MA 2B RECITATION 01/19/11

1. Comments

Most of you are doing well on the homeworks, which is good to know. However, there were some common errors that I think should be addressed.

- First, make sure that you write words and justify your reasoning. If the grader can't understand what you're doing, it is going to be marked wrong, and that goes double for the midterm. For instance, if you're Poisson approximation, state that you are using and say why it is appropriate for your situation (e.g. approximating many binomially distributed trials with low probability of success).
- Second, don't misuse the word "random." The textbook is a bit sloppy with this, and seems to assume that "random" events are what we call IID, which stands for "independent and identically distributed," and uniformly distributed. Of course, this is the case for dice rolls and coin flips and other simple examples in discrete probability, but you will get in trouble if you keep assuming these things, especially when we move to continuous probability. In these notes, I make an effort to be careful in the examples and state the conditions carefully, and I suggest that you do the same.
- Finally, make sure to *state your assumptions*. For instance, many people got points off on the last problem because they did not state that they assumed that the events were independent, and even when the problem explicitly said to state your assumptions. I'll say it again, *do not, as a rule, assume events are independent* unless the problem explicitly states it, or your are doing a simple discrete probability example like coin flips or dice rolls.

Again, I am generally happy with how people are doing on the homeworks, but if you made any of the above errors, it's better to correct it now when the stakes are low to get it wrong on an exam.

The primary difficulty with this next part of probability is just getting comfortable with the language of random variables and distributions and with operations that we'll perform on random variables. Get some practice and become comfortable with them, because the jump from discrete to continuous tends to a point where many people get lost.

2. Random Variables

Random variables are objects in probability used to represent the value of some event. It's tempting to give some kind of wishy-washy explanation of what a random variable "is," but for your first exposure, it's probably best to just see an example.

Example 1. If X is the random variable representing a fair coin flip, then

$$P(X = \text{heads}) = \frac{1}{2}.$$

Date: January 19, 2011.

Example 2. Suppose X is the random variable representing a poker hand. Then

$$P(X = \text{flush}) = \frac{4 \cdot \binom{13}{5}}{\binom{52}{5}}$$
$$P(X = \text{royal flush}) = \frac{4}{\binom{52}{5}}.$$

At first glance, it seems like we're just saying commonsense things in a abstract way, but it turns out that doing this sort of thing actually makes many things clearer once we move beyond toy examples.

The definition I gave above is necessarily vague, and we need a little more mathematical framework to give the actual definition. It's a little abstract, but it's worthwhile to understand what's really going on "behind the scenes," because knowing the precise definition of what you are studying will be helpful, especially once begin to use random variables in interesting ways.

2.1. **What is a Random Variable, Really?** Why are random variables defined in such a particular way? Why do probabilists think that this is the best way to express probabilistic statements? What's so great about this formalism? I'll give a brief glimpse as to how the experts think about this stuff, and hopefully you will be able to understand the motivation for each of the structures involved.

We begin in an abstract setting. It is often useful to have a common notion of "size" of a subset of a space. For instance, in the real line $\mathbf{R}$, a good notion of "size" of an interval $(a, b) \subset \mathbf{R}$ is $b - a$. Similarly, a good notion of "size" in $\mathbf{R}^2$ and $\mathbf{R}^3$ are area and volume.

Any set can be made into a "measure space" by defining some "size function" on subsets of the space. This size function is called a **measure**. For instance, a common size function (which mathematicians call the "standard measure" or "Lebesgue measure" on $\mathbf{R}$) μ on $\mathbf{R}$ is the one described above, that takes intervals (a, b) to $b-a$. Ditto for closed or half-closed intervals. Another example, defined for any set S, is a function ν that takes any subset $A \subset S$ and returns its cardinality. For instance, if $S = \{1, 2, 3, 4, 5\}$, then $\nu(\{2, 3, 5\}) = 3$. Therefore, a measure space is defined to be some pair (S, μ) where S is a set and μ is just some way of measuring the "bigness" of subsets of S.

Quick quiz. What's the (standard) measure of $(-5, -3) \subset \mathbf{R}$? It's $-3-(-5) = 2$. What about $[1, 3] \cup [4, 5)$? The intervals are disjoint, so it's $2 + 1 = 3$.

What's the measure of $3 \in \mathbf{R}$? It's 0. What about the set $\{3, 4\}$? Still zero. What about the integers $\mathbf{Z} \subset \mathbf{R}$? Still zero. What about all the rational numbers $\mathbf{Q} \subset \mathbf{R}$? Believe it or not, it's zero. As you can see, continuous probability is nonintuitive, because infinite sets, and even sets that are "everywhere" in $\mathbf{R}$ like $\mathbf{Q}$ (the mathematical definition is that $\mathbf{Q}$ is "dense" in $\mathbf{R}$) are not just small, but actually have no size at all! In other words, every number you've ever seen in your whole life doesn't even hint at enormous hidden mass that lies beyond the surface. This is why it's good to learn have some formalism, some kind of abstract way of dealing with probability, because it allows us to get a handle on things that are difficult and nonintuitive as the size of infinite subsets of infinite sets.

