CS 70 Discrete Mathematics and Probability Theory
Spring 2016 Walrand and Rao

HW 12

Due Thursday April 21at 10PM

Before you start your homework, write down your team. Who else did you work with on this homework? List names and student ID's. (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.

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.

   Solution: Let X be the number of flips that come up heads. Then $X \sim \text{Binomial}(10, 0.5)$.

$$\begin{aligned}\Pr[(0 \le X \le 3) \vee (7 \le X \le 10)] &= \binom{10}{0}0.5^{10} + \binom{10}{1}0.5^{10} + \binom{10}{2}0.5^{10} + \binom{10}{3}0.5^{10} + \\ &\quad \binom{10}{7}0.5^{10} + \binom{10}{8}0.5^{10} + \binom{10}{9}0.5^{10} + \binom{10}{10}0.5^{10} \\ &= \frac{11}{32} \\ &= 0.34375.\end{aligned}$$

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

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

Solution: p is the probability that the number of toast pieces landing right-side up is between 0 and 3 or between 7 and 10, all ranges inclusive. If we think of the toast as a coin and the outcome of right-side up as Heads, then p is exactly the probability computed in part (a). Hence, $p = 0.34375$.

(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.

Solution: In part (b), we found that $p = 0.34375$. This says that the outcome of 7 right-side up and 3 right-side down pieces of toast, or 7 right-side down and 3 right-side up pieces of toast could be produced by an unbiased rig with probability 0.34375. Since this probability is larger than 0.05, the observed data doesn't give us enough evidence to reject the hypothesis that the rig was unbiased at the 95% confidence level. Adam jumped to conclusions too quickly.

2. **Law of Large Numbers**

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

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

In class, we saw that the Law of Large Numbers holds for $S_n = X_1 + \cdots + X_n$, 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.)

Solution: Intuitively, what is this problem asking? In the context of LLN, this is asking as I increase my n, does the fraction of successful packets sent approach $1-p$, the success probability.

(a) **Yes** or **No**: Each packet is sent on a completely different route.

Solution: Yes. Define X_i to be 1 if a packet is sent successfully on route i. Then $X_i, i = 1,\ldots,n$ is 0 with probability p and 1 otherwise. Since we have individual routes for each packet, we have a total of n routes. The total number of successful packets sent is hence $S_n = X_1 + \cdots + X_n$. Since S_n is a sum of i.i.d. Bernoulli random variables, $S_n \sim \text{Binomial}(n, 1-p)$.

Now similar to notation in the lecture notes, we define $A_n = \frac{S_n}{n}$ to be the fraction of successful packets sent, out of the n packets. Moreover, for each X_i,

$$E[X_i] = 1-p$$

and

$$Var[X_i] = p(1-p).$$

Using Chebyshev's inequality,

$$\begin{aligned} &Pr[\,|A_n - E[A_n]| > \varepsilon] \\ =&Pr[\,|A_n - (1-p)| > \varepsilon] \leq \frac{\text{Var}[A_n]}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

(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).

Solution: Yes. Now we need $\frac{n}{2}$ routes for each pair of packets. Similarly to the previous question, we define $X_i, i = 1,\ldots,\frac{n}{2}$ to be 0 with probability p and 2 (packets) otherwise. Now the total number of packets is $S_n = X_1 + \cdots + X_{\frac{n}{2}}$ and the fraction of received packets is $A_n = \frac{S_n}{n}$.

Now for each $i = 1,\ldots,\frac{n}{2}$

$$E[X_i] = 2(1-p)$$

and

$$Var[X_i] = 4p(1-p).$$

Thus,

$$E[A_n] = \frac{E[X_1] + \ldots + E[X_{\frac{n}{2}}]}{n} = \frac{1}{n} \cdot \frac{n}{2} \cdot 2(1-p) = 1-p$$

and

$$\text{Var}[A_n] = \frac{1}{n^2}\left(\text{Var}[X_1] + \ldots + \text{Var}[X_{\frac{n}{2}}]\right) = \frac{1}{n^2}\cdot\frac{n}{2}4p(1-p) = \frac{2p(1-p)}{n}.$$

Finally, we get

$$\begin{aligned} &Pr[\,|A_n - E[A_n]| > \varepsilon] \\ =&Pr[\,|A_n - (1-p)| > \varepsilon] \leq \frac{2p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

(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.

