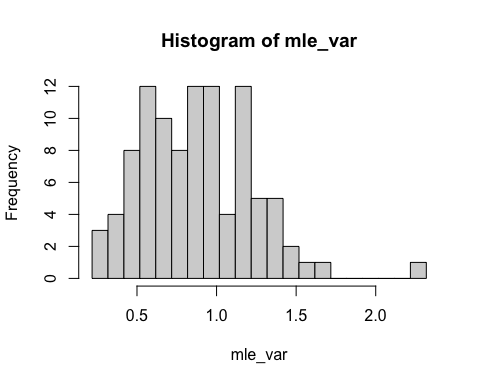
### Question b

Repeat question a but with the normal distribution where mu = 0 and variance = 1. Add mle estimate of variance.

# produces sample with n = 10 experimental trails and normal dist  
y <- rnorm(10)  
len <- length(y)  
# calculate mle estimates for mu and var  
mle\_mu <- c(sum(y / 10))  
# multiply variance by len-1/len so variance has denominator n  
mle\_var <- c(var(y) \* ((len-1)/len))  
# repeat 99 more times to get 100 samples   
for (i in 1:99) {  
 x <- rnorm(10)  
 y <- rbind(y, x)  
 mle\_mu <- append(mle\_mu, (sum(y[i + 1,]) / 10))  
 mle\_var <- append(mle\_var, (var(y[i + 1,]) \* ((len-1)/len)))  
}  
# plot 100 mu estimator values  
# set x-tick marks  
interval <- seq(min(mle\_mu), max(mle\_mu) + .1, by=.1)  
hist(mle\_mu, interval, right=FALSE)



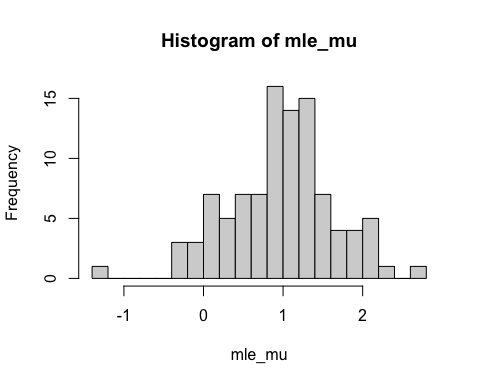
# plot 100 var estimator values   
interval <- seq(min(mle\_var), max(mle\_var) + .1, by=.1)  
hist(mle\_var, interval, right=FALSE)



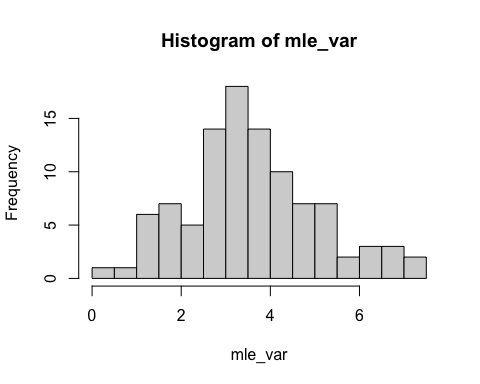
### Question d

Repeat the exercise in part c but assume that mu = 1 and variance = 2.

# produces sample with n = 10 experimental trails and normal dist  
y <- rnorm(10, mean = 1, sd = 2)  
len <- length(y)  
# calculate mle estimates for mu and var  
mle\_mu <- c(sum(y / 10))  
# multiply variance by len-1/len so variance has denominator n  
mle\_var <- c(var(y) \* ((len-1)/len))  
# repeat 99 more times to get 100 samples   
for (i in 1:99) {  
 x <- rnorm(10, mean = 1, sd = 2)  
 y <- rbind(y, x)  
 mle\_mu <- append(mle\_mu, (sum(y[i + 1,]) / 10))  
 mle\_var <- append(mle\_var, (var(y[i + 1,]) \* ((len-1)/len)))  
}  
# plot 100 mu estimator values  
# set x-tick marks  
hist(mle\_mu, breaks = 15)

