MATH 2B RECITATION 1/12/12

ALDEN WALKER

1. Random thoughts

Bayes' rule can be used in cryptanalysis. For example, suppose that we want to decipher a cryptogram C (what you get if you substitute a=z, b=g, etc), and we want to evaluate a potential monoalphabetic cipher p that we think might have produced the cryptogram from some plaintext. In other words, we want to know $P(p|C)$. A priori, this is quite difficult, but we can be tricky to get a handle on it. By Bayes' rule, this is $\frac{P(C|p)P(p)}{P(C)}$. Assuming that all ciphers and cryptograms are equally likely, we can disregard those, so $P(p|C) = D \times P(C|p)$ for some number D which is constant for all ciphers. Therefore, calculating $P(C|p)$ gives us a good score for a cipher. Then we can pick the cipher which scores the best, decrypt, and we have the plaintext.

But what is $P(C|p)$, and why is is easier to calculate than $P(p|C)$? This is the probability that what you get when you decrypt C with p is the plaintext we started from. Of course, we don't know the plaintext, but say we know that it is English. In that case, maybe most ciphers decrypt the cryptogram to gibberish, so $P(C|p)$ may actually tell us something.

Thus, we have reduced the problem to deciding whether a string of letters is English. We can use probability again! One easy way to do this is to look at a large volume of text and compute all the conditional probabilities like $P(\text{e}|\text{th})$ = "having seen the pair 'th', what is the probability that the next letter is 'e'?" Then, if the string of letters is $s_1 \cdots s_n$, you compute

$$P(s_1 \cdots s_n \text{is English}) \approx P(s_1)P(s_2|s_1)P(s_3|s_1s_2) \cdots P(s_n|s_{n-2}s_{n-1})$$

This gives us a score. Then we just find the cipher which scores the highest. This part is definitely nontrivial, but it doesn't really have to do with the course. Check out the "software" section of my website for a toy implementation of this strategy.

Incidentally, this model for English is called a Markov chain. Markov chains have a huge number of applications in applied math. You could also use the model to generate a string of letters that "looks like English". Here is a snippet generated by a Markov chain generator when given *Alice's adventures in wonderland* as the text for computing probabilities:

> Alice in a large eyes like they're surprise was she said, "Alice guessed to the bread and the Duck."
>
> The race-course, in a whisper, had vanished.

2. The Binomial Distribution

A very natural question is "what is the probability that we will see 48 heads if we flip a coin 100 times?" You can answer this with what you already know; it is $\binom{100}{48}(1/2)^{52}(1/2)^{48}$ because there are $\binom{100}{48}$ ways to have 48 heads show up, and the probability of each such string is $(1/2)^{52}(1/2)^{48}$. In general, if you have n independent trials, with probability p of success on each trial (let $q = (1-p)$), then the probability of (exactly) k successes is:

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k q^{n-k}$$

The set of these numbers as k varies is the (n, p) binomial distribution. The values n and p are called the parameters of the distribution.

The **mean** of this distribution (the number of successes we expect) is np. The **mode** of this distribution (the most likely number of successes) is $\lfloor np + p \rfloor$. The reason these two numbers are different is that the distribution is discrete, so on any given toss of a $p = 3/4$ coin, we expect $3/4$ of a head, but the most likely number of heads is 1.

2.1. **Example (HW 1,2).** Suppose that you must pass 7 out of 10 exams to pass a class. You have a probability of 0.9 of passing any particular exam (and they are independent). What are your chances of failing the first two exams and passing the class?

Let $p = 0.9$ be the chance of passing. Let X be the event above that we are interested in. It is clear that

$$\begin{aligned} P(X) &= P(\text{fail exam 1} \cap \text{fail exam 2} \cap \text{fail 0 or 1 exams in 8}) \\ &= (1-p)^2(P(\text{fail 0 in 8}) + P(\text{fail 1 in 8})) \\ &= (1-p)^2\left(\binom{8}{0}p^8 + \binom{8}{1}p^7(1-p)\right) \\ &= 0.00478 \end{aligned}$$

So a little unlikely. If you are asking what your chances are given that you have failed the first two, then you are interested in $P(X|\text{fail first 2}) = P(X \cap \text{fail first 2})/P(\text{fail first 2}) = 0.00478/0.1^2 = 0.478$, so not such a bad probability.

3. Drawing Without Replacement

You have done a few problems in which you drew people or things from populations, but you always assumed that the probability of drawing a particular blood type, for example, stayed constant no matter what you draw. Of course, this isn't true in real life. Let us suppose that we have 10 balls, 3 of which are red. If we draw 5 balls, what is the probability of getting at exactly one red ball? The probability of doing it on the first draw is:

$$\frac{3}{10}\frac{7}{9}\frac{6}{8}\frac{5}{7}\frac{4}{6} = \frac{1}{12}$$

This is obviously different from getting exactly one success out of 5 when the probability of success is 3/10 (which is $5(3/10)(7/10)^4 \approx 0.3602$). It turns out that if you have a sample of size N with G good and B bad, we have

$$P(g \text{ good and } b \text{ bad}) = \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}$$

4. The Normal Distribution (HW 3,5)

4.1. **What's going on?** The binomial distribution is simultaneously ubiquitous and difficult to compute. Therefore, we approximate with other functions. These functions happen to be distributions which appear naturally in the study of probability. You will study them in detail in the second half of the course. For now, it suffices to just basically understand what is going on, as long as you know how to compute approximations to the binomial using the normal and Poisson distributions.

4.2. **What do these words mean?** A distribution refers to the probability that a random variable is a number or between two other numbers. It is often used in a semi-vague way to refer to the probability density function or the cumulative distribution function. The cumulative distribution function for the normal distribution means a function Φ such that $P(X \leq x) = \Phi(x)$. A probability density function ϕ is a a function such that $\Phi(z) = \int_{-\infty}^{z} \phi(x)dx$.

4.3. **The PDF for the Normal Distribution.** The probability density function for the normal distribution is $\phi(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$, where μ, the mean, and σ, the standard deviation, are the parameters, just like n and p for the binomial. The standard normal curve has $\mu = 0$ and $\sigma = 1$ (from now on, Φ will be the cumulative distribution for the standard normal distribution). In practice, only the standard normal curve is used, because for any normal distribution, you can find the probability that a random variable lands in between a and b as $\Phi((b-\mu)/\sigma) - \Phi(a-\mu)/\sigma)$.

Note that doing the relavant integral to compute Φ is hard, but we can do it numerically. Note there is a table in the back of your book, or in Mathematica $\Phi(x)$ is `CDF[NormalDistribution[], x]`.

4.4. **Approximating the Binomial with the Normal.** If we have an (n, p) binomial distribution, then

$$P(\text{between } a \text{ and } b \text{ successes}) \approx \Phi\left(\frac{b+\frac{1}{2}-\mu}{\sigma}\right) - \Phi\left(\frac{a-\frac{1}{2}-\mu}{\sigma}\right)$$

Where μ is the mean np and σ is the standard deviation $\sqrt{npq}$. The fact that this is the right number for the standard deviation will be covered later.

4.5. **Example.** Caltech undergrads show up to certain class independently from each other with probability 0.7. If there are 230 students in the class and 175 seats in the lecture hall, what is the probability that fire laws are being broken?

Actually, I realized that question is exactly like one of your homework problems, so I won't do it, but I thought it was a good question, so I'm leaving it there. Let's suppose that the class is 100 students and that if fewer than 20 students show up, everybody goes out to lunch, and if more than 70 students show up, the class can't fit and everybody goes home and takes a nap. What's the probability that people eat or sleep rather than learn?

Here we have a $(100, 0.7)$ binomial distribution, so the mean is $(100)(0.7) = 70$ and the standard deviation is $\sqrt{(100)(0.7)(0.3)} = 4.58$. Using a normal approximation, we find that the probability that between 0 and 19 students show up is approximately $\Phi\left(\frac{19+\frac{1}{2}-70}{4.58}\right) - \Phi\left(\frac{0-\frac{1}{2}-70}{4.58}\right) = 1.5 \times 10^{-28}$, so that's negligible, and that the probability that between 71 and 100 students show up is $\Phi\left(\frac{100+\frac{1}{2}-70}{4.58}\right) - \Phi\left(\frac{71-\frac{1}{2}-70}{4.58}\right) = 0.457$, so you've got about even odds of napping. Actually, if you consider the probability of napping in class, it seems like it would be much higher.

For comparison, we can compute the relevant exact probability with the binomial distribution, that is, $\sum_{k=71}^{100} \binom{100}{k}(0.7)^k(0.3)^{100-k} = 0.462$

Note that these things are very dependent on small changes. Changing the probability of showing up to class to 0.6 changes the probability of napping to 0.01.

4.6. **Important Note.** Suppose that we scale the problem (change the problem to, say, 50 students, and we need more than 35 to nap). We will NOT get the same answer, because the standard deviation has a square root in it. That is, dividing n and the range that we're interested in (a and b) all by 2 will not give the same probability, because we will divide μ by 2 but not σ because of the square root. Keep this in mind on your homework. If you're interested in whether a probability goes up or down with grouping, go to the extreme (one big group) to develop intuition.

