Homework 1

Jonah Bacon

2/1/2022

Problem 1

```r
# Load data. *Make sure .csv data files are in your working directory
df.50 <- read.csv("Length Sample 50.csv", header = TRUE)
df.250 <- read.csv("Length Sample 250.csv", header = TRUE)

# Visualize data structures
head(df.50)
```

```
##   X        x
## 1 1 31.64691
## 2 2 27.69437
## 3 3 32.70690
## 4 4 26.39351
## 5 5 11.42020
## 6 6 47.69497
```

```r
head(df.250)
```

```
##   X        x
## 1 1 35.42882
## 2 2 21.24048
## 3 3 36.44932
## 4 4 19.21074
## 5 5 56.91008
## 6 6 45.58293
```

```r
## I don't like the column names, so I'll rename them with more intuitive labels:
colnames(df.50) <- c("ID","length")
colnames(df.250) <- c("ID","length")

# Sample size = 50
y.bar.50 <- mean(df.50$length)
var.50 <- var(df.50$length)
CV.50 <- sqrt(var.50)/y.bar.50
var.hat_y.bar.50 <- ( (1000-50)/1000 )* var.50/length(df.50$length)

print(y.bar.50); print(var.50); print(CV.50); print(var.hat_y.bar.50)
```

```
## [1] 39.96425
```

```
## [1] 353.1413
```

```
## [1] 0.4702216
```

```
## [1] 6.709685
```

```r
# Sample size = 250
y.bar.250 <- mean(df.250$length)
var.250 <- var(df.250$length)
CV.250 <- sqrt(var.250)/y.bar.250
var.hat_y.bar.250 <- ( (1000-250)/1000 ) * var.250/length(df.250$length)

print(y.bar.250); print(var.250); print(CV.250); print(var.hat_y.bar.250)
```

```
## [1] 44.46145
```

```
## [1] 638.1551
```

```
## [1] 0.5681716
```

```
## [1] 1.914465
```

In the sample size of 50, the mean length was 39.96 mm with a variance of 353. Converting the variance in mean length to a relative scale results in a CV of 0.47. The variance in our mean length estimate (y_bar) was 6.71.

For the sample size of 250, the calculated mean length was longer (44.46 mm) with a larger variance as well (638). The standardized variability estimate CV was 0.57, indicating that the data was more variable than our sample with 50 individuals. Finally, the estimated variance in the mean estimate was much lower (1.91) owing to the fact that our sample size was larger.

Problem 2

```r
# Create data frame:
samp.ID <- seq(1,15,1)
flower.count <- c(10,15,1,23,18,12,19,9,5,8,3,26,20,12,31)
flower.df <- cbind(samp.ID,flower.count)

# Sample statistics:
flwr.y.bar <- mean(flower.count)
flwr.var <- var(flower.count)
flwr.CV <- sqrt(flwr.var)/flwr.y.bar
var.hat_flwr.y.bar <- ( (6000 - 15)/6000) * flwr.var/length(flower.count) # For this, we need N, which

print(flwr.y.bar); print(flwr.var); print(flwr.CV); print(var.hat_flwr.y.bar)
```

```
## [1] 14.13333
## [1] 74.8381
```

```
## [1] 0.6120921
```

```
## [1] 4.976733
```

```
# Population estimates:
samp.space <- 50*120 # Determine the total sample space (i.e. N, or how many 1x1 quadrats are in the fi
flower.pop.est <- samp.space * flwr.y.bar
flower.pop.est.var <- samp.space^2 * flwr.var
flower.pop.est.CV <- sqrt(flower.pop.est.var)/flower.pop.est

print(flower.pop.est); print(flower.pop.est.var); print(flower.pop.est.CV)
```

```
## [1] 84800
```

```
## [1] 2694171429
```

```
## [1] 0.6120921
```

From our sample, the average number of flower per quadrat was 14.133. Scaling this up to the size of the field allows us to estimate the total population size. Thus, we multiplied the average flowers/quadrat times the number of quadrat areas within the field. Our population estimate was 84,800 flowers in the field. The variance of our sample was 74.84 and the variance for our population estimate was massive (2694171429). However, what is interesting, is that the CV of our sample was identical to our CV for the population estimate. Intuitively this makes sense because our population estimate is based off of the distribution of our sample. Since our sample was an unbiased estimator of our population, the scaled variance doesn't change.

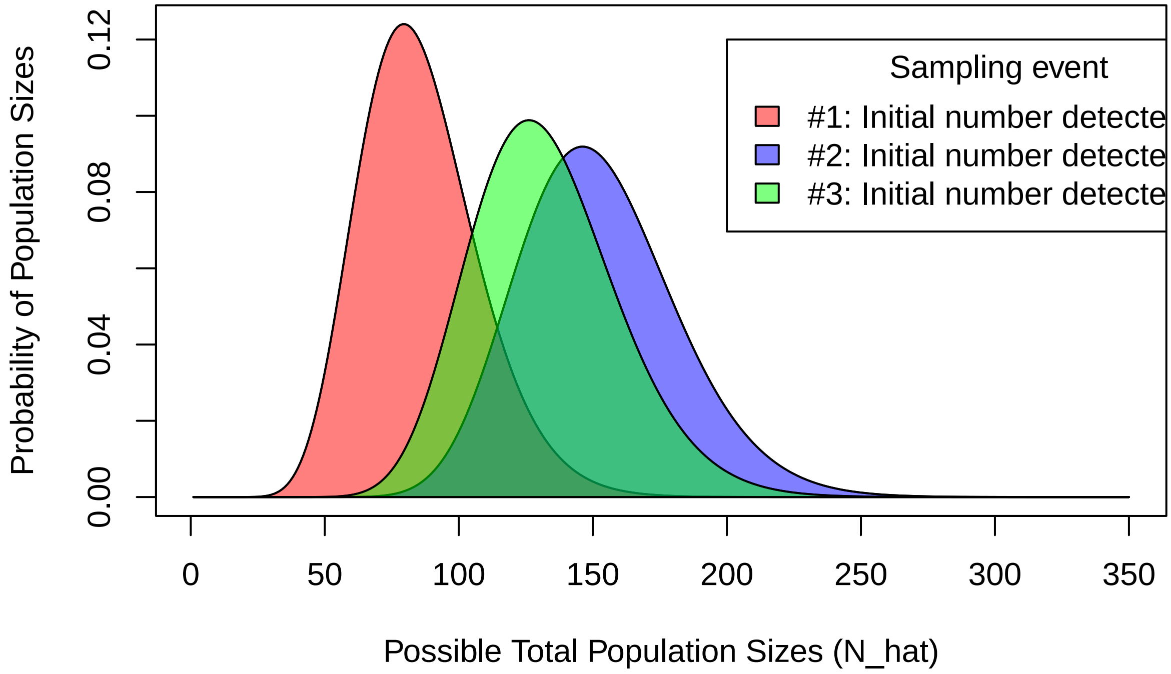
Problem 3

```
# Sampling event 1:
n.detect.1 <- 12 # Number of voles detected in this sample event
p <- 0.15 # Probability of detection = proportion of apartment visible

N_hat.est.1 <- n.detect.1/p
N_hat.trial <- 1:350
probs.N_hat.1 <- dbinom(x=n.detect.1, size=N_hat.trial, prob=p)

# Sampling event 2:
n.detect.2 <- 22 # Number of voles detected in this sample event

N_hat.est.2 <- n.detect.2/p
probs.N_hat.2 <- dbinom(x=n.detect.2, size=N_hat.trial, prob=p)

# Sampling event 3:
n.detect.3 <- 19 # Number of voles detected in this sample event

N_hat.est.3 <- n.detect.3/p
probs.N_hat.3 <- dbinom(x=n.detect.3, size=N_hat.trial, prob=p)
```

```
# Plot the 3 profiles:
plot(x=N_hat.trial, y=probs.N_hat.1, type="l",
     xlab="Possible Total Population Sizes (N_hat)",
     ylab="Probability of Population Sizes") +
  lines(x=N_hat.trial, y=probs.N_hat.2) +
  lines(x=N_hat.trial, y=probs.N_hat.3)
```

```
## integer(0)
```