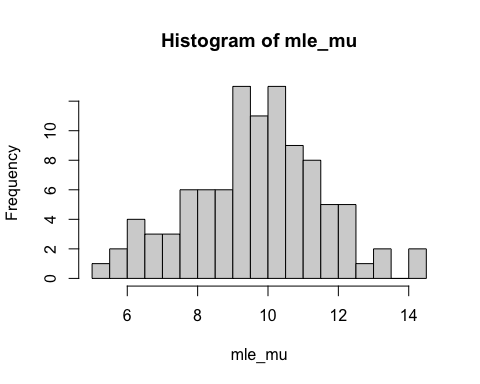


# plot 100 var estimator values   
hist(mle\_var, breaks = 15)

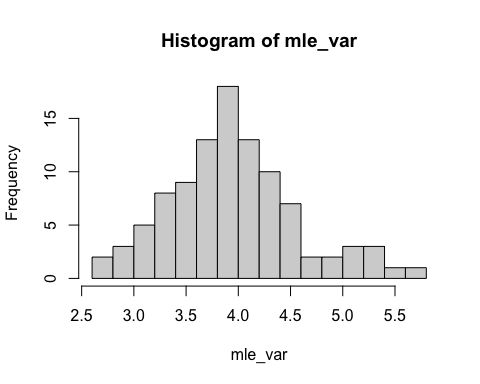


Repeat the exercise in part c but assume that mu = 1 and variance = 2 and n =100.

# produces sample with n = 10 experimental trails and normal dist  
y <- rnorm(100, mean = 1, sd = 2)  
len <- length(y)  
# calculate mle estimates for mu and var  
mle\_mu <- c(sum(y / 10))  
# multiply variance by len-1/len so variance has denominator n  
mle\_var <- c(var(y) \* ((len-1)/len))  
# repeat 99 more times to get 100 samples   
for (i in 1:99) {  
 x <- rnorm(100, mean = 1, sd = 2)  
 y <- rbind(y, x)  
 mle\_mu <- append(mle\_mu, (sum(y[i + 1,]) / 10))  
 mle\_var <- append(mle\_var, (var(y[i + 1,]) \* ((len-1)/len)))  
}  
# plot 100 mu estimator values  
# set x-tick marks  
hist(mle\_mu, breaks = 15)



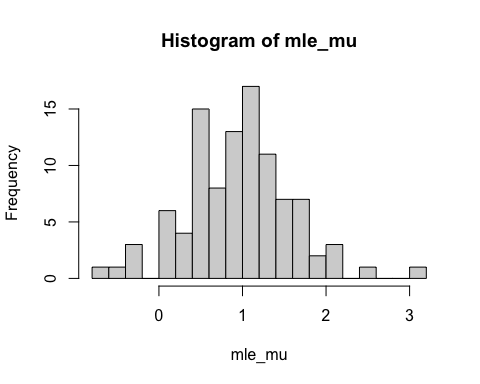
# plot 100 var estimator values   
hist(mle\_var, breaks = 15)



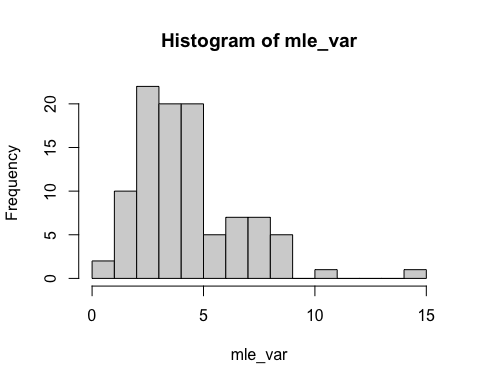
### Question e

Calculate the unbiased estimator of variance. mu = 1, var = 2, n = 10, unbiased variance

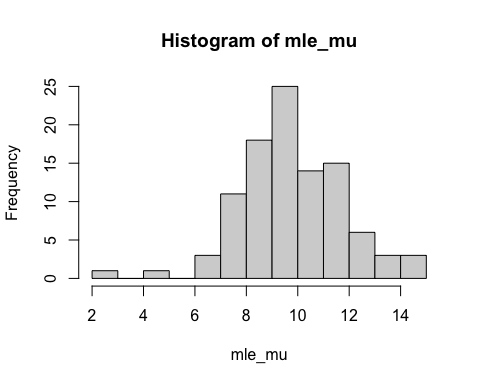
# produces sample with n = 10 experimental trails and normal dist  
y <- rnorm(10, mean = 1, sd = 2)  
len <- length(y)  
# calculate mle estimates for mu and var  
mle\_mu <- c(sum(y / 10))  
# calculate unbiased variance  
mle\_var <- c(var(y))  
# repeat 99 more times to get 100 samples   
for (i in 1:99) {  
 x <- rnorm(10, mean = 1, sd = 2)  
 y <- rbind(y, x)  
 mle\_mu <- append(mle\_mu, (sum(y[i + 1,]) / 10))  
 mle\_var <- append(mle\_var, (var(y[i + 1,])))  
}  
# plot 100 mu estimator values  
# set x-tick marks  
hist(mle\_mu, breaks = 15)



