CAAP Statistics - Lec07
R Session3

Jul 15, 2022

Review

- Definition of Probability
  - Frequentist vs Bayesian
  - Law of Large Numbers
- Probability Distribution
  - Independent vs Disjoint vs Complement
  - Product Rule: $P(A \cap B) = P(A) \times P(B)$
  - Addition Rule:
    $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Sampling from a small population
  - Sampling with & without replacement

Learning Objectives

- Graphical Summary of Data(continued)
- Simulation: Law of Large Numbers
- Simulation
  - Sampling with replacement
  - Sampling without replacement
- Simulation
  - Product Rule
  - General Addition Rule

Simulation

Load packages

```
library(openintro)
library(tidyverse)
library(ggplot2)
```

Law of Large Number

```
coin = c("H","T")
N = 10000
set.seed(0715) # to reproduce the outcomes
tosses = sample(coin, size = N, replace = TRUE)
```

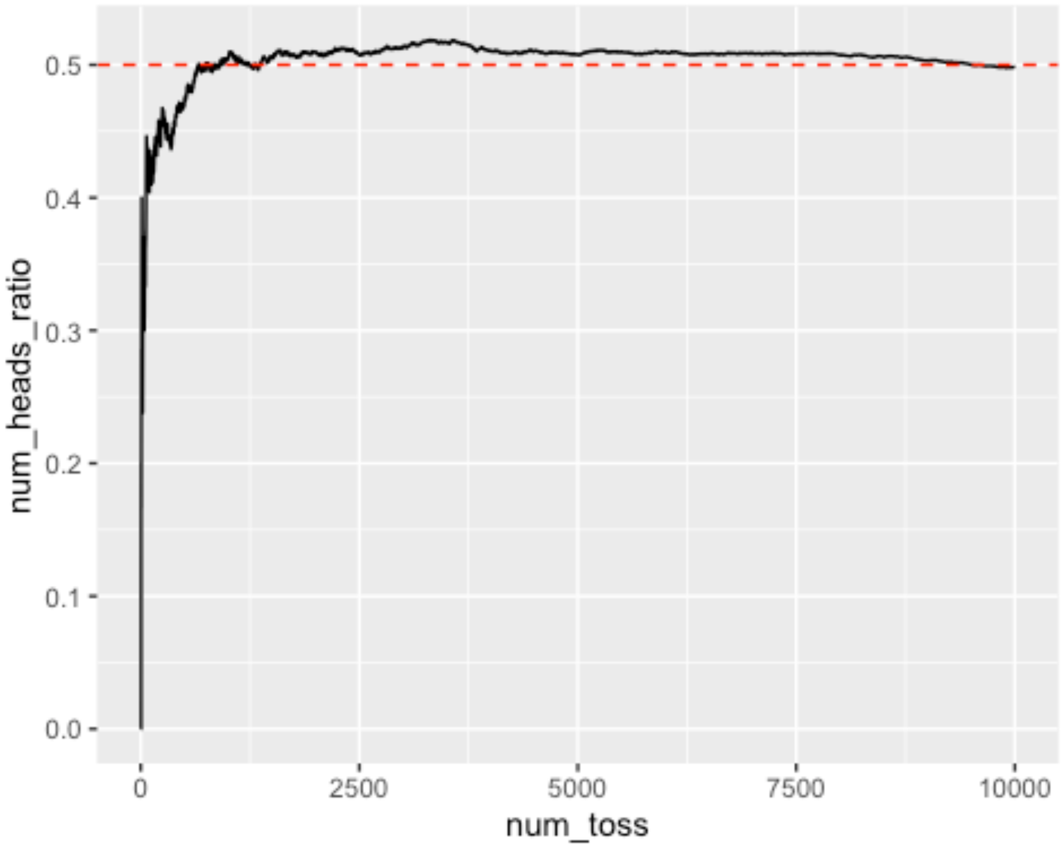
Law of Large Number - Fair Coin

```
head(tosses)
## [1] "T" "T" "T" "T" "H" "T"
table(tosses)
## tosses
```

