Notes for Recitation 11 (11/20/98)

This recitation gives some more practice with analyzing probabilities using trees. It contains properties and examples involving conditional probability and independence. It also looks at some properties of random variables.

1 Predicting a diamond

Consider the following game using a deck of 10 cards with 5 diamonds and 5 clubs. The dealer shuffles the deck and starts dealing them out one at a time, face up. The player is supposed to try to predict when a diamond will appear. That is, at any time as the cards are being dealt he is allowed to say "stop!". Then if the next card is a diamond he wins, otherwise he loses.

What is the best strategy for the player? What is the probability that the player wins?

It is easy to find a strategy that wins half the time – just predict that the first card is a diamond. But is it possible to do better by waiting and observing what has happened? For example, we might wait until the number of clubs dealt is larger than the number of diamonds dealt, and then predict that the next card is a diamond. Of course this may never happen for some arrangements of the deck and in such cases this strategy might lose.

We will argue that no strategy the player might use can make any difference – the probability is always $\frac{1}{2}$, no matter how the player chooses!

To argue this rigorously, we must introduce the following additional condition: if there is only one card left, the player must say "stop". This rules out the stategy that the player never predicts that any card will be a diamond, and thus always loses.

Let's generalize to an arbitrary number n of cards, with c clubs and d diamonds. One way to describe the experiment is by a 2-stage tree: In stage 1, the dealer shuffles the deck, which leads to a random permutation of the n cards, with probability $\frac{1}{n!}$ for each permutation. In stage 2, the player chooses which card to bet on. Since the player sees the cards one by one, his strategy is constrained as follows:
"If in permutation p he bets on card i, $1 \leq i \leq n$, and if permutation q has exactly the same cards in positions $1, \ldots, i-1$, then he will also bet on card i in permutation q." This constraint produces a lot of symmetry in the tree.

Another way to describe the same experiment is by a $(2n-1)$-stage tree. At every odd-numbered stage, the player says whether to "stop" or "go". At every even-numbered stage, a card is dealt, with uniform probability from the ones remaining. A particular strategy for the player represents a particular choice of "stop" or "go" at every node of the tree where the player has a choice.

The main difference in these two trees is that in the first one, the permutation is determined at the beginning, but the player is constrained to "know" only the first $i-1$ cards when he decides whether to bet on the ith card. In the second tree, there is only one node for each "prefix" of the sequence of cards, which makes it obvious that the player's choice must depend only on that prefix. Another difference is that the second tree models the situation where the next card is chosen randomly at each stage rather than the whole permutation being chosen at the beginning. This is like reshuffling the remaining deck after each card is dealt. This repeated shuffling does not change any of the relevant probabilities or conditional probabilities.

We show:

Theorem 1.1 *For every $c, d, n \in N$, where $n \geq 1$ and $c + d = n$, and every player strategy, the probability of guessing a diamond correctly is exactly $\frac{d}{n}$.*

Proof. We can prove this by (ordinary) induction on n.

Base: If $n = 1$ then there is only one card, and one strategy, which is to look at that card. The probability of winning is exactly 1 if $d = 1$, and 0 if $d = 0$, which is as needed.

Inductive step: Assume the statement for n and prove it for $n+1$. Thus, consider an $n+1$-card deck with c clubs and d diamonds. There are two possibilities at the first stage – the player can choose "stop" or "go". We argue that either one of these lead to the same probability of winning, namely, $\frac{d}{n+1}$, which is what we want.

Case 1: He chooses "stop". Then he wins exactly if the first card is a diamond. Since this card is chosen randomly with uniform probability, and there are d out of $n+1$ diamonds in the deck, this yields probability $\frac{d}{n+1}$.

Case 2: He chooses "go". Then the first card is chosen randomly, and is a diamond with probability $\frac{d}{n+1}$ and a club with probability $\frac{c}{n+1}$.
If it is a diamond, then what the player is faced with is a reduced game, with n cards, of which $d-1$ are diamonds and c are clubs. On the other hand, if it is a club, then the player is faced with a different reduced game, with n cards, d diamonds, and $c-1$ clubs. We can use the inductive hypothesis to conclude that the winning probabilities in those two reduced games are $\frac{d-1}{n}$ and $\frac{d}{n}$, respectively.

