EECS 70 Discrete Mathematics and Probability Theory
Spring 2014 Anant Sahai

Discussion 13B

1. **Warmup**

   a) Calculate the variance of a random variable, X, where X represents the value of a standard 6-sided dice.

   b) Calculate the variance of a random variable defined by the binomial distribution.

2. **Homework Polling - Revisited**

   Suppose Prof. Sahai wants to poll the CS 70 students *again* about whether the homeworks have been too hard recently. But now 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. Prof. Sahai 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.

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

   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?

3. **How Many Coupons**

Consider the coupon collecting problem covered in note 17. 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 $n \log n$ 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. Write X as a sum of X_i's. Argue that each X_i is an independent random variable with a geometric distribution.

b) We wish to use Chebyshev's inequality to bound the probability we have to buy substantially more than $n \ln n$ boxes. In order to do this, we need to compute the variance of a geometric random variable. There are multiple ways of doing this, including a recursive trick like that used by Lily's Lottery.

Another approach is to use series techniques. We use the following lemma in our proof: $\sum_{k=1}^{\infty} k(k+1)(1-p)^{k-1} = \frac{2}{p^3}$. Prove this lemma. (Hint: what is the sum of the geometric series $\sum_{k=0}^{\infty}(1-p)^k$. Take the derivative of both sides, what happens?)

c) If $X_i \sim Geom(p)$, show that $\mathbb{E}[X_i^2] = \sum_{k=1}^{\infty} k(k+1)p(1-p)^{k-1} - \sum_{k=1}^{\infty} kp(1-p)^{k-1}$

d) Use your lemma and the fact that $\mathbb{E}[X_i] = 1/p$ to simplify part c) to show $\mathbb{E}[X_i^2] = \frac{2}{p^2} - \frac{1}{p}$

e) Show that variance of a geometric variable with parameter p is $\frac{1-p}{p^2}$. We will later use the simpler upper bound, $Var[X_i] < \frac{1}{p^2}$.

f) Make use of the fact that $\sum_{i=1}^{\infty} \frac{1}{i^2}$ is a positive constant $\frac{\pi^2}{6} \leq 2$ to show that the $Var[X] \leq 2n^2$

g) 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 $n \to \infty$. (Hint: Recall that we estimated in the note that $\mathbb{E}[X] \approx n(\ln n + \gamma) \approx n \ln n$.)