STAT430 Homework #1: Due Friday, February 1, 2019.

Name: Oliver Shanklin

Date: 1 February 2019

---

0. This homework assignment is written in `R Markdown` so that you can open this document in `RStudio`, knit the document to run the `R` code provided, answer the questions, add your own `R` code as necessary, and turn in one final, knitted, seamless document with all code, numerical output, graphical output, and documentation in one place. Read (or re-read) Section 6.7 of Wackerly, Mendenhall and Schaeffer. Also read Sections 7.1-7.3.

Example 6.16 via simulation This example from the book shows that the minimum of two exponentially distributed random variables is itself exponential. Let's explore this result via simulation. We can simulate exponential random variables in `R` using the function `rexp`. Take a look at the `help` for `rexp`:

```r
help(rexp)
```

Notice that this exponential is parameterized with `rate`, where the rate is 1/(expected value). Let's simulate two *realizations* (simulated random values) of an exponential with `rate = 1 / 100`:

```r
Y <- rexp(2, rate = 1 / 100)
Y
```

```
## [1] 85.469030  1.662596
```

The minimum order statistic would be

```r
min(Y)
```

```
## [1] 1.662596
```

To conduct a simulation experiment, we want to simulate two realizations and take their minimum—then repeat many, many times. We will use the `apply` function to do so. To illustrate, suppose we want 5 realizations of the minimum. We'll need 10 exponentials, but let's put them into a matrix with 2 rows and 5 columns:

```r
num_realizations <- 5
Y <- rexp(2 * num_realizations, rate = 1 / 100)
Y
```

```
##  [1]  90.651587 214.800101  89.213401 223.021999  84.912758 639.517377
##  [7]  13.603300 284.151603   1.529638 227.686608
```

```r
YY <- matrix(Y, 2, num_realizations)
YY
```

```
##            [,1]     [,2]      [,3]     [,4]        [,5]
## [1,]   90.65159  89.2134  84.91276  13.6033    1.529638
## [2,]  214.80010 223.0220 639.51738 284.1516  227.686608
```

Now we want the minimum of each column, conveniently obtained using the `apply` function:

```r
min_Y <- apply(YY, MAR = 2, FUN = "min") # 1 for rows, 2 for columns
min_Y
```

```
## [1] 90.651587 89.213401 84.912758 13.603300  1.529638
```
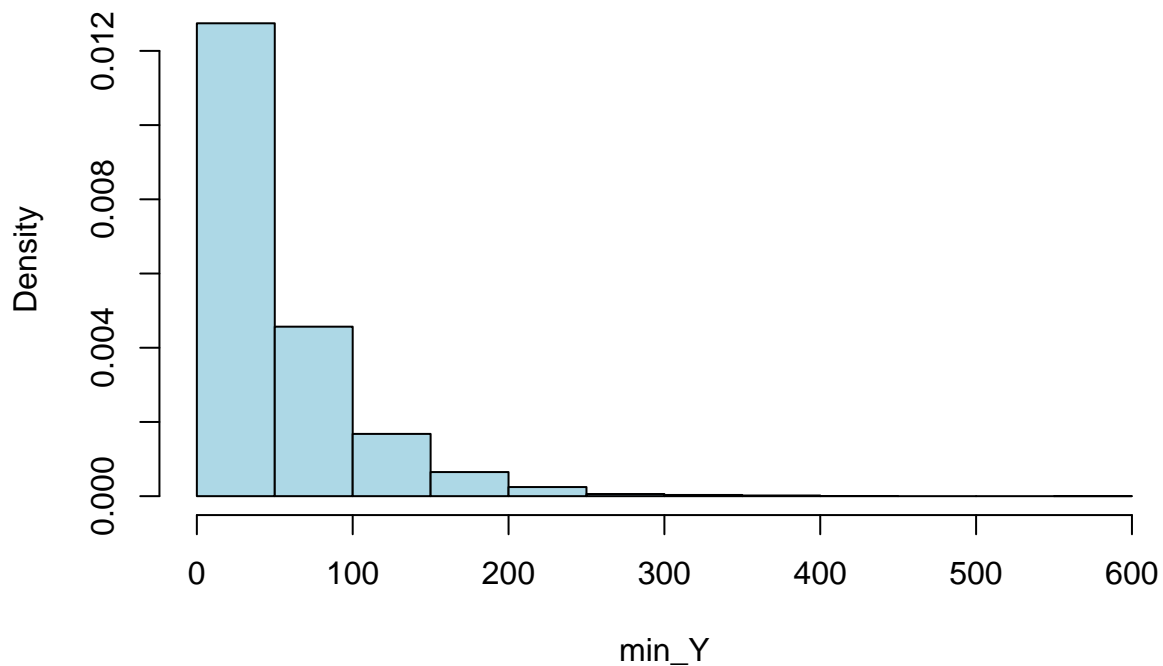
Let's crank up the number of realizations to 10000. Further, let's set our random number seed to 4302019 to make our results reproducible:

```r
set.seed(4302019)
num_realizations <- 10000
Y <- rexp(2 * num_realizations, rate = 1 / 100)
YY <- matrix(Y, 2, num_realizations)
min_Y <- apply(YY, MAR = 2, FUN = "min") # 1 for rows, 2 for columns
hist(min_Y, col = "LightBlue", freq = FALSE)
```