The probability of winning can then be calculated using conditional probabilities and products, as:

$\frac{d}{n+1}\frac{d-1}{n} + \frac{c}{n+1}\frac{d}{n}$
$= \frac{d}{n+1}\frac{d-1+c}{n}$
$= \frac{d}{n+1}\frac{n}{n}$ (using $d + c - 1 = n$)
$= \frac{d}{n+1}$. ■

2 Conditional probability and independence: some properties

Theorem 2.1 *If events A and B are independent, then*

1. *A and $\overline{B}$ are independent.*
2. *$\overline{A}$ and B are independent.*
3. *$\overline{A}$ and $\overline{B}$ are independent.*

Proof. I show just the first property, that A and $\overline{B}$ are independent (the others all follow). That is, I show that $Pr(A \cap \overline{B}) = Pr(A)Pr(\overline{B})$. We have that: $Pr(A \cap \overline{B}) = Pr(A) - Pr(A - \overline{B})$ by a law given in the previous recitation,
$= Pr(A) - Pr(A \cap B)$ by basic set theory,
$= Pr(A) - Pr(A)Pr(B)$ by independence,
$= Pr(A)(1 - Pr(B))$,
$= Pr(A)(Pr(\overline{B}))$, as needed. ■

A version of Theorem 2.1 holds for mutual independence of any number of events. Here is a fact about conditional probabilities:

Theorem 2.2 *If events A, B, and C are mutually independent, then $Pr(A \cap B \mid C) = Pr(A \mid C)Pr(B \mid C)$.*

Proof. $Pr(A \cap B \mid C) = \frac{Pr(A \cap B \cap C)}{Pr(C)}$ by definition of conditional probability,
$= \frac{Pr(A)Pr(B)Pr(C)}{Pr(C)}$ by mutual independence,
$= \frac{Pr(A)Pr(C)}{Pr(C)}\frac{Pr(B)Pr(C)}{Pr(C)}$,
$= \frac{Pr(A \cap C)}{Pr(C)}\frac{Pr(B \cap C)}{Pr(C)}$ by mutual independence,
$= Pr(A \mid C)Pr(B \mid C)$ by definition of conditional probability. ■

We can combine facts like those above, e.g.:

Theorem 2.3 *If events A, B and C are mutually independent, then*

1. $Pr(A \cap B \mid C) = Pr(A \mid C)Pr(B \mid C)$.
2. $Pr(A \cap \overline{B} \mid C) = Pr(A \mid C)Pr(\overline{B} \mid C)$.
3. $Pr(\overline{A} \cap B \mid C) = Pr(\overline{A} \mid C)Pr(B \mid C)$.
4. $Pr(\overline{A} \cap \overline{B} \mid C) = Pr(\overline{A} \mid C)Pr(\overline{B} \mid C)$.
5. $Pr(A \cap B \mid \overline{C}) = Pr(A \mid \overline{C})Pr(B \mid \overline{C})$.
6. $Pr(A \cap \overline{B} \mid \overline{C}) = Pr(A \mid \overline{C})Pr(\overline{B} \mid \overline{C})$.
7. $Pr(\overline{A} \cap B \mid \overline{C}) = Pr(\overline{A} \mid \overline{C})Pr(B \mid \overline{C})$.
8. $Pr(\overline{A} \cap \overline{B} \mid \overline{C}) = Pr(\overline{A} \mid \overline{C})Pr(\overline{B} \mid \overline{C})$.

You might want to prove one or two of these.

3 Whom do you trust?

Suppose that, in a course (not 6.042, naturally), 10% of the problems on quizzes contain errors. If you ask a TA whether a problem has an error, he or she will answer correctly 80% of the time. But if you ask a lecturer, he/she will answer correctly 75% of the time. The correctness of the lecturers' answers, the correctness of the TAs' answers, and the existence of an error are all (mutually) independent.

