Gov 2001: Problem Set 2 Answers

Important Reminder:

The deadline for choosing a paper to replicate is **Wednesday, February 24, 2016 at 5pm**. On or before February 26, 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. 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

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

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)

$$\begin{aligned} E(X) &= \int_0^\infty x\lambda e^{-\lambda x} dx \\ Var(X) &= E(X^2) - [E(X)]^2 \\ &= \int_0^\infty x^2 \lambda e^{-\lambda x} dx - \left[\int_0^\infty x\lambda e^{-\lambda x} dx \right]^2 \end{aligned}$$

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

The expected value of X is 0.5 and the variance is 0.25.

```
ex.func <- function(x, lambda){
  x * lambda * exp(-lambda * x)
}
e.x <- integrate(Vectorize(ex.func), lower=0, upper=Inf, lambda=2)$value

# seems to be the same without Vectorize
integrate(ex.func, lower=0, upper=Inf, lambda=2)$value


# write a function for E(X^2) and integrate
ex.sq.func <- function(x, lambda){
  x^2 * lambda * exp(-lambda * x)
}

e.x.sq <- integrate(Vectorize(ex.sq.func), lower=0, upper=Inf, lambda=2)$value

# seems to be the same, don't have to use Vectorize
integrate(ex.sq.func, lower=0, upper=Inf, lambda=2)

# Var(X) = E(X^2) - [E(X)]^2
var.x <- e.x.sq - e.x^2
var.x
```

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?
The simulated expected value is 0.5055 and the simulated variance is 0.2688, which are very close to the empirical values we calculated above.

```
set.seed(02138)
exp.draws <- rexp(100000, rate=2)
mean(exp.draws)
var(exp.draws)
```

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).
The simulated mean of y is 1.147 and the simulated variance is 1.507.

```
y.draws <- log(exp.draws) / (exp.draws^2 - exp(exp.draws))
mean(y.draws)
var(y.draws)
```

Probability Distributions

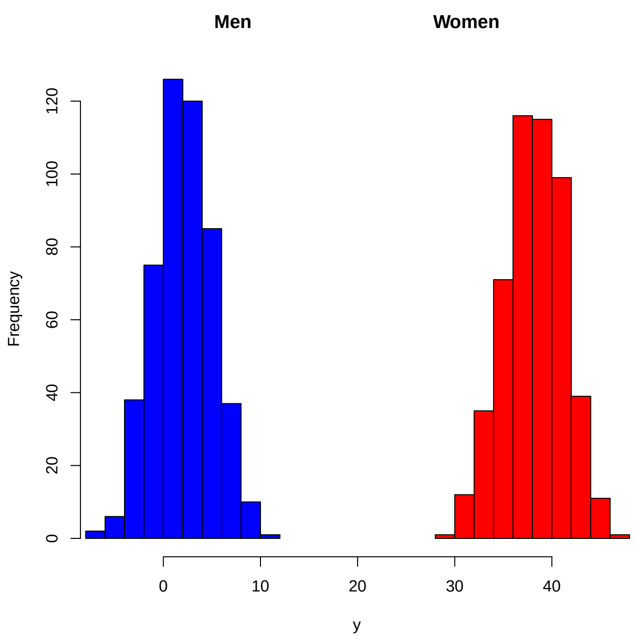
Problem 2

You have the following model:

$$Y_i \sim f_N(\mu, \sigma^2 = 9)$$
$$\mu = 2 + 36x_i$$

Let x_i be a covariate for gender, where $x_i = 0$ for men and $x_i = 1$ for women. Create the gender covariate using the code `c(rep(0,500),rep(1,500))`. Generate one set of simulated outcomes according to the model outlined above. Your simulated data should contain 1000 observations. 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.
The unconditional distribution of y is bimodal. Conditional on the gender covariate, though, the distribution appears to be normal.

2.C) Does your data violate an assumption of the traditional Ordinary Least Squares regression model? Explain your answer.
No it does not. Although y has a bimodal unconditional distribution, when we condition on the gender covariate the distribution of y is normal. Another way to put this is that

the distribution of the residuals is normal, which is the assumption you must make for OLS.