Histogram of min_Y


```r
# freq = FALSE so that histogram is normalized to be a density, with area 1
```

Further, let's add the theoretical probability density function to our histogram. First, we'll need a fine grid of x-values on which to evaluate the function. Let's use 1000 grid points (this is quite arbitrary as long as the number is large enough: too few and the plot will look piecewise linear; too many is not a problem, though we won't get visual improvement in the plot.)

```r
X_grid <- seq(from = 0, to = max(min_Y), length = 1000)
f_X <- (1 / 50) * exp(-X_grid / 50)
hist(min_Y, col = "LightBlue", freq = FALSE)
lines(X_grid, f_X, lwd = 3, col = "magenta")
```

Histogram of min_Y
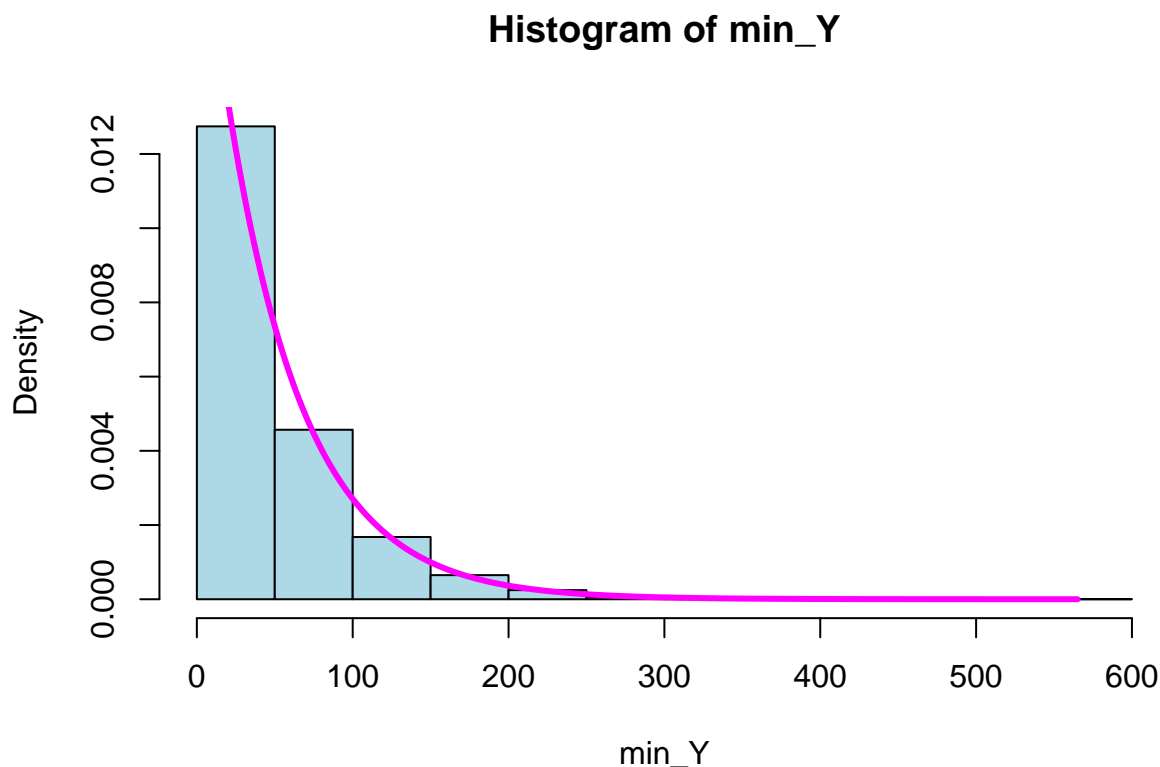
Finally, we can evaluate the theoretical mean and variance of the minimum order statistic, and compare to the empirical values from our simulation. Since the theoretical distribution is exponential with theoretical mean equal to $E(Y_{(1)}) = 50$, its variance is $V(Y_{(1)}) = 2500$. For completeness, and an example of writing up the answer in LaTeX, let's do those calculations analytically: The theoretical mean of $Y_{(1)} = \min(Y_1, Y_2)$ is

$$E(Y_{(1)}) = \int_0^\infty y \frac{1}{50} e^{-y/50} \, dy = 50,$$

which is close to the simulation value of 49.55. The theoretical second moment is

$$E(Y_{(1)}^2) = \int_0^\infty y^2 \frac{1}{50} e^{-y/50} \, dy,$$

which is a constant times the integral of a Gamma density:

$$\Gamma(3)(50)^2 \int_0^\infty \frac{y^{3-1} e^{-y/50}}{\Gamma(3)(50)^3} \, dy = (2)(50)^2 = 5000.$$

Hence the theoretical variance is $5000 - (50)^2 = 2500$, which is close to the simulation value of 2503.24.

1. **Example 6.17**. Modify the code above to reproduce the analysis of Example 6.17, for the maximum order statistic of two exponentials, instead of the minimum. Reset your random number seed to 4302019, then (a) simulate 10000 realizations of the maximum, (b) plot their histogram, (c) add the theoretical probability density function (page 335 of the textbook) to your histogram, (d) compute the empirical mean and variance of the maximum order statistic from your simulation. Finally, (e) compute the theoretical mean and variance of the maximum order statistic, and compare to your simulation values from (d).
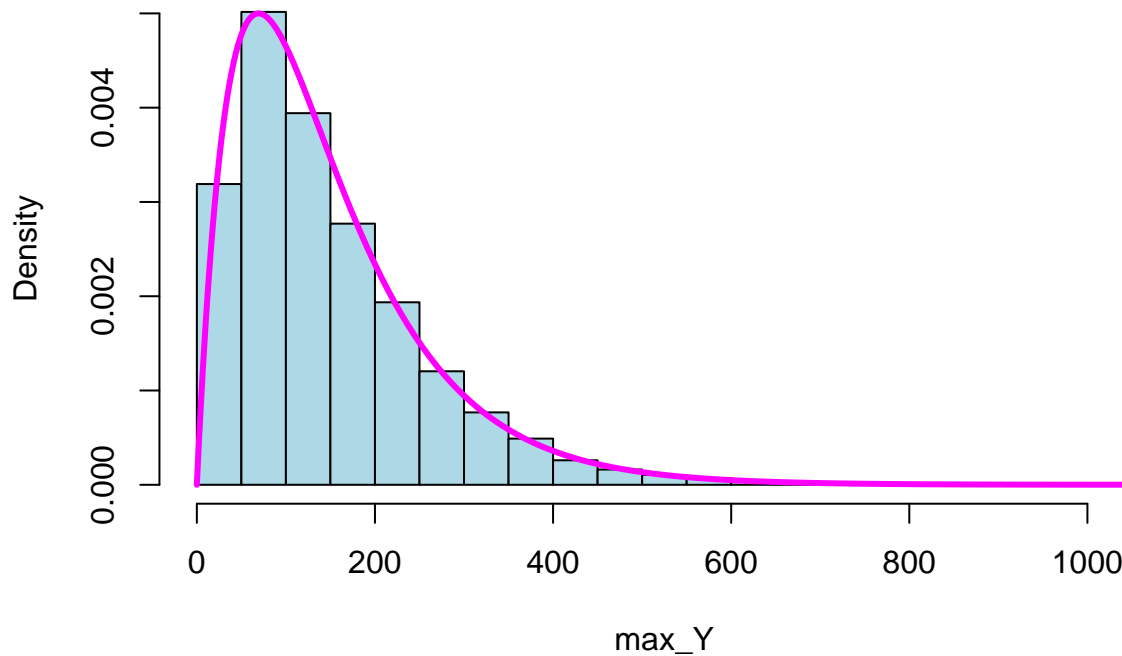
Answer:

Put your answer in this answer block. Some starting R code is provided.

```r
set.seed(4302019)
num_realizations <- 10000
### a)


Y <- rexp(2 * num_realizations, rate = 1 / 100)
YY <- matrix(Y, 2, num_realizations)
max_Y <- apply(YY, MAR = 2, FUN = "max") # 1 for rows, 2 for columns
#hist(max_Y, col = "LightBlue", freq = FALSE)

### b) and c)
X_grid <- seq(from = 0, to = max(max_Y), length = 1000)
f_X <- (1 / 50) * (exp(-X_grid / 100) - exp(-X_grid / 50))
hist(max_Y, col = "LightBlue", freq = FALSE)
lines(X_grid, f_X, lwd = 3, col = "magenta")
```

Histogram of max_Y

Parts a, b, and c are answered in the code block above.

Part d)

```r
mean(max_Y)
```

```
## [1] 148.8218
```

```
var(max_Y)
```

```
## [1] 12318.6
```

Part e) This would be the equation for the expected value of the maximum order statistic.

$$E(Y_{(n)}) = \int_0^\infty y \frac{1}{50}(e^{-y/100} - e^{-y/50})\, dy = 150$$

Here is the second moment of the maximum order statistic.

$$E(Y_{(n)}^2) = \int_0^\infty y^2 \frac{1}{50}(e^{-y/100} - e^{-y/50})\, dy = \frac{\Gamma(3)(100)^3}{50} - \Gamma(3)(50)^2 = 35000$$

So the variance is,

$$Var(Y_{(n)}) = E(Y_{(n)}^2) - \mu_{Y_{(n)}}^2 = 35000 - 150^2 = 12500$$

The empirical mean and variance are pretty similar to the theroitical mean and variance. Which makes sense because we had a large number of realizations.
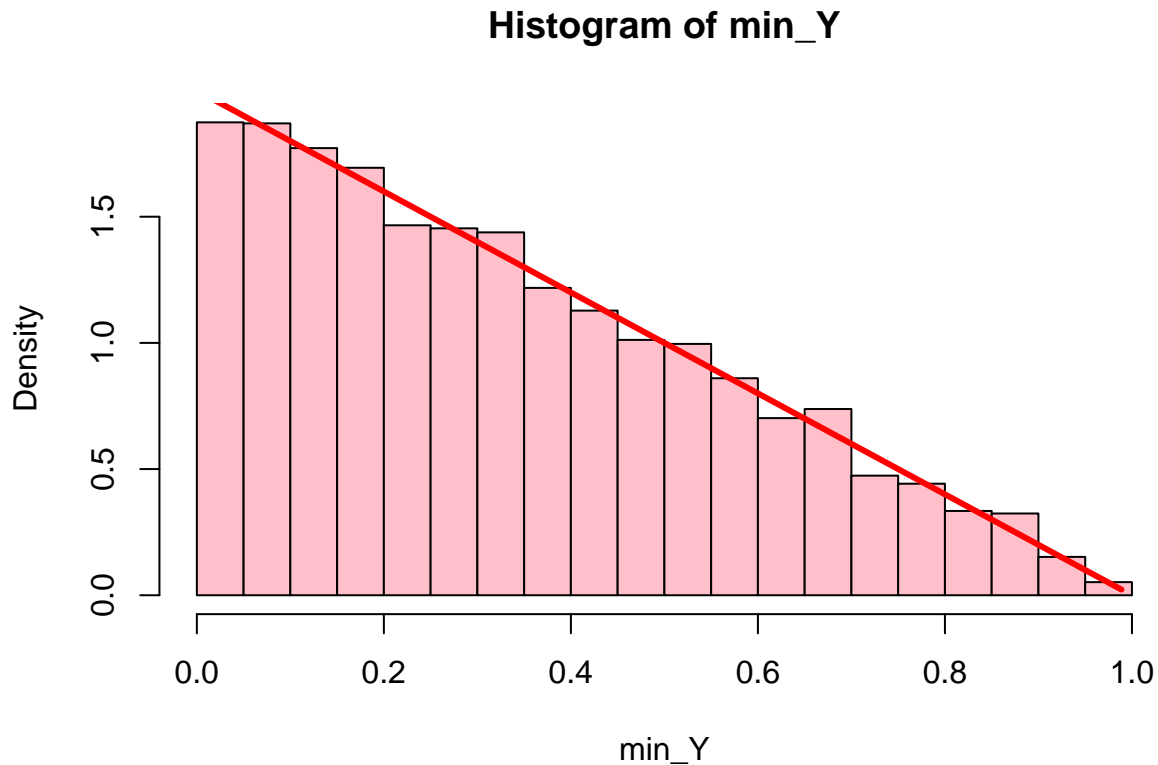
---

2. Complete **Exercise 6.72** of the text. Assess your results by simulation as follows. Reset your random number seed to 4302019, then (a) simulate 10000 realizations of the minimum of two Uniform(0,1) random variables (use `runif` in R), (b) plot their histogram, (c) add the theoretical probability density function that you derived in 6.72(a) to your histogram, (d) compute the empirical mean and variance of the minimum order statistic from your simulation. Finally, compare the theoretical values of $E(U_1)$ and $V(U_1)$ that you computed in 6.72(b) to your simulation values from (d).

