1. More Bounding

Imagine that X is the number of customers that enter a bank at a given hour. To simplify everything, in order to serve n customers you need at least n tellers. One less teller and you won't finish serving all of the customers by the end of the hour. You are the manager of the bank and you need to decide how many tellers there should be in your bank so that you finish serving all of the customers in time. You need to be sure that you finish in time with probability at least 95%.

a. Assume that from historical data you have found out that $\mathbf{E}[X] = 5$. How many tellers should you have?

b. Now assume that you have also found out that $\text{Var}(X) = 5$. Now how many tellers do you need?

2. How Many Coupons?

Consider the coupon collecting problem covered in note 19. There are n distinct types of coupons that we wish to collect. Every time we buy a box, there is one coupon in it, with equal likelihood of being any one of the types of coupons. We want to figure out how many boxes we need to buy in order to get one of each coupon. For this problem, we want to bound the probability that we have to buy lots of coupons — say substantially more than $n \ln n$ coupons.

a. We represent X, the number of boxes we have to buy, as a sum of other random variables. Let X_i represent the number of boxes you buy to go from $i-1$ to i distinct coupons in your hand. The let $X = \sum_{i=1}^{n} X_i$. Argue that each X_i is an independent random variable with a geometric distribution.

b. Prove that $\mathbb{E}[X] \approx n \ln n$. Remember that the expectation of $\text{Geom}(p)$ is $\frac{1}{p}$.

c. We wish to use Chebyshev's inequality to bound the probability we have to buy substantially more than $\mathbb{E}[X] = n \ln n$ boxes. In order to do this, we need to compute the variance of a geometric random variable. We know from lecture that the variance of $\text{Geom}(p)$ is $\frac{1-p}{p^2} \leq \frac{1}{p^2}$. Prove that $Var[X] \leq 2n^2$. (Note that $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$)

d. This means that the standard deviation for X scales like n and not like the expectation $n \ln n$.
Use Chebyshev's inequality to show that $Pr[X \geq \alpha n \ln n]$ tends to zero for any $\alpha > 1$ as $n \to \infty$.

3. Poission

a. A textbook has on average one misprint per page. You may assume that misprints are "rare events" that obey the Poisson distribution. What is the chance that you see exactly 4 misprints on page 1?

b. Suppose the box has 6 brown balls and 4,000 purple balls. A random sample of size n is selected with replacement and X= "number of brown balls selected". Write the distribution of X. Can this be closely approximated with a Poisson distribution? If so, write the approximate distribution. If not, explain why not. If so, but only under certain conditions, explain these conditions and write the approximate distribution.