Lab4

Exercises

Irimie Fabio

Contents

Exercise 1

A

Plot the Probability Mass Function for the Binomial distribution with $n = 18$ and $p = \frac{1}{3}$. Calculate:

1. $P(X = 3)$

```
dbinom(3, 18, 1 / 3)
## [1] 0.06901723
```

2. $P(X \geq 3)$

```
1 - pbinom(2, 18, 1 / 3)
## [1] 0.9673521
# or
pbinom(2, 18, 1 / 3, lower.tail = FALSE)
## [1] 0.9673521
```

3. $P(1 \leq X > 5)$

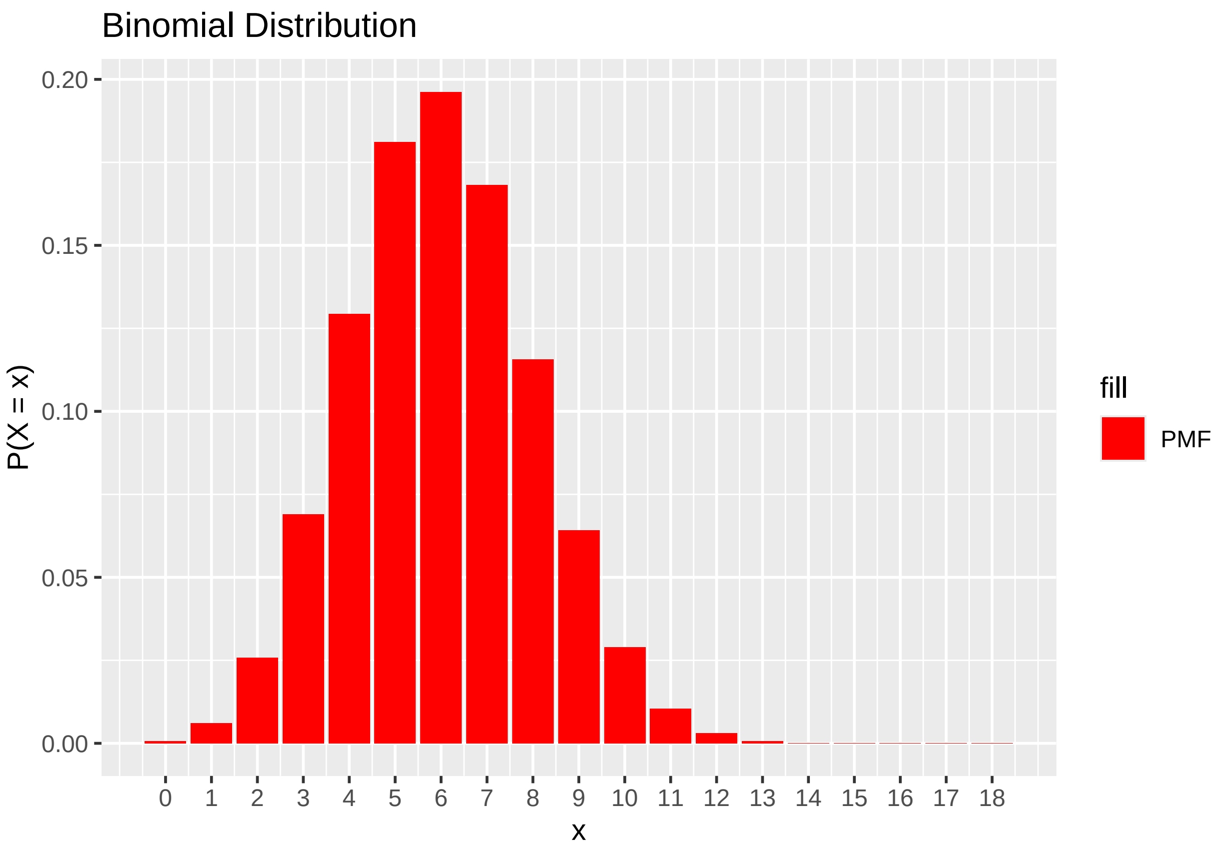
```
pbinom(4, 18, 1 / 3) - pbinom(0, 18, 1 / 3)
## [1] 0.2303957
```

4. $P(X \geq 15)$

```
pbinom(14, 18, 1 / 3, lower.tail = FALSE)
## [1] 1.852509e-05
```

```
library(ggplot2)
df <- data.frame(x = 0:18, y = dbinom(0:18, 18, 1 / 3))
```

```r
ggplot(df, aes(x = x, y = y, fill = "PMF")) +
  geom_col() +
  scale_x_continuous(breaks = 0:18) +
  scale_fill_manual(values = "red") +
  labs(
    title = "Binomial Distribution",
    x = "x",
    y = "P(X = x)"
  )
```



B

Plot the Cumulative Distribution Function for the Poisson distribution with $\lambda = 3$. Calculate:

1. $P(X = 3)$

```r
lambda <- 3
dpois(3, lambda)
## [1] 0.2240418
```

2. $P(X \geq 3)$

```r
ppois(2, lambda, lower.tail = FALSE)
## [1] 0.5768099
```

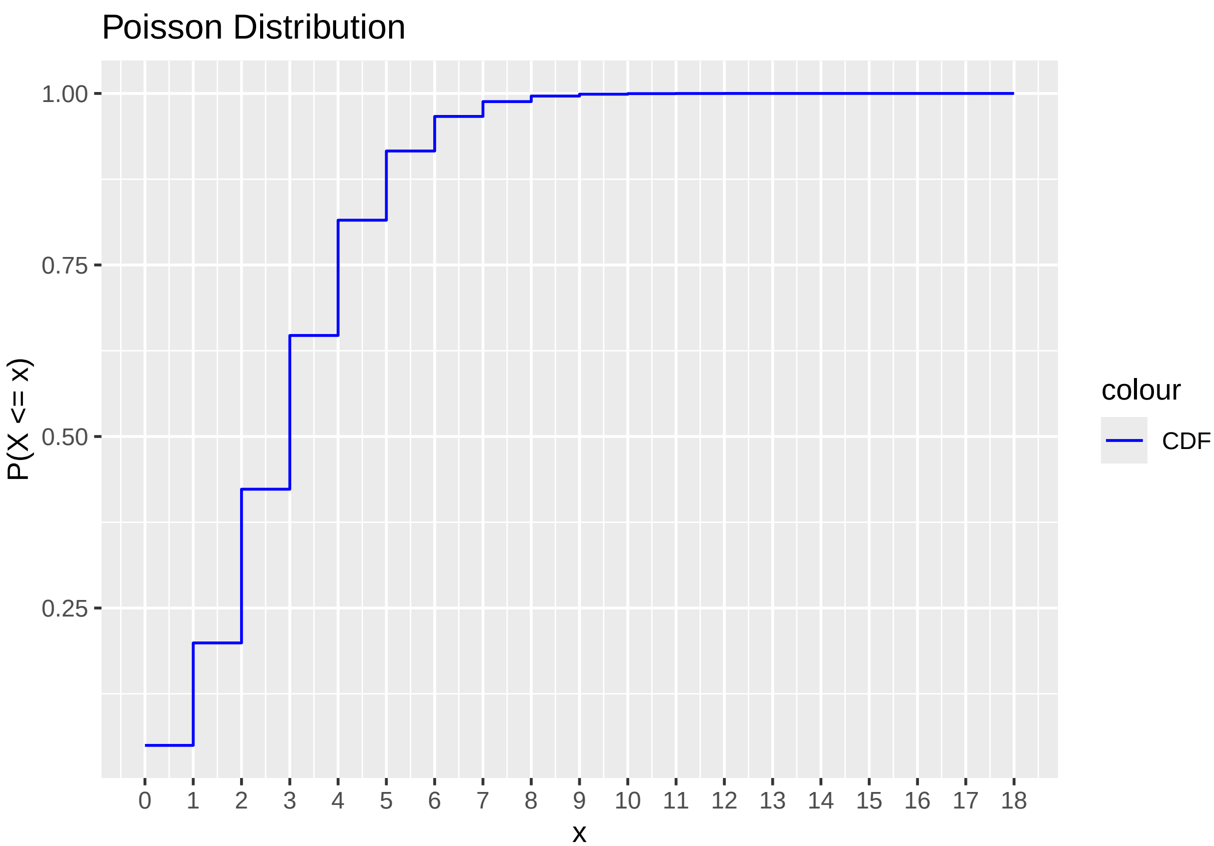
3. $P(1 \leq X > 5)$

```r
ppois(4, lambda) - ppois(0, lambda)
## [1] 0.7654762
```

4. $P(X \geq 15)$

```r
ppois(14, lambda, lower.tail = FALSE)
## [1] 6.703859e-07
```

```r
df <- data.frame(x = 0:18, y = ppois(0:18, lambda))

ggplot(df, aes(x = x, y = y, color = "CDF")) +
  geom_step() +
  scale_x_continuous(breaks = 0:18) +
  scale_color_manual(values = "blue") +
  labs(
    title = "Poisson Distribution",
    x = "x",
    y = "P(X <= x)"
  )
```



Exercise 2

Demonstrate that a Poisson r.v. may be used as an approximation for a binomial r.v.

A

```r
n <- c(20, 30, 40, 100)
p <- c(1 / 4, 1 / 6, 1 / 8, 1 / 20)

pmf <- matrix(NA, nrow = 21, ncol = 4)
```

```r
for (i in 1:4) {
  pmf[, i] <- dbinom(0:20, n[i], p[i])
}

pmf <- as.data.frame(pmf)

colnames(pmf) <- paste("Binomial", n, round(p, 2), sep = "_")

pmf$Poisson <- dpois(0:20, n * p)

pmf$X <- 0:20
```

B

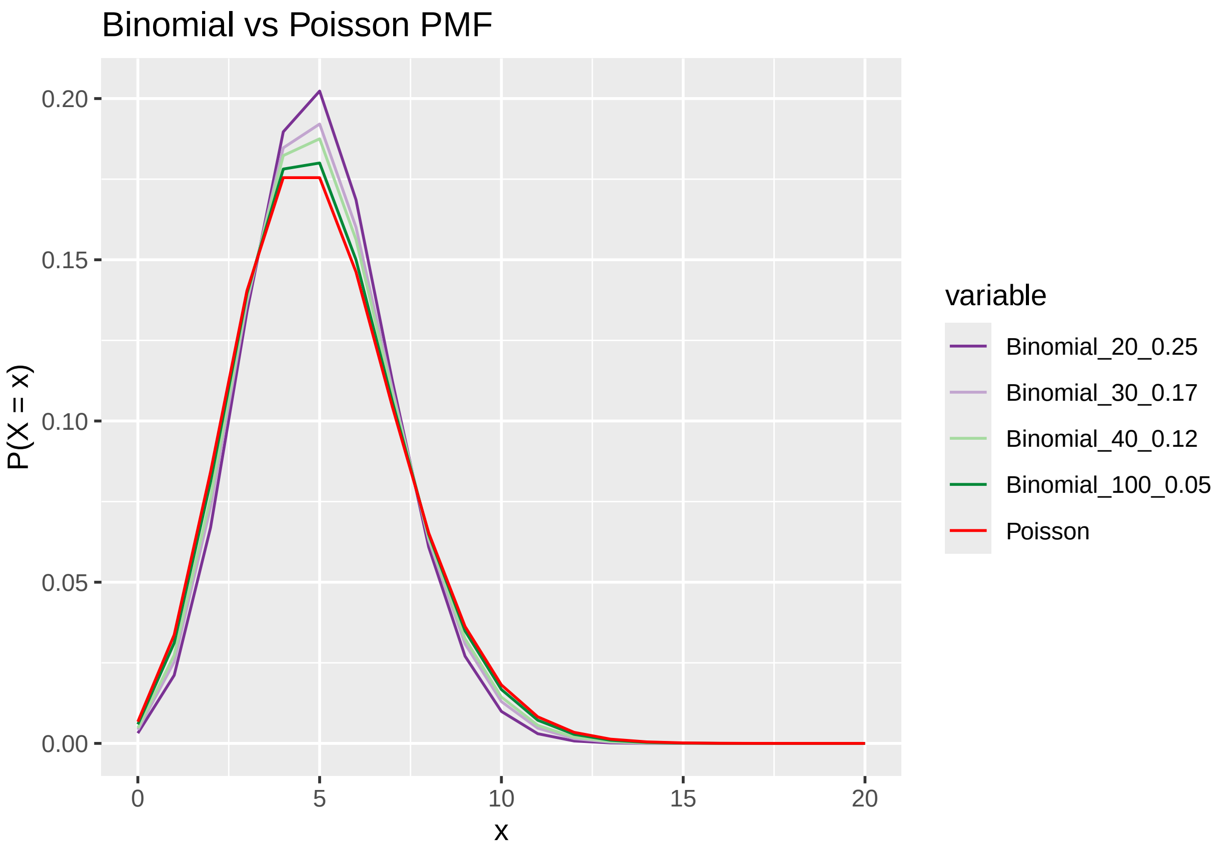
```r
library(reshape2)

df_plot <- melt(pmf, id.vars = "X")
df_plot
```

```
##      X          variable        value
## 1    0  Binomial_20_0.25 3.171212e-03
## 2    1  Binomial_20_0.25 2.114141e-02
## 3    2  Binomial_20_0.25 6.694781e-02
## 4    3  Binomial_20_0.25 1.338956e-01
## 5    4  Binomial_20_0.25 1.896855e-01
## 6    5  Binomial_20_0.25 2.023312e-01
## 7    6  Binomial_20_0.25 1.686093e-01
## 8    7  Binomial_20_0.25 1.124062e-01
## 9    8  Binomial_20_0.25 6.088669e-02
## 10   9  Binomial_20_0.25 2.706075e-02
## 11  10  Binomial_20_0.25 9.922275e-03
## 12  11  Binomial_20_0.25 3.006750e-03
## 13  12  Binomial_20_0.25 7.516875e-04
## 14  13  Binomial_20_0.25 1.541923e-04
## 15  14  Binomial_20_0.25 2.569872e-05
## 16  15  Binomial_20_0.25 3.426496e-06
## 17  16  Binomial_20_0.25 3.569266e-07
## 18  17  Binomial_20_0.25 2.799425e-08
## 19  18  Binomial_20_0.25 1.555236e-09
## 20  19  Binomial_20_0.25 5.456968e-11
## 21  20  Binomial_20_0.25 9.094947e-13
## 22   0  Binomial_30_0.17 4.212720e-03
## 23   1  Binomial_30_0.17 2.527632e-02
## 24   2  Binomial_30_0.17 7.330133e-02
## 25   3  Binomial_30_0.17 1.368292e-01
## 26   4  Binomial_30_0.17 1.847194e-01
## 27   5  Binomial_30_0.17 1.921081e-01
## 28   6  Binomial_30_0.17 1.600901e-01
## 29   7  Binomial_30_0.17 1.097761e-01
## 30   8  Binomial_30_0.17 6.312124e-02
## 31   9  Binomial_30_0.17 3.085927e-02
## 32  10  Binomial_30_0.17 1.296090e-02
## 33  11  Binomial_30_0.17 4.713053e-03
```

```
## 34  12   Binomial_30_0.17 1.492467e-03
## 35  13   Binomial_30_0.17 4.132985e-04
## 36  14   Binomial_30_0.17 1.003725e-04
## 37  15   Binomial_30_0.17 2.141280e-05
## 38  16   Binomial_30_0.17 4.014899e-06
## 39  17   Binomial_30_0.17 6.612776e-07
## 40  18   Binomial_30_0.17 9.551787e-08
## 41  19   Binomial_30_0.17 1.206542e-08
## 42  20   Binomial_30_0.17 1.327196e-09
## 43   0   Binomial_40_0.12 4.789852e-03
## 44   1   Binomial_40_0.12 2.737058e-02
## 45   2   Binomial_40_0.12 7.624663e-02
## 46   3   Binomial_40_0.12 1.379701e-01
## 47   4   Binomial_40_0.12 1.823176e-01
## 48   5   Binomial_40_0.12 1.875267e-01
## 49   6   Binomial_40_0.12 1.562722e-01
## 50   7   Binomial_40_0.12 1.084338e-01
## 51   8   Binomial_40_0.12 6.389849e-02
## 52   9   Binomial_40_0.12 3.245638e-02
## 53  10   Binomial_40_0.12 1.437354e-02
## 54  11   Binomial_40_0.12 5.600080e-03
## 55  12   Binomial_40_0.12 1.933361e-03
## 56  13   Binomial_40_0.12 5.948803e-04
## 57  14   Binomial_40_0.12 1.638956e-04
## 58  15   Binomial_40_0.12 4.058367e-05
## 59  16   Binomial_40_0.12 9.058855e-06
## 60  17   Binomial_40_0.12 1.826996e-06
## 61  18   Binomial_40_0.12 3.334992e-07
## 62  19   Binomial_40_0.12 5.516529e-08
## 63  20   Binomial_40_0.12 8.274793e-09
## 64   0 Binomial_100_0.05 5.920529e-03
## 65   1 Binomial_100_0.05 3.116068e-02
## 66   2 Binomial_100_0.05 8.118177e-02
## 67   3 Binomial_100_0.05 1.395757e-01
## 68   4 Binomial_100_0.05 1.781426e-01
## 69   5 Binomial_100_0.05 1.800178e-01
## 70   6 Binomial_100_0.05 1.500149e-01
## 71   7 Binomial_100_0.05 1.060255e-01
## 72   8 Binomial_100_0.05 6.487089e-02
## 73   9 Binomial_100_0.05 3.490130e-02
## 74  10 Binomial_100_0.05 1.671588e-02
## 75  11 Binomial_100_0.05 7.198228e-03
## 76  12 Binomial_100_0.05 2.809834e-03
## 77  13 Binomial_100_0.05 1.001075e-03
## 78  14 Binomial_100_0.05 3.274191e-04
## 79  15 Binomial_100_0.05 9.880016e-05
## 80  16 Binomial_100_0.05 2.762505e-05
## 81  17 Binomial_100_0.05 7.184222e-06
## 82  18 Binomial_100_0.05 1.743539e-06
## 83  19 Binomial_100_0.05 3.960394e-07
## 84  20 Binomial_100_0.05 8.441893e-08
## 85   0           Poisson 6.737947e-03
## 86   1           Poisson 3.368973e-02
## 87   2           Poisson 8.422434e-02
```

```
## 88   3           Poisson 1.403739e-01
## 89   4           Poisson 1.754674e-01
## 90   5           Poisson 1.754674e-01
## 91   6           Poisson 1.462228e-01
## 92   7           Poisson 1.044449e-01
## 93   8           Poisson 6.527804e-02
## 94   9           Poisson 3.626558e-02
## 95  10           Poisson 1.813279e-02
## 96  11           Poisson 8.242177e-03
## 97  12           Poisson 3.434240e-03
## 98  13           Poisson 1.320862e-03
## 99  14           Poisson 4.717363e-04
## 100 15           Poisson 1.572454e-04
## 101 16           Poisson 4.913920e-05
## 102 17           Poisson 1.445271e-05
## 103 18           Poisson 4.014640e-06
## 104 19           Poisson 1.056484e-06
## 105 20           Poisson 2.641211e-07
```

```r
library(ggplot2)
library(RColorBrewer) # Color palettes

ggplot(df_plot, aes(x = X, y = value, color = variable)) +
  geom_line() +
  scale_color_manual(values = c(brewer.pal(4, "PRGn"), "red")) +
  labs(
    title = "Binomial vs Poisson PMF",
    x = "x",
    y = "P(X = x)"
  )
```



Exercise 3

A

Generate N=1000 random numbers from a binomial distribution with n=9 trials and p=0.8. Thus each of the 1000 random numbers will be an integer between 0 and 9.

```
set.seed(123)
n <- 9
p <- 0.8
t <- 1000

rbinom(t, n, p)
```

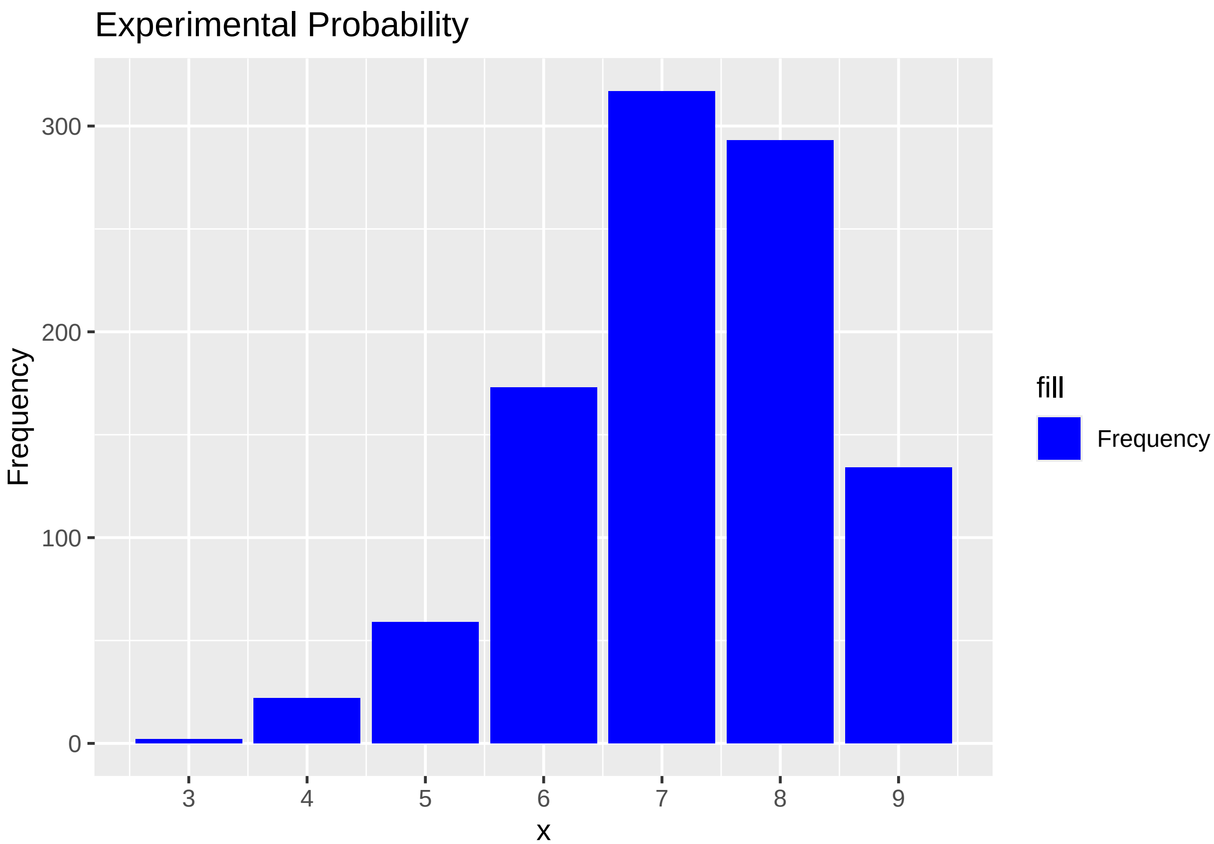
```
##    [1] 8 6 8 6 5 9 7 6 7 7 5 7 7 7 9 6 8 9 8 5 6 7 7 4 7 7 7 7 8 8 5 6 7
##   [34] 6 9 7 6 8 8 8 8 8 8 8 8 8 8 7 8 6 9 7 6 9 7 8 9 6 6 8 7 9 8 8 6 7
##   [67] 6 6 6 7 6 7 7 9 7 8 8 7 8 9 8 7 8 6 9 8 4 6 6 8 9 7 8 7 8 8 6 9 7
##  [100] 7 7 8 7 5 7 6 5 7 8 8 5 8 9 5 7 8 7 5 7 8 7 8 8 8 8 4 8 9 8 7 7 6
##  [133] 7 7 7 7 6 6 5 7 8 8 9 8 6 8 8 9 8 7 6 7 8 8 9 8 7 8 7 8 7 8 7 8 8
##  [166] 7 6 8 8 8 7 8 6 6 7 7 8 7 6 7 6 8 7 8 7 7 8 7 6 6 8 8 4 7 5 7 8 7
##  [199] 8 7 8 5 7 7 8 6 8 8 8 8 7 8 8 7 9 7 8 8 6 5 8 5 7 7 9 8 7 7 7 5 7
##  [232] 8 7 9 8 8 8 6 8 6 7 7 8 7 8 7 8 5 5 7 8 8 7 8 7 6 8 8 7 6 5 6 7 5
##  [265] 7 7 8 8 9 7 6 9 9 8 6 7 5 7 9 7 6 8 8 8 9 8 9 8 9 7 8 9 9 6 6 6 4
##  [298] 9 9 6 6 9 6 7 7 7 8 9 7 7 8 7 7 9 8 6 6 8 8 6 6 8 8 7 9 9 3 9 8 5
##  [331] 7 8 7 6 8 7 7 7 7 5 8 9 7 8 7 8 4 8 8 7 8 5 7 8 7 4 7 8 8 6 8 8 6
##  [364] 8 8 6 8 9 8 7 8 8 6 7 7 5 6 8 8 5 7 6 8 6 7 6 8 8 9 8 5 8 7 6 9 8
##  [397] 8 9 9 4 4 8 6 7 8 7 7 9 8 7 8 6 8 7 8 8 5 8 7 9 7 7 7 8 6 7 8 8 8
```

```
