STAT 513/413, ASSIGNMENT 3

1. Prove theoretically the identity we observed on pages 26 and 27 of Lecture 12: if $Y_i, Y_2, \ldots, Y_N$ are random variables attaining only values 0 and 1, and $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ is their mean, then

$$\bar{Y}(1-\bar{Y}) = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 .$$

2. Let U is a random variable with uniform distribution on $[0,1]$.

(a) Show that U and $1-U$ have the same distribution. (How is this done? Well, there are several ways: one of them is to show that they have *the same* cumulative distribution function.)

(b) Calculate $\mathrm{Var}(e^U)$. What is $\mathrm{Var}(e^{1-U})$, in view of (a)?

(c) Calculate $\mathrm{Cov}(e^U, e^{1-U})$.

(d) Calculate $\mathrm{Var}(e^U + e^{1-U})$.

(e) On the basis of your results above, evaluate numerically (no R code required, just the result)

$$\frac{\mathrm{Var}(e^U + e^{1-U})}{\mathrm{Var}(e^U) + \mathrm{Var}(e^{1-U})} .$$

3. This is the `s12` data vector used in Lecture 11. The most expedient way to get it into R is just to copy, paste, and execute the following command

```
> s12 <- c(3.755103,11.78924,4.129652,0.9881743,1.172208,1.113155,7.531846,
+ 4.969411,0.6259583,2.388653,1.957482,1.932437,2.404424,0.980145,2.445655,
+ 4.915729,0.7597842,0.5466244,0.7560977,1.544395)
```

In Lecture 11, we looked at the normal qqplot to assess visually whether the vector is likely to look as one following normal distribution. In fact, there is a statistical test for that (one of those), based on the quantity called *skewness*, defined for any vector $x_1, x_2, \ldots, x_n$ to be

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}} , \qquad \text{where} \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

(a) Implement the function for the skewness as given above (no packages!) and compute the *skewness* of `s12`.

(b) Using the same function, find out whether 0 lies within 3 standard deviations of the just computed value – that is, whether 0 is contained in $(skewness - 3sd, skewness + 3sd)$. To this end, you need to figure out first what goes in for *sd*: it is the standard deviation of the skewness *as computed from the particular data vector s12*; to get its numerical value, the likely best way is to estimate it using one of the methods explained in Lecture 11. (Do I need

to tell you which one it is? The ability to decide that is the most important thing you may learn in this course!)

Once done, list all code: the implemented function as well as the (cleaned and commented!) transcript of any session obtaining the result. (And do not forget to initialize the random number generator by your personal number.)

4. We are going to compute the integral

$$\int_a^b \mathrm{e}^{-x^2}\,dx$$

by the *simple Monte Carlo method*, as defined in Lecture 13, using the uniform distribution over $[a, b]$.

(a) Figure out via the Hoeffding inequality (in general, mathematically), the minimal N you need to achieve precision to 2 decimal digits (that is, 2 digits *after* decimal point), with probability 0.95 or better. Compute (numerically) the resulting N for the following special cases: (i) $a = -4, b = 4$, (ii) $a = -2, b = 0$, (iii) $a = 0, b = 1$.

We are going to try it now; however, to also learn something from it, we first do the following.

(b) Use the rectangular method as described in Lecture 13 (page 15 or so) to calculate the "true" values of the integral for the special cases (i)–(iii) given in (a). You are free to use the implementation found there or develop your own – only make sure that you modify the constants appropriately. Regarding K, the number of intervals in the subdivision, find it experimentally so that it appears to give you at least *first three* "correct" digits.

(c) And now, implement the simple Monte Carlo for the given integrand, as outlined above, and run it on the special cases (i)–(iii) using the minimal N you derived in (a). Compare with the results you obtained in (b), and report whether the objective of 2 decimal digits was met.

Again, list all code you used, rectangular and Monte Carlo methods, and how they produced the results; do not, however, report on all your experimentation in (b), just show the final results. (Yeah, and use your personal number to initialize the random number generator, as always.)

Due Wednesday, February 24, 2020