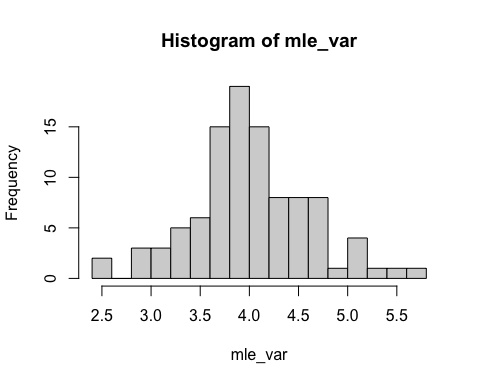
# plot 100 var estimator values   
hist(mle\_var, breaks = 15)

 mu = 1, var = 2, n = 100, unbiased variance

# produces sample with n = 10 experimental trails and normal dist  
y <- rnorm(100, mean = 1, sd = 2)  
len <- length(y)  
# calculate mle estimates for mu and var  
mle\_mu <- c(sum(y / 10))  
# calculate unbiased variance  
mle\_var <- c(var(y))  
# repeat 99 more times to get 100 samples   
for (i in 1:99) {  
 x <- rnorm(100, mean = 1, sd = 2)  
 y <- rbind(y, x)  
 mle\_mu <- append(mle\_mu, (sum(y[i + 1,]) / 10))  
 mle\_var <- append(mle\_var, (var(y[i + 1,])))  
}  
# plot 100 mu estimator values  
# set x-tick marks  
hist(mle\_mu, breaks = 15)



# plot 100 var estimator values   
hist(mle\_var, breaks = 15)



### Question f

What is the critical threshold for this statistic for a Type I error = α = 0.05

# one tailed distribution  
crit\_thres\_one <- qnorm(.95, mean = 0, sd = (1/sqrt(20)))  
print(paste0("the critical threshold for one-sided test is ", crit\_thres\_one))

## [1] "the critical threshold for one-sided test is 0.367800452290057"

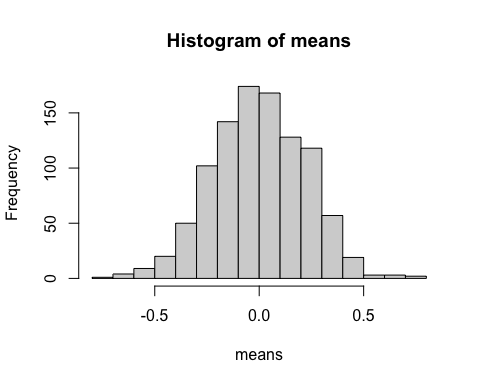
# two tailed distribution  
crit\_thres\_two <- qnorm(.975, mean = 0, sd = (1/sqrt(20)))  
print(paste0("the critical threshold for two-sided test is ", crit\_thres\_two))

## [1] "the critical threshold for two-sided test is 0.438261270288291"

### Question g

Simulate k = 1000 samples under the null hypothesis in part [f], calculate the mean for each sample, and plot a histogram of the means.

means <- c()  
for (i in 1:1000){  
 dist <- rnorm(20, mean = 0, sd = 1)  
 sample\_mean <- mean(dist)  
 means <- append(means, sample\_mean)  
}  
hist(means)



### Question h

What is the power of this test for the null hypothesis in part [f] with a Type I error = α = 0.05?

pow\_one <- pnorm(crit\_thres\_one, mean = .5, sd = (1/sqrt(20)))  
print(paste0("the power for one-sided test with alpha = .05 is ", pow\_one))

