CS 228, Winter 2008
Solutions to Problem Set #0: Probability Review

---

1. After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

   Answer: We are given the following information:

$$\begin{aligned} P(test_1 | disease_1) &= 0.99 \\ P(test_0 | disease_0) &= 0.99 \\ P(disease_1) &= 0.0001 \\ test_1 & \end{aligned}$$

   where $test_1$ means that the test is positive. What the patient is concerned about is $P(disease_1|test_1)$. Roughly speaking, the reason it is a good thing that the disease is rare is that $P(disease_1|test_1)$ is proportional to $P(disease_1)$, so a lower prior for $disease$ will mean a lower value for $P(disease|test)$. Roughly speaking, if 10,000 people take the test, we expect 1 to actually have the disease, and most likely test positive, while the rest do not have the disease, but 1% of them (about 100 people) will test positive anyway, so $P(disease|test)$ will be about 1 in 100. More precisely, using Bayes' rule:

$$\begin{aligned} P(disease_1|test_1) &= \frac{P(test_1|disease_1)P(disease_1)}{P(test_1|disease_1)P(disease_1) + P(test_1|disease_0)P(disease_0)} \\ &= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} \\ &= .009804 \end{aligned}$$

   The moral is that when the disease is much rarer than the test accuracy, a positive test result does not mean the disease is likely. A false positive reading remains much more likely.

2. It is quite often useful to consider the effect of some specific propositions in the context of some general background evidence that remains fixed, rather than in the complete absence of information. The following questions ask you to prove more general versions of the product rule and Bayes' rule, with respect to some background evidence E:

   (a) Prove the conditionalized version of the general product rule:

$$P(A, B|E) = P(A|B, E)P(B|E)$$

   (b) Prove the conditionalized version of Bayes' rule:

$$P(A|B, C) = \frac{P(B|A, C)P(A|C)}{P(B|C)}$$

Answer: The basic axiom to use here is the definition of conditional probability:

(a) We have

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

and

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

hence

$$P(A, B|E) = P(A|B, E)P(B|E)$$

(b) The derivation here is the same as the derivation of the simple version of Bayes' Rule on page 426 in the textbook. First we write down the dual form of the conditionalized product rule, simply by switching A and B in the above derivation:

$$P(A, B|E) = P(B|A, E)P(A|E)$$

Therefore the two right-hand sides are equal:

$$P(B|A, E)P(A|E) = P(A|B, E)P(B|E)$$

Dividing through by $P(B|E)$ we get

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

3. This problem investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations.

(a) Suppose we wish to calculate $P(H|E_1, E_2)$, and we have no conditional independence information. Which of the following sets of numbers are sufficient for the calculation?

i. $P(E_1, E_2), P(H), P(E_1|H), P(E_2|H)$.

ii. $P(E_1, E_2), P(H), P(E_1, E_2|H)$.

iii. $P(E_1|H), P(E_2|H), P(H)$.

Answer: Bayes' Rule gives

$$P(H|E_1, E_2) = \frac{P(E_1, E_2|H)P(H)}{P(E_1, E_2)}$$

Thus the information in (ii) is sufficient—in fact, we don't need $P(E_1, E_2)$ because we can use normalization. Intuitively, the information in (i) is insufficient because $P(E_1|H)$ and $P(E_2|H)$ provide no information about correlations between E_1 and E_2 that might be induced by H. More specifically, we can reduce the $P(E_1, E_2|H)$ in the equation above to $P(E_1|H, E_2)P(E_2|H)$. Obviously, without extra information such as independence assumptions, $P(E_1|H)$ doesn't give us enough information to compute $P(E_1|H, E_2)$.

Mathematically, suppose H has m possible values and E_1 and E_2 have n_1 and n_2 possible, respectively. $P(H|E_1, E_2)$ contains $(m-1)n_1n_2$ independent parameters, whereas $P(E_1, E_2)$ contains $(n_1n_2 - 1)$ independent parameters, $P(H)$ conains $(m-1)$ independent parameters, $P(E_1|H)$ contains $m(n_1 - 1)$ independent parameters, and $P(E_2|H)$ contains $m(n_2 - 1)$ independent paramaters. Hence, the information in (i) contains

$(n_1 n_2 - 1) + (m-1) + m(n_1 - 1) + m(n_2 - 1)$ numbers—insufficient for large m, n_1, and n_2. By the same token, the information in (iii) is also insufficient.

A common error is to compare the number of independent parameters in the information given, to the number of parameters needed to specify the joint, rather than to the generally smaller number of parameters needed to specify $P(H|E_1, E_2)$. Requiring that we have enough parameters to specify the joint is excessive and may lead us to discard actual solutions. As a simple example, suppose we want to compute $P(E_1, H)$, where all the variables are boolean. $P(E_1 \mid H)$ and $P(H)$ are obviously sufficient, although together they only contain three independent parameters between them, less than the seven required to specify the joint $P(E_1, E_2, H)$.

(b) Suppose we know that E_1 and E_2 are conditionally independent given H. Now which of the above three sets are sufficient?

Answer: If $P(E_1|H, E_2) = P(E_1|H)$, then E_1 and E_2 are conditionally independent given H, and the equation above simplifies to

$$P(H|E_1, E_2) = \frac{P(E_1|H)P(E_2|H)P(H)}{P(E_1, E_2)}$$

Obviously, both (i) and (ii) are now sufficient. Moreover, (iii) has also become sufficient because we can use normalization to compute $P(E_1, E_2)$. More specifically, for any values of e_1, e_2 of E_1 and E_2, respectively, the marginal probability of this pair is

$$\begin{aligned} P(e_1, e_2) &= \sum_{h \in \mathtt{dom}(H)} P(e_1, e_2|h)P(h) \\ &= \sum_{h \in \mathtt{dom}(H)} P(e_1|h)P(e_2|h)P(h). \end{aligned}$$

We already have each of the values needed for this computation.

4. Express the statement that X and Y are conditionally independent given Z as a constraint on the joint distribution entries for $P(X, Y, Z)$.

Answer: When dealing with joint entries, it is usually easiest to get everything into the form of probabilities of conjunctions, since these can be expressed as sums of joint entries. Beginning with the conditional independence constraint

$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

we can rewrite it using the definition of conditional probability on each term to obtain

$$\frac{P(X, Y, Z)}{P(Z)} = \frac{P(X, Z)}{P(Z)} \frac{P(Y, Z)}{P(Z)}$$

Hence we can write an expression for joint entries:

$$P(X, Y, Z) = \frac{P(X, Z)P(Y, Z)}{P(Z)} = \frac{\sum_y P(X, Y = y, Z) \sum_x P(X = x, Y, Z)}{\sum_{x,y} P(X = x, Y = y, Z)}$$