MA 2B RECITATION 01/12/11

1. Administrivia

The homework is now due at noon, and on Tuesday due to the holiday.

2. Warmup: Bayes's Rule

This is not explicitly on the homework due to some scheduling conflicts, but it is one of the most important things to take away from this class. Bayes's rule is a simple consequence of the definition of conditional probability, but is used almost everywhere probability is applied.

Theorem 2.1. ***(Bayes's Rule)*** *Let A and B be events, then*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

There's also a more general form of Bayes' rule in the textbook.

Example 1. You have three pancakes: one is burnt on both sides, one is burnt on one side, and one is perfectly cooked. You randomly stack the pancakes on a plate (assuming independence and identical distribution). While carrying it to the table, you look down and notice that the visible side of the top pancake is burnt. What is the probability that other side of this top pancake is burnt?

Solution. Let A be the event that the top pancake is burnt on both sides. Let B be the event that it is burnt on at least one side. To solve this, we apply Bayes's rule. We have

$$P(B|A) = 1$$

We have $P(A) = \frac{1}{3}$, because the that is the probability that we pick the doubly burnt pancake. We then compute

$$\begin{aligned} P(B) &= P(B|\text{unburnt})P(\text{unburnt}) + P(B|\text{one side})P(\text{one side}) + P(B|\text{two sides})P(\text{two sides}) \\ &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

Therefore, by Bayes's rule, we obtain

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{1 \cdot \frac{1}{3}}{2/3} = \frac{1}{2}.$$

This should sound funny to you. Where did I make the mistake? It's in the way I that I defined event A. The event "burnt on at least one side" is not the same as "topmost side of topmost pancake is burnt." Thus, we see that thoroughly understanding the problem and setting up the probability formalism properly is essential for arriving at the correct answer.

The right answer, which you can verify by explicitly writing out all the cases or defining the events properly, is $\frac{2}{3}$.

Date: January 12, 2011.

Here's another incorrect method. Since we are only interested in the behavior of pancakes that have at least one burnt side, disregard the unburnt pancake. Thus, we are now considering just two pancakes, one with probability $\frac{1}{2}$ of being burnt side up and one with probability 1 of being burnt side up. Thus, we have

$$P(\text{half burnt on top} \cap \text{burnt side visible}) = \frac{1}{4}.$$

In other words,

$$P(\text{half burnt on top} \cap \text{burnt side visible ORburnt on top}) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} = \frac{3}{4}.$$

What is wrong here? The mistake here is an incorrect application of conditional probability. By conditioning our probability on a certain event, we are effectively restricting the sample space to the subset where the event holds, but this is different from simply discarding the irrelevant events from the sample space. Watch out for these kinds of errors when you work on problems that are less "canned," because these kinds of principles are misapplied surprisingly often. □

3. Binomial Distribution

The First Question of Probability is "How likely will something happen?" or phrased in the language of probability, "What is a probability that an event A occurs?" The natural next question and the more important to us in applications is to ask "How likely will something happen again?" This question has multiple interpretations in probability, and we will essentially spend the remainder of the course learning different ways to interpret and answer this question.

Let's begin with the simplest possible version of this question. Let's study the case when we have an experiment with a random outcome and two possible outputs ("success" and "failure"). Such an experiment is called a **(Bernoulli) trial**. For instance, flipping a coin is a Bernoulli trial if we set "success" as "heads."

Our first interpretation is to study the probability that something happens *exactly* k times in n *independent* trials.

Theorem 3.1. *(Binomial Distribution) If you have n independent trials, each with equal probability p of success, then the probability of exactly k successes is*

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

where $q = 1 - p$.

The distribution of these probabilities as k varies is called the (n, p) binomial distribution, where the number of trials n and the probability of success p are the parameters.

You actually know how to solve these things already by using combinations.

Example 2. A multiple choice test has 10 questions, each with four answers. Assume that the correct answers are distributed independently and identically. If you randomly guess an answer every question, what is the probability that you get exactly 7 questions correct?

The answer is

$$\binom{10}{7} \left(\frac{1}{4}\right)^7 \left(\frac{3}{4}\right)^3,$$

because there are $\binom{10}{7}$ ways to answer 7 out of 10 questions correctly and the probability of such a string is $\left(\frac{1}{4}\right)^7\left(\frac{3}{4}\right)^3$. As you can see, this directly translates from our formula.

What is the probability that you get 80% or higher?

$$\binom{10}{8}\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^2+\binom{10}{9}\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)^1+\binom{10}{10}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^0.$$

Remark 3.2. The binomial distribution has a **mean** (expected number of successes) of np and a **mode** (most likely number of successes) of $\lfloor np+p\rfloor$. These two averages are distinct because the Binomial distribution is discrete.

For instance, if we flip an unfair coin, say, if the probability of getting heads was 0.55, then we expect 0.55 of a head, but since you cannot get fractions of a head on a single flip, the most likely number of heads is 1.

When do the mean and the mode coincide? When np is an integer and $p < 1$.

4. Drawing Without Replacement

Example 3. Suppose that we have 10 balls, 3 of which are red. If we draw five balls without replacement, what is the probability of getting exactly one red ball? The probability of getting a red ball on the first draw and non-red balls on the rest is

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

This is obviously different than if we did it with replacement.

