ANLY-511 Homework Problems

Explain your work and give concise reasoning. Attach `R` *code with comments if applicable. Using Markdown is the best way to d this. Do not print out any data or any detailed results of simulations. Your solutions for each homework set should fit on six to eight pages.*

1. (2 points) Let a be the 10th digit to the right of the decimal point of $\sin 1.23$. Let $b = \sqrt{a^2 + a^3}$. Let c be the number of digits to the left of the decimal point of $e^{(\ln b)^3}$. Let $d = \sum_{j=1}^{c} j^3$. Let $e = d \mod 77$ Compute a, b, c, d, e. Use `R` wherever it is appropriate, or read off the results from numerical output.

2. (2 points) Consider the random variable defined by counting the number of tries until the first success, for independent trials with success probability p. Given p between 0 and 1, the `R` commands `myattempts(p)` and `1 + rgeom(1,p)` both simulate this random variable. Compare the times for using these functions, for at least five different values of p, including values close to 0 and close to 1. Summarize and explain what you see.

3. (2 points) This is a continuation of the previous problem.
The standard deviation of a geometric distribution is known to equal $\frac{\sqrt{1-p}}{p}$. Verify this for at least five different values of p, with suitable simulations. The `R` function for computing the standard deviation of a vector is `sd()`.

4. (2 points) Problem 1.5 in ch. 1 of **Dalgaard**.

5. (2 points) In the statistical computing community, there is an ongoing debate on the comparison of `R` or SAS. Much of this debate is happening on the Internet. Look up some arguments for and against both SAS and `R` and summarize them in no more than half a page.

6. (5 points) Problem 6 in ch. 1 of **Chihara / Hesterberg**. *Read the chapter and review basic probability. Use* `R` *to calculate all probabilities. For part b, first compute the probability that you will not be in any single sample. For part c, use trial and error. It is sufficient to give the answer as a multiple of 100,000.*

7. (5 points) For a random variable X with a geometric distribution (as in problem 2 and 3) it is known that the expected value of its square is

$$\mathcal{E}(X^2) = \frac{2-p}{p^2}\,.$$

(You don't have to show this). This equation can be solved for p, and the result is

$$p = \frac{\sqrt{1 + 8\mathcal{E}(X^2)} - 1}{2\mathcal{E}(X^2)}$$

(You don't have to show this either). Since we can use data to estimate $\mathcal{E}(X^2)$, this suggests another plug-in estimator for p, namely

$$\hat{p} = \frac{\sqrt{1 + 8\overline{x^2}} - 1}{2\overline{x^2}}$$

where $\overline{x^2}$ is the average of the squares of the observations.
Use simulations to assess its bias, if $p = .1$, $p = .3$, $p = .7$ and you are given samples

of size $n = 4$ from a geometric distribution. Include commented `R` code in your solution.

8. (5 points) Consider the following random experiment: draw a uniformly distributed random number X_1 from the interval $(0, 1)$. Next, draw a uniformly distributed random number X_2 from the interval $(0, 2X_1)$, a uniformly distributed random number X_3 from the interval $(0, 2X_2)$ and so on until X_{10}. What is the expected value of X_{10}? Use a simulation to give an approximate answer. Report your commented `R` code. *The `R` command for drawing a uniformly distributed random number from the interval* $(0, b)$ *is* `runif(1,min = 0, max = b)`.