Using conditional relationships between events to make inferences

Jay McClelland
SymSys 100, Spring 2009

In this document we will consider probabilistic relationships between events and the use of knowledge of such relationships to make inferences. We begin with some important definitions. This document tries to unpack some of the material in the *Wikipedia* article on *Bayesian inference*, which can be consulted if you'd like to read about these ideas a second time from a slightly different angle. Also try *http://yudkowsky.net/rational/bayes*

Independence of events: probabilities and conditional probabilities

We think of two events as independent if knowing something about one tells you nothing about another. For example, suppose I have two dice, and I roll them one after the other. Intuitively the outcome of the role of the second die is independent of the outcome of the role of the first. Learning what number came up on the first tells me nothing about what number might come up on the second (and *vice versa*).

The following 2x2 contingency table contains data that illustrates a situation in which two events appear to be independent, based on (imaginary) data collected over a quarter about two students who happened to be taking the same classes. For each class meeting over the whole quarter, we note whether John showed up and we note whether Mary showed up. The imaginary data obtained from 100 class meetings are as follows:

	E1: John showed up for class	~E1: John did not show up for class	Row totals
E2: Mary showed up for class	60	15	75
~E2: Mary did not show up for class	20	5	25
Column totals	80	20	100

First of all we define the probabilities of various events. We have single events *E1, ~E1* (read 'Not *E1*'), *E2,* and *~E2*. And we have four possible compound events, *(E1&E2), (E1&~E2),* etc. For example, the compound event *(E1&~E2)* is the compound event 'John showed up for class and Mary did not'. The probability of events of each type is just the number of events of that type divided by the total number of events:

$$P(E1) = N(E1)/N(Total)$$
$$P(E2) = N(E2)/N(Total)$$
$$P(E1\&E2) = N(E1\&E2)/N(Total)$$

We can define other probabilities, such as $P(\sim E1)$, $P(\sim E1\&E2)$, etc.

Strictly speaking the numbers in the table allow us only to estimate probabilities, not to measure them exactly. True probabilities are, however, difficult to observe, and subtle to define, and so for our purposes we will continue to treat estimated probabilities as if they were real ones.

It is useful to be able to address such questions as ‘When John showed up for class, what was the probability that Mary also showed up’. This can be called a *conditional* probability: The probability that Mary showed up, conditional on John having showed up.

Conditional probabilities are written $P(a|b)$ and pronounced ‘P of a given b’. The conditional probability $P(a|b)$ is equal to probability of a and b occurring together, divided by the probability of b. Hopefully this seems natural – it is the fraction, out of all the times that b occurred, that a also occurred. For example the probability of John showing up given that Mary showed up would be estimated with this formula:

$$P(E1|E2) = P(E1\&E2)/P(E2)$$

The probability of Mary showing up given that John showed up would be

$$P(E2|E1) = P(E1\&E2)/P(E1)$$

Now, we can give a formal definition for independence of events:

$E1$ and $E2$ are independent if and only if $P(E1\&E2) = P(E1)P(E2)$

Note what follows from this and from the definition of conditional probability: If E1 and E2 are independent, then the conditional probability of E1 given E2 is equal to the simple unconditional probability of E1:

$$P(E1|E2) = P(E1\&E2)/P(E2) = P(E1)P(E2)/P(E2) = P(E1)$$

And similarly, of course:

$$P(E2|E1) = P(E2)$$

As with our dice example, when E1 and E2 are independent, knowing something about E1 tells us nothing about E2.

Q1. Are $E1$ and $E2$ independent in the table above? Are $\sim E1$ and $\sim E2$ independent? Show calculations that support your answers.

Non-independence and conditional independence

The proportions in the table above are consistent with E1 and E2 being independent. But in the table below, something different is happening.

	E1: John showed up for class	~E1: John did not show up for class	Row totals
E2: Mary showed up for class	70	5	75
~E2: Mary did not show up for class	10	15	25
Column totals	80	20	100

Now, *P(E1&E2)* is greater than it should be.

P(E1&E2) = .7
P(E1)P(E2) = .6

There are 100 instances in total, so we expected to see 60 (0.6* 100) instances where both Mary and John showed up for class (*P(E1)P(E2)*). However, we have observed 70 instances, so Mary and John have shown up to class together in 10 more cases than expected.

