Gov 2001: Problem Set 2

Due Wednesday, February 10th by 6pm

Important Reminder:

The deadline for choosing a paper to replicate is **Wednesday, February 24, 2016 at 5pm**. On or before February 24, send a pdf of your proposed paper to Prof. Gary King and CC both TFs. We will approve the paper (or tell you to look for a new one). Include within your email a brief note (approx. 2-5 sentences) explaining why this paper is a good choice and also how you think you might be able to make a contribution if you can replicate it. You are also required to have one of your classmates sign off on your article choice after checking that your article meets all the criteria listed in *Publication, Publication.*

Instructions

You should submit your answers, R code, and any work you did to get to your answers using the Quizzes section on Canvas.

Simulation

Problem 1

Let X be a continuous random variable with pdf $f(x|\lambda)$ and parameter λ with support over the interval $[0, \infty)$, where

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

1.A) What are the equations for $E(X)$ and $\text{Var}(X)$? You don't need to simplify. (e.g. $\int_a^b$ something dx)

1.B) Suppose $\lambda = 2$. Find $E(X)$ and $\text{Var}(X)$ (without simulating) with the help of the `integrate()` function in R.

1.C) It turns out that X follows what is known as the exponential distribution where λ is the rate parameter. Suppose again that $\lambda = 2$. Simulate 10,000 draws of X from the exponential distribution using `rexp()`. Find $E(X)$ and $\text{Var}(X)$ via simulation (use `set.seed(02138)`). Do your answers match your answers in Part B?

1.D) Suppose we have another random variable Y, where

$$Y = \frac{\ln X}{X^2 - e^X}$$

Find $E(Y)$ and $\text{Var}(Y)$ via simulation (hint: there is no need to redraw random quantities).

Probability Distributions

Problem 2

You have the following model:

$$Y_i \sim f_N(\mu_i, \sigma^2 = 9)$$

$$\mu_i = 2 + 36x_i$$

Let x_i be a covariate for gender, where $x_i = 0$ for men and $x_i = 1$ for women. Generate one set of 1000 simulated outcomes according to the model outlined above. Create the gender covariate using the code `c(rep(0, 500), rep(1, 500))`. Set the 'random' number generator in R using the `set.seed(02138)` command before drawing any random quantities.

2.A) Plot a histogram of your simulated data and attach the plot as a PDF or image file.

2.B) Describe the distribution of your simulated data.

2.C) Does your data violate an assumption of the traditional Ordinary Least Squares regression model? Explain your answer.

Problem 3

You have the following model:

$$Y_i \sim f_N(\mu = 18, \sigma_i^2)$$

$$\sigma_i^2 = 2 + 36x_i$$

Let x_i be a covariate for employment, where $x_i = 0$ for unemployed and $x_i = 1$ for employed. Generate one set of 1000 simulated outcomes according to the model outlines above. Create the employment covariate using the code `c(rep(0, 500), rep(1, 500))`. Set the 'random' number generator in R using the `set.seed(02138)` command before drawing any random quantities.

3.A) Create one histogram of the y variable for employed observations and another histogram of the y variable for the unemployed observations. Submit the two histograms as side-by-side graphs in one figure in a PDF or image file. Be sure to label your two graphs and make them look professional.

3.B) What is the mean of the simulated data for employed observations?

3.C) What is the mean of the simulated data for unemployed observations?

3.D) Explain why this model is useful and what this model tells us about the distribution of Y.

Problem 4

Let Y be a continuous random variable with pdf $f(y|\mu, \sigma)$ and parameters μ and σ with support over the interval $(-\infty, \infty)$ (Y can take on any value in this interval), where

$$f(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}|y-\mu|}{\sigma}}$$

4.A) Write a function in R for the pdf of this distribution. Your function should take three inputs: a vector of data (y) and the two parameters (μ and σ).

4.B) Using your function, plot the pdf of Y with $\mu = 0$ and $\sigma = 2$. (You do not need to simulate data for this.)

4.C) Integrate this function over the support of Y (all possible values Y can take) with $\mu = 0$ and $\sigma = 2$. What does this tell us about $f(y|\mu, \sigma)$?

4.D) Using your function, find $P(-2 < Y < 1.75)$ if $\mu = 0$ and $\sigma = 2$.

R code

Please submit all your code for this assignment as a .R file. Your code should be clean, commented, and executable without error.