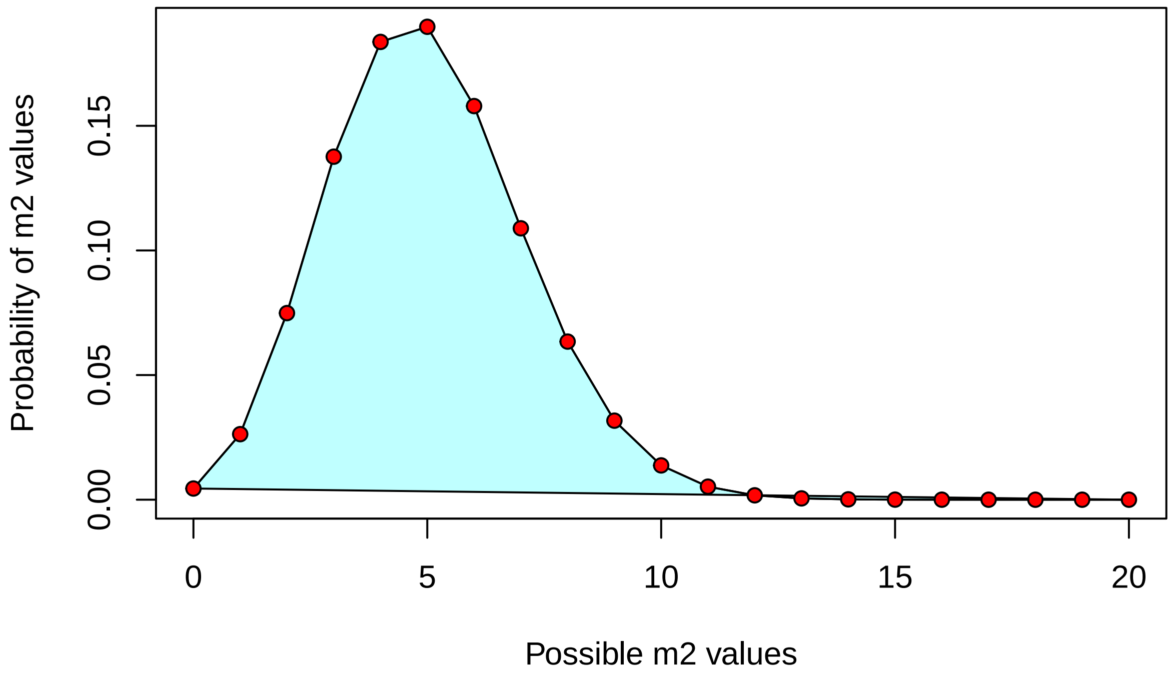
```
legend(x = 200, y = 0.12, legend = c("#1: Initial number detected = 12", "#2: Initial number detected =
       fill = c(rgb(1,0,0, alpha = 0.5),rgb(0,0,1, alpha = 0.5),rgb(0,1,0, alpha = 0.5)), title = "Sampl
polygon(x=N_hat.trial, y=probs.N_hat.1, col=rgb(1,0,0, alpha = 0.5))
polygon(x=N_hat.trial, y=probs.N_hat.2, col=rgb(0,0,1, alpha = 0.5))
polygon(x=N_hat.trial, y=probs.N_hat.3, col=rgb(0,1,0, alpha = 0.5))
```



The estimated N_hat (population size) from the three sampling events was 80, 147, and 127 for the first, second, and third sampling events, respectively. The plot visualizes the distributions resulting from the three sampling events. Event 1 resulted in a distribution with relatively lower variation centered on a lower N_hat estimate. Event 2 had the highest N_hat estimate and the highest variation around that estimate. Event 3 was intermediate between the two in terms of N_hat estimate and variation.

Problem 4

```r
N <- 1000
n1 <- 50
n2 <- 100

# Number of marked individuals expected to be observed:
m2_hat <- n1*n2 / N

# Calculate probability of observing different numbers of marked individuals and plot:
trial.m2 <- 0:20
trial.prob.m2 <- dhyper(x=trial.m2, m=50, n=950, k=100)

plot(x=trial.m2, y=trial.prob.m2, type="l",
     xlab="Possible m2 values",
     ylab="Probability of m2 values")
polygon(x=trial.m2, y=trial.prob.m2, col=rgb(0,1,1, alpha=0.25))
points(x=trial.m2, y=trial.prob.m2, pch=21, bg="red")
```



```r
# Prove original estimator by summing probabilities*m2:
est.m2 <- sum(trial.m2*trial.prob.m2)
est.m2
```

```
## [1] 5
## Varying n2
n2 <- 50
m2_hat.2 <- n1*n2 / N
m2_hat.2
```

```
## [1] 2.5
```

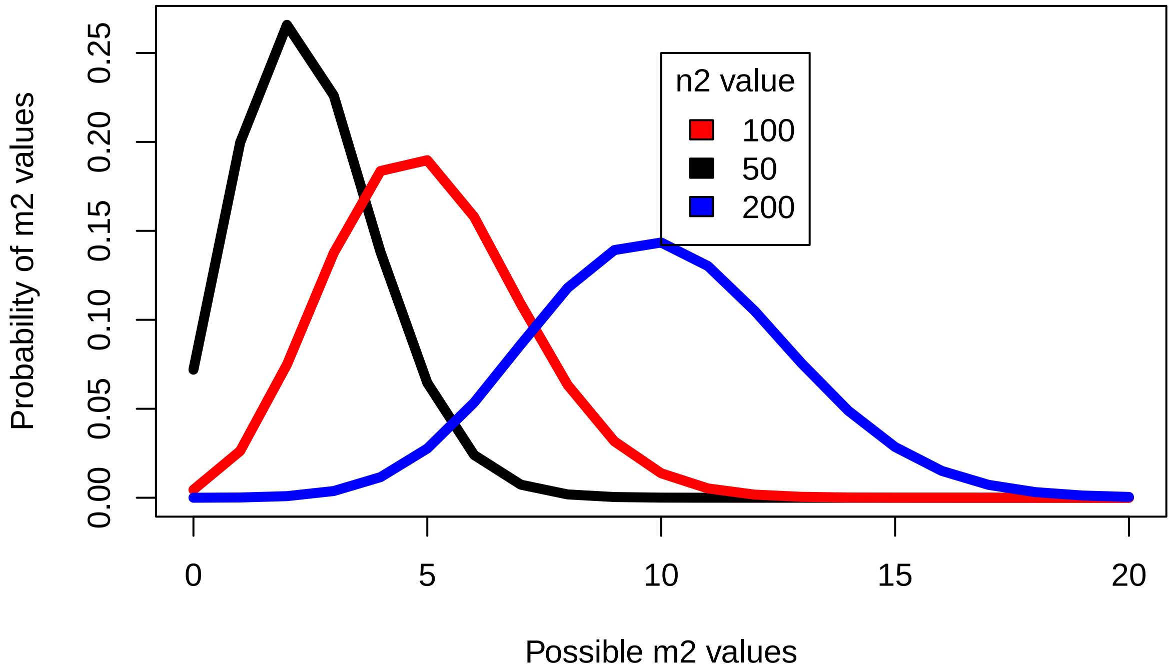
```
n2 <- 200
m2_hat.3 <- n1*n2 / N
m2_hat.3
```

```
## [1] 10
```

```
trial.m2 <- 0:20
trial.prob.m2.2 <- dhyper(x=trial.m2, m=50, n=950, k=50) # Change k (n2) to 50
trial.prob.m2.3 <- dhyper(x=trial.m2, m=50, n=950, k=200) # Change k (n2) to 200

plot(x=trial.m2, y=trial.prob.m2.2, type="l", lwd = 5,
     xlab="Possible m2 values",
     ylab="Probability of m2 values") +
  lines(x=trial.m2, y=trial.prob.m2, col="red", lwd = 5) +
  lines(x=trial.m2, y=trial.prob.m2.3, col = "blue", lwd = 5)
```

```
## integer(0)
```

```
legend(x = 10, y = 0.25, legend = c("100", "50", "200"),
       fill = c("red", "black", "blue"), title = "n2 value")
```



Altering the number of whales inspected for marks (n2) results in different distributions for the number of marked individuals we expect to observe. When we do not put much effort into our resample (n2 = 50), our m2 distribution is shifted to the left with an m2=2.5 having the maximum likelihood. Under this scenario, an m2=0 has a probability of 0.072, which is pretty high in my opinion. Not observing any marked individuals at t2 (m2 = 0) would certainly cause a researcher headaches attempting to estimate N_hat. Under a high t2 sampling effort (n2 = 200), the expected m2 increases to 10 and there is a low probability of observing very few or zero marked individuals. With an intermediate t2 sampling effort (n2 = 100), the expected m2 is unsurprisingly intermediate (E(m2) = 5) and there is a decent chance that very few marked individuals may be observed. Under this scenario, I would try for high effort at t2 to ensure we don't have too few m2 observations to make an accurate N_hat estimate.

```
## Varying N
n2 <- 100

N <- 500
m2_hat.4 <- n1*n2 / N
m2_hat.4
```

```
## [1] 10
```

