Probability Review for IEOR 166

Shehzad Wadalawala

Spring

oduc io

$\cup$ = union = "or"; $A \cup B$ = all outcomes in A or B (or both)
$\cap$ = intersection = "and"; $A \cap B$ = all outcomes in both A and B

What is $\Pr\{A \cup B\}$? Is it always $\Pr\{A\} + \Pr\{B\}$?

Consider the following events
A: the stock market goes up at least 1 percent
B: the stock market goes up at least 2 percent
Note that $A \cup B = A$, so $\Pr\{A \cup B\} = \Pr\{A\}$;

Suppose $\Pr\{A\} = 0.2$, $\Pr\{B\} = 0.3$; if we add them, we get 0.5 which is not $\Pr\{A \cup B\} = \Pr\{A\}$.
Even worse, suppose $\Pr\{A\} = 0.8$, $\Pr\{B\} = 0.9$; if we add them, we get a "probability" over 1 !

What went wrong? We double-counted some outcomes, such as the market going up 3 percent.

Which outcomes did we count twice? The ones in both A and B: $A \cap B$.

We'll fix it by subtracting out one copy of them:

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}$$

2 Co di io al P obabili y

How can we calculate $\Pr\{A \cap B\}$? Well for $A \cap B$ to happen, we need A to happen, with probability $\Pr\{A\}$.

Then, knowing that A happens, our probability that B happens is $\Pr\{B|A\}$.

Multiply these two [they are independent] to get

$$\Pr\{A \cap B\} = \Pr\{A\} \cdot \Pr\{B|A\}$$

Divide to get another useful formula:

$$\Pr\{B|A\} = \frac{\Pr\{A \cap B\}}{\Pr\{A\}}$$

Let's try an example:
What is your probability that Person X's birthday is in December? 1/12 or 31/365, right?
Suppose I told you it's in the last quarter of the year. What is your probability now? 1/3 (or $31/(31+30+31)$)
Let A be the event "birthday in December", and B be the event "birthday in last quarter" ($p = 1/4$ or so)

Then

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\} = 1/12}{\Pr\{B\} = 1/4} = 1/3$$

3 Pa i io i g

To calculate a difficult probability, we often assume we have more info than we really do, then weight the result based on the probability of that info being true.
My roommate is going to the video store; what are the chances he'll pick a movie I like?
Well, he might pick an action flick, comedy, drama, or "other". Here are my probabilities:

	Pick	Match
a	0.3	0.5
c	0.5	0.2
d	0.1	0.9
o	0.1	0.5

Multiply them to get $0.15 + 0.10 + .09 + .05 = 0.39$
Make sure the partitioning includes everything, exactly once!

4 Bayes' Rule

$$\Pr\{C|B\} = \Pr\{B|C\}\frac{\Pr\{C\}}{\Pr\{B\}}$$

It's almost like a conversion: inches = cm * (inches/cm)
Consider an alcohol test for drivers that has a 5% false-positive rate, and a 10% false-negative rate. Suppose 15% of drivers are drunk. Given that you fail the test, what is the probability that you were drunk? Let B be failing the test, and C be drunkenness.
It looks like we need to calculate $\Pr\{B\}$: we do this by partitioning (aka conditioning):

$$\Pr\{B\} = \Pr\{\text{drunk}\}\Pr\{\text{fail}|\text{drunk}\} +$$
$$\Pr\{\text{sober}\}\Pr\{\text{fail}|\text{sober}\}$$
$$= .15 \cdot .90 + .85 \cdot .05 = .1775$$

Now we have

$$\Pr\{\text{drunk}|\text{failed}\} = \Pr\{\text{fail}|\text{drunk}\} \cdot \Pr\{\text{drunk}\}/0.1775$$

$$0.90 \cdot 0.15/0.1775 = 0.7606$$

Is this enough to convict? Is there "reasonable doubt"?
Is the probability of being drunk really 0.15? It depends on how you were stopped: randomly, or for driving poorly. Does this make sense? Consider 10000 drivers. Separate them into drunk and sober:

```
8500 sober      1500 drunk
 425 fail       1350 fail
```

Given that you're in the "fail" group, it does look like about 3 out of 4 of them are drunk.

