CS 70 Discrete Mathematics and Probability Theory
Fall 2015 Jean Walrand HW 13

1. Statistical hypothesis testing
   On one of the Mythbusters episodes[1], the Mythbusters decided to run an experiment to test whether toast tends to land buttered side down.

   At the beginning of the episode, Adam and Jamie built a first attempt at a mechanical rig to drop toast in a controlled fashion. When they tested it on 10 unbuttered pieces of toast as a sanity check, 7 pieces fell upside down and 3 pieces fell right-side up. Adam concluded based upon these numbers that this first rig was obviously biased, so he threw it away in disgust and they built a new rig. Was Adam right, or is this just another case where he jumps to conclusions too quickly?

   Let p denote the probability that, if we drop 10 pieces of unbuttered toast from an unbiased rig (i.e., a rig where each unbuttered piece of toast has a 50% chance of falling upside down and a 50% chance of falling right-side up), 7 or more of the pieces of toast land the same way. In other words, p is the probability of the event that at least 7 pieces land right-side up, or at least 7 pieces land upside down, when dropping from an unbiased rig.

   (a) As a warmup, compute the exact probability that if we flip a fair coin 10 times, we see 0, 1, 2, 3, 7, 8, 9, or 10 heads.

   (b) Now, back to the Mythbusters. With p defined as above, calculate p exactly.

   (c) Use p to decide whether the rig appears biased, using the following rules:

   - If $p > 0.05$, conclude that we cannot rule out the possibility that the rig is unbiased. The rig might be perfectly good as it is.
     (The intuition is: Oh man, that totally could've happened by chance.)
   - If $p \leq 0.05$, with 95% confidence we can conclude that the rig appears to be biased.
     (Sure, it's possible that this rule could lead us astray. Even if our calculations show $p \leq 0.05$, it's in principle *possible* that the rig is unbiased and the observations were just a big coincidence. However, this would require assuming that an event of probability 0.05 or less happened, which is by definition pretty rare. Put another way, if we conclude that the rig is biased whenever $p \leq 0.05$, then we'll wrongly throw away a perfectly good rig at most 5% of the time. This seems good enough.)

   To put it another way, this decision rule gives us a way to test the hypothesis that the rig is unbiased: if $p \leq 0.05$, we reject the hypothesis (with 95% confidence), otherwise if $p > 0.05$ we are unable to reject it (at 95% confidence level).

   Using your value of p and this decision rule, decide whether Adam was right to conclude that his first rig was biased, or whether he jumped to conclusions too quickly.

[1] Season 3, episode 4, air date: March 9, 2005.

2. Messing with Chernoff

One night, word spreads in the 2nd floor labs that there are 4 dozen free donuts upstairs. The 25 people in 271 are working on a looming midnight 61C deadline, and are each 30% likely to stay put and ignore the donuts. The 40 people in the other labs aren't as stressed, but are each 10% likely to have already filled up on pizza (per midterm 2). Assume that everyone makes their decision independently.

(a) A Chernoff bound for binomial variables that one can derive from the lecture states that for the sum of independent indicator variables with expectation μ and $\alpha \geq 1$ that $Pr[X \geq \alpha\mu] \leq e^{\alpha\mu-\mu-\alpha\mu\ln\alpha}$. Use this fact to bound the probability that there'll be enough donuts for everyone (assuming, unrealistically, that no one takes seconds).

(b) Part a uses a form of Chernoff bound derived by applying the Markov bound to $\alpha^{X_1+\ldots+X_n}$. What happens in this problem if you use the same procedure but start by applying the Markov bound to $2^{X_1+\ldots+X_n}$? Hint: You should use the inequality $e^x \geq x+1$.

3. It Catches Up With You

Let $X_1,\ldots,X_n$ be independent Bernoulli random variables that each take value 1 with probability p and 0 with probability $1-p$. You have learned how to use Chebyshev's inequality to say things about the probability that the sum $S = X_1 + X_2 + \cdots + X_n$ deviates from its mean (pn). In this question you will derive another bound called Chernoff's inequality that is much stronger in most cases.

(a) As an example to help you understand the setting better, assume that X_i is the outcome of a coin flip (that is $X_i = 1$ if the coin flip results in heads and otherwise $X_i = 0$). Then $p = 1/2$ and S is the number of heads you observe. Assume that $n = 100$ is the number of coin flips. The expected number of heads you see is $pn = 50$. The exact probability that $S \geq 80$ is $5.5795 \cdot 10^{-10}$. Now using Chebyshev's inequality find an upper bound for this probability. Is your upper bound much larger than the value you computed?

(b) Back to the general setting, prove that if $f : \{0,1\} \to \mathbb{R}$ is any function, then $f(X_1),\ldots,f(X_n)$ are independent. Hint: write down the definition of independence. If f takes the same value at 0 and 1 then everything should be obvious. It remains to prove it in the case where $f(0) \neq f(1)$.

(c) Now if we fix a number t and let $f(x) = e^{tx}$, then $f(X_i) = e^{tX_i}$. Compute the expected value of $f(X_i) = e^{tX_i}$ and write it in terms of p and t.

(d) The following is a famous inequality about real numbers: $1 + x \leq e^x$. Another variant of the inequality (which can be derived by replacing x by $x-1$) is the following: $x \leq e^{x-1}$. Apply the latter inequality with x being the expected value you computed in the previous step in order to get an upper bound on $E[f(X_i)]$. (You don't need to prove either of these inequalities.)

