Stat 509 / Econ 580 - Homework 2

This Homework is due in GradeScope by Friday, October 16 at 8AM.

Remember to upload all `R` (or Python) code used to Canvas; a link will be provided.

(Note: Code and output should *also* be included in the pdf uploaded to GradeScope.)

1. A gambler plays the following game: a fair coin is tossed repeatedly. On the k-th toss, if the coin shows tails, she receives 0, if it shows heads she receives $2/(3^k)$, Let X_k be the gambler's total winnings after k stages.

   (a) Find the pmf and cdf for X_1.

   (b) Find the pmf and cdf for X_2.

   (c) Using simulation in `R`, or otherwise, find either approximately or exactly the pmf and cdf for X_5. Express your answers as plots.

   Hint: you may find the `R` *function* `plot.ecdf` *useful.*

2. (a) Simulate samples of size 500 from an Exponential random variable with parameter $\lambda = 2$ in `R` using `rexp`. Construct an empirical CDF from the data using the `plot.ecdf` command in `R`.

   Recall that the empirical CDF instead of plotting $P(X \leq x)$ for a random variable X, it plots the proportion of observations in the data that are $\leq x$.

   (b) Use the `rexp` or `qexp` functions to add the CDF for an exponentially distributed variable with parameter $\lambda = 2$ to your plot.

   Your solution should contain the few lines of `R` code that you used, together with a single plot containing the empirical CDF and the Exponential CDF superimposed.

OVER

3. Suppose that we have a dataset consisting of n numbers. Let the numbers, in order from smallest to largest, be $x_{(1)}, \ldots, x_{(n)}$ and assume that the numbers lie between 0 and 1 and there are no ties so that $0 < x_{(1)} < x_{(2)} < \cdots < x_{(n)} < 1$.

   (a) Let $\mathbb{F}(t)$ be the empirical CDF for this dataset. Express $\mathbb{F}(t)$ in terms of $x_{(1)}, \ldots, x_{(n)}$.
   Hint: count the number of observations that are $\leq t$.

   (b) If a random variable X has the Rectangular distribution on $[0, 1]$, find the CDF $F(t)$ for X. *Hint: integrate the pdf for X.*

   Suppose somebody gives you the following set of $n = 10$ observations:

   $$0.03, 0.11, 0.42, 0.44, 0.47, 0.66, 0.75, 0.88, 0.89, 0.90$$

   This person claims that this is a set of independent samples from a Rectangular $[0, 1]$ distribution (that they have ordered). If you were skeptical about this claim, it would be natural to compare the empirical CDF $\mathbb{F}(t)$ and the (theoretical) CDF $F(t)$.

   (c) Plot the empirical CDF $\mathbb{F}(t)$ and the (theoretical) CDF $F(t)$ on a single plot. *Hint: your plot can be contained inside the unit square.*

   (d) One way to measure the difference between these CDFs is to look for the largest difference between them:

   $$K = \sup_t |\mathbb{F}(t) - F(t)|\,.$$

   Note: sup is used here because $\mathbb{F}(t)$ is discontinuous.

   Find the value of K for the dataset given above. Also report the value of t where the supremum is attained.

   (Aside) K as defined here is the Kolmogorov-Smirnov statistic for testing whether a set of observations are a random sample from a Rectangular $[0, 1]$ distribution.

4. Goldberger question 2.12.

OVER

5. In this exercise the goal is to simulate data from a given distribution in order to verify the formulae given in Goldberger Table 3.1

   (a) Show via simulation that if X has a Poisson distribution with $\lambda = 2$ then $E(X) = V(X) = 2$. *Hint: Use the* `rpois` *function in* `R`.

      By simulating a few different sized datasets show that as the number of simulations increase, the mean and variance for your simulated data both become closer to 2.

   (b) Show via simulation that if X has a Binomial distribution with $n = 400, p = 0.3$ then $E(X) = np = 120$ and $V(X) = np(1 - p) = 84$. *Hint: Use the* `rbinom` *function in* `R`.

      Again by simulating different sized datasets show that as the number of simulations increase, the mean and variance for your simulated data approach their respective population values.

6. Suppose that you have a (possibly) biased coin that gives Heads with probability p, where p is unknown $(0 < p < 1)$.

   (a) Describe a way to *simulate* flipping a fair coin using only the (possibly) biased coin. *Hint: By considering a sequence of* **pairs** *of flips of the possibly biased coin, find an event that has probability exactly* $1/2$ *(for all values of* p*).*

   (b) Find the expected number of times that you are required to flip the possibly biased coin before it is determined that the event in (a) either has or has not happened.

   (c) Implement the procedure from (a) in `R` with a coin for which $p = 0.2$. Use the command `rbinom(1,1,p)` as a way of 'flipping' this biased coin once. Run your procedure 1000 times. Report your code.

      Hint: you are being asked to run your **procedure** 1000 *times: each 'run' of your procedure may require several (simulations of) flips of the biased coin.*

   (d) Report the average number of the biased coin flips per run of the procedure over these 1000 runs.

7. Goldberger Qu. 3.2