ANLY-511 Homework 2 - Assigned on 09/08/17

9. (2 points) Suppose X has a Gamma distribution with shape parameter $r = 2.5$ and scale parameter $\rho = 5$. Use `R` to compute the following quantities: $Prob(X \leq 10)$, $Prob(X > 5$, $Prob(|X - 8| < 3$, and z such that $Prob(X < z) = .1$.

10. (2 points) Plot the cumulative distribution functions of a binomial distribution, $B(20, 1/3)$, and of a hypergeometric distribution with parameters $k = 20, M = 40$, $M + N = 120$ in the same figure. Use a staircase plot (type = 's'). These two distributions are supposed to be close. Can you confirm that?

11. (2 points) Plot the cumulative distribution function of a binomial distribution, $B(40, .3)$, and of a normal distribution, $N(12, 2.9)$ in the same figure, using a staircase plot. These two distributions are supposed be close. Is that approximately right? What are the main differences between the two distributions?

12. (2 points) A graphical technique for checking whether a sample has an approximate normal distribution is a "quantile-quantile" plot. The `R` command is `qqnorm(x)`, where x is the vector containing the sample values. If the plot is approximately a straight line, then this suggests that the sample comes from a normal distribution. Practice this by making `qqnorm` plots of samples of size 10, 20, 40, 100, 1000 from a standard normal distribution. How close to a straight line are the plots in each case? Where in the plot are the main deviations from being a straight line?

13. (2 points) Refer to the `qqnorm` plot as explained in the previous problem. Make `qqnorm` plots of random samples from a $B(40, .3)$ distribution, for sample sizes between 20 and 1000. How their plots differ from straight lines? Can you explain why?

14. (5 points) Exercise 3.1 in Dalgaard.

15. (5 points) If X has a continuous distribution with cumulative distribution function F, then the new random variable $U = F(X)$ has a uniform $U(0, 1)$ distribution. Verify this with simulations for three different continuous distributions of your choice, by making a random sample of sufficient size, sorting it, plugging it into the cdf F, and plotting the result. Then prove it using algebra for the case where X has an exponential distribution with arbitrary parameter λ.

16. (5 points) Suppose $X = X_1 + X_2$ is the sum of two exponentially distributed random variables with the same parameter λ. Then X^α is very nearly normally distributed for a suitable choice of α. Determine an approximate value for α, using a simulation and `qqnorm` plots.