probability-exercises

March 26, 2019

1 13. Quantifying Uncertainity

13.1 Show from first principles that $P(a \mid b \land a) = 1$.

13.2 [sum-to-1-exercise]Using the axioms of probability, prove that any probability distribution on a discrete random variable must sum to 1.

13.3 For each of the following statements, either prove it is true or give a counterexample.

1. If $P(a \mid b, c) = P(b \mid a, c)$, then $P(a \mid c) = P(b \mid c)$
2. If $P(a \mid b, c) = P(a)$, then $P(b \mid c) = P(b)$
3. If $P(a \mid b) = P(a)$, then $P(a \mid b, c) = P(a \mid c)$

13.4 Would it be rational for an agent to hold the three beliefs $P(A) = 0.4$, $P(B) = 0.3$, and $P(A \lor B) = 0.5$? If so, what range of probabilities would be rational for the agent to hold for $A \land B$? Make up a table like the one in Figure Section **??**, and show how it supports your argument about rationality. Then draw another version of the table where $P(A \lor B) = 0.7$. Explain why it is rational to have this probability, even though the table shows one case that is a loss and three that just break even. (*Hint:* what is Agent 1 committed to about the probability of each of the four cases, especially the case that is a loss?)

13.5 [exclusive-exhaustive-exercise]This question deals with the properties of possible worlds, defined on page Section **??** as assignments to all random variables. We will work with propositions that correspond to exactly one possible world because they pin down the assignments of all the variables. In probability theory, such propositions are called **atomic event**. For example, with Boolean variables X_1, X_2, X_3, the proposition $x_1 \land \neg x_2 \land \neg x_3$ fixes the assignment of the variables; in the language of propositional logic, we would say it has exactly one model.

1. Prove, for the case of n Boolean variables, that any two distinct atomic events are mutually exclusive; that is, their conjunction is equivalent to $false$.
2. Prove that the disjunction of all possible atomic events is logically equivalent to *true*.
3. Prove that any proposition is logically equivalent to the disjunction of the atomic events that entail its truth.

13.6 [inclusion-exclusion-exercise]Prove Equation (Section **??**) from Equations (Section **??**) and (Section **??**).

13.7 Consider the set of all possible five-card poker hands dealt fairly from a standard deck of fifty-two cards.

1. How many atomic events are there in the joint probability distribution (i.e., how many five-card hands are there)?
2. What is the probability of each atomic event?
3. What is the probability of being dealt a royal straight flush? Four of a kind?

13.8 Given the full joint distribution shown in Figure Section **??**, calculate the following:

1. $\mathbf{P}(toothache)$.
2. $\mathbf{P}(Cavity)$.
3. $\mathbf{P}(Toothache \mid cavity)$.
4. $\mathbf{P}(Cavity \mid toothache \vee catch)$.

13.9 Given the full joint distribution shown in Figure Section **??**, calculate the following:

1. $\mathbf{P}(toothache)$.
2. $\mathbf{P}(Catch)$.
3. $\mathbf{P}(Cavity \mid catch)$.
4. $\mathbf{P}(Cavity \mid toothache \vee catch)$.

13.10 [unfinished-game-exercise] In his letter of August 24, 1654, Pascal was trying to show how a pot of money should be allocated when a gambling game must end prematurely. Imagine a game where each turn consists of the roll of a die, player E gets a point when the die is even, and player O gets a point when the die is odd. The first player to get 7 points wins the pot. Suppose the game is interrupted with E leading 4–2. How should the money be fairly split in this case? What is the general formula? (Fermat and Pascal made several errors before solving the problem, but you should be able to get it right the first time.)

13.11 Deciding to put probability theory to good use, we encounter a slot machine with three independent wheels, each producing one of the four symbols bar, bell, lemon, or cherry with equal probability. The slot machine has the following payout scheme for a bet of 1 coin (where "?" denotes that we don't care what comes up for that wheel):

bar/bar/bar pays 20 coins

bell/bell/bell pays 15 coins

lemon/lemon/lemon pays 5 coins

cherry/cherry/cherry pays 3 coins

cherry/cherry/? pays 2 coins

cherry/?/? pays 1 coin

1. Compute the expected "payback" percentage of the machine. In other words, for each coin played, what is the expected coin return?

2. Compute the probability that playing the slot machine once will result in a win.

3. Estimate the mean and median number of plays you can expect to make until you go broke, if you start with 10 coins. You can run a simulation to estimate this, rather than trying to compute an exact answer.

13.12 Deciding to put our knowledge of probability to good use, we encounter a slot machine with three independently turning reels, each producing one of the four symbols bar, bell, lemon, or cherry with equal probability. The slot machine has the following payout scheme for a bet of 1 coin (where "?" denotes that we don't care what comes up for that wheel):

bar/bar/bar pays 21 coins

bell/bell/bell pays 16 coins

lemon/lemon/lemon pays 5 coins

cherry/cherry/cherry pays 3 coins

cherry/cherry/? pays 2 coins

cherry/?/? pays 1 coin

1. Compute the expected "payback" percentage of the machine. In other words, for each coin played, what is the expected coin return?

2. Compute the probability that playing the slot machine once will result in a win.

3. Estimate the mean and median number of plays you can expect to make until you go broke, if you start with 8 coins. You can run a simulation to estimate this, rather than trying to compute an exact answer.

13.13 We wish to transmit an n-bit message to a receiving agent. The bits in the message are independently corrupted (flipped) during transmission with ϵ probability each. With an extra parity bit sent along with the original information, a message can be corrected by the receiver if at most one bit in the entire message (including the parity bit) has been corrupted. Suppose we want to ensure that the correct message is received with probability at least $1-\delta$. What is the maximum feasible value of n? Calculate this value for the case $\epsilon = 0.001$, $\delta = 0.01$.

13.14 We wish to transmit an n-bit message to a receiving agent. The bits in the message are independently corrupted (flipped) during transmission with ϵ probability each. With an extra parity bit sent along with the original information, a message can be corrected by the receiver if at most one bit in the entire message (including the parity bit) has been corrupted. Suppose we want to ensure that the correct message is received with probability at least $1-\delta$. What is the maximum feasible value of n? Calculate this value for the case $\epsilon = 0.002$, $\delta = 0.01$.

13.15 [independence-exercise]Show that the three forms of independence in Equation (Section **??**) are equivalent.

13.16 Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present, but has a 10% false positive rate (indicating that the virus is present, when it is not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus. The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus. Which test returning positive is more indicative of someone really carrying the virus? Justify your answer mathematically.

13.17 Suppose you are given a coin that lands *heads* with probability x and *tails* with probability $1 - x$. Are the outcomes of successive flips of the coin independent of each other given that you know the value of x? Are the outcomes of successive flips of the coin independent of each other if you do *not* know the value of x? Justify your answer.

13.18 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 when you do have the disease is 0.99, as is the probability of testing negative when you don't have the disease). The good news is that this is a rare disease, striking only 1 in 10,000 people of your age. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

13.19 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 when you do have the disease is 0.99, as is the probability of testing negative when you don't have the disease). The good news is that this is a rare disease, striking only 1 in 100,000 people of your age. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

13.20 [conditional-bayes-exercise]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**:

1. Prove the conditionalized version of the general product rule:

$$\mathbf{P}(X, Y \mid \mathbf{e}) = \mathbf{P}(X \mid Y, \mathbf{e})\mathbf{P}(Y \mid \mathbf{e}) \, .$$

2. Prove the conditionalized version of Bayes' rule in Equation (Section **??**).

13.21 [pv-xyz-exercise] Show that the statement of conditional independence

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

is equivalent to each of the statements

$$\mathbf{P}(X \mid Y, Z) = \mathbf{P}(X \mid Z) \quad \text{and} \quad \mathbf{P}(Y \mid X, Z) = \mathbf{P}(Y \mid Z) \, .$$

13.22 Suppose you are given a bag containing n unbiased coins. You are told that $n - 1$ of these coins are normal, with heads on one side and tails on the other, whereas one coin is a fake, with heads on both sides.

1. Suppose you reach into the bag, pick out a coin at random, flip it, and get a head. What is the (conditional) probability that the coin you chose is the fake coin?