Problem 3

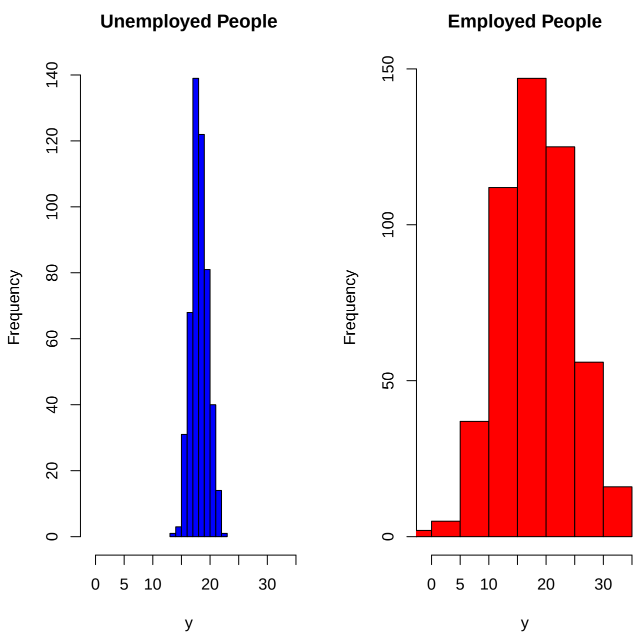
You have the following model:

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

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

Let x_i be a covariate for employment, where $x_i = 0$ for unemployed and $x_i = 1$ for employed. Create the employment covariate using the code `c(rep(0,500),rep(1,500))`. Your simulated data should each contain 1000 observations. Set the 'random' number generator in R using the `set.seed(02138)` command before drawing each set of 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?
The mean for employed people is 17.35.

3.C) What is the mean of the simulated data for unemployed observations?
The mean for unemployed people is 17.91.

3.D) Explain why this model is useful and what this model tells us about the distribution of Y.
The model allows us to study the effect of covariates on the variation in our outcome variable. The model tells us that the variance of distribution of Y is different for different levels of x.

Problem 4

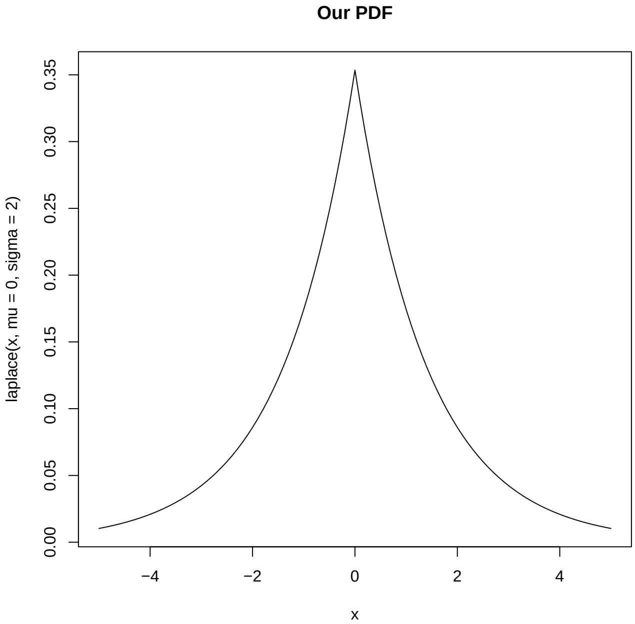
Let Y be a continuous random variable with pdf $f(y|\mu, \sigma)$ and parameters μ and σ with support over the interval $(-\infty, \infty)$, where

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

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).
This PDF is a reparameterized version of the Laplace distribution.

```
laplace <- function(y, mu, sigma){
  exp(-sqrt(2) * abs(y - mu) / sigma) / (sqrt(2) * sigma)
}
```

4.B) Using your function, plot the pdf of Y. (You do not need to simulate data for this.)



4.C) Integrate this function over the support of Y. What does this tell us about the $f(y)$?
Because the distribution integrates to 1 across its support, it is a valid PDF.

```
integrate(laplace, -1000, 1000, mu = 0, sigma = 2)
# also, doesn't really matter what values you pick for mu and sigma,
# as long as your upper and lower bounds are big enough

1 with absolute error < 1e-06
```

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

```
integrate(laplace, -2, 1.75, mu = 0, sigma = 2) #0.7334

0.7333737 with absolute error < 2.4e-05
```

R code

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