Homework 3

Rui Gao

2021-10-17

```
library(ggplot2)  # you'll need this package for the R section

# get `play()` function for #1
source("https://gist.githubusercontent.com/cwickham/abe3b4c4ba5319e8e1dd5102541f2117/raw")
```

Instructions

Use the R Markdown version of this file to complete and submit your homework. Items in **bold** require an answer. Make sure you change the author in the header to your own name.

Conceptual Questions

1. Describe why we do not usually know the population mean. What statistic do we usually use to estimate the population mean and why?

If we want to know a real population mean, we need the data of the the population and there is no need to estimate. We estimate it by taking the average of a sample' values.Due to the expectation of it is the mean of the population, so the larger amount we take, the more accurate result we would get.

2. Consider two hypothetical histograms:

   i) a histogram of a sample of size n from the population and,
   ii) a histogram of k sample means from samples of size n from the population.

   a) For large values of n, which of the above histograms would give you a good estimate of the population distribution?

   Histogram of i)

   b) Which of the above histograms is an estimate of the "sampling distribution of the sample mean for samples of size n"?

   Histogram of ii)

   c) Describe ii) in relation to the population distribution. E.g. how do its center, spread and shape compare to the population distribution?

It has the same center with the population. As the n increases, the shape will getting narrower and more similar to the Normal distribution.

```
d) Consider these two true statements:
   "For large values of $n$ the sampling distribution of the sample mean approaches a Normal distributi

  "For large values of $k$ the histogram in ii) approaches the true sampling distribution of the sample

  One is a consequence of the "Central Limit Theorem" and the other is a consequence of the "Law of Lar


Answer: The first is the Central Limit Theorem, the second one is the Law of Large Numbers.
```

R questions

1. Consider this game: You roll one die, and lose $50 if you roll a 1, but win $15 if you roll anything else. I've written a function for you, `play()` that plays this game (the line starting `source(` in the code chunk at the top of this document gets this function for you).

   You can play by calling the `play()` function

```r
play()
```

```
## You rolled a 3. Your payout is $15.
```

```
## [1] 15
```

   The function `play()` returns a numeric value of either `-50` or `15` depending on your roll, and prints out the result. When you simulate many games, the print out will be time consuming, so use `play(silent = TRUE)` to play without printing results.

   a) **Use simulation to estimate your expected win/loss value for one roll.** Hint: simulate many plays of the game and take the average of the outcomes.

```r
a <- replicate(10000, play(silent = TRUE))
a_1 <- mean(a)
a_1
```

```
## [1] 4.5155
```

   The estimate income is $4.52.

   b) **How many times did you play to find your estimate? How precise do you think your answer is?**

   10000 times. To find how precise it is, I should try it for many times and check the variance of it. The result is down to 3.72 and up to 4.65, so I think it's not that accurate.