- `##    H    T`
- `## 4983 5017`

Law of Large Number - Fair Coin(visualization)

```
num_heads = cumsum(tosses=="H")
heads = data.frame(num_toss = 1:N, num_heads, num_heads_ratio = num_heads/1:N)
ggplot(heads, aes(x=num_toss, y = num_heads_ratio))+
  geom_line()+
  geom_hline(yintercept=0.5, linetype="dashed", color = "red")
```



Law of Large Numbers - Unfair Coin

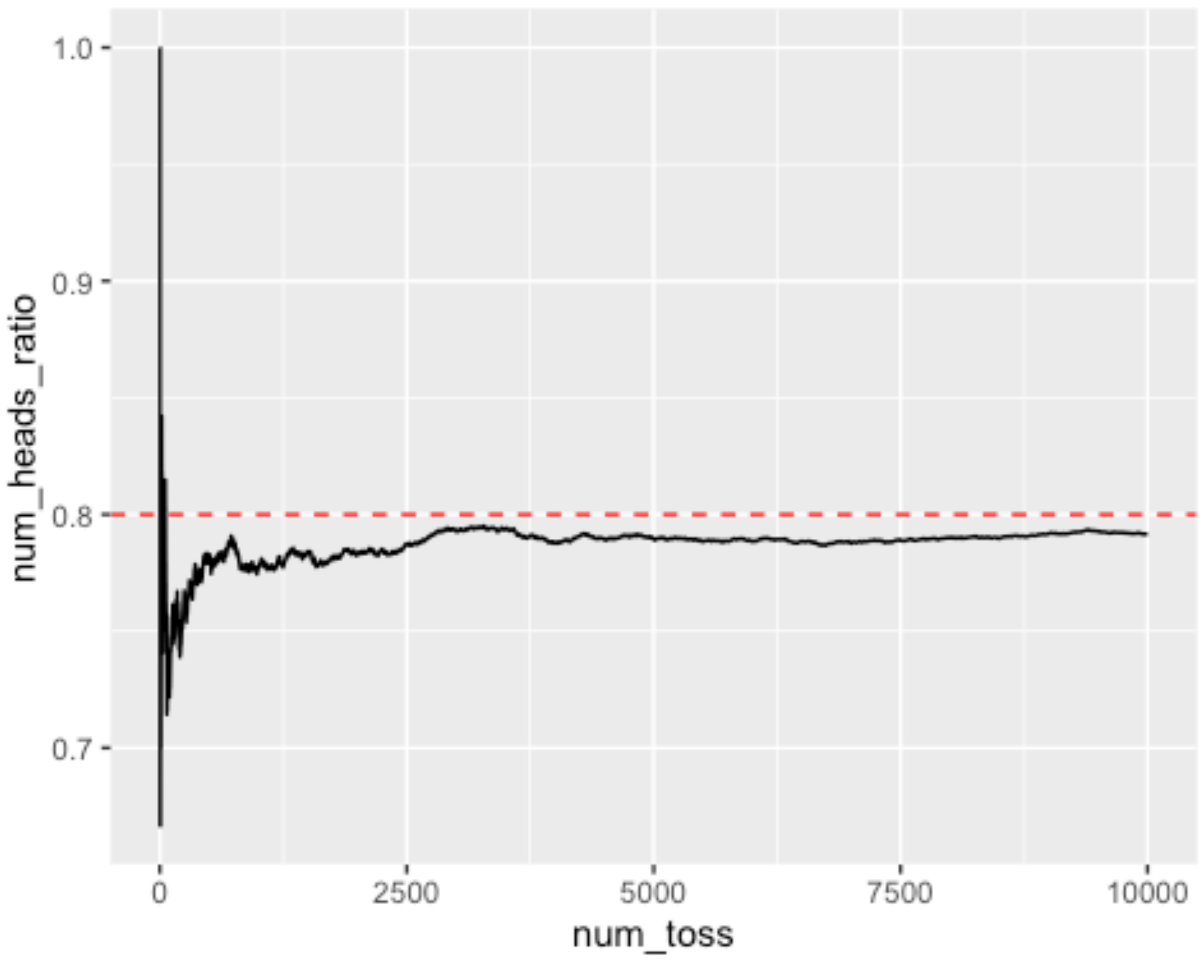
```r
set.seed(0715)
unfair_toss = sample(coin, size = N, prob = c(0.8, 0.2), replace = TRUE)
# Specify the probabilities
```

```r
table(unfair_toss)
```

```
## unfair_toss
##    H    T
## 7918 2082
```

