

1 Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of hw party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

2 Problems

1. Those 3407 Votes

In the aftermath of the 2000 US Presidential Election, many people have claimed that unusually large number of votes cast for Pat Buchanan in Palm Beach County are statistically highly significant, and thus of dubious validity. In this problem, we will examine this claim from a statistical viewpoint.

The total percentage votes cast for each presidential candidate in the entire state of Florida were as follows:

Gore	Bush	Buchanan	Nader	Browne	Others
48.8%	48.9%	0.3%	1.6%	0.3%	0.1%

In Palm Beach County, the actual votes cast (before the recounts began) were as follows:

Gore	Bush	Buchanan	Nader	Browne	Others	Total
268945	152846	3407	5564	743	781	432286

To model this situation probabilistically, we need to make some assumptions. Let's model the vote cast by each voter in Palm Beach County as a random variable X_i , where X_i takes on each of the six possible values (five candidates or "Others") with probabilities corresponding to the Florida percentages. (Thus, e.g., $\Pr[X_i = \text{Gore}] = 0.488$.) There are a total of $n = 432286$ voters, and their votes are assumed to be mutually independent. Let the r.v. B denote the total votes cast for Buchanan in Palm Beach County (i.e., the number of voters i for which $X_i = \text{Buchanan}$).

- Compute the expectation $\mathbf{E}[B]$ and the variance $\text{Var}(B)$.
- Use Chebyshev's inequality to compute an *upper bound* b on the probability that Buchanan receives at least 3407 votes, i.e., find a number b such that

$$\Pr[B \geq 3407] \leq b.$$

Based on this result, do you think Buchanan's vote is significant?

- Suppose that your bound b in part (b) is exactly accurate, i.e., assume that $\Pr[X \geq 3407]$ is exactly equal to b . [*In fact the true value of this probability is much smaller*] Suppose also that all 67 counties in Florida have the same number of voters as Palm Beach County, and that all behave independently according to the same statistical model as Palm Beach County. What is the probability that in *at least one* of the counties, Buchanan receives at least 3407 votes? How would this affect your judgment as to whether the Palm Beach tally is significant?
- Our model assumes that all voters behave like the fabled "swing voters," in the sense that they are undecided when they go to the polls and end up making a random decision. A more realistic model would assume that only a fraction (say, about 20%) of voters are in this category, the others having already decided. Suppose then that 80% of the voters

in Palm Beach County vote deterministically according to the state-wide proportions for Florida, while the remaining 20% behave randomly as described earlier. Does your bound b in part (b) increase, decrease or remain the same under this model? Justify your answer.

2. Statistical hypothesis testing

On one of the Mythbusters episodes¹, 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.

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

3. Law of Large Numbers

Recall that the *Law of Large Numbers* holds if, for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr\left[\left|\frac{1}{n}S_n - \mathbb{E}\left(\frac{1}{n}S_n\right)\right| > \epsilon\right] = 0.$$

In class, we saw that the Law of Large Numbers holds for $S_n = X_1 + \dots + X_n$, where the X_i 's are i.i.d. random variables. This problem explores if the Law of Large Numbers holds under other circumstances.

Packets are sent from a source to a destination node over the Internet. Each packet is sent on a certain route, and the routes are disjoint. Each route has a failure probability of p and different routes fail independently. If a route fails, all packets sent along that route are lost. You can assume that the routing protocol has no knowledge of which route fails.

For each of the following routing protocols, determine whether the Law of Large Numbers holds when S_n is defined as the total number of received packets out of n packets sent. Answer **Yes** if the Law of Large Number holds, or **No** if not, and give a brief justification of your answer. (Whenever convenient, you can assume that n is even.)

- (a) **Yes** or **No**: Each packet is sent on a completely different route.
- (b) **Yes** or **No**: The packets are split into $n/2$ pairs of packets. Each pair is sent together on its own route (i.e., different pairs are sent on different routes).
- (c) **Yes** or **No**: The packets are split into 2 groups of $n/2$ packets. All the packets in each group are sent on the same route, and the two groups are sent on different routes.
- (d) **Yes** or **No**: All the packets are sent on one route.