Now suppose you have a doubt about a problem and ask a TA about it, and he/she tells you that the problem has an error. To double-check, you ask a lecturer, who says that the problem is correct. Now what is the probability that there is an error? In particular, is this probability greater than the initial probability of 10%?

We formulate this as an experiment of choosing one problem randomly and asking a particular TA and lecturer about it. Let E be the event "the problem has an error", T be the event "the TA says the problem has an error", and L be the event "the lecturer says the problem has an error".

The assumptions above tell us: $Pr(E) = 10\%$, $Pr(T \mid E) = Pr(\overline{T} \mid \overline{E}) = 80\%$, $Pr(L \mid E) = Pr(\overline{L} \mid \overline{E}) = 75\%$, and E, L, and T are mutually independent.

We want to compute $Pr(E \mid T \cap \overline{L})$. We use Bayes' rule:

$$Pr(E \mid T \cap \overline{L}) = Pr(E)\frac{Pr(T \cap \overline{L} \mid E)}{Pr(T \cap \overline{L})}. \tag{1}$$

Since E, L, and T are mutually independent, we get (using Theorem 2.3 above): $Pr(T \cap \overline{L} \mid E) = Pr(T \mid E)Pr(\overline{L} \mid E)$ and $Pr(T \cap \overline{L} \mid \overline{E}) = Pr(T \mid \overline{E})Pr(\overline{L} \mid \overline{E})$.

Now let's compute the probabilities we need: $Pr(T \cap \overline{L}|E) = Pr(T|E)Pr(\overline{L}|E)$ by Theorem 2.3 $= (8/10)(25/100) = 2/10$,
$Pr(T \cap \overline{L}|\overline{E}) = Pr(T|\overline{E})Pr(\overline{L}|\overline{E}) = (2/10)(75/100) = 15/100$ by Theorem 2.3,
and

$$\begin{aligned} Pr(T \cap \overline{L}) &= Pr(T \cap \overline{L}|E)Pr(E) + Pr(T \cap \overline{L}|\overline{E})Pr(\overline{E}) \\ &= (2/10)(1/10) + (15/100)(9/10) \\ &= \frac{155}{1000} \end{aligned}$$

Substituting these values in equation 1 we get:

$$Pr(E \mid T \cap \overline{L}) = \frac{1}{10} \cdot \frac{\frac{2}{10}}{\frac{155}{1000}} = .12903$$

Thus, the contradictory information has increased the probability that there is an error from 10% to almost 13%.

Now suppose that the situation is reversed: the TA says the problem does not contain an error, but the lecturer says that it does. Now what is the probability of error? Is it lower or higher than before? Conceivably it will be lower since you trust the TA better than the lecturer, but will it be lower than the original 10%?

The calculation is similar this time: now we want to $Pr(E \mid \overline{T} \cap L)$. By Bayes' rule

$$Pr(E \mid \overline{T} \cap L) = Pr(E)\frac{Pr(\overline{T} \cap L \mid E)}{Pr(\overline{T} \cap L)}. \tag{2}$$

By the independence assumptions, we have: $Pr(\overline{T} \cap L|E) = Pr(\overline{T}|E)Pr(L|E)$ and $Pr(\overline{T} \cap L|\overline{E}) = Pr(\overline{T}|\overline{E})Pr(L|\overline{E})$.

