Exploring Distributions

STA 310: Homework 2

Instructions

- Write all narrative using full sentences. Write all interpretations and conclusions in the context of the data.
- Be sure all analysis code is displayed in the rendered pdf.
- If you are fitting a model, display the model output in a neatly formatted table. (The `tidy` and `kable` functions can help!)
- If you are creating a plot, use clear and informative labels and titles.
- Render and back up your work reguarly, such as using Github.
- When you're done, we should be able to render the final version of the Rmd document to fully reproduce your pdf.
- Upload your pdf to Gradescope. Upload your Rmd, pdf (and any data) to Canvas.

These exercises come from BMLR or are adapted from BMLR, Chapter 3.

Exercises

Exercise 1

At what value of p is the variance of a binary random variable smallest? When is the variance the largest? Back up your answer empirically or mathematically.

Recall that the variance of a binary random variable $X \sim \text{Bernoulli}(p)$ is $p(1-p)$. To get the maximum and minimum, let's calculate the derivative of the variance with respect to p and set it to 0:

$$\frac{d}{dp}p(1-p) = 0 \Rightarrow 1 - 2p = 0 \Rightarrow p = 0.5$$

Note that when $p = 0.5$, $Var(X) = 0.25$, which is clearly the maximum. Additionally, since $p \in (0, 1)$, then $p(1-p) \leq 0$, which means that the minimum possible for the variation is 0, what can be obtained when $p = 0 \vee p = 1$.

Therefore, the variance is the largest ($Var(X) = 0.25$) when $p = 0.25$, and the variance is the smallest ($Var(X) = 0$) when $p = 0 \vee p = 1$.

Exercise 2

How are hypergeometric and binomial random variables different? How are they similar?

Exercise 3

How are exponential and Poisson random variables related?

Exercise 4

How are geometric and exponential random variables similar? How are they different?

Exercise 5

A university's college of sciences is electing a new board of 5 members. There are 35 applicants, 10 of which come from the math department. What distribution could be helpful to model the probability of electing X board members from the math department?

Exercise 6

Chapter 1 asked you to consider a scenario where *"The Minnesota Pollution Control Agency is interested in using traffic volume data to generate predictions of particulate distributions as measured in counts per cubic feet."* What distribution might be useful to model this count per cubic foot? Why?

Exercise 7

Chapter 1 also asked you to consider a scenario where *"Researchers are attempting to see if socioeconomic status and parental stability are predictive of low birthweight. They classify a low birthweight as below 2500 g, hence our response is binary: 1 for low birthweight, and 0 when the birthweight is not low."* What distribution might be useful to model if a newborn has low birthweight?

Exercise 8

Chapter 1 also asked you to consider a scenario where *"Researchers are interested in how elephant age affects mating patterns among males. In particular, do older elephants have greater mating success, and is there an optimal age for mating among males? Data collected includes, for each elephant, age and number of matings in a given year."* Which distribution would be useful to model the number of matings in a given year for these elephants? Why?

Exercise 9

Describe a scenario which could be modeled using a gamma distribution.

Exercise 10

Beta-binomial distribution. We can generate more distributions by mixing two random variables. Beta-binomial random variables are binomial random variables with fixed n whose parameter p follows a beta distribution with fixed parameters α, β. In more detail, we would first draw p_1 from our beta distribution, and then generate our first observation y_1, a random number of successes from a binomial (n, p_1) distribution. Then, we would generate a new p_2 from our beta distribution, and use a binomial distribution with parameters n, p_2 to generate our second observation y_2. We would continue this process until desired.

Note that all of the observations y_i will be integer values from $0, 1, \ldots, n$. With this in mind, use `rbinom()` to simulate 1,000 observations from a plain old vanilla binomial random variable with $n = 10$ and $p = 0.8$. Plot a histogram of these binomial observations. Then, do the following to generate a beta-binomial distribution:

a. Draw p_i from the beta distribution with $\alpha = 4$ and $\beta = 1$.

b. Generate an observation y_i from a binomial distribution with $n = 10$ and $p = p_i$.
c. Repeat (a) and (b) 1,000 times ($i = 1, \ldots, 1000$).
d. Plot a histogram of these beta-binomial observations.

Compare the histograms of the "plain old" binomial and beta-binomial distributions. How do their shapes, standard deviations, means, possible values, etc. compare?

Grading

Total	**33**
Ex 1	5
Ex 2	5
Ex 3	5
Ex 4	5
Ex 5	2
Ex 6	2
Ex 7	2
Ex 8	2
Ex 9	2
Ex 10	5
Workflow & formatting	3

The "Workflow & formatting" grade is to based on the organization of the assignment write up along with the reproducible workflow. This includes having an organized write up with neat and readable headers, code, and narrative, including properly rendered mathematical notation. It also includes having a reproducible Rmd or Quarto document that can be rendered to reproduce the submitted PDF.