

Artificial Intelligence

Academic Year: 2019/2020

Instructor: Giorgio Fumera

Exercises on elements of probability theory and on Bayesian networks

1 Elements of probability theory

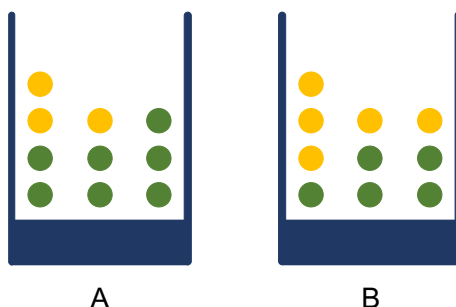
1. Define the *sample space* Ω (i.e., the set of *elementary events*, that is, all the possible, mutually disjoint outcomes) of the following experiments:
 - (a) Tossing a coin
 - (b) Throwing a dice
 - (c) Tossing two coins
 - (d) Throwing two dice
 - (e) Picking a card from a well shuffled deck of 52 cards
 - (f) Number of car accidents in Italy in 2020
 - (g) Number of hours a lightbulb burns before burning out (assume that only integer values are of interest, i.e., 1 hour, 2 hours, etc.)
 - (h) Number of rain days in Cagliari during December 2019 and total rainfall (in cm)
2. Considering the sample spaces defined in the previous exercise, compute the probabilities of the following events (i.e., subsets of the corresponding sample space) using the *classical* definition of probability:
 - (a) Getting $\heartsuit \heartsuit$ after tossing two dice
 - (b) Getting a total of 6 after tossing two dice
 - (c) Drawing the ace of hearts $\heartsuit A$ from a well shuffled deck
 - (d) Drawing an ace (A) or a spade (\spadesuit) from a well shuffled deck
3. Assume that the following axioms are used to define the probability function $P[\cdot]$ (remember that its argument is any event, i.e., any subset of the sample space Ω):
 - (i) $P[A] \geq 0$ for every event A
 - (ii) $P[\Omega] = 1$
 - (iii) if A_1 and A_2 are mutually exclusive (i.e., $A_1 \cap A_2 = \emptyset$), then $P[A_1 \cup A_2] = P[A_1] + P[A_2]$

Prove the following theorems using the above axioms and basic set theory:

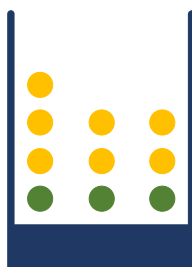
- (a) $P[\emptyset] = 0$
- (b) $P[\overline{A}] = 1 - P[A]$ (\overline{A} denotes the complement of A , that is $\Omega - A$)
- (c) for any events A and B , $P[A] = P[A \cap B] + P[A \cap \overline{B}]$
- (d) for any events A and B , $P[A \cup B] = P[A] + P[B] - P[A \cap B] \leq P[A] + P[B]$
- (e) for any events A and B , if $A \subseteq B$ then $P[A] \leq P[B]$

4. Compute the following conditional probabilities using the classical definition of probability:

- (a) A total greater than 6 is obtained after tossing two dice, given that the outcome for one of them is \square
- (b) Consider the two boxes in the figure below, labelled A and B, which contain a certain number of yellow and green balls. What is the probability that a randomly drawn ball from a randomly chosen box is green? What is the probability that a ball randomly drawn from box A is green?



5. Consider again the two boxes of the previous exercise, and assume that a ball randomly drawn from a randomly selected box turns out to be green: what is the probability that the ball comes from box A?
6. A doctor knows that 50% of the patients suffering from meningitis also have a stiff neck. She also knows that meningitis affects one out of 50,000 people, whereas stiff neck affects one out of 20 people. What is the probability that a given patient who has a stiff neck also has meningitis?
7. (a) Consider the experiment of tossing one dice, and let A denote the event that the outcome is \square or \square , and B the event that the outcome is an even number. Are A and B independent?
- (b) Consider now the experiment of tossing two dice, and let A denotes the event of an odd sum, B the event \square on the first dice, and C the event of a total of 7. Are A and B independent? Are A and C independent? Are B and C independent?
- (b) Consider the experiment of randomly drawing two balls *with replacement* from the box in the figure below (i.e., after the first draw the ball is returned to the box). What is the probability to draw two green balls?