   c) **How much would you be willing to pay to play this game?**

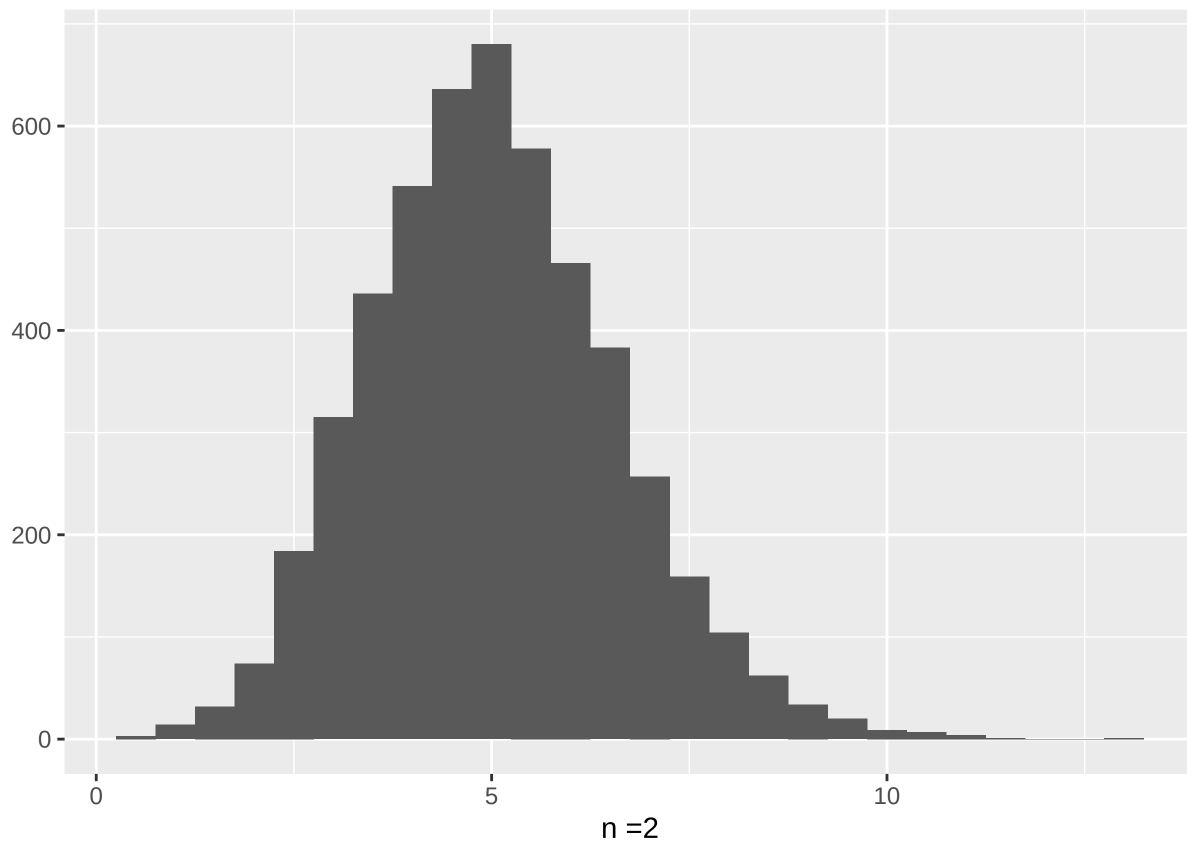
   $100 for fun.

2. In lab you explored the Central Limit Theorem when the population distribution was a Gamma(5, 1). The amazing thing about the Central Limit Theorem is that it applies no matter the shape of the distribution (as long as the distribution has an expected value, and a finite variance). For this question, **choose one of the following distributions, and replicate the exploration from the lab with sample sizes of 2, 10, 50 and 100**:

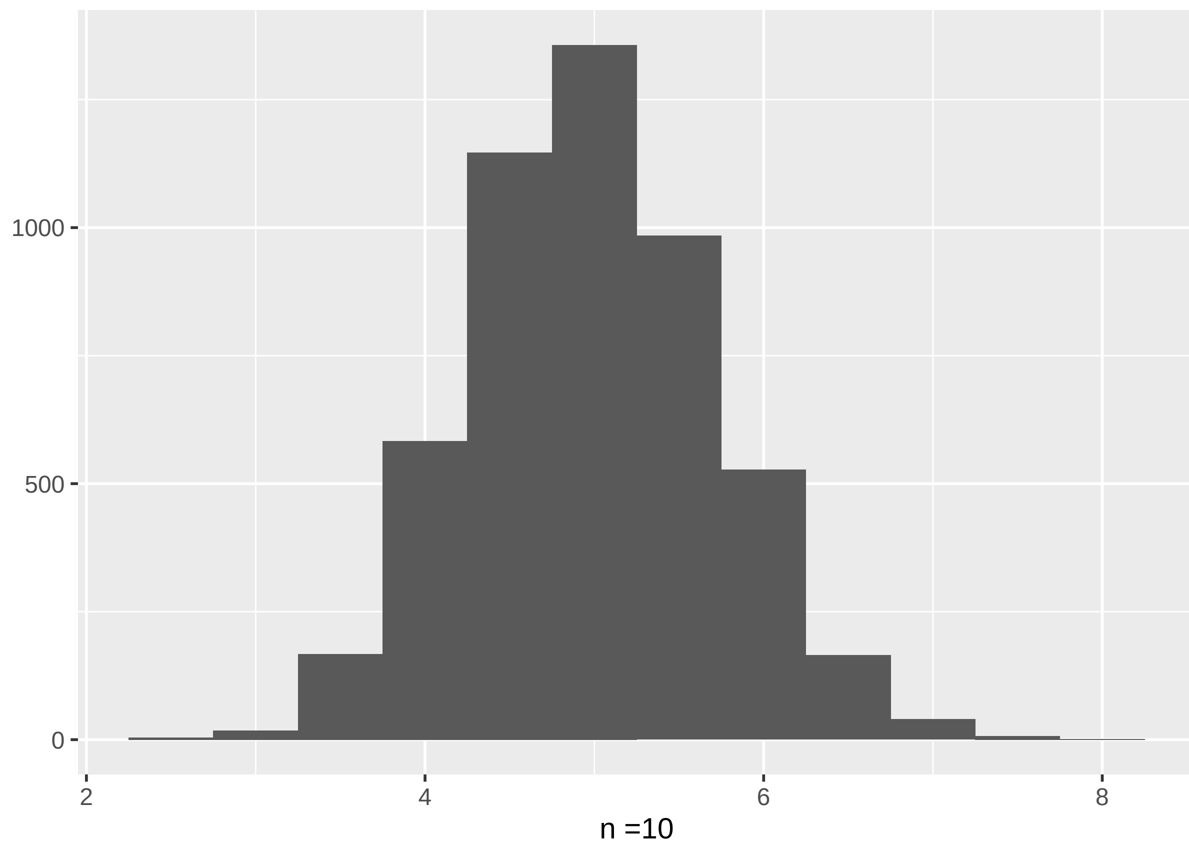
- Continuous uniform on (0, 1), see `?runif`
- Discrete uniform on $1, \ldots, 10$, use `sample()`
- A poisson distribution with your choice of parameter, see `?rpois`
- Beta distribution with both parameters set to 0.5, see `?rbeta`

