1. Written Assignment 6

1.1. **Problem 1.** A random variable defined on $(1, 3)$ has the probability density function (PDF)

$$f(x) = \frac{k}{x} \quad 1 < x < 3$$

where k is a constant.

a) What is the value of k?

b) What is the cumulative distribution function (CDF) $F(x)$ for this probability distribution?

c) What is the expected value of this random variable $E(X)$?

d) What is the variance of this random variable $V(X)$?

1.2. **Problem 2.** The p^{th} percentile ϕ_p of a distribution has the property that

$$F(\phi_p) = P(X \leq \phi_p) = p$$

so that

$$\phi_p = F^{-1}(p)$$

where F^{-1} is the function inverse of the CDF of X and $0 \leq p \leq 1$.

a) For the distribution in problem 1, what is $F^{-1}(x)$?

b) When $p = .5$, ϕ_p is called the **median** of the distribution. What is the median of the distribution in problem 1?

c) When $p = .25$, ϕ_p is called the **first quartile** of the distribution. What is the first quartile of the distribution in problem 1?

d) The quantity $\phi_{75} - \phi_{25}$ is called the **interquartile range** of the distribution. What is the interquartile range of the distribution in problem 1?

1.3. **Problem 3.** Suppose X is a continuous random variable having density function $f(x)$ with support on the interval $[c, d]$, and a and b are constants.

a) Show that

$$E(aX + b) = aE(X) + b$$

b) Show that

$$V(X + b) = V(X)$$

c) Show that

$$V(aX + b) = a^2 V(X)$$

1.4. **Problem 4.** Techniques that use a computer to simulate the behavior of random variables are called **Monte Carlo** methods.

Most statistical software packages have routines for generating a vector of values that simulate independent observations from a uniform distribution on $(0, 1)$. In R, the command `runif(10000)` generates such an array with 10,000 elements.

R has routines to generate all of the common probability distributions (`rbinom` for binomial, `rnorm` for normal, `rexp` for exponential). If you want to produce simulated data for a continuous probability distribution with CDF $F(x)$, you can use

$$F^{-1}(u)$$

where u has a uniform (0,1) distribution.

It is particularly easy to do this in R because R allows you to supply arrays as arguments to functions the same way you supply scalars, so if you want an array `x` containing 10,000 values of $\ln(u)$ corresponding to the natural logs of independent uniform random variables, you only need one line of code:

```
x=log(runif(10000))
```

Use R to simulate 10,000 observations from the distribution in problem 1 (just replace `log()` in the previous expression with the appropriate CDF inverse). You can access the online reference manual for R by entering `help.start()`. Most of the functions we will be using in this course are in the **stats** package, which you can find under the **packages** link. For help with a specific function like `mean`, type `?mean`. You can use the command `hist(x)` to display a histogram of the sample.

a) Use the command `mean(x)` to get the mean of the simulated sample. What is the difference between this value and your answer to 1c?

b) Use the command `var(x)` to get the variance of the simulated sample. What is the difference between this value and your answer to 1d?

c) Use the command `median(x)` to get the median of the simulated sample. What is the difference between this value and your answer to 2b?

d) Use the `quantile` command to get the 25^{th} and 75^{th} percentiles and the interquartile range for the sample. What is the difference between this value and your answer to 2d?