Law of Large Number - Unfair Coin(visualization)

```
num_heads_unfair = cumsum(unfair_toss=="H")
heads_unfair = data.frame(num_toss = 1:N, num_heads_unfair, num_heads_ratio =
num_heads_unfair/1:N)
ggplot(heads_unfair, aes(x=num_toss, y = num_heads_ratio))+
  geom_line()+
  geom_hline(yintercept=0.8, linetype="dashed", color = "red")
```



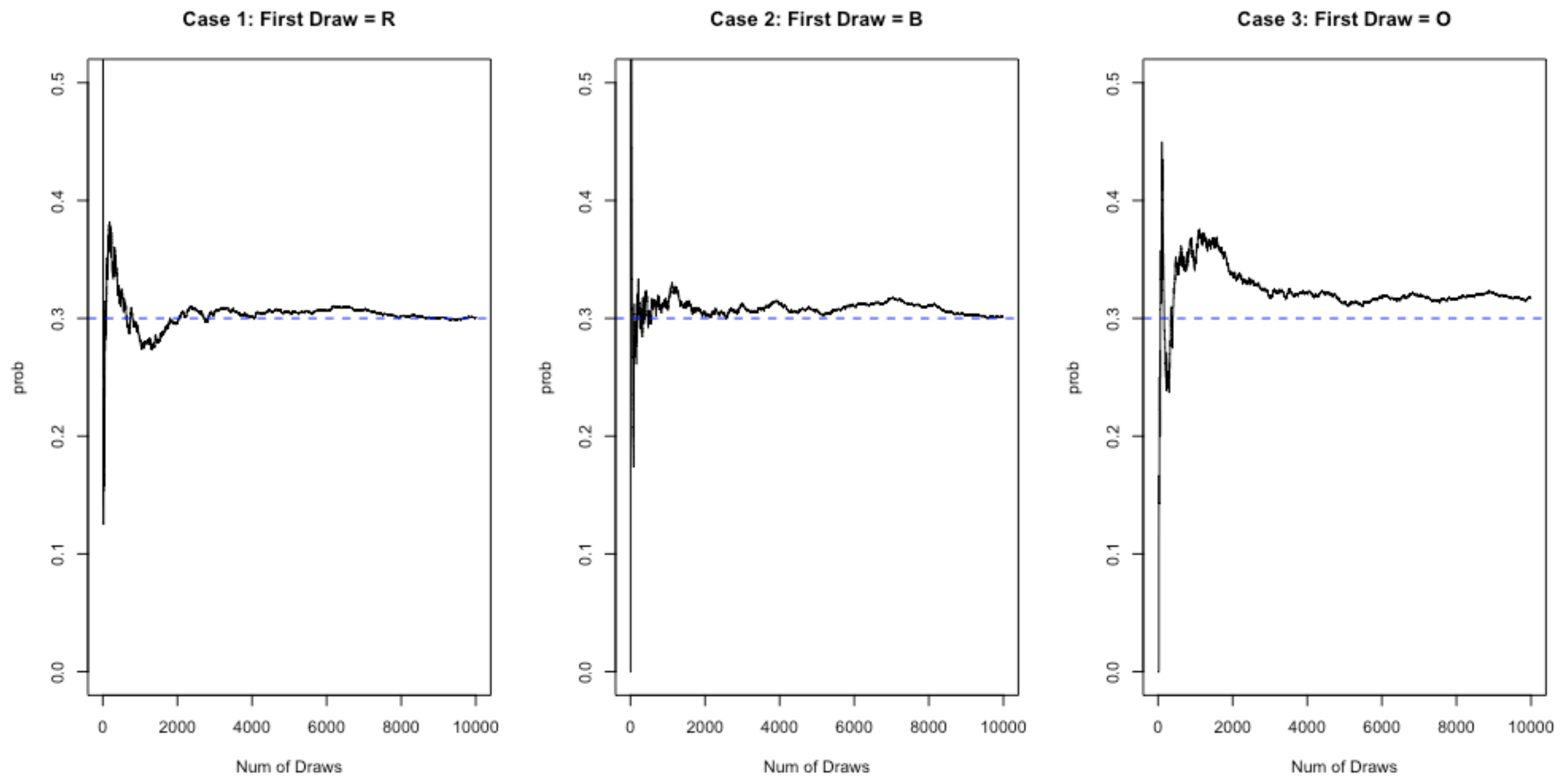
Sampling with & without replacement

```
bag_of_ball = c("R","R","R","R","R","B","B","B","O","O")
set.seed(0715)
nexp = 10000
swr = matrix(0, nrow = nexp, ncol = 2)
swor = matrix(0, nrow = nexp, ncol = 2)
for (i in 1:nexp){
  swr[i,] = sample(bag_of_ball, size = 2, replace = TRUE)
}

for(i in 1:nexp){
 swor[i,] = sample(bag_of_ball, size = 2, replace = FALSE)
}
```

