1. Let A be a 2×2 random Gaussian matrix, that is, the entries are iid samples from $N(0,1)$. Define

$$M = \frac{A + A^T}{2}$$

Let λ_1, λ_2 be the two eigenvalues of M. Define the distance between the two eigenvalues as

$$D = |\lambda_1 - \lambda_2|$$

a. Prove that the pdf $f_D(d)$ of D is given by

$f_D(d) = \frac{d}{2} e^{-\frac{d^2}{4}} I_{d>0}$, where I is an indicator function.

b. Obtain the quantile function of D.

c. Simulate 1000 iid replicates of M. For each replicate compute $(\lambda_{1,b}, \lambda_{2,b}), b = 1,2, \ldots, 1000$ and hence obtain $D_b, b = 1,2, \ldots, 1000$. Obtain the empirical cdf of D_b. Does there exist sufficient evidence to suggest that the observed distribution of D_b agrees with the theoretical distribution derived in part a? Justify.

d. Now construct another 2×2 random matrix, A^* whose three elements are $N(0,1)$ and the fourth one is from $N(0,10^2)$. All entries are mutually independent. Generate 1000 replicates of A^* and the corresponding $M_b^*, D_b^*, b = 1,2, \ldots, 1000$. Obtain the empirical cdf of D_b^*. Is there a statistically significant difference between the observed distribution of D_b^* as compared to the theoretical one derived in part a?

2. A graduate student working in poultry science comes to you for statistical advice in designing her experiment. She wants to compare two types of housing systems on the occurrence of keel bone fractures in laying hens. The two types of housing systems are a colony system (60 birds/cage) and open floor system (200 birds/floor pen). The hens will be sacrificed and their keel bones will be examined for fractures after 52 weeks in their assigned cage system. The concern is that the proposed open floor system may increase the incidence of keel bone fractures. The graduate student wants to determine the number of reps she needs in order to detect at least a 5% difference in the percentage of hens that experience keel bone fractures between the colony system and the open floor system. From looking at the literature, approximately 60% of hens would be expected to have keel bone fracture by 26 weeks in colony system. She would like to have at least 80% power to detect a difference and a Type I error rate of 0.05.

a. For the first meeting, which is at her favorite coffee shop, you do not have access to a computer, only a statistics book and a calculator, so you need to figure out a ballpark estimate of the sample size that she would need for each treatment in order to have the power she desires. Show all formulas and calculations that you use. Also discuss the distributional assumptions that you have to make using this formula and what are some possible problems with using this formula.

b. The graduate student comes back to you for a second meeting at your office, where you have access to a computer and SAS. She has found out that she is going to be limited in the number of experimental units that she will have available.

    a. She has access to four commercial egg producers. Each producer has three of each housing system available for her to use. They are also willing to let her use these units for more than one year. She would like to know if it is possible to complete her experiment in one year. You need to determine estimates for the variances for both the farms and the years. Because there are no literature estimates for these sources of variability, you need to use a “six-sigma” approximation”. From your conversation with her you are given the following:

        i. Highest plausible probability for keel bone fractures for a farm – 0.80
        ii. Lowest plausible probability for keel bone fractures for a farm – 0.50
        iii. Highest plausible probability for keel bone fractures for a year - .75
        iv. Lowest plausible probability for keel bone fractures for a year - .50

    Using the “six-sigma” trick, determine the variances to use for farm and year.

    b. She wants to know the number of experimental units she needs per housing system if she expects the following proportions of hens per experimental unit to experience keel bone fractures (0.60 for colony system, 0.65 for open floor system). Again, she would like a power of .80 and a Type I error rate of 0.05. Make sure you include your SAS program and output.

    c. Write a short report that includes the following: (Remember, she is not a statistician, so write it as clearly as possible.)

        i. One paragraph for the researcher about the distributional assumptions necessary for the sample sizes calculated in Parts a and b (in other words, describe the distributions used).
        ii. One paragraph describing the “six sigma” approximation and how you use it when the data are binomially distributed.
        iii. One paragraph summarizing the results of the sample size calculations.