Cheryl E. Fitzgerald
PHI 600 Phil of Probability
Dr. Timothy McGrew
Exam 2

Probability Lattices and the Chain Rule

1. The Rule of Conjunction: $P(A\&B) = P(A)\ P(B|A)$
With this formula we can ascertain the probability of a conjunction containing four conjuncts:

$$P(A\&B\&C\&D) = P(A)\ P(B|A)\ P(C|A\&B)\ P(D|A\&B\&C)$$

This an application of the Chain Rule. However, if we assume that each conjunct is independent of each other, both individually and in aggregate, we can simplify the formula using the simplified Conjunction Rule for independent terms:

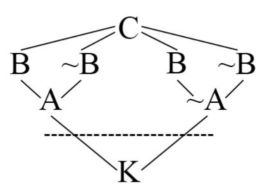
$$P(A\&B\&C\&D) = P(A)\ P(B)\ P(C)\ P(D)$$

When each conjunct is wholly independent of the others both individually and in aggregate, all of the conditional probabilities will be equal to the non-conditional probabilities.

When $P(A) = .9$, $P(B) = .9$, $P(C) = .8$, and $P(D) = .7$,

$$P(A\&B\&C\&D) = .9 \times .9 \times .8 \times .7 = .4536$$

2. Below is a probability lattice for the $P(C|K)$, where K represents unproblematic background knowledge, taking into consideration two other propositions, namely, A and B.



There are four paths we can take from K up to C:

$P(A|K)\ P(B|K\&A)\ P(C|K\&A\&B)$
$P(A|K)\ P({\sim}B|K\&A)\ P(C|K\&A\&{\sim}B)$
$P({\sim}A|K)\ P(B|K\&{\sim}A)\ P(C|K\&{\sim}A\&B)$
$P({\sim}A|K)\ P({\sim}B|K\&{\sim}A)\ P(C|K\&{\sim}A\&{\sim}B)$

Since each of these paths represents one possible way to get from K to C, then

$$P(C|K) = P(A|K)\ P(B|K\&A)\ P(C|K\&A\&B) + P(A|K)\ P({\sim}B|K\&A)\ P(C|K\&A\&{\sim}B) + P({\sim}A|K)\ P(B|K\&{\sim}A)\ P(C|K\&{\sim}A\&B) + P({\sim}A|K)\ P({\sim}B|K\&{\sim}A)\ P(C|K\&{\sim}A\&{\sim}B)$$

2.' It should be quite clear that one of the consequences of the probability lattice is an illustrated version of the Theorem on Total Probability. The last formula stated just *is* the Theorem on Total Probability, only in its conditional probability form, and with two propositions taken into consideration instead of just one. The Theorem on Total Probability stated in its conditional form is:

P(A|K) = P(B|K) P(A|K&B) P(~B|K)P(A|K&~B)

Adding a second proposition to consider produces a formula that looks exactly like the one stated from the lattice. The lattice is able to do this because it does map out each possible way for P(C|K) to be true in regards to two other propositions, which is precisely what the Theorem on Total Probability does. There are four different ways for P(C|K) to be true in regards to the two considered propositions: when A is true and B is true; when A is true, and B is false; when A is false, and B is true; and when A is false and B is false. Thus, the probability lattice acts as a visual representation of the Theorem on Total Probability.

Bayes's Theorem

3. Given that we have the probabilities for P(H), P(E|H), and P(E|~H), we are able to figure out P(H|E) by using the Explicit Form of Bayes's Theorem:

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

Substituting for the numbers, P(H) = a, P(E|H) = b, and P(E|~H) = c, the formula looks like:

$$P(H|E) = \frac{a \times b}{(a \times b) + [(1-a) \times c]}$$

The term (1−a) represents P(~H), since the probability of the negation of a proposition is simply the probability of the proposition subtracted from 1, for the two probabilities form a partition, since one is merely the negation of the other, and the probabilities of a partition sum up to 1.

3.' Given the information above, we can understand P(E) to be the denominator of the above formula,

$$P(E) = (a \times b) + [(1-a) \times c]$$

which can be derived from the Theorem on Total Probability. In other words:

$$P(E) = P(H)\,P(E|H) + P(\sim H)\,P(E|\sim H)$$

since E can be true both when H is true and when H is false. The above formula can be more clearly understood to be derived from the conjunctive probabilities of E and H, and E and ~H, i.e.,

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

And the Conjunction Rule tells us that $P(E\&H) = P(H)\, P(E|H)$, and similarly for $P(E\&{\sim}H)$.