The real definition of a random variable X is that it is a *function* from a measure space Ω to a measure space M:

$$X : \Omega \to M$$

The destination space is usually what you think of as "event space" or "sample space." The probability that a random variable X is in some subset $B \subset M$ is the measure of the preimage of B in Ω, that is,

$$P(X \in B) := \text{measure}(\{x \in \Omega \mid X(x) \in B\}).$$

It's a little hard to take this all in at once if you're never seen anything like this before, but we're just saying something totally obvious in a mathematically rigorous way. It's probably best to see this illustrated in an example.

One thing to note is that since we're working in probability, we need define the spaces and measures and functions so that our definitions make sense.

Example 3. Let X represent the event of randomly picking a number from 1 to 10 with equal probability. How would we represent this space?

What should our sample space be? Consider the set $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

How should we define our measure? Counting measure! But wait, then $P(X \in S) = 10$, which doesn't make sense. Thus, a good way to define the probability here is take the "normalized" counting measure given by taking the value given by the usual counting measure and dividing it by the total size of the sample space.

If we do this, things start to make sense.

$$P(X \in \{1\}) = \frac{1}{10}, P(X \in \text{even}) = \frac{5}{10} = \frac{1}{2}, P(X \in \{2, 5, 7\}) = \frac{3}{10}.$$

Thus, we see that the idea of random variables, despite being a little scary when spoken in rigorous mathematical language, is not that difficult once you strip it down and see what's really going.

3. Joint Distribution and Functions

A **joint distribution** is just the distribution of a pair (X, Y) of random variables. Note that we can recover the original distribution of X or Y by summing over all possibilities for the other variable, just like we did for previous examples in discrete probability. We can also take distributions of more than two variables in a similar fashion.

We can also consider functions of random variables. For instance, if X is the number rolled on a die, what is the distribution of X^2? We have some useful facts about functions of random variables.[1]

- Functions of independent random variables are independent.
- Disjoint blocks of independent random variables are independent. For instance, if X_1, X_2, X_3, X_4 are independent, then (X_1, X_2) and (X_3, X_4) are independent.
- Functions of disjoint blocks of independent random variables are independent, e.g. if X_1, X_2, X_3, X_4 are independent, then X_1X_2 and X_3X_4 are independent.

Example 4. (Indicator functions) Let X_i be the outcome of flipping a coin on the ith trial, where $X_i = 1$ if the ith coin is heads and $X_i = 0$ if the ith coin is tails. Then the number of heads in five flips is given by

$$Y = X_1 + \cdots + X_5.$$

[1] Pitman, p.154

Observe that the distribution of Y is the Binomial(5, 0.5) distribution.

Example 5. (Joint distribution) Let X and Y random variables representing integers chosen at random from 1 to 100 inclusive. What is the probability that $X \times Y \leq 50$?

We are asking about a probability dealing with the ordered pair (X, Y), so we want to consider the joint distribution of X and Y. Let A be the set of all ordered pairs (X, Y) such that $X \cdot Y \leq 50$. Note that we can write our answer as

$$P((X,Y) \in A) = \sum_{k=1}^{100} P((X,Y) \in A | Y = k) P(Y = k).$$

We have

$$P((X,Y) \in A | Y = k) = P(X \leq 50/k) = \frac{\lfloor 50/k \rfloor}{100}.$$

Therefore,

$$P((X,Y) \in A) = \sum_{k=1}^{100} \frac{1}{100} \frac{\lfloor 50/k \rfloor}{100} \approx 0.0207.$$

4. Expectation

The expectation of a random variable is exactly what it sounds like, it is the value that the (numerically-valued) random variable is expected to take. For a random variable X, the expectation is

$$E(X) = \sum_x x P(X = x).$$

Recall that the mean of a distribution P is given by

$$\mu = \sum_x x P(x).$$

Therefore, if a random variable represents an event with a given distribution, the expected value of the variable is the mean of the distribution. But note that this *does not mean that expectation is the same as probability.* We will see an example of this later.

Expectation has some nice properties. Here are some basic ones. There are more properties on pages 180–181 of Pitman.

- Expectation is a linear function. That is,

  $$E(X_1 + \cdots + X_n) = \sum_i E(X_i).$$

  You don't even need to assume that the X_i are independent!
- If X is an indicator function for an event A, then $E(X_1) = P(A)$.
- (Tail sum formula) If X takes values in $\{0, 1, \ldots, n\}$, then

  $$E(X) = \sum_{j=1}^{n} P(X \geq j).$$

- If g is a function of a random variable X, then

  $$E(g(X)) = \sum_x g(x) P(X = x).$$

Remark 4.1. Warning! It is NOT generally true that $E(g(X)) = g(E(X))$.