## [1] "the power for one-sided test with alpha = .05 is 0.277188404310798"

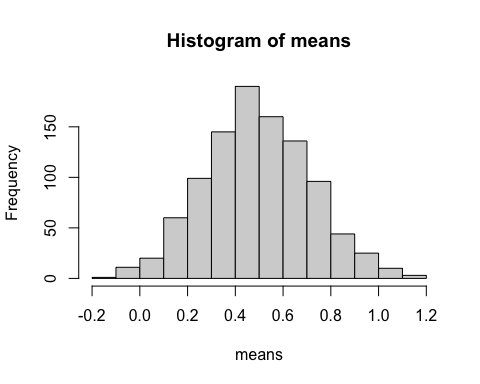
pow\_two <- pnorm(crit\_thres\_two, mean = .5, sd = (1/sqrt(20)))  
print(paste0("the power for two-sided test with alpha = .05 is ", pow\_two))

## [1] "the power for two-sided test with alpha = .05 is 0.391234096951103"

### Question i

Simulate k = 1000 samples under the true distribution (as used in part [h]), calculate the mean for each sample, and plot a histogram of the means.

means <- c()  
for (i in 1:1000){  
 dist <- rnorm(20, mean = .5, sd = 1)  
 sample\_mean <- mean(dist)  
 means <- append(means, sample\_mean)  
}  
hist(means)

 ### Question j Repeat parts [f-i] but consider n = 100.

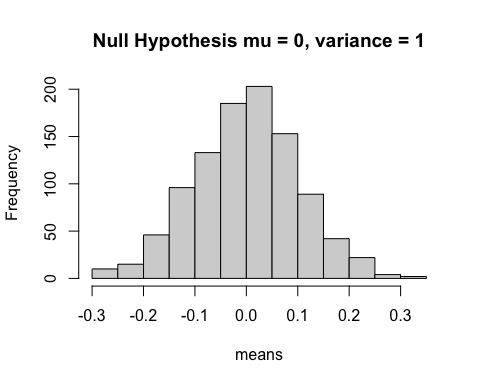
# critical threshold for this statistic for a Type I error = α = 0.05  
# one tailed distribution  
crit\_thres\_one <- qnorm(.95, mean = 0, sd = (1/sqrt(100)))  
print(paste0("the critical threshold for one-sided test is ", crit\_thres\_one))

## [1] "the critical threshold for one-sided test is 0.164485362695147"

# two tailed distribution  
crit\_thres\_two <- qnorm(.975, mean = 0, sd = (1/sqrt(100)))  
print(paste0("the critical threshold for two-sided test is ", crit\_thres\_two))

## [1] "the critical threshold for two-sided test is 0.195996398454005"

# Simulate k = 1000 samples under the null hypothesis  
means <- c()  
for (i in 1:1000){  
 dist <- rnorm(100, mean = 0, sd = 1)  
 sample\_mean <- mean(dist)  
 means <- append(means, sample\_mean)  
}  
hist(means, main = "Null Hypothesis mu = 0, variance = 1")



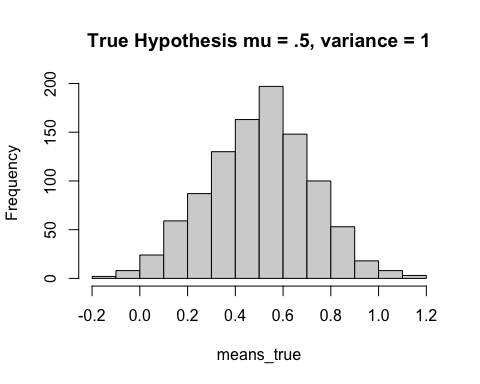
# calculate power  
pow\_one <- pnorm(crit\_thres\_one, mean = .5, sd = (1/sqrt(100)))  
print(paste0("the power for one-sided test with alpha = .05 is ", pow\_one))

## [1] "the power for one-sided test with alpha = .05 is 0.00039661500452629"

pow\_two <- pnorm(crit\_thres\_two, mean = .5, sd = (1/sqrt(100)))  
print(paste0("the power for two-sided test with alpha = .05 is ", pow\_two))

## [1] "the power for two-sided test with alpha = .05 is 0.00118274929819736"

# simulate k = 1000 samples under true distribution  
means\_true <- c()  
for (i in 1:1000){  
 dist <- rnorm(20, mean = .5, sd = 1)  
 sample\_mean <- mean(dist)  
 means\_true <- append(means\_true, sample\_mean)  
}  
hist(means\_true, main = "True Hypothesis mu = .5, variance = 1")



### Question 3

that we obtain for where   
This is derived from:  
  
  
  
  
  
For the Bernoulli distribution we get the same result: