CS 67800, Spring 2017/18

Problem Set 0: Probability Review

Submission Date : Monday 9/4/2018, 23:59

---

1. A fair coin is tossed 4 times. Define: X - the number of Heads in the first 2 tosses; Y - the number of Heads in all 4 tosses.
   - Calculate the table of the joint probability $P(X, Y)$
   - Calculate the tables of marginal probabilities $P(X)$ and $P(Y)$.
   - Calculate the tables of conditional probabilities $P(X|Y)$ and $P(Y|X)$.
   - What is the distribution of $Z = Y - X$?

2. 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 1 in 10,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

3. Show that the statement
$$P(A, B|C) = P(A|C)P(B|C)$$
is equivalent to the statement
$$P(A|B, C) = P(A|C)$$
and also to
$$P(B|A, C) = P(B|C)$$
(you need to show both directions).

4. 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:
$$P(A|B, C) = \frac{P(B|A, C)P(A|C)}{P(B|C)}$$

Note: Here P is a joint probability distribution over the specified variables. For example, $P(A, B|C)$ is a joint distribution over the variables A and B given the variable C. Thus, for every possible value of C, it specifies a probability distribution over all combinations of values for A and B.

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

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