8. Consider the *random variable* X defined as the total obtained after tossing two dice. What is the domain of X ? What is the *probability density function* of X , $f_X(x)$? (Give a graphical representation of this function.)
9. Consider again the experiment of tossing two dice.
 - (a) Let the random variables X and Y be defined respectively as the outcome of the first dice and the highest value among the two dice. What is the domain of the two-dimensional random variable X, Y ? What is the *joint probability density function* $f_{X,Y}(x, y)$? (Use the classical definition of probability to compute it.)

- (b) Let the random variables X and Y be defined respectively as the total of the two dice and the outcome of the first dice. What is the joint probability density function $f_{X,Y}(x,y)$? Compute the *marginal* density function $f_X(x)$ from the joint density function.
10. An instructor observed from past oral examinations that students knew the correct answer to 70% of questions. Now she observes that in a multiple-choice test with five alternatives, with only one of them correct, a student got the correct answer. How can she compute the probability that the student actually knew the correct answer?
11. A cheater is known to use in about 30% of his games of cards a stacked card deck in which the cards $\diamondsuit 2$, $\heartsuit 2$, $\clubsuit 2$, $\spadesuit 2$ are replaced with $\diamondsuit A$, $\heartsuit A$, $\clubsuit A$, $\spadesuit A$ (in other words, the stacked card deck contains eight aces – two for each suit – and no 2's). Assume that at the beginning of a game, after the deck has been shuffled, one of the other players randomly draws a card and gets a king: how can that player evaluate the probability that the deck is stacked? What if the drawn card was an ace?
12. Consider an ordinary, well shuffled deck of cards, and assume that three cards are consecutively drawn from it *without replacement* (i.e., after each draw the drawn card is *not* returned to the deck).
- Denoting with X_1 , X_2 and X_3 the random variables corresponding to the first, second and third card drawn, rewrite their joint probability density function using the *chain rule*.
 - Evaluate each of the probabilities appearing in the expression obtained from the chain rule, when $X_1 = A\diamondsuit$, $X_2 = J\heartsuit$, $X_3 = 2\spadesuit$.
 - Discuss whether any conditional independence assumption can be made on the conditional probabilities involved in the same expression.

2 Bayesian networks

1. Consider the following Boolean random variables related to the state of a given car:

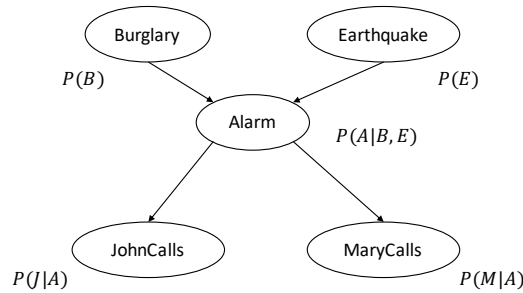
- *Battery*: is the battery charged?
- *Fuel*: is fuel tank empty?
- *Ignition*: does the ignition system work?
- *Moves*: does the car move?
- *Radio*: can the radio be switched on?
- *Starts*: does the engine fire?

Rewrite the joint probability density function of such variables using the chain rule, after defining a proper set of causal relationships between them, and represent it using a Bayesian network.

2. Two astronomers, in different parts of the world, look at the same region of the sky using their telescopes and count the number of stars it contains. Their counts may be inaccurate for several reasons, including the fact that their telescopes can occasionally (with a small probability) be out of focus. Draw a Bayesian network to represent the above information, after defining a proper set of causal relationships between the random variables involved.
3. Consider four random variables A , B , C and D . Rewrite their joint probability density function using the chain rule, by considering them sorted from D to A . Assuming that *no* conditional independence assumption can be made, draw the corresponding Bayesian network. What observations can be made on the resulting network structure? Assuming all the random variables to be Boolean, how many probability values need to be provided to define the joint probability density function using the derived expression of the chain rule, that is the corresponding Bayesian network?

4. Any probability which does not correspond to any node of a Bayesian network can in principle be estimated by suitably applying general rules like the definition of conditional probability, Bayes' formula, the *product* and *sum* rules, etc. to the available probabilities.