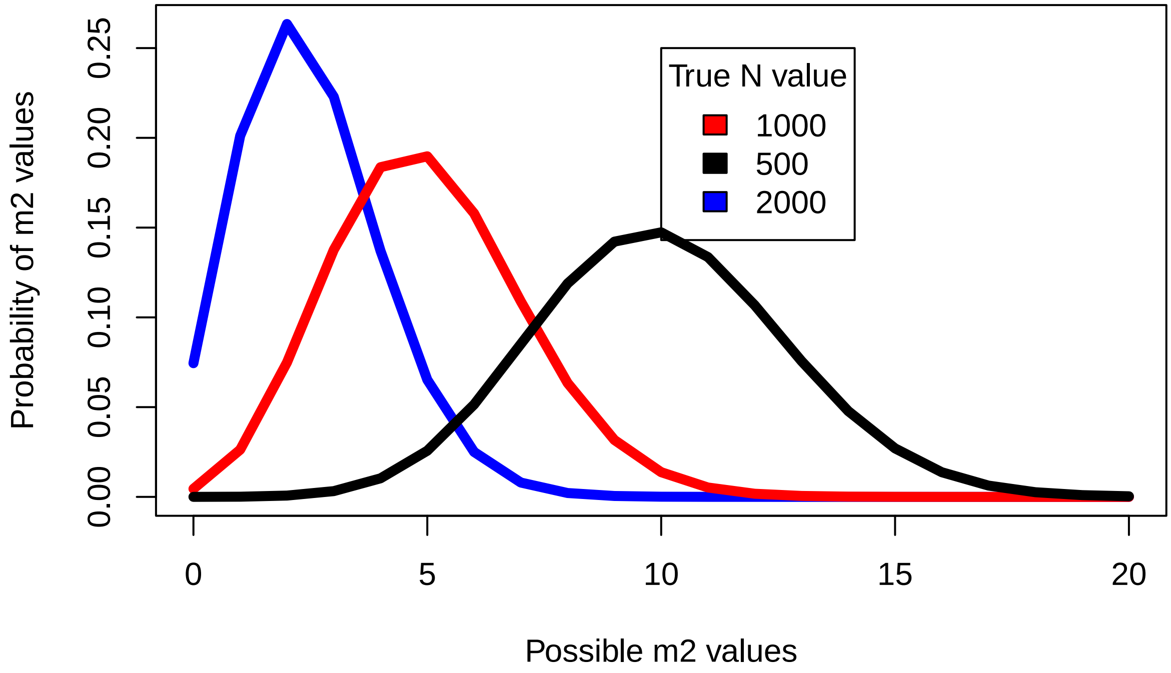
```
N <- 2000
m2_hat.5 <- n1*n2 / N
m2_hat.5
```

```
## [1] 2.5
```

```
trial.m2 <- 0:20
trial.prob.m2.4 <- dhyper(x=trial.m2, m=50, n=450, k=100) # Change n (= N-m) to 450
trial.prob.m2.5 <- dhyper(x=trial.m2, m=50, n=1950, k=100) # Change n (= N-m) to 1950

plot(x=trial.m2, y=trial.prob.m2.5, type="l", lwd = 5, col = "blue",
     xlab="Possible m2 values",
     ylab="Probability of m2 values") +
  lines(x=trial.m2, y=trial.prob.m2, col="red", lwd = 5) +
  lines(x=trial.m2, y=trial.prob.m2.4, col = "black", lwd = 5)
```

```
## integer(0)
```

```
legend(x = 10, y = 0.25, legend = c("1000", "500", "2000"),
       fill = c("red", "black", "blue"), title = "True N value")
```



When the true abundance (N) varies, our expected m2 varies as well. When N is high (N = 2000) we expect to observe fewer marked individuals at t2 as compared to if N were low (N = 500). Under these two scenarios, the expected m2 (E(m2)) is 2.5 and 10, respectively. When N is intermediate (N = 1000), the E(m2) is intermediate as well (E(m2) = 5). If we expected N to be relatively large, we would expect low m2 under constant t2 effort, and conversely if we expected N to be relatively small, we would expect higher m2 values under constant t2 effort.

Problem 5

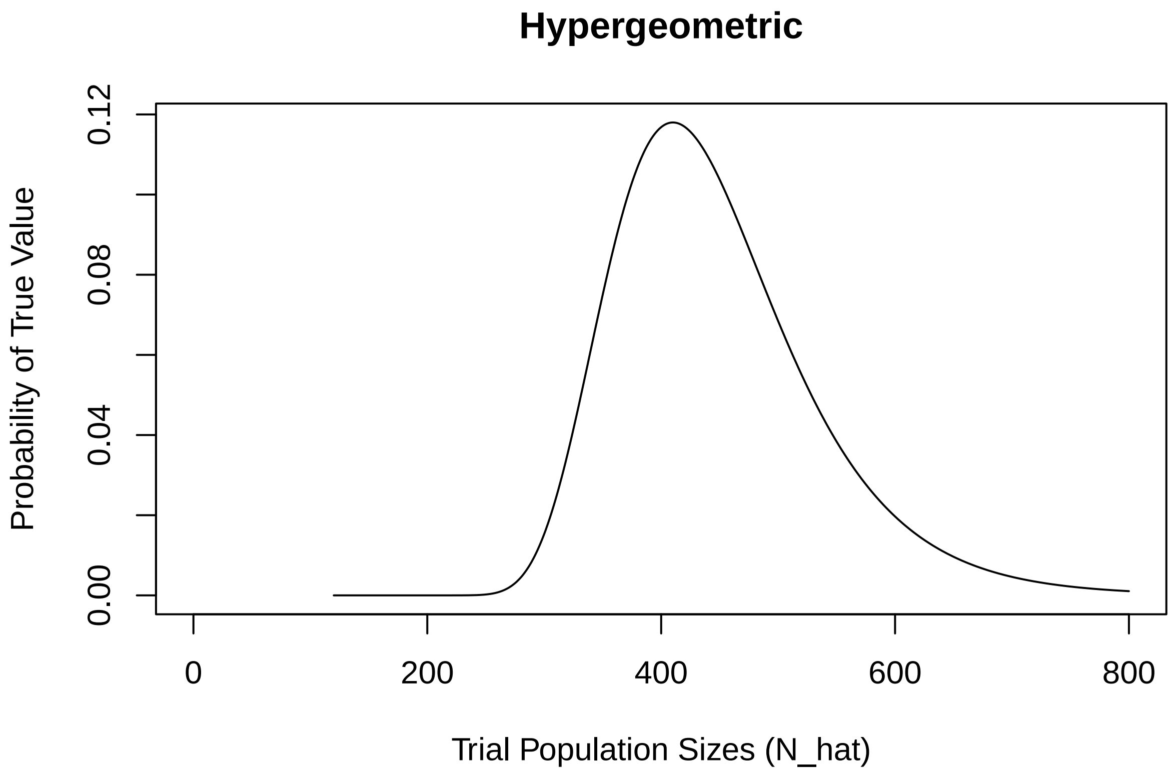
Chapman abundance estimation

May

```
n1 = 65
n2 = 120
m2 = 19

chap.may.est <- ((n1+1)*(n2+1))/(m2+1) - 1
chap.may.var <- ((n1+1)*(n2+1)*(n1-m2)*(n2-m2)) / ((m2+1)^2 * (m2+2))
chap.may.CV <- sqrt(chap.may.var)/chap.may.est

trial.N <- 0:800
may.trial.N.prob <- dhyper(x=m2, m=n1, n=trial.N-n1, k=n2)
```

```
## Warning in dhyper(x = m2, m = n1, n = trial.N - n1, k = n2): NaNs produced
```

```
plot(x=trial.N, y=may.trial.N.prob, type="l",
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Hypergeometric")
```