Write up your exploration in a way that a reader can follow without having to understand your code.

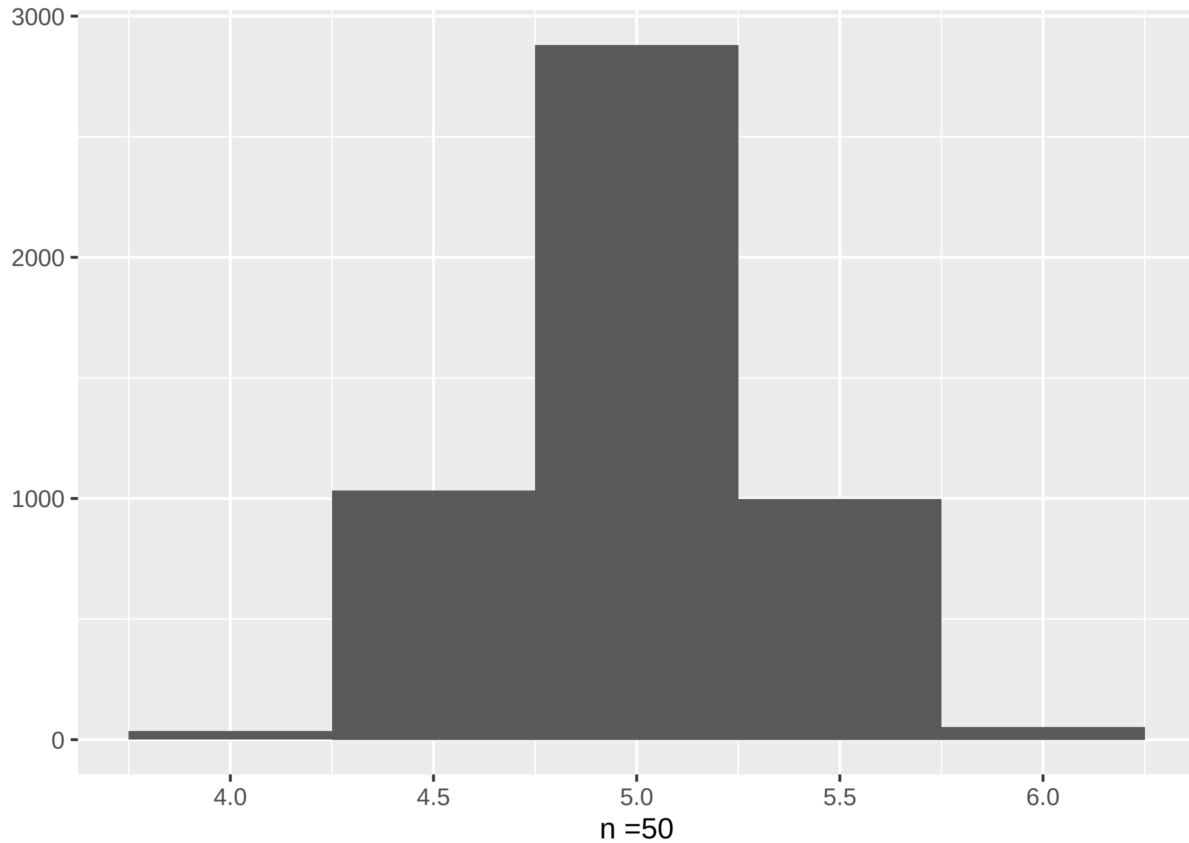
```
q_2 <- replicate(5000, mean(rpois(2, 5)))
q_10 <- replicate(5000, mean(rpois(10, 5)))
q_50 <- replicate(5000, mean(rpois(50, 5)))
q_100 <- replicate(5000, mean(rpois(100, 5)))



qplot(q_2, binwidth = 0.5, xlab = "n =2")
```



