MATH 21-01 (Introductory Statistics), HW 10 (100 points). Due: 11/09/2016 in class.

Not from textbook (100 pts)

Please show your work and state the reasoning behind your computations for each of the following problems.

- (A - 25 pts) Suppose a given population is very large and has a standard deviation $\sigma = 20$, but it's mean is not known. Utilize the CLT to find the probability that during estimation of the population mean using a sample mean from a sample of size $n = 100$ (with the sample being a simple random sample from the population), the error will be less than 3 (i.e. the sample mean $\bar{x}$ will be within 3 units of the population mean μ, on either side).

- (B - 10 pts) Suppose a discrete random variable X has the following probability distribution:

$$P(X = 0) = 0.20, \quad P(X = 1) = 0.20, \quad P(X = 2) = 0.40, \quad P(X = 3) = 0.20$$

  Find the mean and standard deviation of X.

- (C - 25 pts) Suppose a vegetable company producing fresh spinach in bags reports the following probability distribution for dead bugs found per bag:

$$P(X = 0) = 0.80, \quad P(X = 1) = 0.12, \quad P(X = 2) = 0.05, \quad P(X = 3) = 0.03$$

  Where $P(x = k)$ refers to the probability of finding k dead bugs in one bag. Suppose 50 bags of the vegetables are bought. Using the CLT, what is the probability of finding more than 100 dead bugs amongst all the spinach in all the bags?

- (D - 20 pts) Suppose we toss 300 fair coins. Using R, find the probability that we get more than 160 heads and that we get less than 120 heads. How would you estimate this probability using a normal distribution and the poisson distribution? Compute the estimated probabilities in R, using the respective distributions. Please show your work and supply the R code you used.

- (E - 20 pts) Model the following problem using R. Please supply the R code you used. Suppose we take samples of size n from a large population distributed as the binomial distribution with success probability 0.3. Use 5000 trials (each trial consists of a sample of size n) and compute for each trial the sample mean and median of the samples. Produce histograms of the sample means and sample medians using samples of size $n = 5$, $n = 30$, $n = 500$ and compare to the population mean. What can you say about the shape of your histograms as n increases? Judge the quality of the estimators (sample mean and sample median) as estimators of the population mean in terms of bias and standard deviation.