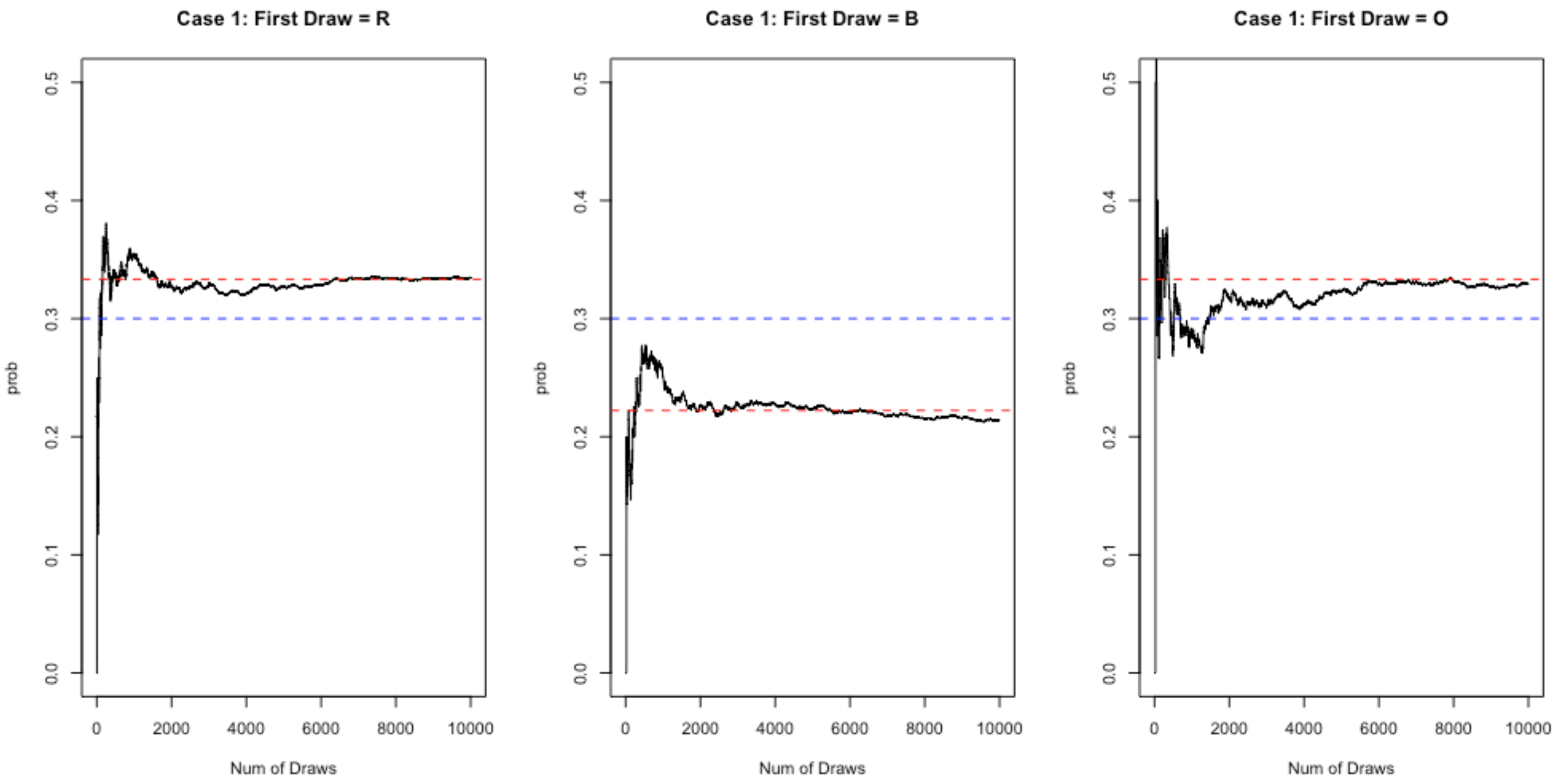
Sampling with Replacement(Independent case)

```
par(mfrow=c(1,3))
plot(cumsum(swr[,1]=="R" & swr[,2]=="B")/cumsum(swr[,1]=="R"), type="l", xlab="Num of Draws", ylab = "prob",
ylim = c(0,0.5), main ="Case 1: First Draw = R")
abline(h=3/10, col="blue", lty = 2)
plot(cumsum(swr[,1]=="B" & swr[,2]=="B")/cumsum(swr[,1]=="B"), type="l", xlab="Num of Draws", ylab = "prob",
ylim = c(0,0.5), main ="Case 2: First Draw = B")
abline(h=3/10, col="blue", lty = 2)
plot(cumsum(swr[,1]=="O" & swr[,2]=="B")/cumsum(swr[,1]=="O"), type="l", xlab="Num of Draws", ylab = "prob",
ylim = c(0,0.5), main ="Case 3: First Draw = O")
abline(h=3/10, col="blue", lty = 2)
```