The general rule is that if we have a sample of size N with G and B bad, we have

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

where $G+B=N$ and $g+b=n$.

5. "Distributions"

This is generally the part of probability where lots of jargon gets thrown your way, so let's take some time to process the things that you are expected to know.

Question 5.1. *What is a distribution?*

In probability, a distribution is a function that describes the probability of a random variable taking certain values. It usually refers to one of two things: the cumulative distribution function or the probability density function.

The cumulative distribution function (CDF) is a function Φ such that $P(X \leq x)=\Phi(x)$. A probability density function (PDF) is a function ϕ such that $\Phi(z)=\int_{-\infty}^{z}\phi(x)\,dx$. Which distribution a problem is referring to is usually clear from context. For example, $P(X \leq x)\to 1$ as $x\to\infty$.

The most important PDF is the one for the normal distribution:

$$\phi(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2/\sigma^2},$$

where the parameters are the mean μ and the standard deviation σ. Graphing this function gives you the normal curve. Indeed, one of the distinguishing marks of a probabilist or statistician is their ability to freehand a normal curve.

The **standard normal distribution** is the normal distribution where $\mu = 0$ and $\sigma = 1$, and Φ will denote the CDF for the standard normal distribution. In practice, we only use the standard normal distribution for calculations, because for any normal distribution, the probability that a random variable lands between a and b (with $a < b$) is

$$\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Remark 5.2. Trying to compute Φ is difficult, but we can approximate it numerically. There are tables in the back of your book and a number of ways to calculate it on the computer. For instance, in Mathematica, the command is `CDF[NormalDistribution[], x]`. However, I recommend that everybody at least learn how to use the normal table in the back of the book, because you probably won't be allowed to use computers for your exams.

6. Normal Approximation of the Binomial Distribution

You want to use the **normal approximation** to the binomial distribution when the number of trials is large. As you may have noticed, if you look at graphs of the binomial distribution with increasing n but fixed p, the curves of the distribution all seem to be a approaching a curve that looks a lot like the normal curve. This because of law of large numbers and the miraculous properties of the normal distribution (in particular, the Central Limit Theorem, which we will discuss in later weeks).

Theorem 6.1. *If we have an (n, p) Binomial distribution, then*

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

where μ is the mean np and σ is the standard deviation $\sqrt{npq}$. We will learn why this is the right number for the standard deviation later.

Example 4. Caltech undergrads show up to a certain class independently from each other with probability 0.7. Suppose that a class has 100 students enrolled. If fewer than 20 students show up to class, 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 is the probability that people eat and sleep rather than learn?

Solution. 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, 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 it's unlikely that you'll get a free lunch out of it. However, the probability that between 71 and 100 students show up is

$$\Phi\left(\frac{100+\frac{1}{2}-70}{4.58}\right) - \Phi\left(\frac{71-\frac{1}{2}-70}{4.58}\right) \approx 0.457,$$

so there's a good chance that you'll be able to take a nap!

For comparison, the exact probability with the binomial distribution is

$$\sum_{k=71}^{100} \binom{100}{k} (0.7)^k (0.3)^{100-k} = 0.462.$$

□

Remark 6.2. *Warning!* Note that these things are very dependent on small changes. Changing the probability of showing up to even 0.6 changes the probability of napping to 0.01.

Remark 6.3. *Important!* The behavior described above *does not scale down well* (for instance for 50 students and 35 students needed to nap the probabilities are completely different), because the standard deviation has a square root in it. Keep this in mind for your homework.

7. Poisson Approximation of the Binomial Distribution

You want to use the **Poisson approximation** when the number of trials is large, and the probability p of success is *small.* (A good mnemonic: "the s in Poisson is for small").

It's better to use the Poisson approximation as opposed to the normal approximation because the probability "bunches up" around 0 (see Pitman, p. 117), because it's not allowed to take values less than 0.

The Poisson distribution is a discrete distribution that only takes in one parameter: the mean μ. It is given by

$$P(k \text{ successes}) = e^{-\mu} \frac{\mu^k}{k!}$$

Events involving the Poisson distribution tend to be counterintuitive, even when you see the math, probably because we are not very used to thinking about very low probability events, so be careful about using your intuition on these types of problems. Not surprisingly, "get-rich-quick" schemes are a fertile source of examples of this type.

Example 5. If you are the average American, your life expectancy is 77.8 years, giving you 28,397 days. If you fill out a lottery ticket every single day of your life and the odds of winning the lottery are 1/15,000,000, what are the odds that you win the lotto at least once in your life?

Solution. There are 28,397 independent trials, each with probability 1/15,000,000 of success, so the mean number of successes is $\mu = np = 0.00189313$. We have

$$P(0 \text{ successes}) = e^{-\mu} = 0.9981,$$

so our chances of winning are about 0.00189. □

Remark 7.1. It's not just a coincidence that the number is approximately μ. Note that the slope of e^{-x} is -1 at 0, so using a linear approximation gives us $1 - x$, so if x is small, then $1 - e^{-x} \approx x$.

Actually, our answer is a lot better than expected. If the jackpot is 250 million, then your average payout in this scheme is at least

$$-28,397 \cdot (0.9981) + 250,000,000 \cdot (1 - 0.0998) \approx 444,492$$

which isn't bad. However, what this should tell you is that the odds of winning must be less or the payout must be less, because otherwise the lottery wouldn't raise money.