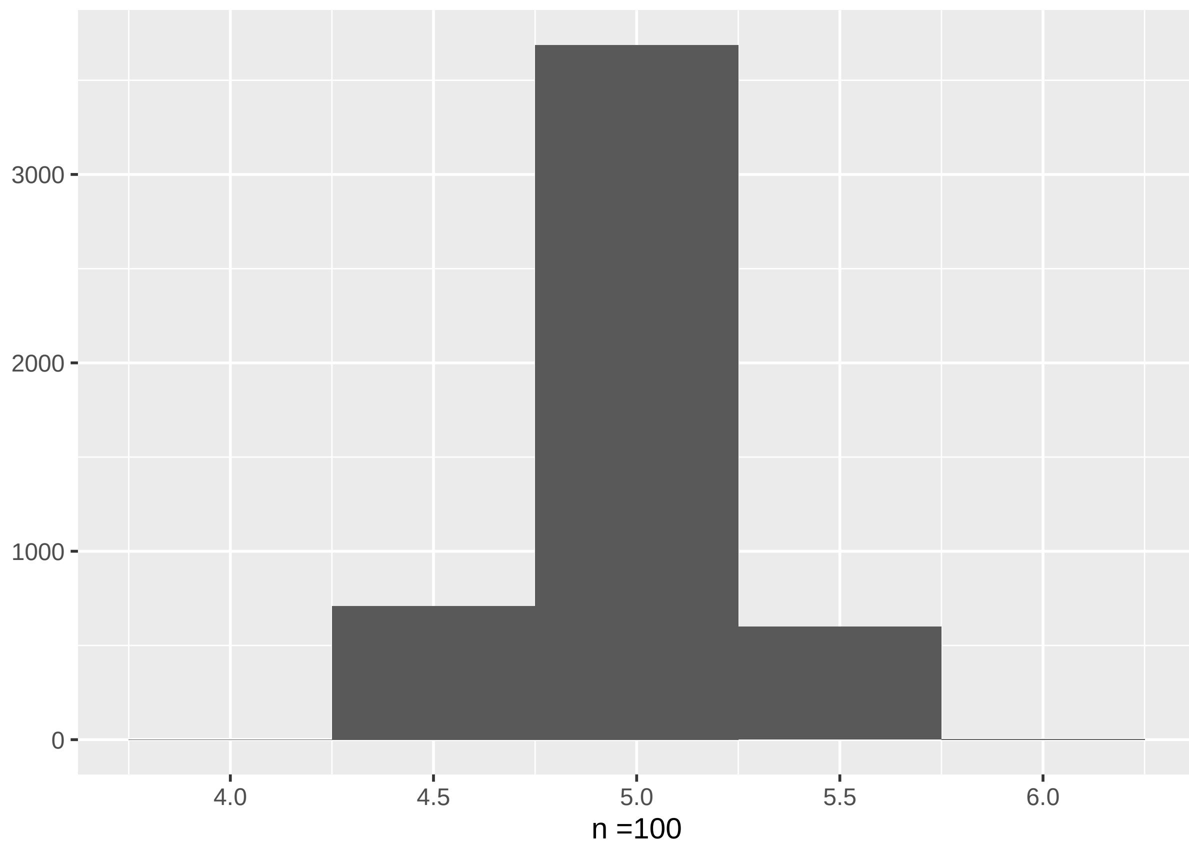

```
qplot(q_10, binwidth = 0.5, xlab = "n =10")
```




```
qplot(q_50, binwidth = 0.5, xlab = "n =50")
```




```
qplot(q_100, binwidth = 0.5, xlab = "n =100")
```



For these samples, all centered at 5.0, and as the n increases, the scale of samples' means decreases. So, the Poisson follows the Central Limit Theorem.