ARTIFICIAL INTELLIGENCE (E016350)
GHENT UNIVERSITY
AY 2024/2025

Prof. Aleksandra Pizurica
Assistant: N. Vercheval
E-mail: ai@lists.ugent.be

Solutions: Probabilistic reasoning – Basics

1. Show from the main principles that $P(a|b \wedge a) = 1$.

Solution:

The "main principles" needed here are the definition of conditional probability, $P(x|y) = P(x \wedge y)/P(y)$, and the definitions of the logical connectives. It is not enough to say that if $b \wedge a$ is "given" then a must be true! From the definition of conditional probability, and the fact that $A \wedge A \Leftrightarrow A$ and that conjunction is commutative and associative, we have

$$P(a|b \wedge a) = \frac{P(a \wedge (b \wedge a))}{P(b \wedge a)} = \frac{P(b \wedge a)}{P(b \wedge a)} = 1$$

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

   (a) How many atomic events are there in the joint probability distribution (i.e., how many five-card hands are there)?

   (b) What is the probability of each atomic event?

   (c) What is the probability of being dealt a royal straight flush? Four of a kind?

Note: A straight flush is a hand that contains five cards of sequential rank, all of the same suit, such as Q♣ J♣ 10 ♣ 9♣ 8♣ (a "queen-high straight flush"). An ace-high straight flush, such as A♠ K♠ Q♠ J♠ 10♠, is called a royal flush or royal straight flush. Four of a kind, also known as quads, is a hand that contains four cards of one rank and one card of another rank, e.g., 5♢ 5♠ 5♡ 5♣ Q♡ ("four of a kind, fives").

Solution:

This is a classic combinatorics question that could appear in a basic text on discrete mathematics. The point here is to know how to apply the relevant axioms of probability and to help students to grasp the concept of the joint probability distribution as the distribution over all possible states of the world.

(a) There are $\binom{52}{5} = (52 \times 51 \times 50 \times 49 \times 48)/(1 \times 2 \times 3 \times 4 \times 5) = 2{,}598{,}960$ possible five-card hands.

(b) By the fair-dealing assumption, each of these is equally likely. Each hand therefore occurs with probability $1/2{,}598{,}960 \approx 3.85 \times 10^{-7}$.

(c) There are four hands that are royal straight flushes (one in each suit). Since the events are mutually exclusive, the probability of a royal straight flush is just the sum of the probabilities of the atomic events, i.e., $4/2{,}598{,}960 = 1/649{,}740 \approx 1.54 \times 10^{-6}$.

For "four of a kind" events, there are 13 possible "kinds" and for each, the fifth card can be one of 48 possible other cards. The total probability is therefore $(13 \times 48)/2{,}598{,}960 = 1/4{,}165 = 2.4 \times 10^{-4}$.

3. 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 $\mathbf{e}$:

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

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

   (b) Prove the conditionalized version of Bayes' rule on evidence $\mathbf{e}$:

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

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

(a) We have

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

and

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

hence

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

(b) The derivation here is the same as the derivation of the simple version of Bayes' Rule. First we write down the dual form of the conditionalized product rule, simply by switching X and Y in the derivation of (a):

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

Therefore the two right-hand sides are equal:

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

Dividing through by $\mathbf{P}(Y \mid \mathbf{e})$ we get

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

4. Let us consider the example from the theory class: a domain consisting of just the three Boolean variables *Toothache, Cavity,* and *Catch* (the dentist's nasty steel probe catches in my tooth). The full joint distribution is a $2 \times 2 \times 2$ table as shown below.

	toothache		$\neg$*tootache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

Calculate the following:

(a) $P(toothache)$;

(b) $\mathbf{P}(Cavity)$;

(c) $\mathbf{P}(Toothache \mid cavity)$;

(d) $\mathbf{P}(Cavity \mid toothache \vee catch)$.

Solution:
The main point of this easy exercise is to understand the various notations of bold $\mathbf{P}$ versus non-bold P, uppercase versus lowercase variable names, and efficient computation of conditional probabilities using normalization, and where needed summing out the hidden variables.

(a) This asks for the probability that *Toothache* is true.

$$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2.$$

(b) This asks for the vector of probability values for the random variable *Cavity*. It has two values, which we list in the order $\langle true, false \rangle$. First add up $0.108 + 0.012 + 0.072 + 0.008 = 0.2$. Then we have $\mathbf{P}(Cavity) = \langle 0.2, 0.8 \rangle$.