Similarly *P(~E1&~E2)* is larger than we would expect if these events were independent:

P(~E1&~E2) = .15
P(~E1)P(~E2) = (25/100)*(20/100) = .05

Again, with 100 instances in total, we expected to see just 5 (0.05*100) instances where both Mary and John were absent (*P(~E1)P(~E2)*). However, we have observed 15 instances – 3 times what we would expect if John's absences were independent of Mary's!

We will call this a violation of *overall independence*. One possibility is that there is some direct influence between John and Mary that affects their attendance in class (the influence could be one way, or bi-directional). On the other hand, it's also possible that something else is independently affecting their behavior. Correlation does not necessarily imply causality!

For example, maybe John and Mary don't know each other and never interact, but both of them have trouble getting up on Monday mornings. Suppose 20 of the classes were on Monday mornings, and for that the data look like this:

	E1: John showed up for class	~E1: John did not show up for class	Totals for E2
E2: Mary showed up for class	1	2	3
~E2: Mary did not show up for class	5	12	17
Column totals	6	14	20

Now, *p(E2&E1)* = 1/20 = .05
P(E1)P(E2) = (6/20)*(3/20) = 18/400 = .045, about the same as .05

For classes at other times in the week, the data look like this:

	E1: John showed up for class	~E1: John did not show up for class	Row totals
E2: Mary showed up for class	69	3	72
~E2: Mary did not show up for class	5	3	8
Column total	74	6	80

P(E2&E1) = 69/80 = .86
P(E1)P(E2) = (74/80)*(72/80) = .83, again, about the same as .86

Moral of the story (everyone should know this whether or not they are interested in minds, brains, or computers): A violation of overall statistical independence can be consistent with conditional independence. In this case, *P(E1)* and *P(E2)* are conditionally independent given the presence or absence of the condition: It's Monday morning.

Q2: Develop your own example in which two events would not be independent overall, but they would be conditionally independent given some other conditioning event. Present a hypothetical overall contingency table in which overall independence is violated but where there could be conditional independence subject to some plausible condition that you can envision (you are not required to make two other tables that can be combined to produce the required table, but you are allowed to make a go of it if you would like). You'll get more points if you can come up with a very different example, while still constructing something plausible. Keep your written answer to 200 words and one contingency table.

Inductive inference

We now consider the problem of inferring whether one event occurred, given evidence about another event. This is a kind of *inductive inference.*

Let's suppose the evidence concerns whether or not the ground is wet, and we want to infer whether it rained. More formally, we will call 'it rained' our 'hypothesis' (*H)* and we want to determine the probability that it is true, under the 'evidence' (*E)* that the ground is wet – i.e. we want to compute $p(H|E)$, the probability that the hypothesis 'it rained' is true given the evidence 'the ground is wet'.

It may help keep you motivated to note that this problem is closely related to the problem of identifying the intended phoneme a person intended to say (was it an 's' or an 'sh') given some actual (and possibly ambiguous) sound that reached your ears. We will apply the analysis we develop here to that example in the second part of this document, to be distributed later.

Now, back to our rain example, let's suppose we have the following data from the last 100 days:

	It rained last night	Did not rain last night	Totals for ground wetness
Ground is wet	20	30	50
Ground is dry	0	50	50
Totals for rain	20	80	100

The data indicate that every time it rains, the ground is wet.

$$P(E|H) = P(E\&H)/P(H) = 20/20 = 1$$

So, what should you conclude if you wake up some morning and we see that the ground is wet? Perhaps you will be tempted to say: Gee, every time it rains, the ground is wet, so therefore, we can be sure that it rained!

However, you are neglecting the possibility that sometimes the ground is wet even when it hasn't rained. For example, you might have a sprinkler system that sometimes comes on and causes the ground to be wet.

What was the actual proportion of times that it rained, when the ground was wet?

$$P(H|E)= P(H\&E)/P(E) = 20/50 = .4$$

In this case, you'd be wrong more often than not if you declared that it had rained every time you noticed that the ground was wet! Such is the situation at my house, where we do have a sprinkler system.

Using Bayes' Rule to Derive Posterior Probabilities

Let's suppose we don't actually have a contingency table over events, but we do have some conditional probability information.

For example, suppose you know:

The chance of rain in March is 20% i.e., $p(H) = .2$
The ground is always wet after it rains i.e., $p(E|H) = 1$.
The ground has a 1/3 chance of being wet even when it hasn't rained, i.e., $P(E|{\sim}H) = .333$

Now, can you tell me, what is the probability that it has rained, given that the ground is wet on the morning of March 17?