Consider now the following Bayesian network (refer to the textbook for its meaning):



Show how the following conditional probability distributions of the kind $P(\text{query}|\text{evidence})$ can be computed from the above network:

- $P(\text{Burglary}|\text{JohnCalls})$
- $P(\text{Burglary}|\text{Alarm})$
- $P(\text{JohnCalls}|\text{Burglary})$
- $P(\text{Burglary}|\text{Alarm}, \text{Earthquake})$
- $P(\text{Alarm}|\text{JohnCalls}, \text{Earthquake})$

Solution

1 Elements of probability theory

- The possible outcomes are “head” and “tail”; denoting them with the symbols H and T, the sample space can be represented as $\Omega = \{H, T\}$.
 - The possible outcomes are six values on the faces of the dice, e.g.: $\Omega = \{\square, \square, \square, \square, \square, \square\}$ or $\Omega = \{1, 2, 3, 4, 5, 6\}$.
 - Each possible outcome of the experiment is the pair of sides of the two coins; note that the two coins are *distinct* objects, and therefore the outcome (H, T) is different from (T, H). Accordingly: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$.
 - Analogously to the previous case, there are 36 possible outcomes, e.g.: $\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), (2, 2), \dots, (6, 6)\}$.
 - The outcomes are obviously 52. Each of them can be represented as a suit-value pair, e.g. (an ace is denoted by A): $\{(\diamond, 2), (\diamond, 3), \dots, (\diamond, A), (\heartsuit, 2), \dots, (\heartsuit, A), (\clubsuit, 2), \dots, (\clubsuit, A), (\spadesuit, 2), \dots, (\spadesuit, A)\}$.
 - In this case the possible outcomes are natural numbers. Since any upper limit would be arbitrary, a proper choice is to set the sample space equal to the (infinite) set of natural numbers $\Omega = \mathbb{N}$ (low probability values can be assigned to high values of the number of accidents).
 - Following a similar reasoning as in the case above, one can set $\Omega = \mathbb{N}$.
 - The sample space can be defined as the set of pairs of values (number of days, total rainfall in cm). Both values are non-negative integers; the number of days is also upper-bounded by 366 (note that 2020 is a leap year), whereas no meaningful upper bound can be defined for the total rainfall. Accordingly, $\Omega = \{(d, r) : d \in \mathbb{N}, r \in \mathbb{N}, d \leq 366\}$.