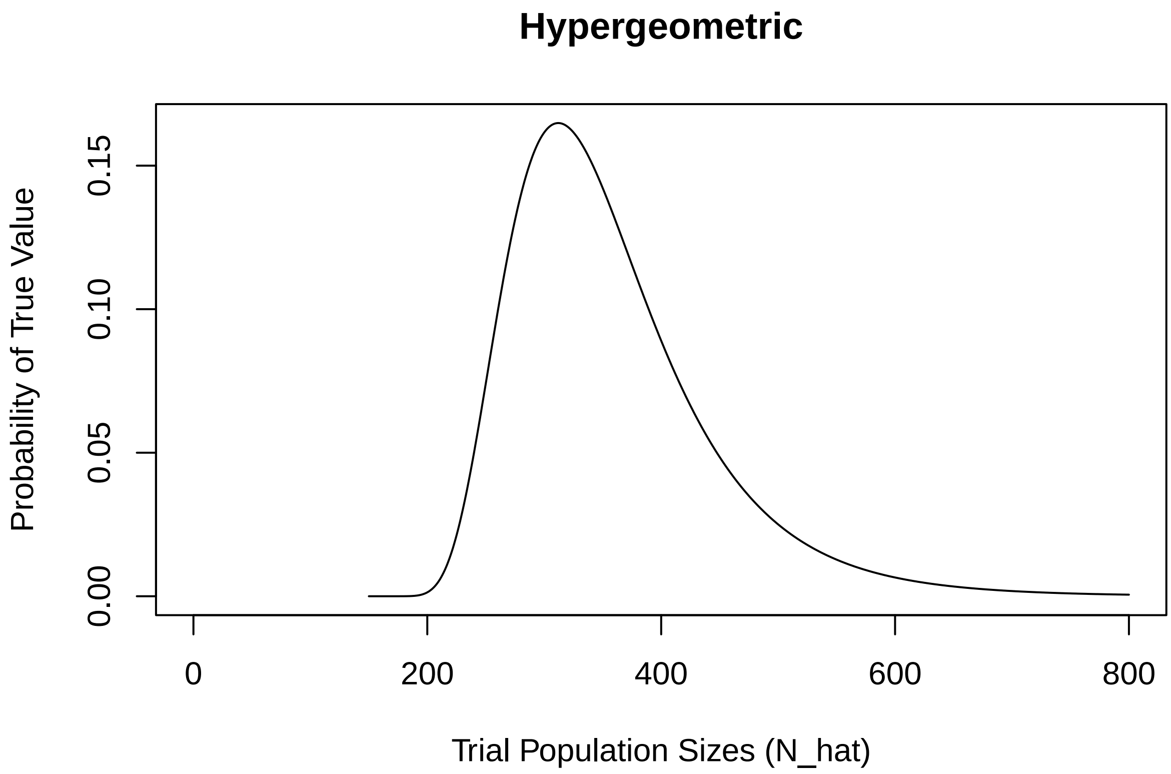


### June

```
n1 = 25
n2 = 150
m2 = 12

chap.june.est <- ((n1+1)*(n2+1))/(m2+1) - 1
chap.june.var <- ((n1+1)*(n2+1)*(n1-m2)*(n2-m2)) / ((m2+1)^2 * (m2+2))
chap.june.CV <- sqrt(chap.june.var)/chap.june.est

june.trial.N.prob <- dhyper(x=m2, m=n1, n=trial.N-n1, k=n2)
```

```
## Warning in dhyper(x = m2, m = n1, n = trial.N - n1, k = n2): NaNs produced
```

```
plot(x=trial.N, y=june.trial.N.prob, type="l",
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Hypergeometric")
```

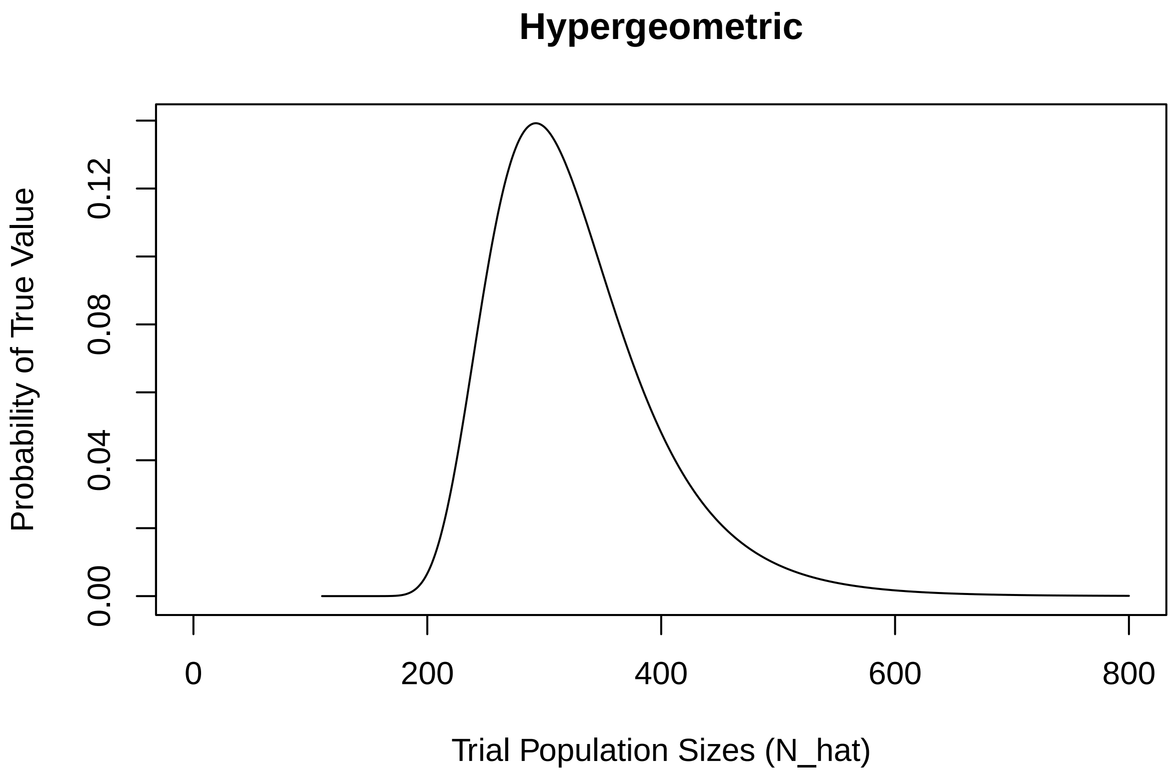


### July

```
n1 = 40
n2 = 110
m2 = 15

chap.july.est <- ((n1+1)*(n2+1))/(m2+1) - 1
chap.july.var <- ((n1+1)*(n2+1)*(n1-m2)*(n2-m2)) / ((m2+1)^2 * (m2+2))
chap.july.CV <- sqrt(chap.july.var)/chap.july.est

july.trial.N.prob <- dhyper(x=m2, m=n1, n=trial.N-n1, k=n2)
```

```
## Warning in dhyper(x = m2, m = n1, n = trial.N - n1, k = n2): NaNs produced
```

```
plot(x=trial.N, y=july.trial.N.prob, type="l",
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Hypergeometric")
```

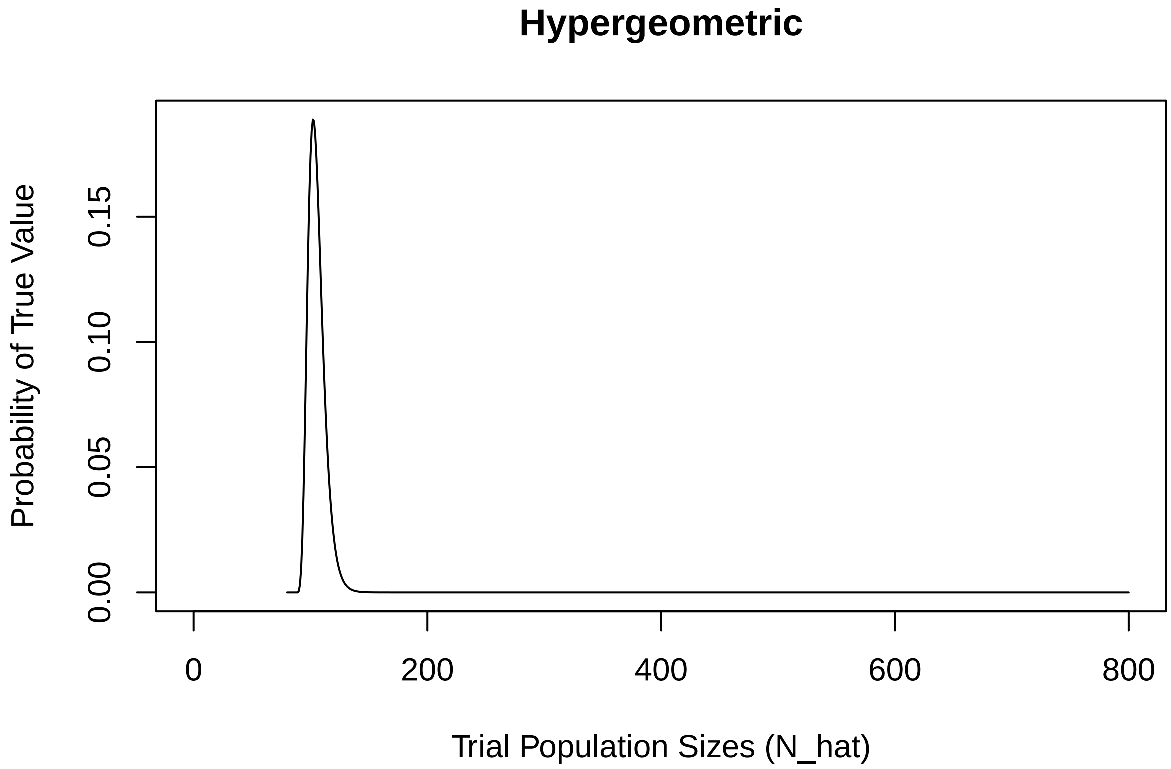


### August

```
n1 = 45
n2 = 80
m2 = 35

chap.august.est <- ((n1+1)*(n2+1))/(m2+1) - 1
chap.august.var <- ((n1+1)*(n2+1)*(n1-m2)*(n2-m2)) / ((m2+1)^2 * (m2+2))
chap.august.CV <- sqrt(chap.august.var)/chap.august.est

august.trial.N.prob <- dhyper(x=m2, m=n1, n=trial.N-n1, k=n2)
```

```
## Warning in dhyper(x = m2, m = n1, n = trial.N - n1, k = n2): NaNs produced
```

```
plot(x=trial.N, y=august.trial.N.prob, type="l",
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Hypergeometric")
```

