EECS 70 Discrete Mathematics and Probability Theory
Spring 2013 Anant Sahai

HW 14

Due May 6

1. Do Anything Enough, and It Eventually Seems Normal

Suppose a standard 6-sided die [with faces 1 through 6] is rolled n times, and let A be the average of the results.

1. How does the Central Limit Theorem help us approximate the distribution of A?
2. Let A' be a random variable drawn from the Gaussian distribution best approximating the distribution of A. If $n = 100$, what are the bounds of an interval $[a,b]$ such that $A' \in [a,b]$ with probability exactly 90%?. [For any computations involving the normal distribution, you may use a computer or table as you like, so long as you spell out the manner in which you used them]
3. Supposing $n = 30$, compute the probability that $3 \leq A' \leq 4$.
4. What is the minimum n for which, with probability at least 99%, we have $3 \leq A' \leq 4$?

2. An Easy Problem

This problem is a continuation of the binary entropy problem on the previous homework.

Consider a distribution which takes on the value p with probability p and the value $1-p$ with probability $1-p$. Let X_1, X_2, ..., X_n be random variables drawn i.i.d. from this distribution.

1. Show that $P(X_1 = x_1, X_2 = x_2, \dots X_n = x_n) = x_1 x_2 x_3 \cdots x_n$. In other words, the probability of a particular string of values is just the product of those values.
2. Show that, for any $\varepsilon > 0$, the probability of "$|\log_2(\frac{\sqrt[n]{X_1 X_2 \dots X_n}}{p^p(1-p)^{1-p}})|$ is larger than ε" approaches zero as n grows large. [Hint: Remember the Law of Large Numbers. This looks complicated, but it's actually very simple.]
3. Now consider Y_i to be i.i.d. Bernoulli(p) random variables and let $\vec{Y}$ be the resulting random n-length binary string.
   Consider the set T_ε^n of n-length binary strings $\vec{y}$ for whom $2^{-n(H(p)+\varepsilon)} < P(\vec{Y} = \vec{y}) < 2^{-n(H(p)-\varepsilon)}$.
   Use the above to show that for every $\varepsilon > 0, 1 > \delta > 0$, there exists an n large enough so that the probability $P(\vec{Y} \in T_\varepsilon^n) > 1 - \delta$.
4. If you have a bag of marbles, and the total weight of the bag is at most 1, and each marble weighs at least $2^{-n(H(p)+\varepsilon)}$, what can you conclude about the number of marbles in the bag?
5. Interpret the above statement in the context of the set T_ε^n. See if you can see this as a law of large numbers corresponding to strings and their probabilities — that for a large enough n, almost all the probability lies in a "small" (well, smaller than 2^n anyway) set of strings that each have exponentially small probabilities that are "close" to each other. Connect this observation to the previous homework problem where you derived the binary entropy function $H(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$ and got it two different ways.

3. Two-to-One Function

1. Can you describe explicitly a function $f(x,y)$ which takes as input two natural numbers [i.e., integers ≥ 0], and produces as output one natural number, in such a way that $f(x_1,y_1) = f(x_2,y_2)$ only when $x_1 = x_2$ and $y_1 = y_2$? [In other words, an injection from $\mathbb{N}^2$ to $\mathbb{N}$]
2. Can you arrange for such an f to be not just an injection, but actually a bijection? Again, give an explicit definition. [This can be the same as in the last answer, if the answer you gave was already a bijection; otherwise, see if you can find a way to modify it a little to accomplish this goal]
3. Can you define such an f using only familiar arithmetic operations[1] (e.g., $f(x,y) = x^2 + y^x$ or $f(x,y) = xy + 2^{x^3 y}$, except those don't work)?

4. It Catches Up With You (carried over from HW13 but part 9 is new)

Let $X_1,\dots,X_n$ 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.

1. 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 $pn = 50$. Using a computer calculate the probability that $S \geq 80$ (note that since the probability is very small, you can't use simulations). Now using Chebyshev's inequality find an upper bound for this probability. Is your upper bound much larger than the value you computed?
2. Back to the general setting, prove that if $f : \{0,1\} \to \mathbb{R}$ is any function, then $f(X_1),\dots,f(X_n)$ 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 $f(0) \neq f(1)$.
3. 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.
4. The following is a famous inequality about real numbers: $1 + x \leq e^x$. Plot $1+x$ and e^x in the same figure and observe that the inequality holds. 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.)
5. Remembering that $f(X_1),\dots,f(X_n)$ are all independent what is $E[f(X_1)f(X_2)\dots f(X_n)]$ in terms of $E[f(X_1)],\dots,E[f(X_n)]$? Use the upper bound you got from the previous step to get an upper bound on $E[f(X_1)f(X_2)\dots f(X_n)]$. 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).
6. Observe that $f(X_1)\dots f(X_n) = e^{t(X_1+\cdots+X_n)} = e^{tS}$. 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.
7. 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

[1] Not that there's anything particularly wrong with non-arithmetic definitions; this is just for fun

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.