Solution: No. In this situation, we have

$$X_i = \begin{cases} 0 & \text{with probability } p \\ \frac{n}{2} & \text{with probability } (1-p) \end{cases}$$

for $i = 1, 2$. Now $S_n = X_1 + X_2$ and $A_n = \frac{X_1+X_2}{2}$.
We have

$$E[X_i] = \frac{n}{2}(1-p)$$

and

$$Var[X_i] = \frac{n^2}{4}p(1-p).$$

Thus,

$$E[A_n] = \frac{E[X_1]+E[X_2]}{n} = \frac{1}{n}n(1-p) = 1-p$$

and

$$\text{Var}[A_n] = \frac{1}{n^2}\left(\text{Var}[X_1] + \text{Var}[X_2]\right) = \frac{1}{n^2}\cdot\frac{n^2}{2}p(1-p) = \frac{p(1-p)}{2}.$$

Finally, we get

$$\begin{aligned} &Pr[\,|A_n - E[A_n]| > \varepsilon] \\ =&Pr[\,|A_n - (1-p)| > \varepsilon] \leq \frac{p(1-p)}{2\varepsilon^2} \end{aligned}$$

that does not converge to 0 as $n \to \infty$, so the Law of Large Numbers does not hold.

(d) **Yes** or **No**: All the packets are sent on one route.

Solution: No. $S_n = X_1$, where $X_1 = n$ with probability $1-p$ and $X_1 = 0$ with probability p. $A_n = \frac{X_1}{n}$.
Thus,

$$E[A_n] = \frac{E[X_1]}{n} = \frac{n(1-p)}{n} = 1-p$$

and

$$\text{Var}[A_n] = \frac{1}{n^2}\text{Var}[X_1] = \frac{1}{n^2}\cdot n^2 p(1-p) = p(1-p).$$

The inequality results in

$$\begin{aligned} &Pr[\,|A_n - E[A_n]| > \varepsilon] \\ =&Pr[\,|A_n - (1-p)| > \varepsilon] \leq \frac{p(1-p)}{\varepsilon^2}. \end{aligned}$$

Same as before, this does not converge to 0 as $n \to \infty$, and the LLN does not hold.

For problems (c) and (d), you should've had the intuition that since the packets are automatically sent through 1 or 2 routes, increasing n does not really help for LLN.

3. **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 $X_i = \text{Buchanan}$).

(a) Compute the expectation $\mathbf{E}[B]$ and the variance $\text{Var}(B)$.

Solution: Let B_i be a random variable representing whether the ith person voted for Buchanan. Then $B_i = 1$ if and only if $X_i = \text{Buchanan}$, so $B_i \sim \text{Bernoulli}(0.003)$. Note that the B_i's are independently and identically distributed, with $\mathbf{E}[B_i] = 0.003$ and $\text{Var}(B_i) = 0.003 \times (1 - 0.003) = 0.002991$. Moreover, by linearity of expectation and independence, we find that $\mathbf{E}[B] = \sum_{i=1}^{n} \mathbf{E}[B_i] = 432286 \times 0.003 \approx 1297$ and $\text{Var}(B) = \sum_{i=1}^{n} \text{Var}(B_i) = 432286 \times 0.002991 \approx 1293$.

(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?

Solution: Chebyshev's inequality says that

$$\Pr[|B - \mathbf{E}[B]| \geq a] \leq \frac{\mathrm{Var}(B)}{a^2}.$$

In our case $\mathbf{E}[B] = 1297$ and $\mathrm{Var}(B) = 1293$, so if we take $a = 2110$, we find that $\Pr[|B - 1297| \geq 2110] \leq 1293/2110^2 \approx 0.0003$. Now note that the condition $|B - 1297| < 2110$ is equivalent to the condition $-813 < B < 3407$, and since B is non-negative, we find that $\Pr[B > 3407] \leq 0.0003$ (roughly), so we can take $b \approx 0.0003$. In other words, receiving 3407 votes for Buchanan in Palm Beach County seems very unlikely to happen by chance, under this simple model. So yes, this is statistically significant.