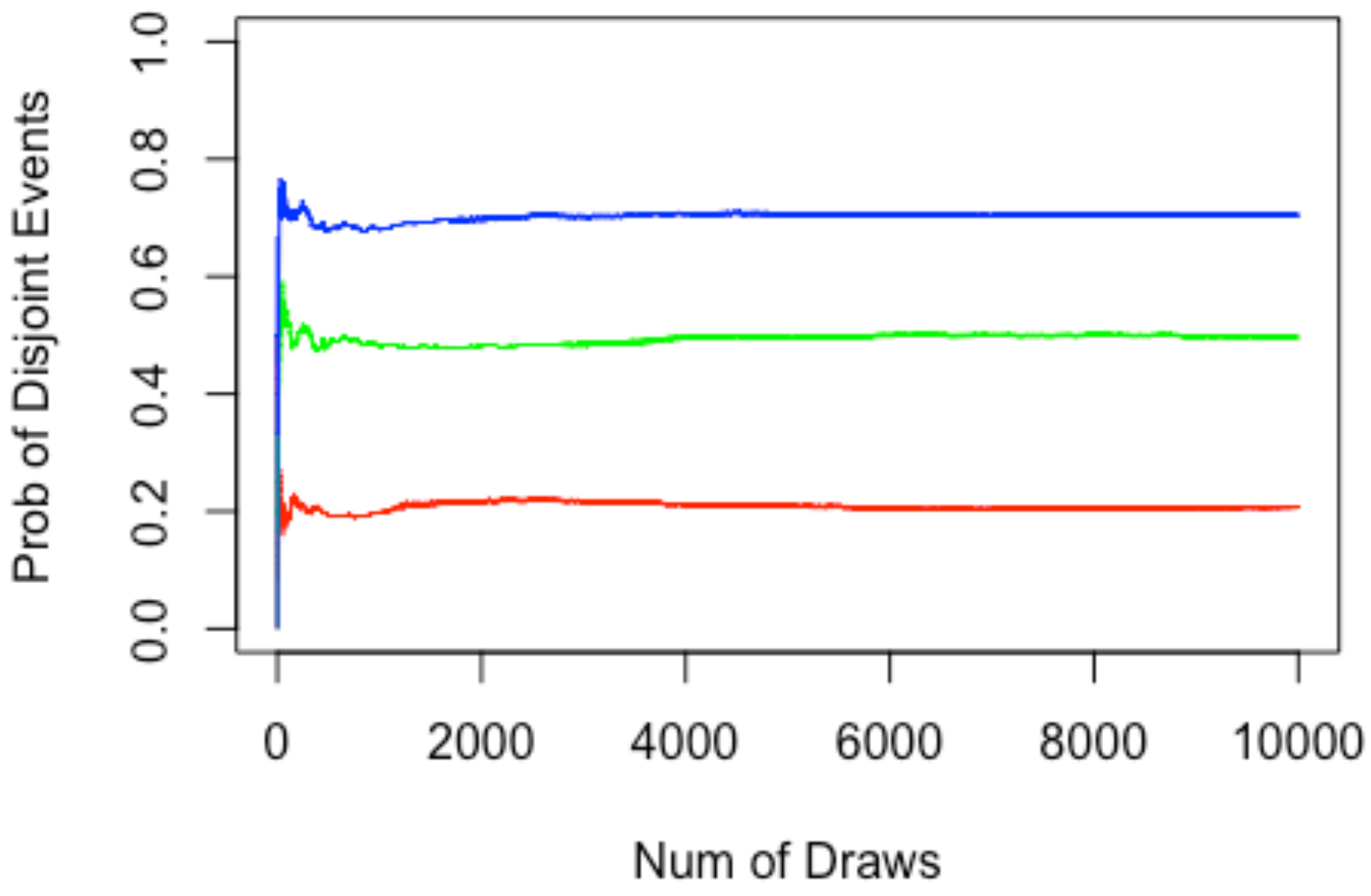
Sampling without Replacement(Independent case)

```
par(mfrow=c(1,3))
plot(cumsum(swor[,1]=="R" & swor[,2]=="B")/cumsum(swor[,1]=="R"), type="l", xlab="Num of Draws",ylab = "prob",
ylim = c(0,0.5), main ="Case 1: First Draw = R")
abline(h=3/10, col="blue", lty=2)
abline(h=3/9, col="red", lty = 2)
plot(cumsum(swor[,1]=="B" & swor[,2]=="B")/cumsum(swor[,1]=="B"), type="l", xlab="Num of Draws",ylab = "prob",
ylim = c(0,0.5), main ="Case 1: First Draw = B")
abline(h=3/10, col="blue", lty=2)
abline(h=2/9, col="red", lty = 2)
plot(cumsum(swor[,1]=="O" & swor[,2]=="B")/cumsum(swor[,1]=="O"), type="l", xlab="Num of Draws",ylab = "prob",
ylim = c(0,0.5), main ="Case 1: First Draw = O")
abline(h=3/10, col="blue", lty=2)
abline(h=3/9, col="red", lty = 2)
```