Answer:

```
set.seed(4302019)
num_realizations <- 10000


Y <- runif(2*num_realizations, 0, 1)
YY <- matrix(Y, 2, num_realizations)
min_Y <- apply(YY, MAR=2, FUN="min")
hist(min_Y, col="Pink", freq=FALSE)

x_grid <- seq(from =0, to=max(min_Y), length.out = 1000)
f_x <- 2*(1-x_grid)
lines(x_grid, f_x, lwd=3 ,col="Red")
```

Histogram of min_Y


```r
mean(min_Y)
```

```
## [1] 0.3350795
```

```r
var(min_Y)
```

```
## [1] 0.05598781
```

$$E[U_{(1)}] = \int_0^1 y2(1-y)\,dy = 1/3$$

$$E[U_{(1)}^2] = \int_0^1 y^2 2(1-y)\,dy = 1/6$$

$$V[U_{(1)}] = E[U_{(1)}^2] - E[U_{(1)}]^2 = 1/6 - (1/3)^2 = 1/18$$

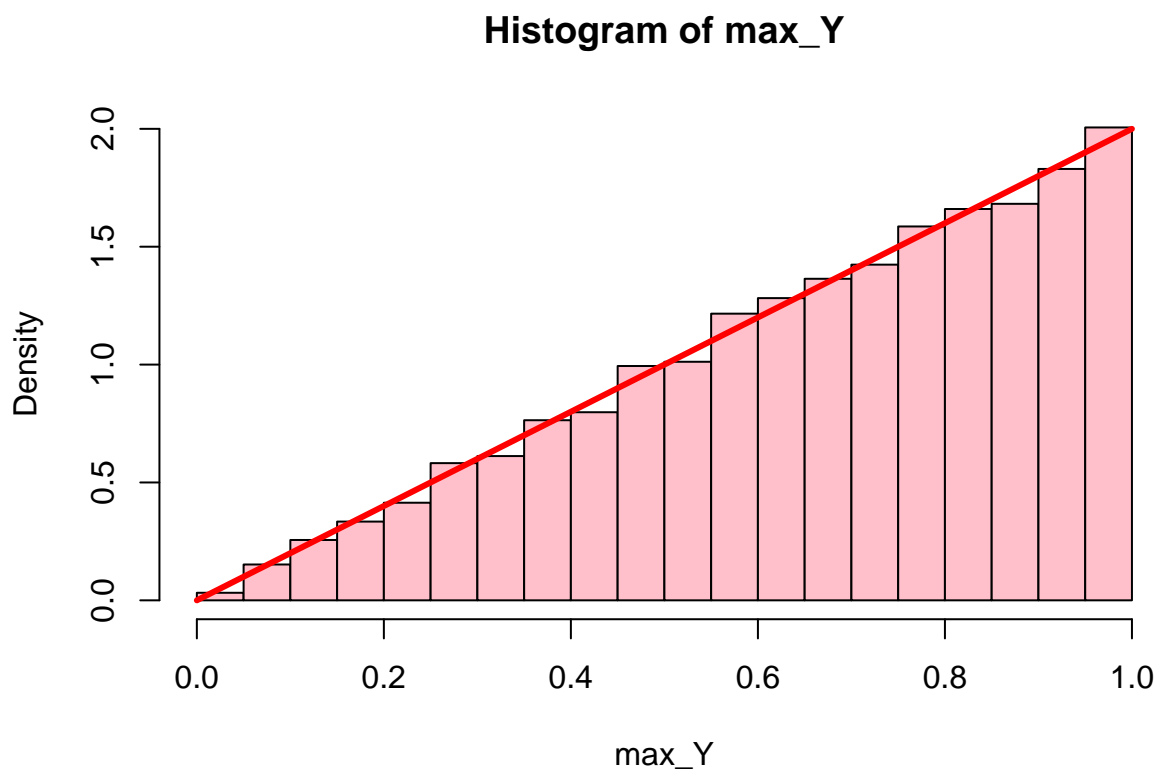The empirical mean and variance are similar to the theoretical mean and variance.

---

3. Complete **Exercise 6.73** of the text. Assess your results by simulation as follows. Reset your random number seed to 4302019, then (a) simulate 10000 realizations of the maximum of two Uniform(0,1) random variables (use `runif` in R), (b) plot their histogram, (c) add the theoretical probability density function that you derived in 6.73(a) to your histogram, (d) compute the empirical mean and variance of the maximum order statistic from your simulation. Finally, compare the theoretical values of $E(U_2)$ and $V(U_2)$ that you computed in 6.73(b) to your simulation values from (d).