5. The Poisson Approximation to the Binomial Distribution (HW 4,8)

Note that if p is very small, then the normal approximation isn't good, because the probability "bunches up" around 0 (cf. p.117). Here, it's better to use a Poisson approximation, which is a discrete distribution with a single parameter μ under which $P(k\text{successes}) = e^{-\mu}\frac{\mu^k}{k!}$.

The maximum possible error of this approximation is about $p/4$.

5.1. **Simple Example.** If you are an average American, your life expectancy is 77.8 years, giving you $28,397$ days. If you buy a lottery ticket every single day of your life (you parents, being conscientious, buy it for you while you are under 18) and the odds of winning the lottery are $1/15,000,000$, what are the odds that you win the lotto at least once?

Let's use the Poisson approximation. There are $28,397$ independent trials, each with a probability of success of $1/15,000,000$, so the mean number of successes is $\mu = 0.00189313$ We want $P(0 \text{ successes}) = e^{-\mu} = 0.9981$, so your chances are about 0.00189 (it's not an accident that that number is about μ–note that the slope of e^{-x} is -1 at 0, so using a linear approximation gives $1 - x$, so if x is small, $1 - e^{-x} \approx x$).

Actually, that's better than I expected. If the jackpot is 250 million, then your average payout in this scheme is at least $-28,397 * (0.9981) + 250,000,000 * (1 - 0.0998) \approx 444,492$, which is nothing to sneeze at. I'm not suggesting that you try this out, though. Actually, what this tells you is that the odds of winning must be less or the payout must be less, because otherwise the state would lose money.

5.2. **How to Win at Roulette.** First, you need an infinite amount of money (this will fail if you have any fixed amount, no matter how large). Bet 1 dollar on the first try. From then on, if you fail, bet double your last bet. If you win, start over with 1 dollar. If at some try k you win, you will have bet $\sum_{i=0}^{k-1} 2^i = 2^k - 1$. When you win, your bet was 2^{k-1}, so you win 2^{k-1}, bringing your total to $2 \times 2^{k-1} = 2^k$, so you earn 1 dollar. Because you have an infinite amount of money, you can always keep betting until you win your dollar.

5.3. **More Complicated Example.** Sometimes, you have to make assumptions about a problem in order to get an approximate solution. Almost always, the assumption is that things are independent, when really they aren't. Here we'll see a problem where this occurs.

Also, I find that problems which end up involving the Poisson approximation tend to be a little hard to wrap my mind around. It's very important to think about what the events are and how you are applying the model.

Let's look at the birthday problem/paradox. You saw in class that with 23 people, there is a probability of more than 0.5 that two of them have the same birthday. Here are a couple ways to think about that.

5.3.1. *Way 1.* Think of the people arriving in the room in order. For the first person, there is a probability of 1 that there is one birthday on every day. For the second person, there is a probability of 364/365 that we have two distinct birthdays, and so one, so the probability that after 23 people we have 23 distinct birthdays is $\prod_{i=1}^{22}(365 - i)/365 = 0.4927$.

5.3.2. *Way 2.* Think of the pairs of people as events, and the chance that any given pair share a birthday is clearly 1/365. There are $\binom{23}{2}$ pairs, so the chance of finding at least one sharing pair in $\binom{23}{2}$ trials is approximately $e^{-\binom{23}{2}/365} = 0.5$ (rounded to the fifth decimal). The actual probability is 0.5072, so that's not too bad. The reason this is an approximation is that we are applying the Poisson approximation to the binomial distribution, but also because these events are not actually independent! If pair (a, b) and (b, c) each share a birthday, there's a pretty good chance that (a, c) shares also. However, this effect is slight.

5.4. **Another Example.** Over the course of a year, you get 2,000 spam emails. Suppose these emails arrive randomly and independently. What is the probability that on a given day you receive no spam?

A good way to think about this is that we have 2000 separate email events, and for each event there is a probability of 1/365 that the email shows up on a given day. The mean number of emails received on a day is then 2000/365 = 5.48. That is, we have 2000 trials, each with a probability of "success" of 1/365. Using the Poisson approximation, the probability of receiving no spam is then $e^{-5.48} = 0.0042$, which should explain why you often get spam.

However, if you consider the days of the year as 365 trials, each of which has a probability of success (no spam) of 0.0042, then the mean number of days on which you don't get email is 1.522, so by the Poisson approximation, the probability of 1 or more successes is $1 - e^{-1.522} = 0.782$, so you'll probably get a good day once in a while.