- If X and Y are *independent*, then

$$E(XY) = E(X)E(Y).$$

Note in particular that the additivity of expectation can simplify a lot of calculations.

Example 6. If I roll 5 dice, what is the expected total? We want to find $E(X_1 + X_2 + X_3 + X_4 + X_5)$ where X_i is the value of the ith die. By the tail sum formula,

$$E(X_i) = \frac{1}{6}\sum_{i=1}^{6} i = 3.5.$$

By the addition formula for expectation,

$$E(X_1 + \cdots + X_5) = E(X_1) + E(X_2) + \cdots + E(X_5) = 5 \cdot 3.5 = 17.5.$$

Example 7. Suppose you flip a coin. The expected number of heads is $\frac{1}{2}$. The probability of getting heads is also $\frac{1}{2}$. These numbers are the same, but *expectation is NOT probability.*

For instance, if you flip a coin twice, what is probability of getting at least one head? Since expectation is additive, the expected number of heads in two coin flips is 1. However, the probability of seeing at least one head is $\frac{3}{4}$.

Be especially careful when expectation is between 0 and 1. If you're sloppy with this distinction, you'll end up making many mistakes and come to some convoluted conclusions.

So most of your homework is consists of practicing many different applications of these concepts. Now, let's look something fun and strange.

Example 8. You saved the life of a rich probabilist by skillfully applying what you learned in your Ma 2b recitation. To thank you, you are presented with two indistinguishable suitcases, each of which contains a positive sum of money.

One suitcase contains twice as much money as another. You can pick one suitcase and keep whatever amount it contains.

You pick a suitcase at random, but before you open it you are offered the possibility to pick the other one instead. What should you do? Should you take the offer and switch or just take the suitcase you chose? Think about it for a little bit before you go on.

An argument for switching? Let A denote the amount of money in my chosen suitcase. The probability that A is the smaller amount is $\frac{1}{2}$ and the probability that A is the larger amount is also $\frac{1}{2}$. The other suitcase contains either $2A$ or $A/2$ dollars. Thus, the expected value of the money X in the other suitcase is

$$E[X] = \frac{1}{2} \cdot 2A + \frac{1}{2} \cdot \frac{A}{2} = \frac{5}{4}A.$$

This is greater than A, so you gain on average by swapping. So you switch.

Now, here's a twist. Suppose that after you pick the other suitcase but before you open it, you are offered the possibility to switch again.

After the first switch, let B denote the amount of money in the suitcase and reason in exactly the same way. The most rational thing to do is to swap again.

Therefore, if the rich probabilist offers me a choice and if I am rational, I will end up swapping suitcases forever, a strange Dante-esque Hell for greedy people. Maybe you should not have switched at all?

Q: Find the flaw in the line of reasoning above.

One explanation is that when calculating the expectation, A is effectively used to denote two different quantities. This error is highlighted if we let M denote the smaller amount and $2M$ the larger amount, then reconsider what happens.

If $A = M$, then the other envelope contains $2A$ (or $2M$).

If $A = 2M$, then the other envelope contains $A/2$ or M.

Each of these steps treats A as a random variable, but calculating the expectation continues to use A as a fixed variable, still equal in every case. In other words, $2A$ is supposed to represent the amount in the envelope if $A = 2M$, and $A/2$ is supposed to represent the value if $A = M$.

We can't continue using the same symbol A under these two incompatible assumptions, because then $A = M = 2M$ which implies that $1 = 2$ if A is nonzero (per our assumption).

OK. But by tweaking our problem slightly, we get something even stranger.

What if you are allowed to look in the first suitcase before being offered to switch? Then A really is a fixed variable, and the explanation above breaks down. In this case, you will not end up switching envelopes indefinitely, since the contents of both envelopes are known after the first switch. However, there is still something strange going on.

Suppose that your twin (or doppelgänger) opens the other suitcase without telling you how much it contains. Thus, our variable is fixed. Nevertheless both you and your twin will find it's better to switch by the same argument as above. This is puzzling since you and your twin can't both win when switching envelopes. What's going on?

Here's another way to look at it. Let's take off our probability glasses and just try and reason logically. The two following arguments lead to conflicting conclusions:

- Let the amount in the suitcase you chose be A. By swapping, the player can gain A or lose $A/2$. Thus, the potential gain is strictly greater than the potential loss.
- Let the amount in the envelopes be Y and $2Y$. Now by swapping, the player may gain Y or lose Y. So the potential gain is equal to the potential loss.

How on earth do you analyze a seemingly simple situation like this? Actually, this question is still an actively researched topic in probability theory, often called the "two envelope problem." So we see that even with the current language of probability, with all its strengths and successes and amazing applications, still has its limitations. There are still mysteries out there in probability land that we don't know how to approach. While probability theory as studied in this class is known and codified and essential for any person to know, it is far from the final word. Who knows? Maybe you could be the one to offer a new way of viewing probability that can be applied to easily analyze problems like this.