Thus, if we had the values for a, b, and c, we would be able to calculate P(E) and be able to use the Simple Form of Bayes's Theorem

$$P(H|E) = \frac{P(H)\, P(E|H)}{P(E)}$$

4. We have the following information regarding a medical condition: the base rate of the population, which is P(C), the true positive rate for a certain test, which is P(P|C), and the false positive rate for that test, P(P|~C).

$P(C) = .0001$
$P(P|C) = .9$
$P(P|{\sim}C) = .05$

If I test positive, then using Bayes's Theorem I can figure out the probability that I have this medical condition.

$$P(C|P) = \frac{P(C)\, P(P|C)}{P(C)\, P(P|C) + P({\sim}C)\, P(P|{\sim}C)}$$

Plugging in the numbers we get:

$$P(C|P) = \frac{.0001 \times .9}{(.0001 \times .9) + (.9999 \times .05)}$$

After calculating the values of the numerator and denominator, we get:

$$P(C|P) = \frac{.00009}{.050895}$$

And this fraction equals approximately .001768. So I don't seem to have much to worry about.

5. If my friend takes a more reliable test that guarantees $P(P|C) = .99$ and $P(P|{\sim}C) = .02$, and tests positive,

$$P(C|P) = \frac{.0001 \times .99}{(.0001 \times .99) + (.9999 \times .02)}$$

After calculating the values of the numerator and denominator, we get:

$$P(C|P) = \frac{.000099}{.020097}$$

And this fraction equals approximately . 00493. So my friend also seems to not have much to worry about.

6. To represent Ted's new probability conditionalized on the evidence that the coin comes up heads 10 times in 10 flips, we have to factor in the value that the evidence carries in Bayes's Theorem. Since Ted would claim that P(H|F) = .5, then in light of the evidence, the second term in the numerator, P(E|F), would represent the probability that the coin comes up heads 10 times in 10 flips if the coin is fair. The value we seek is P(H&H&H&H&H&H&H&H&H&H|F). We can use the Chain Rule, but since each flip is wholly independent of each other flip, and any aggregate of them, we can take the non-conditioned probability of each proposition and multiply them together, which amounts to $(.5)^{10}$, which equals approximately .00098. But we also need to factor in the probability of the evidence if the coin is not fair, P(E|~F), for the last term in the denominator. We're seeking a similar conjunctive probability, that heads comes up 10 times in 10 flips, except that it is conditioned on the coin's being biased. Thus we get $(.75)^{10}$, which equals approximately .0563. Plugging these numbers into Bayes's Theorem we get:

$$P_T(F|E) = \frac{.99 \times .00098}{(.99 \times .00098) + (.01 \times .0563)}$$

After calculating the numerator and denominator, we get:

$$P_T(F|E) = \frac{.00097}{.00153}$$

And this fraction equals approximately .63. Thus, after the evidence, Ted's new probability that the coin is fair drops dramatically from .99 to .63.

As for Ned, we can plug in the numbers for Bayes's Theorem to get:

$$P_N(F|E) = \frac{.2 \times .00098}{(.2 \times .00098) + (.8 \times .0563)}$$

After calculating the numerator and denominator, we get:

$$P_N(F|E) = \frac{.000196}{.045236}$$

And this fractions equals approximately .0043. Thus, after the evidence, Ned's new probability that the coin is fair drops a little from .2 to .0043. This result illustrates that it won't take much to drag down a high prior probability for some proposition in light of a bit of evidence against it. Despite Ted's extremely confident prior probability in the fairness of the coin, after only a mere 10 flips of 10 straight heads, his probability has dropped by about .3.

Over a long series of flips, eventually their probabilities for the fairness of the coin will converge and, depending on how long they keep flipping, either become identical, or very near to it. And it makes no difference how the flips turn out. From the current example, we already saw how dramatically 10 flips dropped Ted's high prior probability; but since Ned's prior probability was already low, it did not bring it down as much as it did Ted's. For each flip of heads, Ted's conditionalized probabilities will drag down lower and lower, in a much more dramatic way than Ned's, for Ned's probabilities are already quite low. So the effect that the evidence of flipping heads has on Ted's probabilities is much more dramatic than on Ned's, i.e., it's much more damning. Conversely, if the series of flips turns out much closer to 50/50 of heads and tails, then this will dramatically increase Ned's probabilities, since they are quite low, and it will only minimally affect Ted's, since they are not nearly as low. Eventually, with enough flips, their probabilities will stabilize at the same value, or very nearly the same value. This demonstrates that if two people dramatically differ in their prior probabilities for some proposition, the observation of evidence will affect their probabilities such that they conform to the evidence, not the other way around, and that if one is to value truth and rationality, then either one person or the other, or both, will be required to change his beliefs about the proposition.

