

Notes for Recitation 20

Problem 1. The following two parts are not related. Try them, to make sure you understand the jargon of random variables, distributions, probability density functions, etc. Ask your TA if you don't understand/remember what some phrase means.

- (a) Suppose $X_1, X_2,$ and X_3 are three mutually independent random variables, each having the uniform distribution

$$\Pr(X_i = k) \text{ equal to } 1/3 \text{ for each of } k = 1, 2, 3.$$

Let M be another random variable giving the maximum of these three random variables. What is the density function of M ?

Solution. This can be easily hashed out by counting the possible outcomes:

$$\begin{aligned} M \text{ is } 1 & \text{ with probability } \frac{1}{27} \\ & 2 \text{ with probability } \frac{7}{27} \\ & 3 \text{ with probability } \frac{19}{27} \end{aligned}$$

- (b) Suppose X, Y are two independent binomial random variables with parameters (n, p) and (m, p) , respectively. What is $\Pr(X + Y = k)$?

Solution. The pdf of X is the probability of tossing k Heads out of n independent flips of a coin with bias p . Likewise for Y and m flips. Since, X and Y are independent, the pdf of $X + Y$ corresponds to $n + m$ independent flips, i.e., $X + Y$ is a Binomial variable with parameters $(n + m, p)$. Hence,

$$\binom{m+n}{k} p^k (1-p)^{m+n-k}.$$

Problem 2. I'm God. Seriously. So, I know everything that everybody thinks. In particular, I know who each one of the 250,000,000 Americans want to vote for in the upcoming elections. I know that a fraction $p = 0.52$ of them want to vote for the current president.

You are mortal. An insignificant dot in space-time. But a quite significant dot among dots. You work close to the president and, within a week, you must answer his agonizing question: "Am I winning?" Or, in math jargon (but with the same agony): "Is $p > \frac{1}{2}$?"

Your *first* idea is to ask me (I'm God). But you haven't talked to me for a long time, so you know I won't tell you. Your *second* idea is to call every American, ask them, then divide the yes's by 250 million. But you soon realize there is not enough time (there is a reason for representative democracy). Your *third* idea... You have no third idea! In your panic as the week is almost over, you start picking Americans at random, call them, and ask!

Amazingly, that's the correct approach. But you should be careful what you are going to say to the president! Let's see.

(a) In your first phone call, you pick 1 American *uniformly at random*, call, and ask whether he/she will vote for the president. What is the probability that the answer is going to be "yes"... (i) from my perspective? (ii) from your perspective? How would you model this in terms of coin flips?

Solution. From my perspective, it's 0.52. From your perspective, it is also 0.52. The only problem is that you don't know that, so you just call it p . Clearly, from your perspective, the first phone call is the same as flipping a coin with an *unknown* bias, which you call p (and I know is 0.52).

(b) In your second phone call, you again pick 1 American *uniformly at random*, call, and ask whether he/she will vote for the president. But wait! When selecting the second voter, shouldn't you exclude the guy that you asked in the first phone call? No. What's bad if you exclude him/her?

Solution. If you do this, you alter the coin that you are flipping. The bias will decrease or increase, depending on whether the first guy said "yes" or "no", respectively. The analysis gets messy, so you don't want to do this.

(c) So, in each one of n phone calls, you pick 1 American *uniformly at random* and ask. Your plan is to eventually divide the number M of positive answers by n to get $P = \frac{M}{n}$. An MIT friend tells you that, as the numerical outcome of a random experiment, this P is a random variable; and that, according to his calculations,

$$\Pr(|P - p| \leq 0.03) \geq 0.95. \quad (1)$$

When you are done calling people, you divide to get P , and it's 0.53. You call the president up and... what do you say?

- Mr. President, $p = 0.53$!
- Mr. President, with probability at least 95%, p is within 0.03 of 0.53!

- Mr. President, either p is within 0.03 of 0.53 or something very strange (less than 5-in-100) has happened.

For each statement answer: (i) Are you justified to claim it? (ii) Is it true?

Solution. The first statement is clearly off the mark.

- (i) Since you haven't asked all Americans, you can only make probabilistic statements about p .
- (ii) The statement is also false, since $p = 0.52$. However, with a different choice of voters, it could have been true. Of course, even in that case, you wouldn't be justified to claim it.

The second statement is also wrong.

- (i) The *unknown constant* p is *either* within 0.03 of 0.53 *or* more than 0.03 away of 0.53. In the first case, the probability of it being as you claim is 1; in the second case, it is 0. The crucial point is that you don't know which case holds. You could make the above claim, only if you knew you were in the first case. Sadly, you don't.
- (ii) The claim is actually true in this case. Since $p = 0.52$, the unknown constant is indeed within 0.03 of 0.53. So the probability that you talk about is 1, and therefore at least 95%. But, as we said, it could be 0 and then the statement would be false.

The third statement is the correct one.

- (i) You are justified to claim this statement. To see why, start with the statement

$$\text{either } |0.53 - p| \leq 0.03 \quad \text{or} \quad |0.53 - p| > 0.03.$$

which is clearly true. Now read it as follows: *Either* p is within 0.03 of 0.53 *or* it is not and therefore my random random variable P took a value from a set that is hit only 5 times in 100. So, clearly *either* p is within 0.03 of 0.53 *or* something strange has happened.