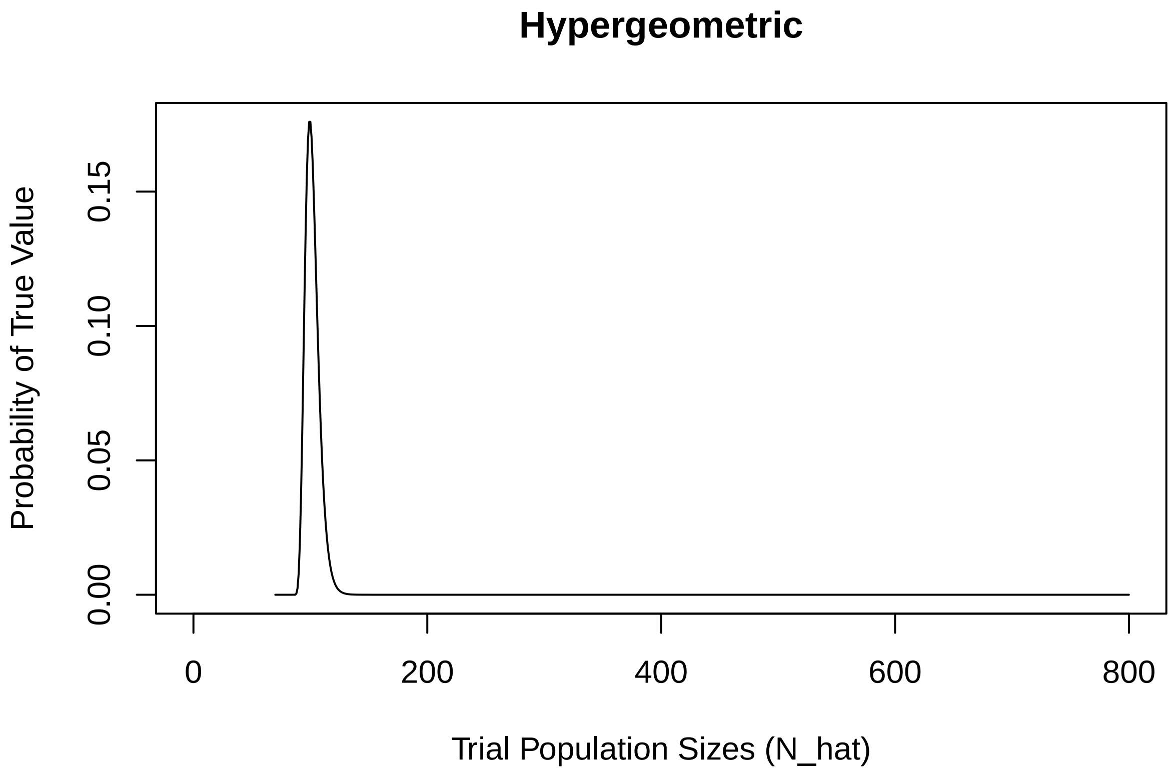


### September

```
n1 = 60
n2 = 70
m2 = 42

chap.september.est <- ((n1+1)*(n2+1))/(m2+1) - 1
chap.september.var <- ((n1+1)*(n2+1)*(n1-m2)*(n2-m2)) / ((m2+1)^2 * (m2+2))
chap.september.CV <- sqrt(chap.september.var)/chap.september.est

september.trial.N.prob <- dhyper(x=m2, m=n1, n=trial.N-n1, k=n2)
```

```
## Warning in dhyper(x = m2, m = n1, n = trial.N - n1, k = n2): NaNs produced
```

```
plot(x=trial.N, y=september.trial.N.prob, type="l",
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Hypergeometric")
```



Combined table & plot

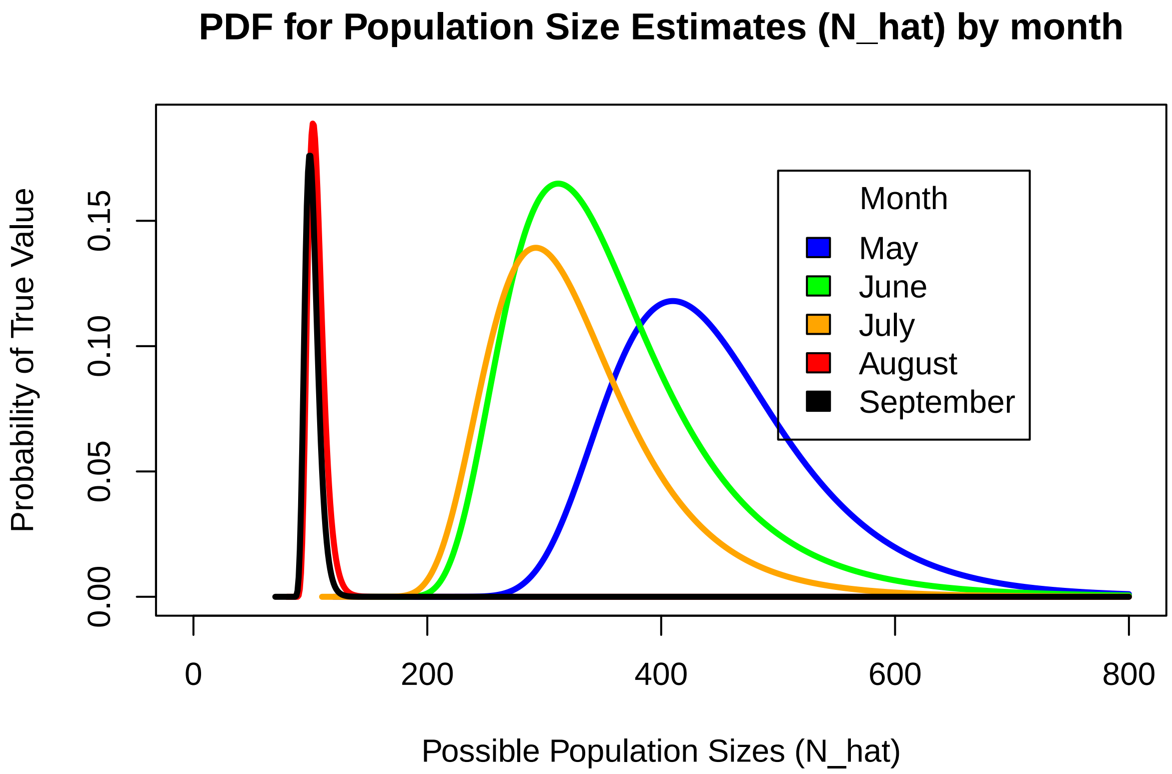
```
chapman_table <- cbind(N_hat <- c(chap.may.est, chap.june.est, chap.july.est, chap.august.est, chap.sep
rownames(chapman_table) <- c("May", "June", "July", "August", "September")
colnames(chapman_table) <- c("N_hat", "var", "CV")
chapman_table
```

```
##                N_hat        var         CV
## May        398.30000 4417.01857 0.16686079
## June       301.00000 2976.85714 0.18126439
## July       283.43750 2483.59949 0.17582615
## August     102.50000   34.96622 0.05768999
## September   99.72093   26.83052 0.05194315
```

```
plot(x=trial.N, y=august.trial.N.prob, type="l", col = "red", lwd = 3,
     xlab="Possible Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="PDF for Population Size Estimates (N_hat) by month") +
  lines(x=trial.N, y=may.trial.N.prob, col="blue", lwd = 3) +
  lines(x=trial.N, y=june.trial.N.prob, col="green", lwd = 3) +
  lines(x=trial.N, y=july.trial.N.prob, col="orange", lwd = 3) +
  lines(x=trial.N, y=september.trial.N.prob, col="black", lwd = 3)
```

```
## integer(0)
```

```r
legend(x = 500, y = 0.17, legend = c("May", "June", "July", "August", "September"),
       fill = c("blue", "green", "orange", "red", "black"), title = "Month")
```



