CS188 Fall 2015 Probability Section 2

1 Marginalization

S	R	$P(S,R)$
s	r	0.05
s	$\neg r$	0.75
$\neg s$	r	0.10
$\neg s$	$\neg r$	0.10

Consider the binary random variables, S and R, from last week. Their joint distribution is given in the table above. Calculate the following quantities:

(a) $\sum_{r' \in (c, \neg c)} P(s, r') = 0.05 + 0.75 = 0.8$

(b) $\sum_{r' \in (c, \neg c)} P(\neg s, r') = 0.1 + 0.1 = 0.2$

(c) $\sum_{s' \in (c, \neg c)} P(r, s') = 0.05 + 0.1 = 0.15$

(d) $\sum_{s' \in (c, \neg c)} P(\neg r, s') = 0.75 + 0.1 = 0.85$

Last week, we calculated marginal probabilities by directly looking at the probability table. What we just did is a more algorithmic way of doing the same thing. In general, for two random variables X and Y, the formula for calculating the marginal probability for X is $P(X) = \sum_{y'} P(X, y')$. Extending this formula to multiple random variables is not difficult.

2 Product Rule and Chain Rule

Suppose that, if we randomly choose a student, the probability that they like to play basketball is 0.01. Now, suppose that, if we randomly choose a student that likes to play basketball, the probability that they are tall is 0.3. In other words, the probability that a student is tall *given that* they like to play basketball is 0.3.

(a) Intuitively, would you expect the probability that a student likes to play basketball *and* is tall to be smaller or larger than 0.01? Why?

We expect the probability to be lower, since 0.01 is the probability that the students like to play basketball, and out of all students that like to play basketball, only a fraction of them are tall.

(b) Now, calculate the probability that a student likes to play basketball and is tall.

We can use the definition of conditional probability here to get:

$$P(tall \mid likes\_basketball) = \frac{P(tall, likes\_basketball)}{P(likes\_basketball)}$$

From filling in known quantities, we find that $P(tall, likes\_basketball) = 0.003$.

Consider two binary random variables, L and T. L takes on values l and $\neg l$, corresponding to whether or not you're late for work. T takes on values t and $\neg t$ and corresponds to whether or not there's a traffic jam. So, l, t means that it's you're late for work and there's a traffic jam. We are given the following probability tables:

T	$P(T)$
t	0.4
$\neg t$	0.6

L	T	$P(L \mid T)$
l	t	0.8
l	$\neg t$	0.25
$\neg l$	t	0.2
$\neg l$	$\neg t$	0.75

(c) Calculate $P(L, T)$ from the tables given.

L	T	$P(L, T)$
l	t	0.32
l	$\neg t$	0.15
$\neg l$	t	0.08
$\neg l$	$\neg t$	0.45

What we've seen so far is two examples of the product rule, which states that, for two random variables X and Y, their joint probability can be calculated as $P(X, Y) = P(X)P(Y \mid X)$. This rule can be derived directly from the definition of conditional probability.

Now suppose that we add another random variable M representing whether or not your boss is mad, and in addition you are given the following table:

L	T	M	$P(M \mid L, T)$
l	t	m	0.95
l	t	$\neg m$	0.05
l	$\neg t$	m	0.75
l	$\neg t$	$\neg m$	0.25
$\neg l$	t	m	0.40
$\neg l$	t	$\neg m$	0.60
$\neg l$	$\neg t$	m	0.10
$\neg l$	$\neg t$	$\neg m$	0.90

(d) Calculate $P(L, T, M)$ from the tables given.

L	T	M	$P(L, T, M)$
l	t	m	0.304
l	t	$\neg m$	0.016
l	$\neg t$	m	0.1125
l	$\neg t$	$\neg m$	0.0375
$\neg l$	t	m	0.032
$\neg l$	t	$\neg m$	0.048
$\neg l$	$\neg t$	m	0.045
$\neg l$	$\neg t$	$\neg m$	0.405

What we did here was treat (L, T) as one random variable, and then the joint probability followed from the product rule. In general, the product rule can be generalized to any arbitrary number of random variables as what is known as the chain rule:

$$P(A_1, \ldots, A_n) = \prod_{i=1}^{n} P(A_i \mid A_1, \ldots, A_{i-1})$$

So the previous example could have been solved by writing:

$$P(M, L, T) = P(T)P(L \mid T)P(M \mid L, T) = P(L, T)P(M \mid L, T)$$

One important property of the product rule and chain rule is that they are both *commutative*, which means that any ordering of the random variables is valid. Generally, you'll find that some orderings are easier than others, based on the information you're given.

3 Normalization Trick

Suppose you have a bag that contains 10 balls of different colors. You pull all of the balls out of the bag and find the following numbers for each color:

Color	Number
red	4
blue	3
green	1
black	2

Now, you want to figure out what the probability of picking a ball of a certain color is. To do this, you fill in the following probability table:

Color	Number
red	0.4
blue	0.3
green	0.1
black	0.2

This may seem trivial, but there's actually two ways you could have done this. One way is to go through each entry in the table, and figure out the probability of that entry individually. The other way is to take the original table, and simply *normalize* it so that the entries sum to one. This is what is known as the *normalization trick*, and it will come in handy when we look more in depth at conditional probabilities. The normalization trick states that, for a table that gives a set of frequencies that don't necessarily sum to one, we can normalize it to get a valid probability distribution.

Let's consider again the probability table from Part 1:

S	R	$P(S, R)$
s	r	0.05
s	$\neg r$	0.75
$\neg s$	r	0.10
$\neg s$	$\neg r$	0.10

Using the normalization trick, calculate $P(S \mid r)$ and $P(S \mid \neg r)$.

S	$P(S \mid r)$
s	0.33
$\neg s$	0.67

S	$P(S \mid \neg r)$
s	0.88
$\neg s$	0.12

Starting from the joint distribution, we can simply grab the first and third rows, and then normalize them to get our answer for part (a). For part (b), we grab the second and fourth rows and normalize.

4 Bayes' rule

Recall the product rule that we saw earlier:

$$P(A, B) = P(A \mid B)P(B)$$

Notice that we can also use commutativity to obtain:

$$\begin{aligned} P(A, B) &= P(B, A) \\ &= P(B \mid A)P(A) \\ \rightarrow P(A \mid B)P(B) &= P(B \mid A)P(A) \end{aligned}$$

This leads us to Bayes' rule, which is written as follows:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

For two random variables A and B. This rule in general can be very useful, because it relates conditional probabilities together and allows us to calculate one from the other.

As an example, let's take a look at one variant of what is commonly known as the false positive paradox. In a population of 1000 people, 2% have a deadly disease. You are administering a test for this disease, which has a false positive rate of 5% (i.e., it tests positive when a person doesn't have the disease 5% of the time) and a false negative rate of 0%. Supposing an individual tests positive for the disease, what is the probability that they actually have the disease?

There are many ways to solve this problem, but let's use Bayes' rule to figure it out. Let T be a random variable indicating whether or not the person tests positive, and D indicate whether or not the person actually has the disease. We are looking for $P(d \mid t)$. Using Bayes' rule, we can write this as:

$$\begin{aligned} P(d \mid t) &= \frac{P(t \mid d)P(d)}{P(t)} \\ &= \frac{1 * 0.02}{P(t)} \end{aligned}$$

So, after applying all of the known quantities, all we need to figure out now is $P(t)$. However, using the law of total probability, i.e., marginalization and the product rule, we can find that:

$$\begin{aligned} P(t) &= \sum_{d' \in (d, \neg d)} P(t, d') \\ &= \sum_{d' \in (d, \neg d)} P(t \mid d')P(d') \\ &= 1 * 0.02 + 0.05 * 0.98 = 0.069 \end{aligned}$$

So, our final answer is $P(d \mid t) = \frac{0.02}{0.069} = 0.29$. This is called a paradox because it's surprising that the probability is so low given that the test, which is quite accurate, comes up positive. But with a simple application of Bayes' rule, we can see how this all works out.