Answer:

```r
set.seed(4302019)
num_realizations <- 10000


Y <- runif(2*num_realizations, 0, 1)
YY <- matrix(Y, 2, num_realizations)
max_Y <- apply(YY, MAR=2, FUN="max")
hist(max_Y, col="Pink", freq=FALSE)

x_grid <- seq(from=0, to = max(max_Y), length.out = 1000)
f_x <- 2*(x_grid)
lines(x_grid, f_x, lwd=3, col="Red")
```

Histogram of max_Y


```r
mean(max_Y)
```

```
## [1] 0.6682412
```

```r
var(max_Y)
```

```
## [1] 0.05471455
```

$$E[U_{(n)}] = \int_0^1 y2y\,dy = 2/3$$

$$E[U_{(n)}^2] = \int_0^1 y^2 2y\,dy = 1/2$$

$$Var[U_{(n)}] = E[U_{(1)}^2] - E[U_{(n)}]^2 = 1/18$$

The empirical mean and variance are similar to the theoretical mean and variance.

---

4. Complete **Exercise 6.80** of the text. Assess your results by simulation for $n = 2$ as follows. Reset your random number seed to 4302019, then (a) simulate 10000 realizations of the maximum of two Beta(2,2) random variables (use `rbeta` in R with `shape1 = 2` and `shape2 = 2`), (b) plot their histogram, (c) add the theoretical probability density function that you derived in 6.80(b) to your histogram, (d) compute the empirical mean and variance of the maximum order statistic from your simulation. Finally, compare the theoretical value of $E(Y_{(2)}$ that you computed in 6.80(c) to your simulation values from (d).


Answer:

```
set.seed(4302019)
num_realizations <- 10000

Y <- rbeta(2*num_realizations, shape1=2, shape2=2)
YY <- matrix(Y,2,num_realizations)
max_Y <- apply(YY, MAR=2, FUN="max")
hist(max_Y, col="Pink", freq=FALSE)

# The theotetical density function

x_grid <- seq(from=0, to = max(max_Y), length.out = 1000)
f_x <- 12*(x_grid)^3*(x_grid - 1)*(2*x_grid - 3) #Density function for the beta functions.
lines(x_grid, f_x, lwd=3, col="Red")
```

Histogram of max_Y
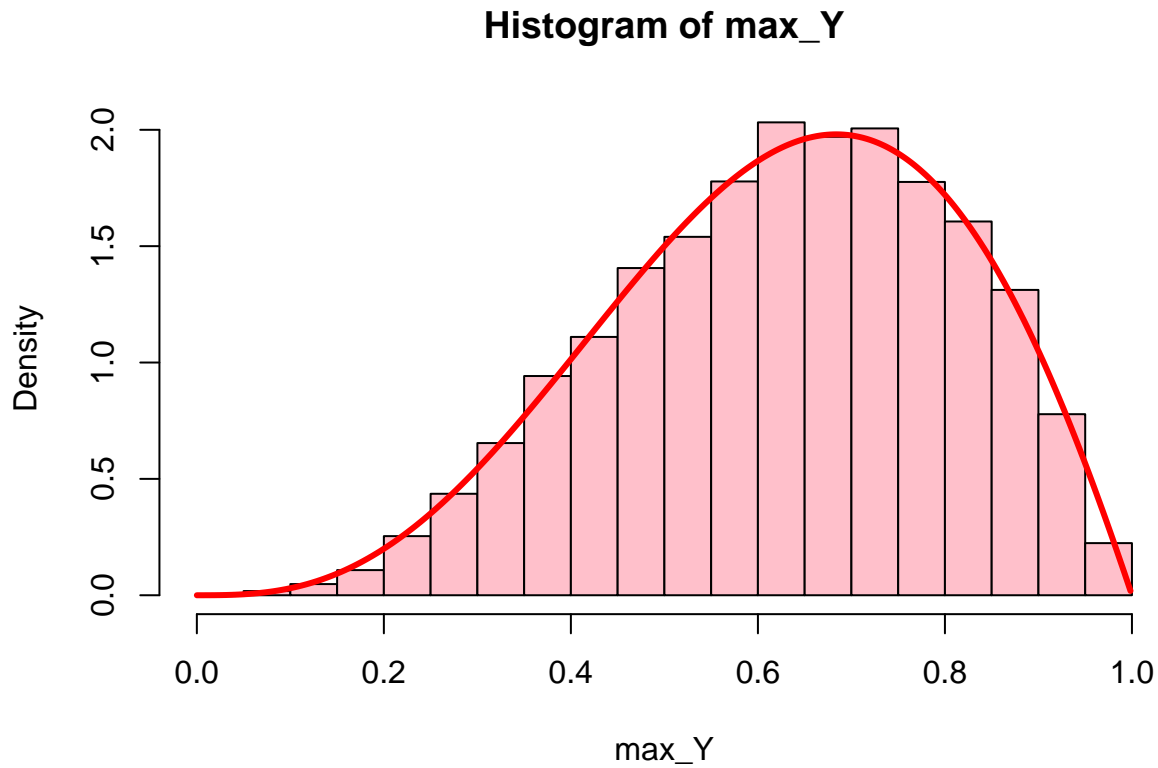

```r
mean(max_Y)
```

```
## [1] 0.6290335
```

```r
var(max_Y)
```

```
## [1] 0.0325877
```