Then: $Pr(\overline{T} \cap L|E) = Pr(\overline{T}|E)Pr(L|E)$
$= (2/10)(75/100) = 3/20$,
$Pr(\overline{T} \cap L|\overline{E}) = Pr(\overline{T}|\overline{E})Pr(L|\overline{E})$
$= (8/10)(25/100) = 1/5$, and

$$\begin{aligned} Pr(\overline{T} \cap L) &= Pr(\overline{T} \cap L|E)Pr(E) + Pr(\overline{T} \cap L|\overline{E})Pr(\overline{E}) \\ &= (3/20)(1/10) + (1/5)(9/10) \\ &= \frac{195}{1000} \end{aligned}$$

Substituting these values in equation 2 we get

$$Pr(E \mid T, \overline{L}) = \frac{1}{10}\frac{\frac{15}{100}}{\frac{195}{1000}} = .076923\ldots$$

So this time, the contradictory information has decreased the probability of an error from 10% to about 7.7%. These calculations confirm the common sense rule that when two people make contradictory statements, we should be influenced more by what the most "authoritative" person (the one that is more often right) says. But note that this does not mean that you should *believe* in what the most authoritative person says. In the first calculation, the TA's opinion raises the probability of an error from 10% to 13%, still way below 50%!

4 Sums of random variables

Sums of random variables arise in many applications. The basic question is: given a pair (or a set) of random variables, what is the distribution of the random variable defined as their sum?

For two random variables, the distribution function for their sum can be obtained by a special formula. Let the random variables be R_1 and R_2 with distribution fuctions f_1 and f_2. Then the distribution function for their sum is

$$g(x) = \sum_{x_1, x_2 : x_1 + x_2 = x} f(x_1, x_2).$$

Here f is the *joint* distribution function of f_1 and f_2, i.e., $f(x)$ is the probability that $R_1 = x_1$ and $R_2 = x_2$.

If R_1 and R_2 are independent, then $f(x_1, x_2) = f_1(x_1) f_2(x_2)$, so

$$g(x) = \sum_{x_1, x_2 : x_1 + x_2 = x} f_1(x_1) f_2(x_2).$$

Do you recognize this? The function g is the *convolution* of f_1 and f_2.

We can represent this nicely in terms of generating functions in the special case when the range of R_1 and R_2 is $\mathbb{N}$. For this we define generating functions of probabilities.

Definition 4.1 *Let R be a random variable $R : S \to \mathbb{N}$, then the* probability generating function (PGF) *for R is*

$$G(z) = \sum_{i=0}^{\infty} Pr(R = i) z^i$$

Examples:

- R is Bernoulli: $Pr(R = 1) = p$, $Pr(R = 0) = 1 - p$.

$$G(z) = (1 - p) + pz$$

- R is uniform on $[0, N-1]$.

$$G(z) = \frac{1}{N} + \frac{1}{N}z + \frac{1}{N}z^2 + \ldots + \frac{1}{N}z^{N-1} = \frac{1}{N}\left(\frac{1-z^N}{1-z}\right)$$

- R is binomial with parameters p and n.

$$G(z) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = [(1-p) + pz]^n$$

Now we can state the product rule for sums of independent random variables. Notice that this is analogous to the product rule when we discussed counting.

Theorem 4.2 *If R_1 and R_2 are independent random variables from S to $\mathbb{N}$ with PGF's $G_1(z)$ and $G_2(z)$, then the PGF of $R = R_1 + R_2$ is*

$$G(z) = G_1(z)G_2(z)$$

Proof. We know that $G(z) = \sum_{n=0}^{\infty} Pr(R = n)z^n$. Now,

$$\begin{aligned} Pr(R = n) &= Pr(R_1 + R_2 = n) \\ &= \sum_{i,j:i+j=n} Pr(R_1 = i \wedge R_2 = j) \\ &= \sum_{i,j:i+j=n} Pr(R_1 = i)Pr(R_2 = j) \end{aligned}$$

The last step uses the fact that R_1 and R_2 are independent. But then, the coefficient of z^n in $G(z)$ is

$$Pr(R = n) = \text{coefficient of } z^n \text{ in } \left(\sum_{i=0}^{\infty} Pr(R_1 = i)z^i\right)\left(\sum_{j=0}^{\infty} Pr(R_2 = j)z^j\right)$$

Hence $G(z) = G_1(z)G_2(z)$. ∎

As an example here is a problem we've seen before, but we'll view it differently. There are n devices in a circuit we are building. Each of the n devices could be faulty. The probability that any one fails is p, and the failures are independent. From this we would like to get the distribution for the number of failures — for example, what is the probability that there are 2 failures?.

