Problem Set 5

In this problem set, you will continue to be asked to work with the gamma distribution and you may find it useful to refer back to the previous problem set and/or solutions. I recommend Wikipedia as a reference, `https://en.wikipedia.org/wiki/Gamma_distribution`.

Keep your rendered `.pdf` to no more than 5 pages. Only provide code in the rendered `.pdf` when it is specifically requested.

1. This problem revisits the multivariate normal distribution, which I still don't expect you to be very familiar with but that doesn't mean you can't work with it!

(a) Write a function that takes n, the length of the vector that you'd like to simulate and ρ, an autocovariance parameter and returns an n dimensional random vector obtained by constructing a covariance matrix C with elements $c_{ij} = \rho^{|i-j|}$, taking its Cholesky decomposition, and multiplying the transpose of its Cholesky decomposition by n independent standard normal random variables (note - order matters). Print the code you use define the function to the rendered `.pdf`.

(b) Write a for loop that generates 2 random vectors of length 50 using your function from (a) with $\rho = 0.99$ and stores them in a 2×50 matrix called `mat`. Print the code you use to create this matrix and implement this for loop to the rendered `.pdf`.

(c) Evaluate the sample means, covariances, and correlation of the bivariate random variables you simulated in (b). Summarize your results in a table made using `kable`. Use your knowledge from previous statistics classes to compute these quantities.

(d) In at most one sentence, describe how the results in the table in (c) relate to what you would expect to see.

2. Consider a truncated gamma random variable x with shape a and rate b that is constrained to the interval $(0, c)$. The density $g(x|a, b, c)$ of this random variable satisfies:

$$g(x|a, b, c) = \frac{f(x|a, b)}{F(c|a, b)},$$

where $f(x|a,b)$ is the density of a gamma random variable with shape a and rate b and $F(c|a,b)$ is the probability that a gamma random variable with shape a and rate b is less than or equal to c.

(a) The CDF of this truncated gamma random variable is $G(x|a,b,c) = \frac{F(x|a,b)}{F(c|a,b)}$. Write a function that returns the inverse CDF of this truncated random variable as a function of a, b, and c. Print the code you use define the function to the rendered `.pdf`.

(b) Use the function from part (a) and uniform random variables to simulate $1,000$ draws from a truncated gamma distribution with shape 1 and rate 1 constrained to the interval $(0,1)$. Make a histogram of the draws using the `hist` function setting the arguments `freq = FALSE` and `br = seq(0, 1, length.out = 10)`.

(c) Write a for loop that generates $1,000$ draws from a gamma distribution with shape 1 and rate 1. What proportion of these draws are less than 1?

(d) Make two histograms next to each other, one being the histogram you constructed in (b) and the other being a histogram of the draws from (c) that are less than 1, using the same arguments for both. Overlay a line for the density of a truncated gamma random variable with shape 1 and rate 1 constrained to be less than 1 on both plots.

(e) In at most one sentence, describe whether or not it seems that both methods produce draws from the same distribution.

(f) In at most one sentence, describe how the number of draws produced by the two methods compares. Is one more efficient than the other?

3. Type `data(mtcars)` to load in the `mtcars` dataset.

(a) How many observations (rows) are in this data and how many variables (columns)?

(b) Create a table using `kable` that summarizes the minimum, mean, maximum, and standard deviation for each of the following variables: `mpg`, `hp`, and `wt`. Give the columns descriptive names e.g. "Miles per Gallon" for `mpg`. You'll want to create a table with 4 rows and 3 columns.

(c) Make a plot of the car weights against their miles per gallon. Make sure that your plot is clearly annotated and readable.