7. In general, there are two main different approaches to measuring the degree of confirmation of a hypothesis by a piece of evidence, the ratio measure and the difference measure. According to the ratio measure, the degree of confirmation is measured by taking the ratio of the conditional probability to the prior, $P(H|E) / P(H)$. When the value of the fraction is less than 1, E does not confirm H; when it is greater than 1, E does confirm H; and when it is equal to 1, E is irrelevant to H. According to the difference measure, the prior is subtracted from the conditional probability, $P(H|E) - P(H)$. When the value of the subtraction is less than 0, E does not confirm H; when it is greater than 0, E does confirm H; and when it is equal to 0, E is irrelevant to H. For the most part, it is not crucial to decide between the two, because they will always give the same answer to the question, "Does E confirm H?", for essentially, we hold the belief that if $P(H|E) > P(H)$ is true, then E does confirm H, if $P(H|E) < P(H)$ is true, then E does not confirm H, and if $P(H|E) = P(H)$ is true, then E is irrelevant to H.

However, Schlesinger believes that the ratio measure has an advantage over the difference measure, which he demonstrates through various examples. I will only discuss the last example, which he claims is the strongest argument against the difference measure.

A scientist desires to confirm his theory T via experiment. But he's limited on funding, so he can only afford one experiment, and thus, he must choose wisely which experiment to perform. His theory T is comprised of a main hypothesis, h, and an auxiliary hypothesis, a; thus, T = (h&a). He has a choice between two different experiments, e and f, but to make things difficult, e is irrelevant to a, and so can confirm only h, and f is irrelevant to h, and so can only confirm a. He needs to decide which experiment will confirm T to the *highest* degree. Since as long as the probability of one conjunct of a conjunction is confirmed by some piece of evidence, the probability of the conjunction is also confirmed, in some sense, and so he needs only one of the experiments to confirm T. Essentially, P(T|e) = P(h&a|e) and P(T|f) = P(h&a|f); the task is determining which probability has the greater value.

It is supposed that P(h) = 1/8, P(h|e) = 1/4, P(a) = 1/3, and P(a|f) = 1/2. According to the ratio measure, the degree of confirmation for h by e is 1/4 / 1/8 = 2, and the degree of confirmation for a by f is 1/2 / 1/3 = 3/2. We can see that h is confirmed by e to a higher degree than a is confirmed by f, for 2 > 3/2. Thus, it would be more worthwhile for our scientist to perform experiment e. However, according to the difference measure, the degree of confirmation for h by e is 1/4 – 1/8 = 1/8, and the degree of confirmation for a by f is 1/2 – 1/3 = 1/6. With these degrees of confirmation, a is confirmed by f to a higher degree than h is confirmed by e, for 1/6 > 1/8, and thus it would be more worthwhile for our scientist to perform experiment f. Clearly, we cannot have it both ways: one of these measures must not be correct.

To discover which is incorrect, we need to consider the degree of confirmation for the *conjunction* of h&a in light of e, and in light of f, and compare these values, for our scientist is concerned with confirming T, which just is (h&a). P(h&a|e) = P(h|e)P(a|e&h). First, e is irrelevant to a, so we can reduce P(a|e&h) to P(a|h). Second, we assume that a and h are independent of each other, and this allows that P(a|h) = P(a). Now the value we seek is P(h|e)P(a) = 1/4 ×1/3 = 1/12. Similarly, P(h&a|f) = P(h|f)P(a|f&h). Since f is irrelevant to h, we can reduce P(h|f) to P(h). And since a and h are assumed to be independent, P(a|f&h) = P(a|f). The value that we seek now is P(h)P(a|f) = 1/8 × 1/2 = 1/16. Comparing these two degrees of confirmation for (h&a), in other words, for T, we can see quite clearly that e confirms T to a higher degree than f, for 1/12 > 1/16. Thus, our scientist should perform experiment e.

So according to this example, we have a clear demonstration where the difference measure fails and the ratio measure provides the correct answer. Thus, claims Schlesinger, the ratio measure should be taken as closer to an accurate measure of the degree of confirmation than the difference measure.