Define the random variables $R_1, \ldots, R_n$ as

$$R_i = \begin{cases} 1 & i^{th} \text{ device fails} \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$\forall i, Pr(R_i = 1) = p$$

Let R be the number of failures overall.

$$R = R_1 + R_2 \ldots + R_n$$

The PGF for each R_i is $(1-p)+pz$. Therefore, by the product rule, the PGF for R is $[(1-p)+pz]^n$. Hence, the probability of k errors is

$$\begin{aligned} Pr(R = k) &= \text{coefficient of } z^k \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

This is also an alternative derivation of the Binomial distribution.

5 Binomial distribution examples

5.1 Transmission on a noisy channel

Suppose that we are transmitting bits across a noisy channel. (For example, say your modem uses a phone line that faintly picks up a local radio station.) Suppose we transmit $10,000$ bits, and each arriving bit is incorrect with probability 0.01. Assume that these errors occur independently (this is not realistic). What is the probability that more than 2% of the bits are erroneous?

We can solve this problem using the bound given in lecture on $F_{n,p}$, the cumulative distribution function for the binomial distribution.

$$\begin{aligned} Pr(> \text{ than } 2\% \text{ errors}) &= Pr(\leq \text{ } 98\% \text{ correct}) \\ &= F_{n,0.99}(0.98n) \\ &\leq 1.98 \frac{2^{-0.005646 \cdot 10,000}}{0.3509\sqrt{10,000}} \\ &\leq 2^{-60} \end{aligned}$$

The probability that more than 2% of the bits are erroneous is incredibly small! This again demonstrates the extreme improbability of outcomes on the tails of the binomial distribution.

5.2 Polling

Another good example of binomial distributions comes up in polling. The Gallup polling service reported that on November 10, only 37% of American adults thought Louise Woodward was guilty.[1] Furthermore, Gallup asserted that since they asked the opinions of 623 adults, "one can say with 95 percent confidence that the error attributable to sampling and other random effects is within plus or minus 4 percentage points". Can we confirm this claim?

We want to determine p, the fraction of Americans that think Louise is guilty. Our plan is to sample the opinion of n people chosen uniformly and at random, *with replacement*. That is, we might poll the same person twice! This may seem inefficient, but it simplifies the analysis. If G is the number of people in our sample that say Louise is guilty, then we will claim that p is about G/n.

The problem is that our sample might not be representative. For example, maybe everyone in the country outside of North Dakota thinks Louise is guilty, but by bad luck the sample contained mostly North Dakotans. Then our poll would give the wrong answer.

Let ϵ be the margin of error we can tolerate, and let δ be the probability that our result lies outside this margin. How many people must we poll so that our result is within ϵ percent of national opinion with probability at least $1-\delta$? For example, Gallup claims that for $\epsilon = 0.04$ and $\delta = 0.05$, polling 623 people is sufficient.

We can define δ, the probability that our poll is off by more than the margin of error ϵ, as follows:

$$\begin{aligned}
\delta &= \underbrace{Pr\left(\frac{G}{n} < p - \epsilon\right)}_{\substack{\text{too many in sample} \\ \text{say “not guilty”}}} + \underbrace{Pr\left(\frac{G}{n} > p + \epsilon\right)}_{\substack{\text{too many in sample} \\ \text{say “guilty”}}} \\
&= Pr(G < (p-\epsilon)n) + Pr(G > (p+\epsilon)n)
\end{aligned}$$

Each term in the definition of δ can be evaluated using the bound given above on $F_{n,p}$. In the second term, we must use the same trick as in the Noisy Channel problem to ensure that $\alpha < p$. We observe that $Pr(G/n > p+\epsilon) = Pr((n-G)/n < 1-p-\epsilon)$, where $(n-G)/n$ is the fraction of people polled who say that Louise is not guilty, and $1-p$ is the fraction of all Americans who say that she is not guilty. This gives:

$$\delta \le F_{n,p}((p-\epsilon)n) + F_{n,1-p}((1-p-\epsilon)n)$$

[1] Well, we could make this sound more current by restating it in terms of the percentage that think Bill Clinton should be impeached...but you get the idea, so we won't bother...