Well, we need to calculate p(H|E), and we know:

$P(H) = .2$; from which it follows that $P({\sim}H) = .8$
$P(E|H) = 1$
$P(E|{\sim}H) = .333$

We need to derive $P(H|E)$ from this given information. How can we do it?

First, let's rely on some of the definitions we introduced earlier

$P(E|H) = P(E\&H)/P(H)$. This implies: $P(E\&H) = P(E|H)P(H)$ (1)
$P(H|E) = P(E\&H)/P(E)$. This implies: $P(E\&H) = P(H|E)P(E)$ (2)

These expressions will make more sense if you unpack them in terms of the example. Try it: "The probability that the ground is wet given that it rained is equal to …". Read the material twice, first explicitly unpacking it into the examples and then reading the expressions with the example content in mind. Continue down the page that way.

We can replace $P(E\&H)$ in (1) with $P(H|E)P(E)$, based on (2). This gives us:

$$P(H|E)P(E) = P(E|H)P(H)$$

Dividing by P(E):

$$P(H|E) = P(E|H)P(H)/P(E)$$

We are getting close to having an expression in which our unknown, $P(H|E)$, is expressed in terms of things that we already know, $P(E|H)$ and $P(H)$… but what about $P(E)$, the probability of the evidence?

The total or overall probability of the evidence $P(E)$, is just the sum of two cases, the case where the evidence is present and the hypothesis is true, which is $P(E\&H)$ plus the case where the evidence is present and the hypothesis is false, $P(E\&{\sim}H)$. In our example, the total probability of the ground being wet is the probability that the ground is wet and it has rained, plus the probability that the ground is wet and it has not rained. Going back to results above, we can substitute $P(E|H)P(H)$ for $P(E\&H)$ and similarly, $P(E|{\sim}H)P({\sim}H)$ for $P(E\&{\sim}H)$, to obtain:

$$P(E) = P(E|H)P(H) + P(E|{\sim}H)P({\sim}H)$$

Thus it follows that:

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E \mid H)P(H) + P(E \mid {\sim} H)P({\sim} H)} \tag{3}$$

The above expression is ***Bayes' Rule***, and we have just gone through the proof of ***Bayes' Theorem***. Make sure you understand it: It is more important than some of the Ten Commandments.

Let's summarize in words in terms of our current example:

> The probability that it rained given that the ground is wet, P(H|E), depends on the probability of the ground being wet when it has rained, P(E|H), times the probability of rain P(H); it also depends on the overall probability of the ground being wet, P(E). P(E) is equal to the probability of the ground being wet after it rains, P(E|H), times the probability of rain P(H*); plus the probability of the ground being wet when it did not rain P(E|~H) times the probability that it did not rain P(~H).*

In our case, since we know all of the quantities in this expression, we can calculate the probability that it rained, given that the ground is wet on the morning of March 17:

$$P(H|E) = 1*.2/[1*.2 + .33*.8] = .4286$$

I hope you've all got this. Let's check your understanding with these questions:

Q3. (a) What is $P(H|E)$ if there is no sprinkler system, so that $P(E|{\sim}H) = 0$?
(b) What is $P(H|E)$ in April, when the chance of rain is only 5%, i.e. $P(H)$ is .05?

Q4: One out of 1000 people over 50 have colon cancer (P = .001). A test called the CRC test yields a positive result in 90% of patients with the illness. It also yields a positive result in 1 out of 100 individuals that do not have the illness. What is the probability that someone over 50 has colon cancer, given that they have a positive result on this test? (a) First, make an estimate, and write it down. (b) Then compute the correct answer using Bayes Rule. Treat 'person has colon cancer' as 'H', 'person has a positive result on the test' as 'E', and show your work.

Q5. Chase, Hertwig & Gigerenzer (1998) discuss in the Box 1 of their article a way of providing information for an example like ***Q4*** in terms of what they call "natural frequencies". (a) Rewrite ***Q4*** in accordance with their suggestions. They present evidence that people are much more likely to get the right answer if the question is posed in terms of natural frequencies. (b) State the reason they give for this, and discuss whether you agree, considering your experience from part (a) of ***Q4*** (~200-300 words).