8. If we try to reformulate Condorcet's formula for testimony into Bayes's Theorem, we will take p to be the prior of the event, P(E), and t to be the probability that the witness reports that E occurred when E did in fact occur, P(R|E). We end up with:

$$P(E|R) = \frac{P(E)P(R|E)}{P(E)P(R|E) + [1 - P(E)][1 - P(R|E)]}.$$

Unfortunately, this does not actually give us Bayes's Theorem: the problem lies in (1 – t), i.e., [1 – P(R|E)] in our reformulation. We know that what we want this term to represent is the probability that the witness reports the event when the event did not occur, in other words, P(R|~E). But [1 – P(R|E)] cannot give us this value, for instead it would give us P(~R|E), via normality. P(R|E) and P(R|~E) do not form a partition, and do not necessarily sum to 1; they simply do not bear the type of relation that allows for normalization. The probability that would provide us with right value using normality would be $P(\sim R|\sim E)$, so that $[1 - P(\sim R|\sim E)]$ does give us $P(R|\sim E)$. $P(\sim R|\sim E)$ is the probability that the witness does not report that E occurred when E did not occur. However, this does not make clear whether the witness does not report anything at all, or whether he does report that E did not occur, for both of these might be just as likely.

The problem is that R does not really represent what it needs to in order for it to play the role we want it to; it matters what is being reported, since a witness can either report that E occurred, report that E did not occur, or report nothing at all. We need to symbolize the witness's report that E occurred as R_E and his report that E did not occur as $R_{\sim E}$. What we want for the last term in our Bayesian reformulation is $P(R_E|\sim E)$. The only way that we can obtain this using normality is by assuming that R_E and $R_{\sim E}$ form a partition; in other words, we must assume that regardless of what happens, the witness will report something, and only report either that E occurred or E did not occur. We assume both that $P(\sim R_E|E) = 0$ and $P(\sim R_{\sim E}|\sim E) = 0$, or we assume both that $P(R_N|E) = 0$ and $P(R_N|\sim E) = 0$, depending on how we want to symbolize that the witness does not give a report. Ultimately, if all we have is the probability that the witness will report that E when E occurs, we do not thereby know the probability that he would report that E did not occur when E occurs, because we do not know the probability that we would report nothing when E occurs. In order for Condorcet's formula to work, we must assume that the witness is forthcoming.

The necessary assumptions are: (a) that the witness will report something, and not nothing, and (b) that the witness with only report either that E or that not E. In other words, our assumptions are: (a) $P(\sim R_E|\sim E) = P(\sim R_{\sim E}|\sim E) = 0$, and (b) that R_E and $R_{\sim E}$ form a partition. With this in mind, we can modify Condorcet's formula to include $t^* = P(R_{\sim E}|\sim E)$:

$$\frac{pt}{pt + (1 - p)(1 - t^*)}.$$

And in our Bayesian formulation:

$$\frac{P(E)P(R_E|E)}{P(E)P(R_E|E) + [1 - P(E)][1 - P(R_{\sim E}|\sim E)}.$$

which gives us

$$\frac{P(E)P(R_E|E)}{P(E)P(R_E|E) + P(\sim E)P(R_E|\sim E)}.$$

And this would be the correct Bayesian formulation. But it only comes with the price of those assumptions.

Combined Evidence

9. The probability that the event occurred given the testimonies of three witnesses under the assumption that these testimonies are fully independent of each other can be represented using the Compound Odds Form of Bayes's Theorem.

$$\frac{P(E|W_1\&W_2\&W_3)}{P(\sim E|W_1\&W_2\&W_3)} = \frac{P(E)}{P(\sim E)} \times \frac{P(W_1|E)}{P(W_1|\sim E)} \times \frac{P(W_2|E)}{P(W_2|\sim E)} \times \frac{P(W_3|E)}{P(W_3|\sim E)}$$

If the prior odds are 100 to 1 against E occurring, and the respective odds that the witnesses' testimonies are truthful are 4 to 1, 5 to 1, and 10 to 1, plugging these odds into the formula we get:

$$\frac{P(E|W_1\&W_2\&W_3)}{P(\sim E|W_1\&W_2\&W_3)} = \frac{1}{100} \times \frac{4}{1} \times \frac{5}{1} \times \frac{10}{1} = \frac{200}{100} = \frac{2}{1}$$

Thus, it is twice as probable, 2 to 1, that the event E occurred than not. In other words,

$P(E|W_1\&W_2\&W_3) \approx .667$

10. We can figure out the probability that the tale is true, given that all seven of the knaves have told this same tale, using the Explicit Form of Bayes's Theorem. If the prior odds that the tale are true are 7 to 1 against, then the probability that the tale is true is P(T) = .125, and thus, P(~T) = .875, seven times that of T. For each knave the probability that he reports T given that T occurred is P(R|T) = .125, and thus, P(~R|T) = .875, again, seven times that of his telling the truth. Since each report, whether it's true or false, is wholly independent of any of the other reports, in order to obtain the probability that all of the knaves tell the truth about T when T occurs, we can use the conjunction rule for independence and simply multiply the probabilities for each knave's reliability. In other words, $P(R_n|T)^7$, since the reliability for each knave is the same. Thus,