(e) Remembering that $f(X_1),\ldots,f(X_n)$ are all independent what is $E[f(X_1)f(X_2)\ldots f(X_n)]$ in terms of $E[f(X_1)],\ldots,E[f(X_n)]$? Use the upper bound you got from the previous step to get an upper bound on $E[f(X_1)f(X_2)\ldots f(X_n)]$. You should be able to express your answer in terms of p, n, and t. Now let $\mu = pn$ be the expected value of S. Re-express your upper bound in terms of μ and t (i.e. remove the occurrences of p and n and rewrite them in terms of μ).

(f) Observe that $f(X_1)\ldots f(X_n) = e^{t(X_1+\cdots+X_n)} = e^{tS}$. Let us call e^{tS} the random variable Y. Does it always take positive values? Let's say we are interested in bounding the probability that $S \geq (1+\alpha)\mu$ where α is a non-negative number. Prove that $S \geq (1+\alpha)\mu$ is the same event as $Y \geq e^{t\mu(1+\alpha)}$. Use Markov's inequality on the latter event to derive an upper bound for $\Pr[S \geq (1+\alpha)\mu]$ in terms of μ, t, and α.

(g) For different values of t you get different upper bounds for the probability that $S \geq (1+\alpha)\mu$. But of course all of them are giving you an upper bound on the same quantity. Therefore it is wiser to pick a t that minimizes the upper bound. This way you get the tightest upper bound you can using this method. Assuming that α is fixed, find the value t that minimizes your upper bound. For this value of t what is the actual upper bound? Your answer should only depend on α and μ. Hint: in order to minimize a positive expression you can instead minimize its ln. Then you can use familiar methods from calculus in order to minimize the expression.

(h) Here we want to compare Chernoff's bound and the bound you can get from Chebyshev's inequality. Assume for simplicity that $p = 1/2$, so $\mu = n/2$.

First compute Chernoff's bound for the probability of seeing at least 80 heads in 100 coin flips (the quantity you bounded in the first part). Compare your answer to that part and see which one is closer to the actual value.

Now back to the setting with general n and α, write down the Chernoff bound as c^n where c is an expression that only contains α and not n. This shows that for a fixed value of α, Chernoff's bound decays exponentially in n. Now write down Chebyshev's inequality to bound $\Pr[|S-\mu| \geq \alpha\mu]$. Show that this is also a bound on $\Pr[S \geq (1+\alpha)\mu]$. Write down this bound as γn^{β} where γ and β are some numbers that do not depend on n. This shows that Chebyshev's inequality decays like n^{β}. In general an exponential decay (which you get from Chernoff's) is much faster than a polynomial decay (the one you get from Chebyshev's).

4. Disguise it

Collecting statistics about sensitive issues (such as the percentage of the population that have a certain STD, etc.) is always a challenge.

Such a situation can arise in 70 if the professor wants to ask people if the homeworks have been too hard recently. People who respond to such questions might be more comfortable answering the truth if the polling mechanism gives them plausible deniability. Suppose that you want to ask a Yes/No question. You ask people to first roll a dice (on their own). If the result is 6 they should report the true answer, but otherwise you ask them to flip a coin and based on that randomly answer Yes/No. The dice roll is kept secret and not revealed to the professor.

(a) First, let's consider the system without the dice-rolling part nor the coin-tossing part. Suppose that exactly $\frac{5}{6}$ of the students are told to give a canned answer and exactly half of the canned answers are Yes and half of the canned answers are No. The remaining $\frac{1}{6}$ of the students sampled give the true answer. Further assume that these $\frac{1}{6}$ students have exactly the same proportion of YES/NO as the whole student population.

The professor doesn't know which students were canned and which were giving real answers.

Suppose that a fraction q of the answers you get are Yes. What fraction p of the population should you assume would answer the original question with a Yes (assuming sensitivity was not an issue)? Express this as a formula in terms of q.

(b) Now, suppose only that coin toss has been introduced back into the problem, but everything else is as before. Exactly $\frac{5}{6}$ of the students toss a coin for their answers while the remaining $\frac{1}{6}$ of students answer honestly. Furthermore, this subset of students has exactly the same proportion of YES/NO as the entire population.

Argue using the law of large numbers that as the number of people asked goes to infinity the formula from the previous part approaches the true fraction with confidence approaching 1.

Use your calculations/simulations to say how big of a class must it be so that we believe that we will get q correct to within ± 0.1 with a confidence of 95%?

(c) Now consider the original scheme with the dice and the coin. Argue using the law of large numbers that as the number of people asked goes to infinity the formula from the previous part approaches the true fraction with confidence approaching 1.

(We are not asking for proofs here because the Laws of Large Numbers are Empirical Facts for now. However, you should try to be precise in your argumentation. Later on in the course, we will be able to prove such statements.)

5. Probabilistically Buying Probability Books

Chuck will go shopping for probability books for K hours. Here, K is a random variable and is equally likely to be 1, 2, or 3. The number of books N that he buys is random and depends on how long he shops. We are told that

$$\Pr[N = n | K = k] = \frac{c}{k}, \qquad \text{for } n = 1, \ldots, k$$

for some constant c.

(a) Compute c.

(b) Find the joint distribution of K and N.

(c) Find the marginal distribution of N.

(d) Find the conditional distribution of K given that $N = 1$.

(e) We are now told that he bought at least 1 but no more than 2 books. Find the conditional mean and variance of K, given this piece of information.

(f) The cost of each book is a random variable with mean 3. What is the expectation of his total expenditure? *Hint:* Condition on events $N = 1, \ldots, N = 3$ and use the total expectation theorem.