This is an expression for the probability that our poll is off by more than the margin of error. The problem is that the expression contains p, the fraction of Americans that think Louise is guilty. This is the number we are trying to determine by polling!

Fortunately, we can upper bound δ by using the following:

Fact 1 *For all ϵ, the maximum value of δ occurs when $p = \frac{1}{2}$.*

We will not prove this fact, though we could. The fact implies that to get an upper bound on δ, we can pretend that half of the people think Louise is guilty. This gives:

$$\begin{aligned}
\delta &\le F_{n,\frac{1}{2}}((\frac{1}{2}-\epsilon)n) + F_{n,1-\frac{1}{2}}((1-\frac{1}{2}-\epsilon)n) \\
&= 2F_{n,\frac{1}{2}}((\frac{1}{2}-\epsilon)n)
\end{aligned}$$

Now suppose that we want a margin of error of 4% as Gallup claimed. Plugging in $\epsilon = 0.04$ gives:

$$\begin{aligned}
\delta &\le 2F_n(0.46n) \\
&\le 2 \cdot 6.75 \cdot \frac{2^{(-0.004622\ldots)\cdot n}}{(1.2492\ldots)\sqrt{n}}
\end{aligned}$$

We want to poll enough people so that δ is less than 0.05. The easiest way is to plug in values for n, the number of people polled:

n = number of people polled	upper bound on probability poll is wrong		
500	9.7%		
600	6.4%		
623	5.9%	$\leftarrow$	Gallup's poll size
650	5.3%		
662	5.0%	$\leftarrow$	our poll size
700	4.3%		

Gallup's poll size is just about right. By our calculation, polling 662 people is sufficient to determine public opinion to within 4% with confidence of 95%. We can be certain that Gallup's poll of 623 people gives 95% confidence with at most a 4.13% margin of error. But the real margin of error may be less, since we made several approximations. Still, our approximations must be quite good, since 4.13% is quite close to Gallup's claim of 4%.

The remarkable point is that the population of the country has no effect on the poll size! Whether there are a thousand people or a billion in the country, polling only a few hundred is sufficient!

6 Team problems

Problem 1. Consider the scenario considered earlier: a quiz problem has an error with probability 10%, the TA says whether or not there is an error, answering correctly with probability 80%, and the lecturer answers correctly with probability 75%. The three events of existence of an error, the TA responding correctly and the lecturer responding correctly are mutually independent.

Is the event T, "the TA says that there is an error", independent of the event L, "the lecturer says that there is an error"?

Problem 2. This is a real problem that came from a research project in computational biology last year.

It is known that atoms occur in different isotopes. For example, Carbon atoms usually have 6 protons and 6 neutrons, but occasionally they have 7 neutrons. In fact, 1.11% of all Carbon atoms have 7 neutrons and 98.89% have 6. From what we know, the appearance of the different isotopes seems to be a random event, unrelated to the functionality of the atom. We can tell which Carbon atoms have an extra neutron only by their atomic weight. The atomic weight of the generic Carbon atom is 12 while the weight of its rarer isotope is 13.

The research team is studying proteins composed of Carbon, Oxygen, Nitrogen and Hydrogen (the four building blocks of life). Their atomic weights and isotopes are listed below. The second column is the atomic weight and the third column is the frequency of the isotope. Assume that the probability of an unlisted isotope ocurring is zero (i.e., no tritium in our model).

Hydrogen	1	$P_{H,1}$
	2	$P_{H,2}$
Carbon	12	$P_{C,12}$
	13	$P_{C,13}$
Nitrogen	14	$P_{N,14}$
	15	$P_{N,15}$
Oxygen	15	$P_{O,15}$
	16	$P_{O,16}$
	17	$P_{O,17}$

Suppose we want to predict the distribution of atomic weights for any protein. Describe a simple algorithm for computing the weight distribution of any protein that contains n_C carbon atoms, n_O oxygen atoms, n_N nitrogen atoms and n_H hydrogen atoms. [Hint: Use generating functions!]