## [430] 6 5 7 8 4 8 7 9 7 9 8 8 7 7 8 6 7 6 8 8 7 7 8 8 9 8 6 6 6 8 9 5 8
## [463] 7 8 7 8 8 7 8 6 9 6 8 6 7 7 7 9 8 7 8 6 8 7 5 9 9 7 8 6 5 9 7 7 9
## [496] 4 8 9 7 6 8 8 8 9 8 8 7 7 5 8 7 9 6 8 7 8 7 6 9 6 8 8 8 9 7 7 6 9
## [529] 4 8 5 7 7 6 9 6 8 6 7 8 6 9 6 8 6 5 9 6 6 8 8 8 7 6 7 7 7 7 7 9 9
## [562] 5 6 8 7 7 8 6 7 6 7 6 7 8 5 7 8 9 7 7 6 6 6 8 6 8 8 7 5 7 7 9 5 9
## [595] 9 6 7 8 6 9 8 7 8 8 8 6 8 6 8 8 7 9 9 4 7 6 6 6 5 9 6 6 6 8 9 8 8
## [628] 6 8 8 6 6 7 7 9 8 5 6 6 8 7 6 6 8 9 7 6 9 8 7 6 5 7 7 7 6 6 9 7 8
## [661] 6 6 9 7 8 9 5 8 7 8 9 7 9 8 8 7 6 8 7 7 7 7 6 6 5 7 7 8 7 8 9 7 7
## [694] 8 8 7 8 9 8 8 6 8 9 6 8 7 7 8 8 8 7 7 7 7 7 8 9 9 4 9 8 7 6 6 7 6
## [727] 6 7 6 9 8 6 7 5 8 8 6 4 8 9 4 8 9 7 8 7 6 9 7 5 8 8 8 9 7 7 7 6 9
## [760] 8 7 8 7 7 7 8 8 8 7 7 5 6 7 6 9 8 8 8 7 6 8 6 9 7 6 7 8 9 8 5 5 7
## [793] 8 7 6 8 9 8 8 8 7 8 9 9 8 5 7 7 6 7 9 7 8 6 6 8 8 9 7 7 7 8 7 4 6
## [826] 9 7 9 8 7 9 6 7 7 7 7 7 7 8 9 5 7 7 9 7 8 6 9 8 9 8 7 4 7 7 8 5 9
## [859] 9 6 8 9 7 7 7 8 8 9 7 7 8 8 6 7 6 7 9 7 6 7 9 5 7 8 6 5 8 8 8 8 7
## [892] 6 7 7 6 6 9 8 9 6 5 7 6 7 7 7 6 6 6 9 9 9 6 7 8 5 7 7 8 8 6 9 8 6
## [925] 7 7 8 7 8 6 6 8 6 7 5 8 7 6 6 9 5 7 8 8 9 8 8 8 8 8 7 8 5 9 8 5 7
## [958] 9 8 8 7 7 7 8 6 7 7 9 6 7 6 9 6 8 8 7 5 9 8 8 7 7 5 7 6 6 8 8 8 8
## [991] 9 8 7 8 7 6 7 8 7 9
```

