Discussion 13W & 14M

1. Vegas

On the planet Vegas, everyone carries a coin. Many people are honest and carry a fair coin (heads on one side and tails on the other), but a fraction p of them cheat and carry a trick coin with heads on both sides. You want to estimate p with the following experiment: you pick a random sample of n people and ask each one to flip his or her coin. Assume that each person is independently likely to carry a fair or a trick coin.

(a) Given the results of your experiment, how should you estimate p?

(b) How many people do you need to ask to be 95% sure that your answer is off by at most 0.05?

2. 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 $n \ln n$ 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. 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.

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 $\mathbb{E}[X] \approx n(\ln n + \gamma) \approx n \ln n$.)

3. Chernoff vs. Chebyshev

Here we want to compare Chernoff's bound and the bound you can get from Chebyshev's inequality.

(a) Consider the experiment of flipping a fair coin 100 times. Let S be the number of heads you obtain in your experiment. The exact probability of $S \geq 80$ is $5.5795 \cdot 10^{-10}$. What estimate does Chebyshev's bound give for the probability of seeing at least 80 heads in 100 coin flips? What about Chernoff's bound? Compare your answers and see which one is closer to the actual value.

(b) Now back to the setting with general n and α, write down Chernoff's bound in the form c^n, where c is an expression that only contains α and not n.

(c) Show that Chebyshev's bound on $\Pr[|S-\mu| \geq \alpha\mu]$ is also a bound on $\Pr[S \geq (1+\alpha)\mu]$.

(d) Using the previous part above, write a bound on $\Pr[S \geq (1+\alpha)\mu]$ of the form γn^{β}, where γ and β are numbers that do not depend on n. What does this tell you about Chebyshev's inequality vs. Chernoff's inequality?

4. Shading areas in probability

Let's say you toss a coin 1000 times, and you calculate the fraction of heads. The central limit theorem tells you that the distribution of the fraction of heads looks approximately like that of a Gaussian random variable. In simpler terms, the shape of the distribution will look like a bell centered at 0.5, and rapidly falling on either side.

(a) Sketch the bell curve above in free hand. You don't have to be too accurate; just get the general shape right.

(b) Shade the area under the curve that represents the outcome of getting more than 800 heads.

(c) Shade the area under the curve that represents the outcome of getting less than 200 heads.

(d) Shade the area under the curve that represents the outcome of getting no more than 50 heads away from the average.

(e) Shade the area under the curve that represents the outcome of getting more than 200 heads away from the average.