Solutions

Problem 1

The answer is no. We will show that the probability that the two events occur is different from the product of the probabilities. Define the events T, L, and E as before.

The probability of T is

$$\begin{aligned} Pr(T) &= Pr(T \cap E) + Pr(T \cap \overline{E}) \\ &= \frac{80}{100}\frac{10}{100} + \frac{20}{100}\frac{90}{100} \\ &= \frac{26}{100} \\ &= .26 \end{aligned}$$

The probability of L is

$$\begin{aligned} Pr(L) &= Pr(L \cap E) + Pr(L \cap \overline{E}) \\ &= \frac{75}{100}\frac{10}{100} + \frac{25}{100}\frac{90}{100} \\ &= \frac{30}{100} \\ &= .3 \end{aligned}$$

The probability of $L \cap T$ is

$$\begin{aligned}
Pr(L \cap T) &= Pr(L \cap T \cap E) + Pr(L \cap T \cap \overline{E}) \\
&= Pr(L \cap T|E)Pr(E) + Pr(L \cap T|\overline{E})Pr(\overline{E}) \\
&= Pr(L|E)Pr(T|E)Pr(E) + Pr(L|\overline{E})Pr(T|\overline{E})Pr(\overline{E}) \\
&= \frac{75}{100}\frac{80}{100}\frac{10}{100} + \frac{25}{100}\frac{20}{100}\frac{90}{100} \\
&= \frac{105}{1000} \\
&= .105
\end{aligned}$$

This is higher than

$$Pr(L)Pr(T) = .26 \times .3 = .078$$

The intuition is that when the TA says that the problem has an error, that it is more likely to have an error, and therefore the lecturer is more likely to say there is an error.

Problem 2

Let R_H be the random variable whose value is the atomic weight of a random Hydrogen atom. Similarly, let R_C, R_N and R_O be the random variables for the atomic weight of Carbon, Nitrogen and Oxygen atoms.

We can write down the generating functions for their distributions:

$$\begin{aligned}
G_H(z) &= P_{H,1}z + P_{H,2}z^2 \\
G_C(z) &= P_{C,12}z^{12} + P_{C,13}z^{13} \\
G_N(z) &= P_{N,14}z^{14} + P_{N,15}z^{15} \\
G_O(z) &= P_{0,15}z^{15} + P_{O,16}z^{16} + P_{O,17}z^{17}
\end{aligned}$$

Then, let R be the random variable representing the atomic weight of a protein with n_C carbon atoms, n_O oxygen atoms, n_N nitrogen atoms and n_H hydrogen atoms. We can write:

$$R = \sum_{i=1}^{n_H} R_{H,i} + \sum_{i=1}^{n_C} R_{C,i} + \sum_{i=1}^{n_N} R_{N,i} + \sum_{i=}^{n_O} R_{O,i}$$

Here the distribution of each hydrogen atom in the protein, $R_{H,i}$, is identical to the distribution of R_H. Similarly, the $R_{C,i}$'s have the same distribution as R_C, the $R_{N,i}$ have the same distribution as R_N and the $R_{O,i}$'s have the same distribution as R_O.

By the product rule, the PGF for R is

$$\begin{aligned} G(z) &= (P_{H,1}z + P_{H,2}z^2)^{n_H}(P_{C,12}z^{12} + P_{C,13}z^{13})^{n_C} \\ &\quad (P_{N,14}z^{14} + P_{N,15}z^{15})^{n_N}(P_{0,15}z^{15} + P_{O,16}z^{16} + P_{O,17}z^{17})^{n_O} \end{aligned}$$

Hence the probability that a protein molecule with n_H, n_C, n_N, n_O atoms of H, C, N, O respectively has atomic weight k is the coefficient of z^k in the above polynomial.