2. Suppose you continue flipping the coin for a total of k times after picking it and see k heads. Now what is the conditional probability that you picked the fake coin?

3. Suppose you wanted to decide whether the chosen coin was fake by flipping it k times. The decision procedure returns $fake$ if all k flips come up heads; otherwise it returns $normal$. What is the (unconditional) probability that this procedure makes an error?

13.23 [normalization-exercise]In this exercise, you will complete the normalization calculation for the meningitis example. First, make up a suitable value for $P(s \mid \neg m)$, and use it to calculate unnormalized values for $P(m \mid s)$ and $P(\neg m \mid s)$ (i.e., ignoring the $P(s)$ term in the Bayes' rule expression, Equation (Section **??**)). Now normalize these values so that they add to 1.

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

1. Suppose we wish to calculate $P(h \mid e_1, e_2)$ and we have no conditional independence information. Which of the following sets of numbers are sufficient for the calculation?

    1. $\mathbf{P}(E_1, E_2)$, $\mathbf{P}(H)$, $\mathbf{P}(E_1 \mid H)$, $\mathbf{P}(E_2 \mid H)$
    2. $\mathbf{P}(E_1, E_2)$, $\mathbf{P}(H)$, $\mathbf{P}(E_1, E_2 \mid H)$
    3. $\mathbf{P}(H)$, $\mathbf{P}(E_1 \mid H)$, $\mathbf{P}(E_2 \mid H)$

2. Suppose we know that $\mathbf{P}(E_1 \mid H, E_2) = \mathbf{P}(E_1 \mid H)$ for all values of H, E_1, E_2. Now which of the three sets are sufficient?

13.25 Let X, Y, Z be Boolean random variables. Label the eight entries in the joint distribution $\mathbf{P}(X, Y, Z)$ as a through h. Express the statement that X and Y are conditionally independent given Z, as a set of equations relating a through h. How many *nonredundant* equations are there?

13.26 (Adapted from Pearl [-@Pearl:1988].) Suppose you are a witness to a nighttime hit-and-run accident involving a taxi in Athens. All taxis in Athens are blue or green. You swear, under oath, that the taxi was blue. Extensive testing shows that, under the dim lighting conditions, discrimination between blue and green is 75% reliable.

1. Is it possible to calculate the most likely color for the taxi? (*Hint:* distinguish carefully between the proposition that the taxi *is* blue and the proposition that it *appears* blue.)

2. What if you know that 9 out of 10 Athenian taxis are green?

13.27 Write out a general algorithm for answering queries of the form $\mathbf{P}(Cause \mid \mathbf{e})$, using a naive Bayes distribution. Assume that the evidence $\mathbf{e}$ may assign values to *any subset* of the effect variables.

13.28 [naive-bayes-retrieval-exercise] Text categorization is the task of assigning a given document to one of a fixed set of categories on the basis of the text it contains. Naive Bayes models are often used for this task. In these models, the query variable is the document category, and the "effect" variables are the presence or absence of each word in the language; the assumption is that words occur independently in documents, with frequencies determined by the document category.

1. Explain precisely how such a model can be constructed, given as "training data" a set of documents that have been assigned to categories.

2. Explain precisely how to categorize a new document.

3. Is the conditional independence assumption reasonable? Discuss.

13.29 In our analysis of the wumpus world, we used the fact that each square contains a pit with probability 0.2, independently of the contents of the other squares. Suppose instead that exactly $N/5$ pits are scattered at random among the N squares other than [1,1]. Are the variables $P_{i,j}$

and $P_{k,l}$ still independent? What is the joint distribution $\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$ now? Redo the calculation for the probabilities of pits in [1,3] and [2,2].

13.30 Redo the probability calculation for pits in [1,3] and [2,2], assuming that each square contains a pit with probability 0.01, independent of the other squares. What can you say about the relative performance of a logical versus a probabilistic agent in this case?

13.31 Implement a hybrid probabilistic agent for the wumpus world, based on the hybrid agent in Figure Section **??** and the probabilistic inference procedure outlined in this chapter.