(c) This asks for the vector of probability values for *Toothache*, given that *Cavity* is true.

$$\mathbf{P}(Toothache \mid cavity) = \alpha \langle 0.108{+}0.012, 0.072{+}0.008 \rangle = \alpha \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle.$$

Here we already knew from (a) that $\alpha = 1/0.2$, so we could have used that immediately as well, but in general we obtain it always simply by normalizing the sum of the probabilities of all outcomes to 1.

(d) This asks for the vector of probability values for *Cavity*, given $toothache \vee catch$, so proportinal to the vector where the first term is sum of the first three entries from the first row of the table ($toothache \vee catch$ when *cavity*) and the second term the sum of the first three entries from the second row of the table ($toothache \vee catch$ when $\neg cavity$):

$$\begin{aligned} \mathbf{P}(Cavity \mid toothache \vee catch) &= \alpha \langle \underbrace{0.108 + 0.012 + 0.072}_{0.192}, \underbrace{0.016 + 0.064 + 0.144}_{0.224} \rangle \\ &= (0.416)^{-1} \langle 0.192, 0.224 \rangle \\ &= \langle 0.4615, 0.5385 \rangle. \end{aligned}$$

5. 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.

Solution:

If the probability x is known, then successive flips of the coin are independent of each other, since we know that each flip of the coin will land heads with probability x. Formally, if $F1$ and $F2$ represent the results of two successive flips, we have

$$P(F1 = heads, F2 = heads \mid x) = P(F1 = heads \mid x)P(F2 = heads \mid x)$$

Thus, the events $F1 = heads$ and $F2 = heads$ are independent.

If we do not know the value of x, however, the probability of each successive flip is dependent on the result of all previous flips. The reason for this is that each successive flip gives us information to better estimate the probability x (i.e., determining the posterior estimate for x given our prior probability and the evidence we see in the most recent coin flip). This new estimate of x would then be used as our "best guess" of the probability of the coin coming up heads on the next flip. Since this estimate for x is based on all the previous flips we have seen, the probability of the next flip coming up heads depends on how many heads we saw in all previous flips, making them dependent.

6. **(Old exam question)** 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.

Solution:
Let the Boolean random variable V denote the virus presence in the patient, where v denotes that the patient has the virus, and $\neg v$ that he/she is virus free. Similarly, let the Boolean random variables A and B denote the possible outcomes of the two tests, each of which can be positive (values a and b, respectively) or negative ($\neg a$ and $\neg b$, respectively). From the problem statement we have:

$$\begin{aligned}
P(v) &= 0.01 \\
P(a \mid v) &= 0.95 \\
P(a \mid \neg v) &= 0.10 \\
P(b \mid v) &= 0.90 \\
P(b \mid \neg v) &= 0.05
\end{aligned}$$

The test whose positive result is more indicative of the virus being present is the one whose posterior probability, $P(v \mid a)$ or $P(v \mid b)$ is largest. One can compute these probabilities directly from the information given, finding that

$$\begin{aligned} P(v \mid a) &= \frac{P(a \mid v)P(v)}{P(a)} = \frac{P(a \mid v)P(v)}{P(a \mid v)P(v) + P(a \mid \neg v)P(\neg v)} \\ &= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.10 \times 0.99} = 0.0876, \end{aligned}$$

and similarly $P(v \mid b) = 0.1538$, so since $P(v \mid b) > P(v \mid a)$ we can conclude that test B is more indicative of someone really carrying the virus.

7. 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?

Solution:
It is given that $P(disease) = 0.0001$ and that $P(test \mid disease) = 0.99$, as well as $P(\neg\, test \mid \neg\, disease) = 0.99$. What we want to know is $P(disease \mid test)$. We will use the Bayes' rule to express this diagnostic probability in terms of what we know:

$$\begin{aligned} P(disease \mid test) &= \frac{P(test \mid disease)P(disease)}{P(test \mid disease)P(disease) + P(test \mid \neg\, disease)P(\neg\, disease)} \\ &= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} = 0.009804 \end{aligned}$$

which is less than 1%. So, even tough you tested positive, the probability that you have the disease is very low!