The theoretical expected value is

$$E[Y_{(n)}] = \int_0^1 y * [12y^3(y-1)(2y-3)]dy = 22/35$$

The theoretical second moment is

$$E[Y_{(n)}^2] = \int_0^1 y^2 * [12y^3(y-1)(2y-3)]dy = 3/7$$

Varience is

$$Var[Y_{(n)}] = E[Y_{(n)}^2] - E[Y_{(n)}]^2 = 41/1225$$

Again the theoretical expected value and varience are very similar to the emperical solutions.

5. In class, we used complete enumeration to study the sampling distribution of the sample mean when drawing an iid sample of size $n = 2$ from a population with

$$P(Y_i = 1) = P(Y_i = 3) = P(Y_i = 5) = P(Y_i = 15) = 1/4.$$

The exact sampling distribution of the sample mean on the integers $1, 2, \ldots, 15$ was shown to be

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	0	0	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	0	0	0	0	$\frac{1}{16}$

Let's compare the exact sampling distribution to an approximation based on simulation:

```
set.seed(4302019)
num_realizations <- 10000
Y <- sample(c(1, 3, 5, 15), size = 2 * num_realizations, replace = TRUE)
YY <- matrix(Y, 2, num_realizations)
Ybar <- apply(YY, MAR = 2, FUN = "mean")
# Compute and plot the true probabilities:
true_prob <- c(1, 2, 3, 2, 1, 0, 0, 2, 2, 2, 0, 0, 0, 0, 1) / 16
plot(1:15, true_prob,
     ylab = "Probability",
     xlab = "Sample Mean",
     type = "b", pch = 16, lwd = 3, col = "magenta")
lines(sort(unique(Ybar)), table(Ybar) / num_realizations, type = "h", lwd = 3, col = "blue")
```
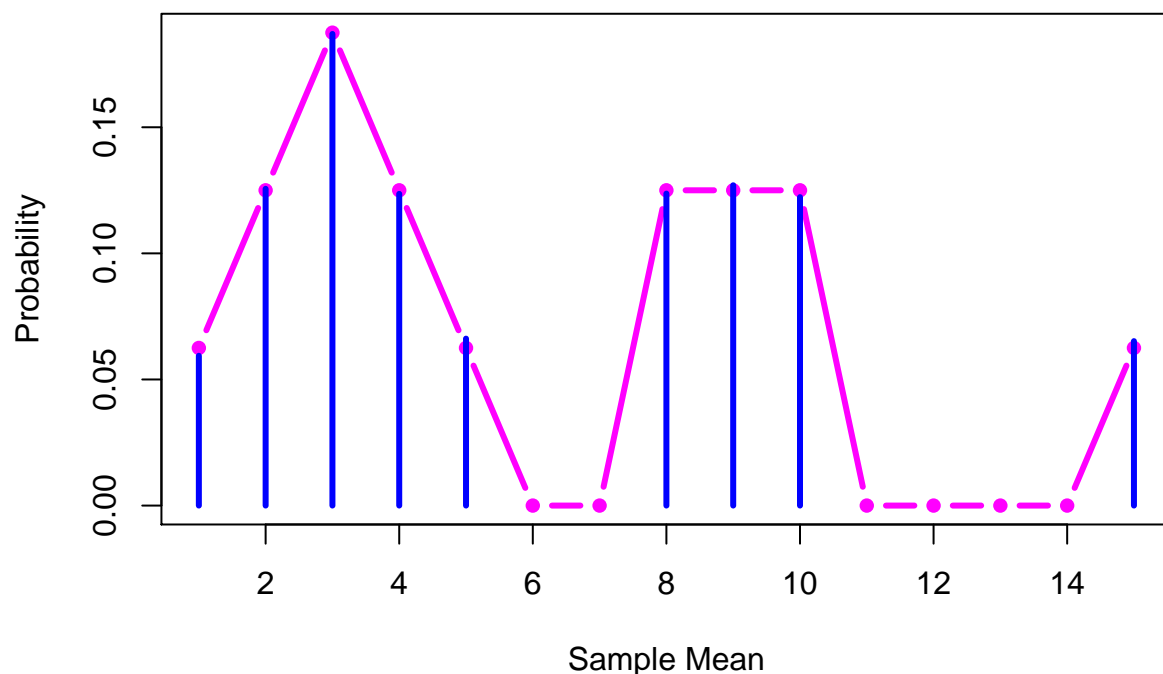