8. 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).

9. **The Plot Thickens:** We now have four different ways to approximate the probability that the average of n independent Bernoulli random variables (each with probability p of success) will differ from its mean value by more than some particular amount ε [call this probability ProbWander(n, p, ε)]: we can calculate it exactly, we can use the approximation (which is a bound) given by the Chebyshev bound, we can use the approximation (which is a bound) given by the Chernoff bound, or we can use the approximation (which is not necessarily a bound) given by the Central Limit Theorem. Pick a value for p and for ε and graph $\log(\text{ProbWander}(n, p, \varepsilon))$ against n, using each of the four different ways to approximate to ProbWander (all on the same graph). Do this again with another choice of p and ε. Describe how the four different approximation methods relate to each other [in particular, any inequalities which do or do not hold between them in general].

5. A Familiar Taste (carried over from HW13; only part 6 needs to be done)

In this question you will have used Stirling's approximation to study the binomial distribution. We flipped n independent coins (each having shown heads with probability $1/2$) and counted the number of times we saw heads. S denoted this random variable. We would haved liked to have studied the distribution of S.

1. You calculated the exact probability $\Pr[S = k]$ for each $0 \leq k \leq n$.
2. You used Stirling's approximation $m! \simeq \sqrt{2\pi m}(m/e)^m$ and replaced all of the factorials in the expression you had found for $\Pr[S = k]$. You simplified your expression so that you didn't have the constant e anymore.
3. Then you assumed that $k = tn$ where $0 \leq t \leq 1$ was a real number. You simplified the expression you had gotten in the previous step and wrote it in terms of t and n. You should have simplified enough that your expression looked like the following

   $$\frac{1}{A\sqrt{n}}B^n$$

   where A and B only depended on t and not n.
4. You plotted the expression you had for $0 \leq t \leq 1$ for these choices of n: 10, 20, 100. You used a log-scale plot, i.e. instead of having plotted $f(t)$, you plotted $\ln(f(t))$ where $f(t)$ was your approximation for $\Pr[S = tn]$. "Did these graphs all look somewhat similar?"ed.

5. You plotted the expression you obtaineded as a function of n for three fixed values of t, $t = 1/4$, $t = 1/3$, and $t = 1/2$ (you should have had three plots in the same figure). Your plot should have gone from $n = 1$ to $n = 100$. Again you plotted using log-scale (you used log-scale only for your approximation, i.e. the y-axis, not the x-axis).

At this point, apparently in a moment of confusion, you moved on to *Problem* 6 rather than Part 6. That was careless, and you will be penalized accordingly. However, you have a chance to rectify the situation now.

6. Note that $\Pr[S \geq tn] \geq \Pr[S = tn]$. For $t > 1/2$ the previous question (Chernoff's inequality) gives you an upper bound on $\Pr[S \geq tn]$. Compute this upper bound in terms of t and n. This upper bound should obviously be greater than $\Pr[S = tn]$. Plot both the approximation for $\Pr[S = tn]$ you got using Stirling's formula, and the upper bound you got using Chernoff's in the log-scale for three different values of n: 10, 20, 100. Let t vary in the interval $[0.5, 1]$ in each plot.

6. The Pun is the Lowest-Hanging Fruit of Ostensible Wit

Let A be the collection of halting computer programs in your favorite general-purpose programming language, and let B be the collection of non-halting programs in that same language. Answer the following questions for A and for B:

1. Is this collection finite, countably infinite, or uncountably infinite? How do you know?
2. Is it possible for there to be a computable surjection from the natural numbers onto this collection? If so, describe one; if not, why not?

7. Your Own Problem

Write your own problem related to this week's material and solve it. You may still work in groups to brainstorm problems, but each student should submit a unique problem. What is the problem? How to formulate it? How to solve it? What is the solution?

8. Your Own Problem/Intriguing Possibilities

CS70 homework is now over. Go live your life. Have fun. Pick something about the world which annoys you, and change it. Come up with an idea that excites you, seek out friends who see its potential, and work together to bring your imagination into reality. Take risks, fail, learn from your failures, succeed, then find new ways to fail. Acknowledge one skill you admire but lack, and acquire it; then another. Now replace "skill" with "experience", and do it again. Plan a road-trip, or join a start-up, or buy a guitar and a quality video camera. Whatever your thing is. Discover what you really like and pursue it, making the most of the limited time you get to spend above the ground.

...Oh, but take the final exam first. That should be your focus for now; you can always do that other stuff later, probably.