CS 70	Discrete Mathematics and Probability Theory	
Summer 2016	Dinh, Psomas, and Ye	Discussion 5D

1. **Chebyshev's Inequality**

   (a) Derive the Chebyshev's Inequality $\Pr[|X-\mu| \geq \alpha] \leq \frac{\mathrm{Var}(X)}{\alpha^2}$ using Markov Inequality $\Pr[X \geq \alpha] \leq \frac{\mathbf{E}[X]}{\alpha}$.
   Refer to Lecture Note 14 Page 2 and 3.

   (b) Let $\mathscr{H}$ be a hat containing slips of paper with numbers written on each slip. Assume that the numbers on the slips of paper need not be distinct, further assume that every slip of paper is equally likely to be chosen. Prove that at least 75 percent of the elements in $\mathscr{H}$ must be within 2 standard deviations of the mean of X. What percentage of the elements in $\mathscr{H}$ must be within 3 standard deviations of the mean of X? What about 4? How do these results justify our use of variance as a measure of spread of a distribution?

   Plug in 2 SD's as α to Chebyshev's inequality. Argument extends without loss of generality.Conceptual Answer: Because we know that given both the mean and the SD of numbers, we have a pretty good sense of where the numbers are.

2. **Tellers**

   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?

   You should apply Markov's. You get $\Pr[X \geq 100] \leq E[X]/100 = 0.05$. So you need 100 tellers.

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

   You should apply Chebyshev's. You get $\Pr[X > 5+10] \leq 5/(10^2) = 0.05$. So you need 15 tellers this time.

3. **Homework Polling**

   Suppose Alex wants to poll the CS 70 students about whether the homeworks have been too hard recently. Suppose everyone is comfortable enough to answer honestly, either no (0) or yes (1). Let the true fraction of students who think the homework is too hard be p. Let the response of the i^{th} polled student be X_i. Alex would like to poll n students, with n large enough that the probability of estimating p to within $\pm 2\%$ is at least 90%.

   a) Let $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the average of the n responses. What is the expectation of M_n?
   What is the event that we are interested in whose probability we would like to be at least 90%? Draw a picture of the distribution of $M_n - p$ and mark the region that corresponds to the event of interest.

$$X_i = \begin{cases} 1 & \text{hw is too hard} \\ 0 & \text{hw is not too hard} \end{cases}$$

X_i is i.i.d Bernoulli r.v. with $\Pr[X_i = 1] = p$. So, $M_n \sim \frac{1}{n} Binom(n, p)$:

$$E[M_n] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \cdot (np) = p$$

We want

$$\Pr[|M_n - p| < 0.02] > 0.9$$

Or equivalently,

$$\Pr[|M_n - p| \geq 0.02] \leq 0.1$$

b) Now use Chebyshev's inequality to find a safe n regardless of what p is.

$$Var[M_n] = \frac{1}{n^2} \sum_{i=1}^{n} Var[X_i] = \frac{p(1-p)}{n}$$

$$\Pr[|M_n - p|] \geq 0.02] \leq \frac{Var[M_n]}{0.02^2} = \frac{p(1-p)}{(0.02)^2 n} \leq 0.1$$

$$n \geq 25000 p(1-p) \leq 6250$$

Here, $p(1-p) \leq 1/4$ when equality holds of $p = 1/2$.

c) What if instead of wanting an accuracy of $\pm 2\%$ we wanted a relative error of 2%. This means that if the true value was p, we want the answer we get to be within $[0.98p, 1.02p]$. Can we pull that off using Chebyshev's inequality? Do you think we could pull that off with a single universal choice of n that does not depend on (the unknown) p?

We want

$$\Pr[0.98p < M_n < 1.02p] > 0.9$$

$$\Pr[|M_n - p| \geq 0.02p] \leq 0.1$$

So, plugging in the $Var[M_n]$ in the Chebyshev's inequality, we have

$$\Pr[|M_n - p| \geq 0.02p] \leq \frac{Var[M_n]}{(0.02p)^2} = \frac{1-p}{0.02^2 np} \leq 0.1$$

Solving this, we get

$$n \geq 25000 \frac{1-p}{p}$$

It depends on p as there is no simple bounds on $\frac{1-p}{p} = 1/p - 1$.