4. Erasures, Bounds, and Probabilities

Alice is sending 1000 bits to Bob. The probability that a bit gets erased is p , and the erasure of each bit is independent of the others.

Alice is using a scheme that can tolerate upto one-fifth of the bits being erased. That is, as long as Bob receives at least 801 of the 1000 bits correctly, he can decode Alice's message.

In other words, Bob becomes unable to decode Alice's message only if 200 or more bits are erased. We call this a "communication breakdown", and we want the probability of a communication breakdown to be at most 10^{-6} .

- (a) Use Markov's inequality to upper bound p such that the probability of a communications breakdown is at most 10^{-6} .
- (b) Use Chebyshev's inequality to upper bound p such that the probability of a communications breakdown is at most 10^{-6} .

5. Binomial CLT (Optional)

In this question we will explicitly see why the central limit theorem holds for the binomial distribution as the number of coin tosses grows.

Let X be the random variable showing the total number of heads in n independent coin tosses.

(a) Compute the mean and variance of X . Show that $\mu = E[X] = n/2$ and $\sigma^2 = \text{Var}[X] = n/4$.

(b) Prove that $\Pr[X = k] = \binom{n}{k} / 2^n$.

(c) Show by using Stirling's formula that $\Pr[X = k] \simeq \frac{1}{\sqrt{2\pi}} \left(\frac{n}{2k}\right)^k \left(\frac{n}{2(n-k)}\right)^{n-k} \sqrt{\frac{n}{k(n-k)}}$.

In general we expect $2k$ and $2(n-k)$ to be close to n for the probability to be non-negligible. When this happens we expect $\sqrt{\frac{n}{k(n-k)}}$ to be close to $\sqrt{\frac{n}{(n/2) \times (n/2)}} = 2/\sqrt{n}$. So replace that part of the formula by $2/\sqrt{n}$.

(d) In order to normalize X , we need to subtract the mean, and divide by the standard deviation. Let $Y = (X - \mu)/\sigma$ be the normalized version of X . Note that Y is a discrete random variable. Determine the set of values that Y can take. What is the distance d between two consecutive values?

(e) Let $X = k$ correspond to the event $Y = t$. Then $X \in [k - 0.5, k + 0.5]$ corresponds to $Y \in [t - d/2, t + d/2]$. For conceptual simplicity, it is reasonable to assume that the mass at point t is distributed uniformly on the interval $[t - d/2, t + d/2]$. We can capture this with the idea of a "probability density" and say that the probability density on this interval is just $\Pr[Y = t]/d = \Pr[X = k]/d$.

Compute k as a function of t . Then substitute that for k in the approximation you have from part 5c to find an approximation for $\Pr[Y = t]/d$. Show that the end result is equivalent to

$$\frac{1}{\sqrt{2\pi}} \left(\left(1 + \frac{t}{\sqrt{n}}\right)^{1 + \frac{t}{\sqrt{n}}} \left(1 - \frac{t}{\sqrt{n}}\right)^{1 - \frac{t}{\sqrt{n}}} \right)^{-n/2}$$

(f) As you can see, we have expressions of the form $(1+x)^{1+x}$ in our approximation. To simplify them, write $(1+x)^{1+x}$ as $\exp(\ln(1+x)(1+x))$ and then replace $\ln(1+x)(1+x)$ by its Taylor series.

The Taylor series up to the x^2 term is $\ln(1+x)(1+x) \simeq x + x^2/2 + \dots$ (feel free to verify this by hand). Use this to simplify the approximation from the last part. In the end you should get the familiar formula that appears inside the CLT:

$$\frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

(The CLT is essentially taking a sum with lots of tiny slices and approximating it by an integral of this function. Because the slices are tiny, dropping all the higher-order terms in the Taylor expansion is justified.)