- (ii) The statement is true. In the particular case, it is true because the first half of it is true.

Fact from lecture. Suppose a coin that comes up heads with probability p is flipped n times. Then for all $\alpha < p$

$$\Pr(\# \text{ heads} \leq \alpha n) \leq \frac{1 - \alpha}{1 - \alpha/p} \cdot \frac{2^{nH(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}} \cdot p^{\alpha n}(1-p)^{(1-\alpha)n}$$

where:

$$H(\alpha) = \alpha \log_2 \frac{1}{\alpha} + (1 - \alpha) \log_2 \frac{1}{1 - \alpha}$$

Problem 3. A coin that comes up heads with probability p is flipped n times. Find an upper bound on

$$\Pr(\# \text{ heads} \geq \beta n)$$

where $\beta > p$. Think about the number of tails and plug into the monster formula above.

Solution.

$$\Pr(\# \text{ heads} \geq \beta n) = \Pr(\# \text{ tails} \leq (1 - \beta)n)$$

Now tails comes up with probability $1 - p$. So the answer is the same as above with α replaced by $1 - \beta$ and p replaced by $1 - p$:

$$\Pr(\# \text{ heads} > \beta n) \leq \frac{\beta}{1 - \frac{1-\beta}{1-p}} \cdot \frac{2^{nH(\beta)}}{\sqrt{2\pi\beta(1-\beta)n}} \cdot p^{\beta n}(1-p)^{(1-\beta)n}$$

Here we're using the fact that $H(1 - \beta) = H(\beta)$.

Problem 4. A Gallup poll in November 2004 found that 35% of the adult population of the United States believes that the theory of evolution is “well-supported by the evidence”. Gallup polled 1016 people and claims a margin of error of 3 percentage points.

Let’s check Gallup’s claim. Suppose that there are m adult Americans, of whom pm believe evolution is well-supported and $(1 - p)m$ do not. Gallup polls n Americans selected uniformly and independently at random. Of these, qn believe that evolution is well-supported and $(1 - q)n$ do not. Gallup then estimates that the fraction of Americans who believe evolution is well-supported is q .

Note that the only randomization in this experiment is in who Gallup chooses to poll. So the sample space is all sequences of n adult Americans. The response of the i -th person polled is “yes” with probability p and “no” with probability $1 - p$ since the person is selected uniformly at random. Furthermore, the n responses are mutually independent.

- (a) Give an upper bound on the probability that the poll’s estimate will be 0.03 or more too low. Just write the expression; don’t evaluate yet!

Solution. We can regard each response as a coin flip that is heads with probability p . In these terms, qn is the total number of heads flipped. So we have:

$$\begin{aligned} & \Pr(qn \leq (p - 0.03)n) \\ & \leq \frac{1 - (p - 0.03)}{1 - (p - 0.03)/p} \cdot \frac{2^{nH(p-0.03)}}{\sqrt{2\pi(p-0.03)(1-(p-0.03))n}} \cdot p^{(p-0.03)n}(1-p)^{(1-(p-0.03))n} \end{aligned}$$

- (b) Give an upper bound on the probability that the poll’s estimate will be 0.03 or more too high. Again, just write the expression.

Solution. Reasoning as before and using the answer to the preceding problem gives:

$$\begin{aligned} & \Pr(qn > (p + 0.03)n) \\ & \leq \frac{p + 0.03}{1 - \frac{1-(p+0.03)}{1-p}} \cdot \frac{2^{nH(p+0.03)}}{\sqrt{2\pi(p+0.03)(1-(p+0.03))n}} \cdot p^{(p+0.03)n}(1-p)^{(1-(p+0.03))n} \end{aligned}$$

- (c) The sum of these two answers is the probability that Gallup’s poll will be off by 3 percentage points or more, one way or the other. Unfortunately, these expressions both depend on p — the unknown fraction of evolution-believers that Gallup is trying to estimate!

However, the sum of these two expressions is maximized when $p = 0.5$. So evaluate the sum with $p = 0.5$ and $n = 1016$ to upper bound the probability that Gallup’s error is 0.03 or more. Pollsters usually try to ensure that there is a 95% chance that the actual percentage p lies within the poll’s error range, which is $q \pm 0.03$ in this case. Is Gallup’s evolution poll properly designed?

Solution. The probability that the error is 0.03 or more is about 0.07, which means that p will lie within the error range of a polled fraction with probability 0.93. So our

estimates suggest Gallup's poll is not quite large enough to meet the claimed 0.95 probability. Since Gallup is a professional, we expect he's got the poll size right, by using a more accurate numerical estimation formula – or he may have considered it legitimate to round a very slightly larger margin of error down to 0.03.

(d) If we accept all of Gallup's polling data and calculations, can we conclude that there is a high probability that the number of adult Americans who believe evolution is well-supported by the facts is 35 ± 3 percent?

Solution. No. This is an question of fact, which is either true or false. We *can* say that either the statement above is true or else a 1-in-20 event occurred during the poll; specifically, Gallup chose an unrepresentative sample. This may convince you that p is "probably" in the range 0.35 ± 0.03 , but there is no way to convert that informal "probably" to a mathematical probability.