(c) 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?

Solution: Let p_j be the probability that the jth county does not receive 3407 votes for Buchanan. We have from part (b) that $p_j = 1 - b \approx 0.9997$. Note that the probability that no county yields at least 3407 votes for Buchanan is $p_1 \times \cdots \times p_{67}$, since the voters in each county behave independently. Thus, the probability that Buchanan does not receive 3407 votes in any county is about $(0.9997)^{67} \approx 0.98$. Consequently, the probability that Buchanan *does* receive at least 3407 votes in some county is about $1 - 0.98 \approx 0.02$. In other words, this seems unlikely to happen by chance.

4. **Uniform Probability Space**

Let $\Omega = \{1,2,3,4,5,6\}$ be a uniform probability space. Let also $X(\omega)$ and $Y(\omega)$, for $\omega \in \Omega$, be the random variables defined as follows:

Table 1: The random variables X and Y.

ω	1	2	3	4	5	6
$X(\omega)$	0	0	1	1	2	2
$Y(\omega)$	0	2	3	5	2	0

(a) Calculate $V = L[Y|X]$;

(b) Calculate $W = E[Y|X]$;

(c) Calculate $E[(Y-V)^2]$;

(d) Calculate $E[(Y-W)^2]$.

[*Hint:* Recall that $L[Y|X]$ and $E[Y|X]$ are functions of X and that you need to specify their value as a function of X.]

Solution:

(a) We find $E[X] = 1, E[Y] = 2, E[XY] = 2$, so that $cov(X,Y) = 0$ and $L[Y|X] = E[Y] = 2$.

(b) We see that $E[Y|X=0] = 1, E[Y|X=1] = 4, E[Y|X=2] = 1$.

(c) $E[(Y-V)^2] = E[(Y-2)^2] = (4+0+1+9+0+4)/6 = 3$.

(d) $E[(Y-W)^2] = (1+1+1+1+1+1)/6 = 1$.

5. **0070**

James Bond is imprisoned in a cell from which there are three possible ways to escape: an air-conditioning duct, a sewer pipe and the door (which is unlocked). The air-conditioning duct leads him on a two-hour trip whereupon he falls through a trap door onto his head, much to the amusement of his captors. The sewer pipe is similar but takes five hours to traverse. Each fall produces amnesia and he is returned to the cell immediately after each fall. Assume that he always immediately chooses one of the three exits from the cell with probability $\frac{1}{3}$. On average, how long does it take before he opens the unlocked door and escapes?

Solution: Due to the memorylessness of the scenario, i.e., what James Bond chose in the previous attempt has nothing to do with his current attempt and all the following attempts, it is like starting all over again, as every attempt is just like his first attempt. Let X be the random variable of the time 007 needs to escape. We can start by considering the outcome of Bond's first attempt, and see how the expected time to escape, $E[X]$, changes as a result of this. We have

$$\begin{aligned} E[X] &= E[X|A]\Pr[A] + E[X|S]\Pr[S] + E[X|D]\Pr[D] \\ &= \frac{1}{3}(E[X|A] + E[X|S] + E[X|D]) \end{aligned}$$

where $E[X|A]$ means the expected time to escape given that Bond went through the AC-duct in his first attempt, and similarly for $E[X|S]$ and $E[X|D]$. Now we take a closer look at these conditional expectations. Clearly, $E[X|D] = 0$ because Mr. Bond escapes immediately. Meanwhile, we have $E[X|A] = 2 + E[X]$ because given that 007 chose AC-duct in his first attempt, he wastes 2 hours and have to try again (which triggers another completely fresh attempt with no memory and thus takes $E[X]$ hours to escape). Similarly, we have $E[X|S] = 5 + E[X]$. So now, we have derived a recursive relation for the expected time to escape

$$E[X] = \frac{1}{3}(2 + E[X] + 5 + E[X] + 0)$$

Solving this for $E[X]$, we get that $E[X] = \boxed{\mathbf{007}}$.