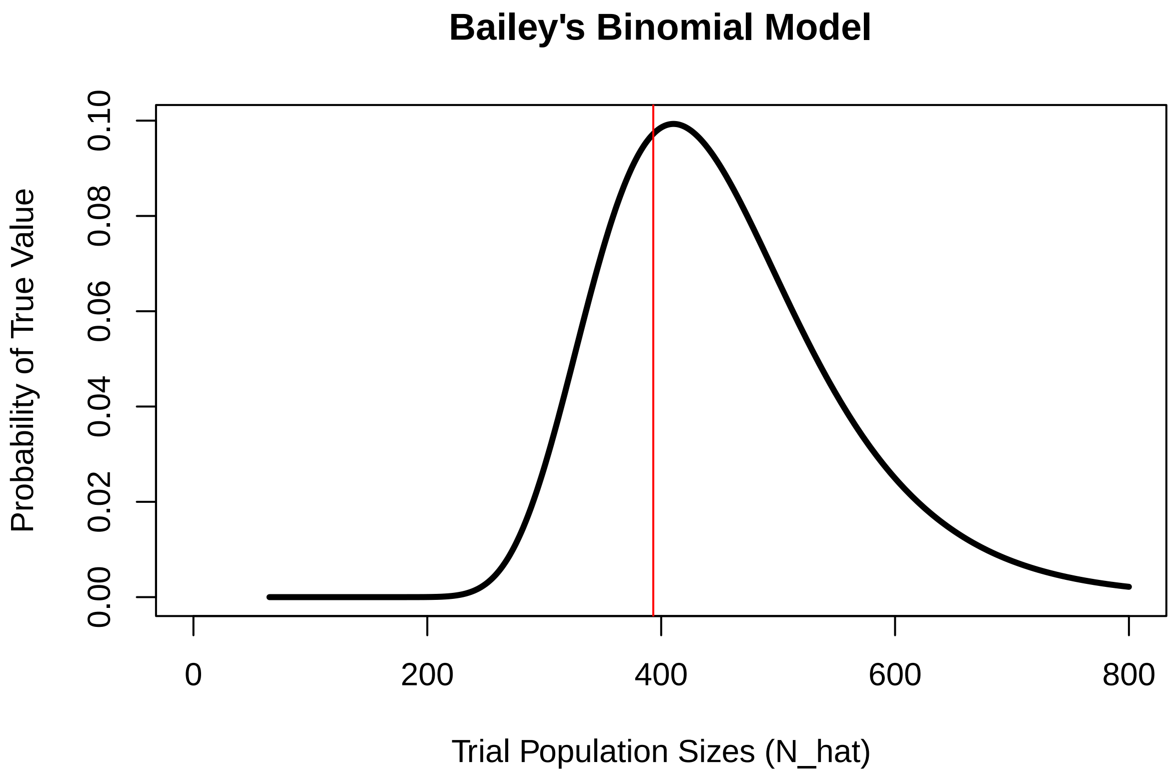
Bailey abundance estimation (sample w/ replacement)

May

```r
n1 = 65
n2 = 120
m2 = 19

bailey.may.est <- (n1*(n2+1)) / (m2+1)
bailey.may.var <- (n1^2 * (n2+1) * (n2-m2)) / ((m2+1)^2 * (m2+2))
bailey.may.CV <- sqrt(bailey.may.var)/bailey.may.est

trial.N <- 0:800
may.trial.N.prob_bailey <- dbinom(x=m2, size=n2, prob=n1/trial.N)
```

```
## Warning in dbinom(x = m2, size = n2, prob = n1/trial.N): NaNs produced
```

```r
plot(x=trial.N, y=may.trial.N.prob_bailey, type="l", lwd = 3,
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Bailey's Binomial Model")
abline(v=bailey.may.est, col="red")
```

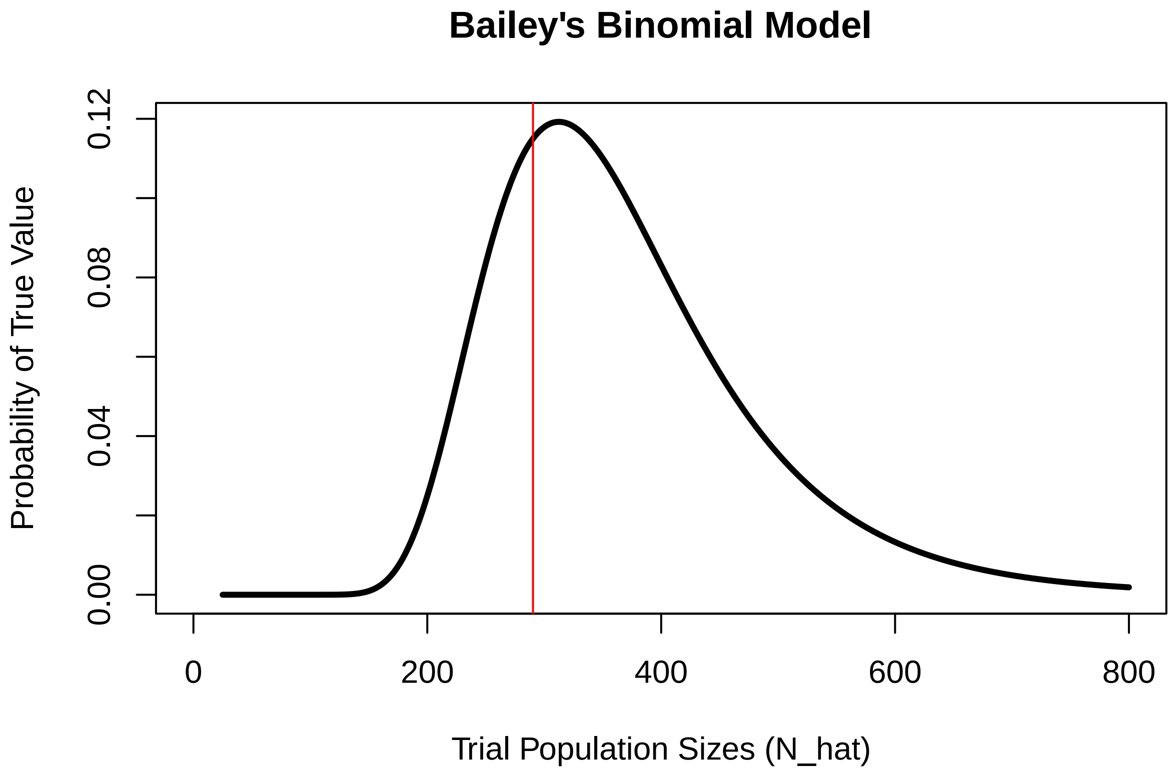


### June

```
n1 = 25
n2 = 150
m2 = 12

bailey.june.est <- (n1*(n2+1)) / (m2+1)
bailey.june.var <- (n1^2 * (n2+1) * (n2-m2)) / ((m2+1)^2 * (m2+2))
bailey.june.CV <- sqrt(bailey.june.var)/bailey.june.est

june.trial.N.prob_bailey <- dbinom(x=m2, size=n2, prob=n1/trial.N)
```

```
## Warning in dbinom(x = m2, size = n2, prob = n1/trial.N): NaNs produced
```

```
plot(x=trial.N, y=june.trial.N.prob_bailey, type="l", lwd = 3,
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Bailey's Binomial Model")
abline(v=bailey.june.est, col="red")
```