B

Plot the experimental probability using the geom_bar() function.

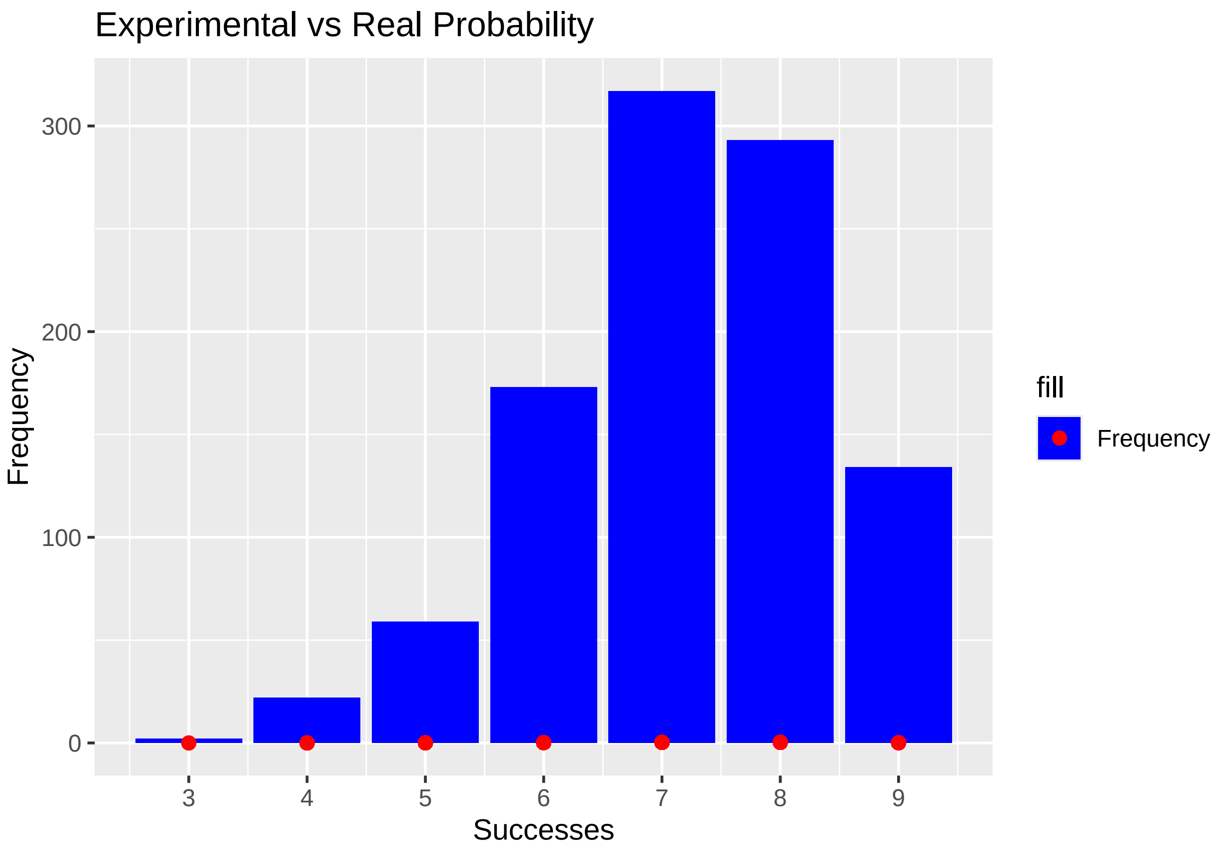
```r
df <- data.frame(x = rbinom(t, n, p))

ggplot(df, aes(x = x, fill = "Frequency")) +
  geom_bar() +
  scale_x_continuous(breaks = 0:9) +
  scale_fill_manual(values = "blue") +
  labs(
    title = "Experimental Probability",
    x = "x",
    y = "Frequency"
  )
```



C

For each value of the x-axis obtained in the previous plot, compute the real probability mass function and add it in the plot as red dots using geom_point().

```
ggplot(df, aes(x = x, fill = "Frequency")) +
  geom_bar() +
  geom_point(
    aes(x = x, y = dbinom(x, n, p)),
    color = "red",
    size = 2
  ) +
  scale_x_continuous(breaks = 0:9) +
  scale_fill_manual(values = "blue") +
  labs(
    title = "Experimental vs Real Probability",
    x = "Successes",
    y = "Frequency"
  )
```


Experimental vs Real Probability
Frequency
300
200
100
0
3
4
5
6
7
8
9
Successes
fill
Frequency