5 Ra dom Va iables

For those times when a random event has a numerical value associated with it, we call it a random variable (RV). We usually use capital letters like X, Y, and Z; I write them as Roman letters to distinguish them from lowercase. Two fundamental kinds: discrete (usually integer) and continuous. For a discrete RV, we describe it by the probability of each value. This is the probability mass function, pmf:

$$p_X(x) = \Pr\{X = x\}$$

Since the RV has to equal *something*, the sum of these values must be 1:

$$\sum_{x \in S} p_X(x) = 1$$

For a continuous RV, we describe it by the probability that the RV will be near a value. We call this the probability density function (pdf):

$$p_X(x)dx \approx \Pr\{x \leq X \leq x + dx\}$$

Again, since the RV must fall somewhere, the "sum" (integral) of the pdf must be one:

$$\int_{-\infty}^{\infty} p_X(x)dx = 1$$

Many of our continuous RV's are non-negative, so we can integrate from 0 instead of $-\infty$.
Both discrete and continous RV's have a Cumulative Distribution Function, CDF. This is

$$P_X(x) = \Pr\{X \leq x\}$$

For a discrete RV, this is a step function. For a continuous RV, this is a continuous function. It goes from 0 at $-\infty$ to 1 at $+\infty$ (if not before).
Expected value: expected long-run average of many samples of the random variable.
Suppose I have N samples of a discrete RV, and I average them:

$$\frac{1}{N}(X_1 + X_2 + \cdots X_N)$$

Let's consider that sum: how many of them will be equal to 1? I'd say $N \cdot p_X(1)$. How many of them will be equal to 2? I'd say $N \cdot p_X(2)$. So, let's group them:

$$X_1 + X_2 + \cdots X_N =$$

$$1 \cdot Np_X(1) + 2 \cdot Np_X(2) + \cdots$$

$$\sum_{k=-\infty}^{\infty} kNp_X(k)$$

When we divide by N, we get that the long-run average is

$$\mathrm{E}[X] = \sum_{k=-\infty}^{\infty} kp_X(k)$$

Similarly, for continuous RV, we do an integral instead of a sum:

$$\mathrm{E}[X] = \int_{k=-\infty}^{\infty} xp_X(x)dx$$

I'm going to write out all the terms in the expected value sum:

$1p(1)$	$2p(2)$	$3p(3)$	$4p(4)$	
$p(1)$	$p(2)$	$p(3)$	$p(4)$	$1 - P(0)$
	$p(2)$	$p(3)$	$p(4)$	$1 - P(1)$
		$p(3)$	$p(4)$	$1 - P(2)$
			$p(4)$	$1 - P(3)$

Now, the usual expression sums down the columns first. But we should get the same answer if we sum along the rows. What's the sum of the first row? It's the probability that X is 1 or greater: $1 - P(0)$. The sum of the second row is the probability that X is 2 or greater: $1 - P(1)$. And so forth. We sum these rows to get

$$\mathrm{E}[X] = \sum_{k=0} (1 - P_X(k))$$

6 Expec ed Value of a Fu c io

To take the expected value of $Y = f(X)$, we can't just take $f(\mathrm{E}[X])$; we have to re-do the summation or integration:

$$\mathrm{E}[Y] = \sum_{k=-\infty} \infty f(k)p_X(k)$$

$$\mathrm{E}[Y] = \int_{x=-\infty} \infty f(x)p_X(x)dx$$

Using these, we can see that

$$\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$$

even if X and Y are dependent. This is very useful.

7 Va ia ce

If we have a random *variable*, how much does it vary? We could measure this by the expected difference from its mean, but that turns out to be zero. Instead, we square the difference:

$$\mathrm{Var}\,(X) = \mathrm{E}\left[(X - E[X])^2\right]$$

Remember, this is not the same as $(\mathrm{E}\,[X - \mathrm{E}\,[X]])^2$.
If we multiply an RV by a constant, the variance changes; if we add a constant, it doesn't:

$$\mathrm{Var}\,(aX + c) = a^2 X$$

If two variables are independent, then

$$\mathrm{Var}\,(X + Y) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y)$$

When consulting, we don't quote the variance to our clients, since it is hard to interpret.
To get a measure of variability that's in the same units as the RV, we take the square root of the variance. This is the "standard deviation".

8 Joi P obabili ies

Suppose someone conducted a study and found pairs of adult brothers and sisters, and recorded their heights. One could make a two-dimensional histogram, plotting how many pairs (by percent) had each height. This would give you an estimate of the probability $\Pr\{X = x, Y = y\}$ where X is the sister's height, and Y is the brother's. This is the Joint Mass Function (or Joint Density Function).
Just like a single-variable PMF, it should sum to 1.
If you know that I am 6 feet tall, what is your pmf for my sister's height? This is a conditional pmf: $p_{X|Y}(x|y)$. It also has to sum (over x) to 1.
The other conditional pmf is: knowing my sister's height, what is mine? This is $p_{Y|X}(y|x)$.

9 Ma gi al Dis ibu io

Given the joint pmf, and not knowing how tall I am, what is your pmf for my sister's height? This is written, using the law of total probability,

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

and this is the marginal pmf for X. Y is similar.
Independence: do you expect these to be independent? If I told you the average of my mom & dad's heights, would my height be independent from my sister's? This is "conditional independence".