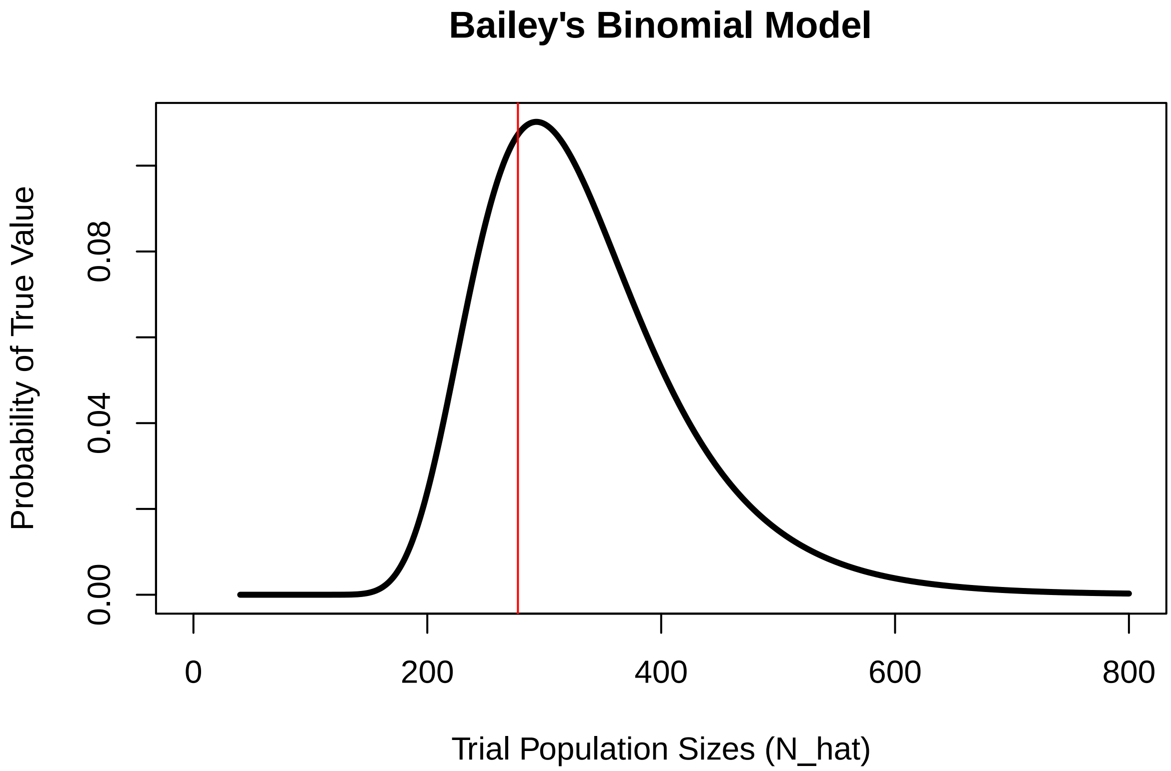


### July

```
n1 = 40
n2 = 110
m2 = 15

bailey.july.est <- (n1*(n2+1)) / (m2+1)
bailey.july.var <- (n1^2 * (n2+1) * (n2-m2)) / ((m2+1)^2 * (m2+2))
bailey.july.CV <- sqrt(bailey.july.var)/bailey.july.est

july.trial.N.prob_bailey <- dbinom(x=m2, size=n2, prob=n1/trial.N)
```

```
## Warning in dbinom(x = m2, size = n2, prob = n1/trial.N): NaNs produced
```

```
plot(x=trial.N, y=july.trial.N.prob_bailey, type="l", lwd = 3,
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Bailey's Binomial Model")
abline(v=bailey.july.est, col="red")
```

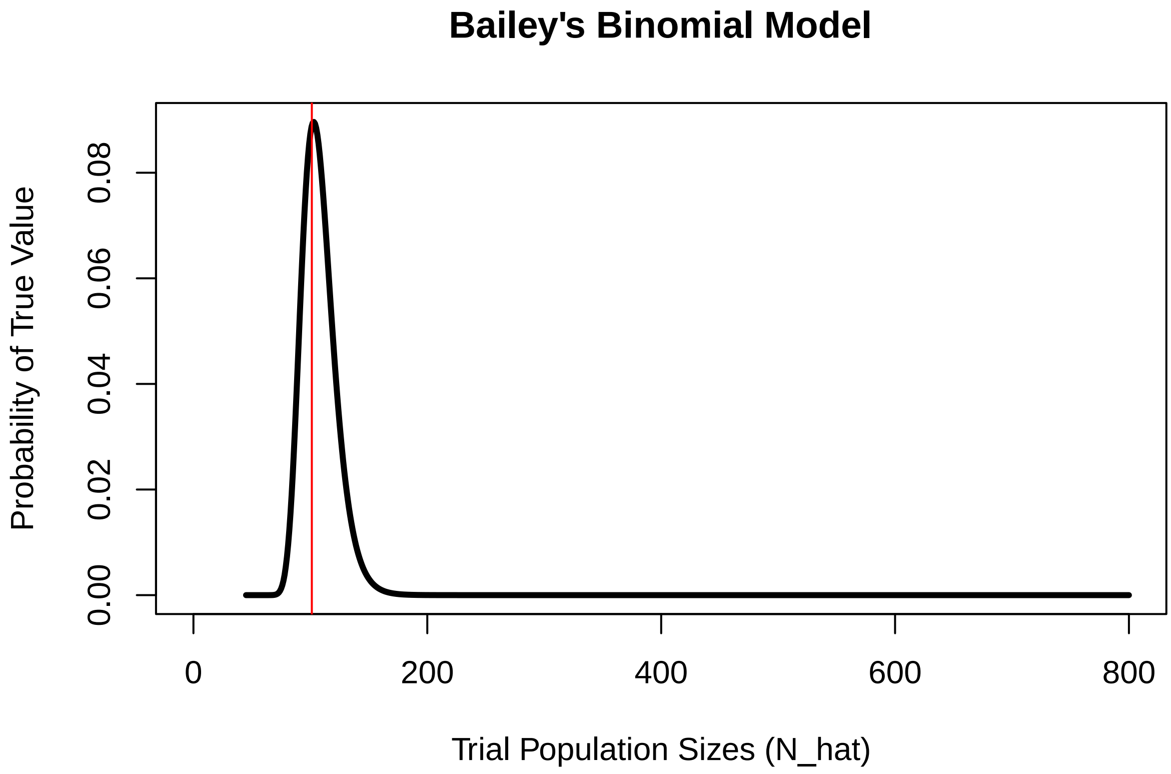


### August

```
n1 = 45
n2 = 80
m2 = 35

bailey.august.est <- (n1*(n2+1)) / (m2+1)
bailey.august.var <- (n1^2 * (n2+1) * (n2-m2)) / ((m2+1)^2 * (m2+2))
bailey.august.CV <- sqrt(bailey.august.var)/bailey.august.est

august.trial.N.prob_bailey <- dbinom(x=m2, size=n2, prob=n1/trial.N)
```

```
## Warning in dbinom(x = m2, size = n2, prob = n1/trial.N): NaNs produced
```

```
plot(x=trial.N, y=august.trial.N.prob_bailey, type="l", lwd = 3,
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Bailey's Binomial Model")
abline(v=bailey.august.est, col="red")
```

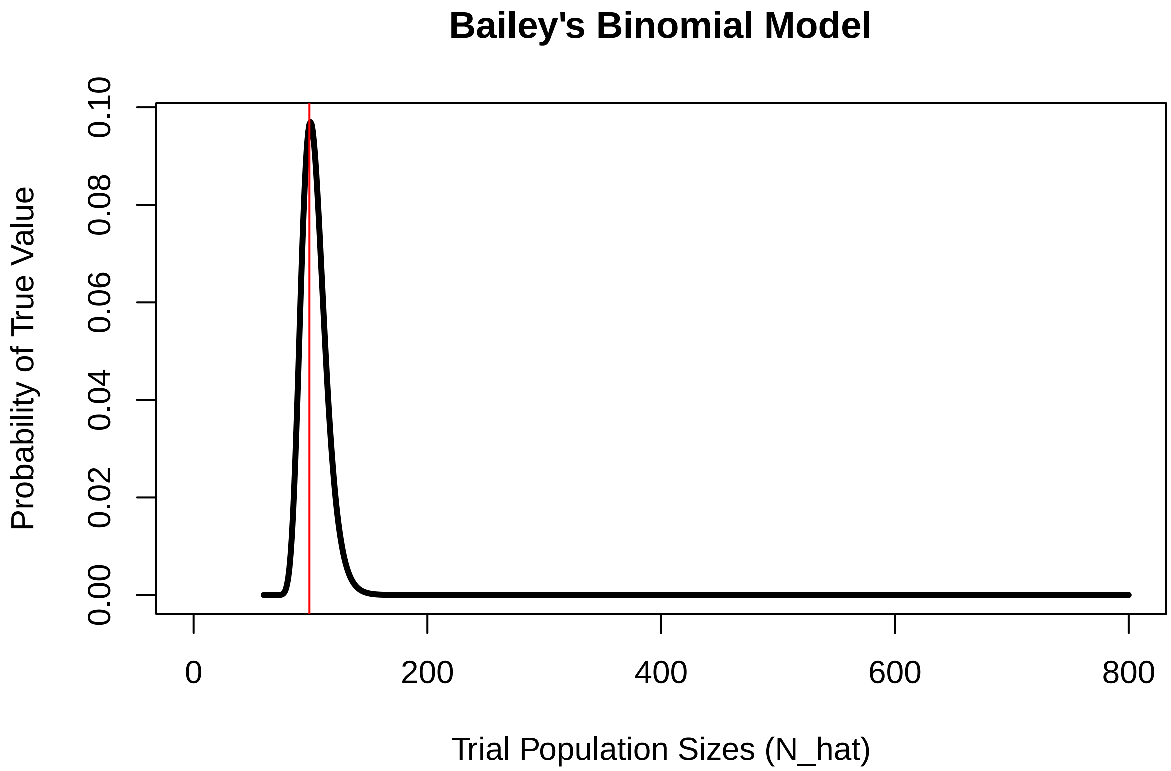


### September

```
n1 = 60
n2 = 70
m2 = 42

bailey.september.est <- (n1*(n2+1)) / (m2+1)
bailey.september.var <- (n1^2 * (n2+1) * (n2-m2)) / ((m2+1)^2 * (m2+2))
bailey.september.CV <- sqrt(bailey.september.var)/bailey.september.est

september.trial.N.prob_bailey <- dbinom(x=m2, size=n2, prob=n1/trial.N)
```

```
## Warning in dbinom(x = m2, size = n2, prob = n1/trial.N): NaNs produced
```

```
plot(x=trial.N, y=september.trial.N.prob_bailey, type="l", lwd = 3,
     xlab="Trial Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="Bailey's Binomial Model")
abline(v=bailey.september.est, col="red")
```