Now adapt this example to study the *range*,

$$|Y_1 - Y_2| = \max\{Y_1, Y_2\} - \min\{Y_1, Y_2\}.$$

Use complete enumeration to give the exact sampling distribution of the range. Then approximate the exact sampling distribution of the range by simulation, drawing 10,000 samples of size $n = 2$ from the population. Compare your exact distribution to your simulation-based approximation with a figure like that given above.

Answer:

```r
set.seed(4302019)
num_realizations <- 10000
Y <- sample(c(1, 3, 5, 15), size = 2 * num_realizations, replace = TRUE)
YY <- matrix(Y, 2, num_realizations)


Y_min <- apply(YY, MAR = 2, FUN = "min")
Y_max <- apply(YY, MAR = 2, FUN = "max")
Y_range <- Y_max - Y_min

# Compute and plot the true probabilities:
true_prob <- c(4, 0, 4, 0, 2, 0, 0, 0, 0, 0, 2, 0, 2, 0, 2) / 16
plot(0:14, true_prob,
     ylab = "Probability",
     xlab = "Sample Range",
     type = "b", pch = 16, lwd = 3, col = "magenta")
lines(sort(unique(Y_range)), table(Y_range) / num_realizations, type = "h", lwd = 3, col = "blue")
```
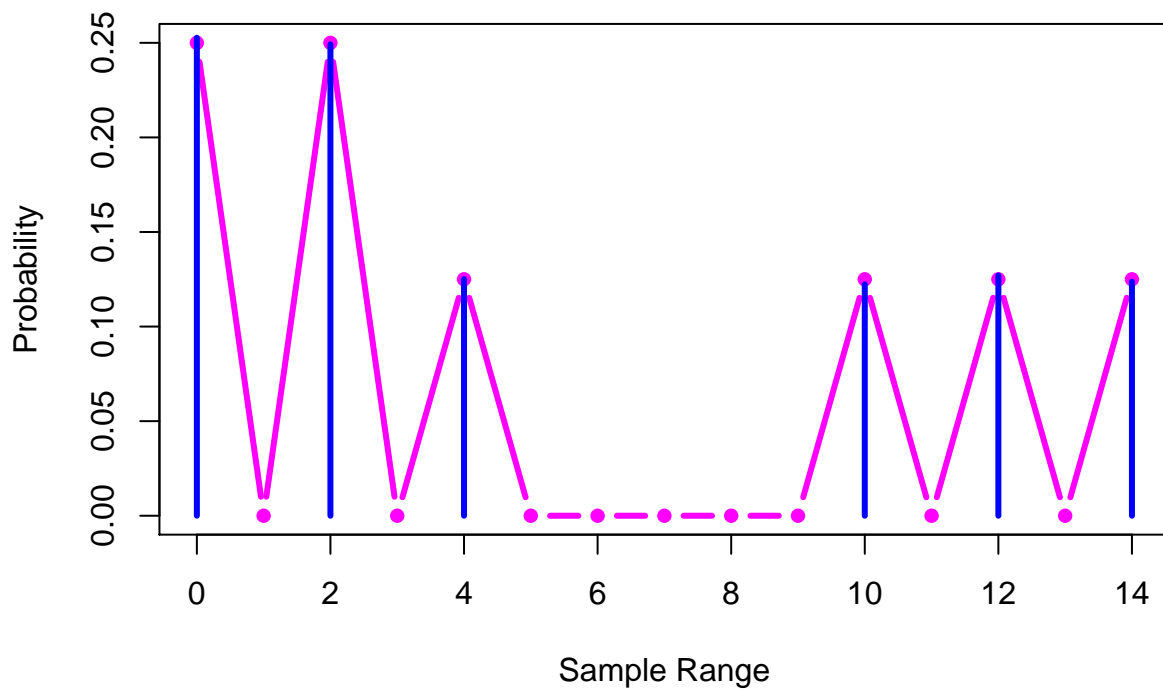

The pmf of the

$$range(Y_1, Y_2)$$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\frac{4}{16}$	0	$\frac{4}{16}$	0	$\frac{2}{16}$	0	0	0	0	0	$\frac{2}{16}$	0	$\frac{2}{16}$	0	$\frac{2}{16}$