$P(R_n|T)^7 = (.125)^7 \approx .0000004768$

If we multiply this by P(T), .125, we get approximately .0000000596, and this gives us the numerator and one term of the denominator. For the second term in the denominator, we need to determine the probability, given T did not occur, i.e., that the tale being told is false, that all seven knaves would have come up with the same tale out of a possible of 70 false tales. But this is not a 1/70 chance, because for each knave, there are 70 possible stories he might tell, and we want the chance that they all happen to tell the same story.

In order to see how this value is determined, it is easiest to compare it with a truth table in bivalent logic. Since we only have true and false, we have two possible stories, either something did or did not occur. Let's say we have two witnesses; what are the chances they will both tell the same tale? This will be the number of ways in which the two witnesses tell the same story divided by the total possible ways in which their telling of the stories can be distributed. The total possible ways in which the stories can be distributed is essentially the number of rows in the truth table, the formula for which is raising 2 to the power of the number of propositions involved, which are our witnesses. Since we have 2 witnesses, we raise 2 to the 2^{nd} power and get 4. Now since there are only two different stories, there are only two ways in which both witnesses can tell the same story: either they both say it's true, or they both say it's false. In this case, the chances are 1/2. Now, let's say we have our seven knaves, but let's keep it simple and stick to the two stories. We raise 2 to the 7^{th} power and we get 128, so that's 128 different possible combinations of the distribution of the two stories amongst these seven knaves. But since there are still only two different stories, there are still only 2 different ways in which all seven knaves can tell the same story, either when they all say it's true or they all say it's false. So now we have a 2/128, or 1/64 chance that all seven knaves tell the same story. And that's only when there are 2 different possible stories!

But in our example, we have 70 different possible stories, which means we must raise 70 to the 7^{th} power to get our total number of possible combinations of the distribution of the stories amongst the knaves, and this gives us 8,235,430,000,000! And the chances that all seven knaves tell the same story is 70/8,235,430,000,000, or 1/117,649,000,000. Thus,

$$P(R_1\&R_2\&R_3\&R_4\&R_5\&R_6\&R_7|{\sim}T) \approx .0000000000085$$

Now we multiply this by P(~T), .875, to get .000000000007435. Thus, the denominator of our formula is

$$.0000000596 + .000000000007435 = .000000059607435$$

And finally, plugging in the numerator and denominator,

$$P(T|R_1\&R_2\&R_3\&R_4\&R_5\&R_6\&R_7) \approx .99988$$

This clearly demonstrates the power of identical testimonies among several witnesses—even when they are all normally pretty unreliable—when there are so many different, equally likely false stories that could be told. In other words, it would be so incredibly improbable that several people, whom we are assuming to be wholly independent of each other, despite being normally unreliable, all come up with the same story when there are many possible stories they could have come up with, assuming they are all equally likely. So given all of their identical testimonies, we'd be incredibly irrational if we chose not to believe them.

10'. The correct form of Bayes's Theorem needed to figure out this problem is the Compound Odds Form. Since the odds that each testimony are all identical, and wholly independent of each other, we can simplify the formula thus:

$$\frac{P(M|T^{10})}{P({\sim}M|T)} = \frac{P(M)}{P({\sim}M)} \times \left[\frac{P(T|M)}{P(T|{\sim}M)}\right]^{10}$$

If the prior odds for M are 1,000,000 to 1 against its occurrence, then the minimum odds needed that each of the testimonies are truthful that will push the odds that M occurs, given these testimonies, above .5 is 4 to 1. With a 4 to 1 ratio of odds raised to the 10^{th} power we get 1,048,576 to 1, which is just greater than the prior odds against M.

The assumption of independence among the testimonies, individually and in aggregate, is absolutely necessary in order to work out the formula this way, for if we had not assumed independence, the formula would take on a much more complicated form, for then the odds of each testimony would not only be conditional on the whether or not M occurs, but also conditional on the other testimonies. A very simple version with only two witnesses would look like:

$$\frac{P(M|T_1\&T_2)}{P(\sim M|T_1\&T_2)} = \frac{P(M)}{P(\sim M)} \times \frac{P(T_1|M)}{P(T_1|\sim M)} \times \frac{P(T_2|M\&T_1)}{P(T_2|\sim M\&T_1)}$$

Without the independence assumption, this formula, with several more witnesses, would not only get monstrously long, but unsolvable.