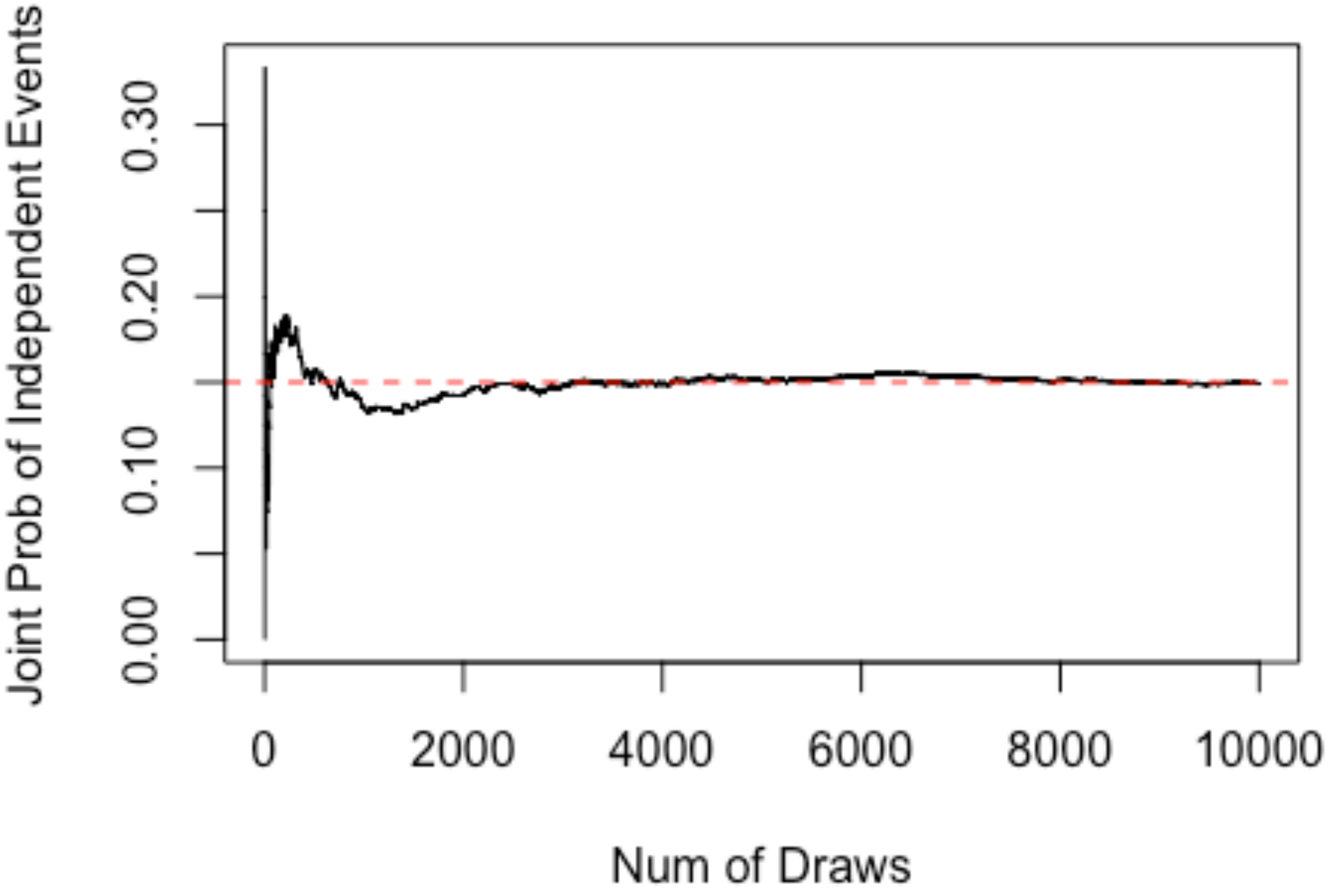
Probability of Union of Disjoint Events

```
plot(1:nexp,cumsum(swr[,1]=="O")/1:nexp, col="red",type="l", ylim = c(0,1), ylab="Prob of Disjoint Events", xlab = "Num of Draws")
lines(1:nexp,cumsum(swr[,1]=="R")/1:nexp, col="green")
lines(1:nexp,cumsum(swr[,1]=="O"|swr[,1]=="R")/1:nexp, col="blue")
```



Joint Probability of Independent Events

```
plot(cumsum(swr[,1]=="R"&swr[,2]=="B")/1:nexp, type="l",ylab="Joint Prob of 
Independent Events", xlab = "Num of Draws")
abline(h=(5/10)*(3/10),col="red",lty=2)
```



First Quiz on Next Thursday

- **Quiz**
  - OpenIntro Chapter 1-4
  - Lecture 1-9
  - Code from R sessions will be on the exam. You need to know the meaning of each code— eg: what does `sample(1:6, size =2)` do?
- Office hour from **7pm** via Zoom.
  - **This week, make sure to participate at least one office hour!!**
  - 5 out of 50 participation pts