Combined table & plot

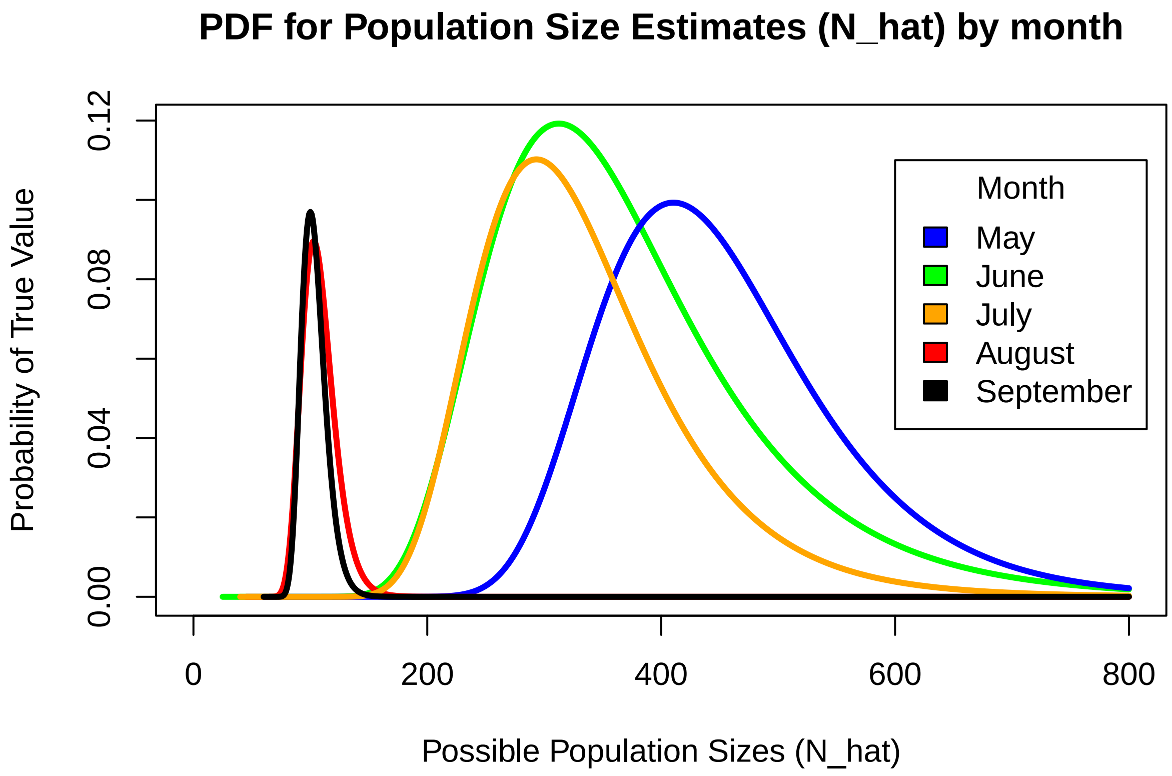
```
bailey_table <- cbind(N_hat <- c(bailey.may.est, bailey.june.est, bailey.july.est, bailey.august.est, ba
rownames(bailey_table) <- c("May", "June", "July", "August", "September")
colnames(bailey_table) <- c("N_hat", "var", "CV")
bailey_table
```

```
##                N_hat        var         CV
## May        393.25000 6146.87202 0.19936933
## June       290.38462 5504.54353 0.25549773
## July       277.50000 3876.83824 0.22437572
## August     101.25000  153.92736 0.12253577
## September   99.06977   87.96893 0.09467243
```

```
plot(x=trial.N, y=june.trial.N.prob_bailey, type="l", col = "green", lwd = 3,
     xlab="Possible Population Sizes (N_hat)",
     ylab="Probability of True Value",
     main="PDF for Population Size Estimates (N_hat) by month") +
  lines(x=trial.N, y=may.trial.N.prob_bailey, col="blue", lwd = 3) +
  lines(x=trial.N, y=august.trial.N.prob_bailey, col="red", lwd = 3) +
  lines(x=trial.N, y=july.trial.N.prob_bailey, col="orange", lwd = 3) +
  lines(x=trial.N, y=september.trial.N.prob_bailey, col="black", lwd = 3)
```

```
## integer(0)
```

```
legend(x = 600, y = 0.11, legend = c("May", "June", "July", "August", "September"),
       fill = c("blue", "green", "orange", "red", "black"), title = "Month")
```

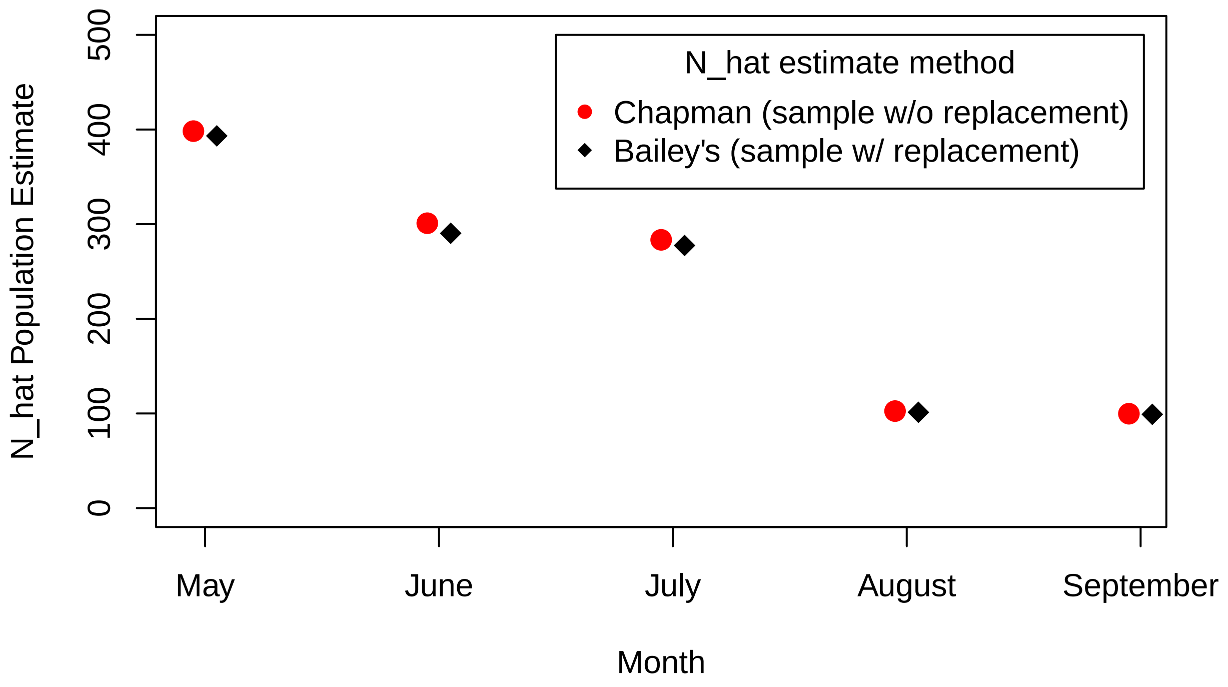


Summary

```
plot(x = 1:5 - 0.05, y = chapman_table[,1], pch = 16, col = "red", cex = 1.5,
     xlab = "Month",
     ylab = "N_hat Population Estimate",
     ylim = c(0,500),
     xaxt = "n") +
  points(x = 1:5 + 0.05, y = bailey_table[,1], pch = 18, col = "black", cex = 1.5) +
  axis(side = 1, at = c(1:5), labels = c("May", "June", "July", "August", "September"))
```

```
## numeric(0)
```

```
legend(x = 2.5, y = 500, legend = c("Chapman (sample w/o replacement)", "Bailey's (sample w/ replacemen
```



From the results of this experiment, it is apparent that the population is declining over the course of the season. These results indicate that management action should be taken to curb fishing pressure to hopefully reduce the current seasonal population decline in the rainbow trout.