2. The classical definition of probability can be applied for all the considered experiments, since all their elementary events (i.e., the elements of the sample space Ω) can be considered *equally likely*.
 - (a) The probability of the (elementary) event (\heartsuit, \heartsuit) can be computed as the ratio between the number of favourable outcomes (one) and the total number of outcomes (the size of the sample space, that is 36): $P[(\heartsuit, \heartsuit)] = \frac{1}{36}$.
 - (b) This event is a subset of five elementary events: $\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$. Therefore there are 5 favourable outcomes out of 36, and the corresponding probability is $\frac{5}{36}$.
 - (c) The considered event is an elementary one, therefore its probability is $\frac{1}{52}$.
 - (d) In an ordinary deck (13 cards for each of the four suits) there are 16 cards (i.e., elementary events of the considered experiment) that are either an ace or a spade, *or both* (note that the ace of spade $\spadesuit A$ must be considered only once). Therefore the corresponding probability is $\frac{16}{52} = \frac{4}{13}$.
3. These are possible proofs:
 - (a) Consider two events $A_1 = \emptyset$ and $A_2 = \emptyset$; since $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 = \emptyset$, from axiom (iii) it follows that $P[\emptyset] = P[\emptyset] + P[\emptyset]$, which can hold only if $P[\emptyset] = 0$.
 - (b) Since $A \cup \bar{A} = \Omega$ and $A \cap \bar{A} = \emptyset$, it follows from axiom (iii) that $P[\Omega] = P[A \cup \bar{A}] = P[A] + P[\bar{A}]$. Since $P[\Omega] = 1$ by axiom (ii), the result follows.
 - (c) For any pair of sets A and B the two following equalities hold: $A = (A \cap B) \cup (A \cap \bar{B})$, and $(A \cap B) \cap (A \cap \bar{B}) = \emptyset$. Therefore, from axiom (iii) one gets $P[A] = P[(A \cap B) \cup (A \cap \bar{B})] = P[A \cap B] + P[A \cap \bar{B}]$.
 - (d) For any pair of sets A and B the two following equalities hold: $A \cup B = A \cup (\bar{A} \cap B)$, and $A \cap (\bar{A} \cap B) = \emptyset$. Axiom (iii) can now be applied again to obtain: $P[A \cup B] = P[A] + P[\bar{A} \cap B]$. Now theorem (c) proven above can be applied to the term $P[\bar{A} \cap B]$ in the above expression (by exchanging A and B) to get $P[\bar{A} \cap B] = P[B] - P[A \cap B]$. By replacing this expression in the previous one, one obtains: $P[A \cup B] = P[A] + P[B] - P[A \cap B]$. Finally, from axiom (i) it follows that $P[A \cap B] \geq 0$, and therefore $P[A \cup B] \leq P[A] + P[B]$.
 - (e) Any set B can be written as $(B \cap A) \cup (B \cap \bar{A})$. If $A \subseteq B$, then $B \cap A = A$, and therefore $B = A \cup (B \cap \bar{A})$, with $A \cap (B \cap \bar{A}) = \emptyset$. From axiom (iii) it follows that $P[B] = P[A] + P[B \cap \bar{A}]$. Finally, since $P[B \cap \bar{A}] \geq 0$ by axiom (i), it follows that $P[A] \leq P[B]$.
4.
 - (a) If the outcome for one of the dice is \heartsuit , then there are 6 favourable outcomes: $(\heartsuit, \heartsuit), (\heartsuit, \clubsuit), (\heartsuit, \diamondsuit), (\heartsuit, \spadesuit), (\clubsuit, \heartsuit)$ and (\spadesuit, \heartsuit) . Note indeed that the conditioning event refers to *one* of the dice, not to the *first* or to the *second* dice. The requested probability is therefore $\frac{6}{36} = \frac{1}{6}$.
 - (b) The sample space of the considered experiment is made up of *twenty* mutually disjoint outcomes, each of which can be described as a box-ball pair:
 - (A, yellow ball): 3 distinct elementary events;
 - (A, green ball): 7 distinct elementary events;
 - (B, yellow ball): 5 distinct elementary events;
 - (B, green ball): 5 distinct elementary events.
 These elementary events can be considered equally likely, and therefore the classical definition of probability can be applied. The favourable outcomes (drawing a green ball) are 12, thus the corresponding probability is $\frac{12}{20} = \frac{3}{5}$.
 If a ball is randomly selected from box A, instead, there are 7 favourable outcomes out of 10, and therefore $P[\text{green ball} \mid \text{box A}] = \frac{7}{10}$.
5. The probability of interest can be denoted by $P[\text{box A} \mid \text{green ball}]$, which can be computed in several ways.

The sample space in this case is made up of the 12 elementary events corresponding to a green ball, and 7 of them are favourable outcomes (balls coming from box A), which leads to $P[\text{box A} \mid \text{green ball}] = \frac{7}{12}$.

The same probability can also be computed using the definition of conditional probability, as:

$$\frac{P[\text{box A} \cap \text{green ball}]}{P[\text{green ball}]}.$$

It is easy to see that the probability at the numerator involves 7 favourable outcomes out of 20, whereas the one at the denominator involves 12 favourable outcomes out of 20, which leads again to $P[\text{box A} \mid \text{green ball}] = \frac{7}{12}$.

One can also apply Bayes' formula to get:

$$P[\text{box A} \mid \text{green ball}] = \frac{P[\text{green ball} \mid \text{box A}]P[\text{box A}]}{P[\text{green ball}]}.$$

The above probabilities can be easily evaluated as $P[\text{box A}] = \frac{1}{2}$, $P[\text{green ball} \mid \text{box A}] = \frac{7}{10}$, and (see above) $P[\text{green ball}] = \frac{12}{20} = \frac{3}{5}$, which leads again to the same result obtained above.

6. The sample space can be defined as the set of individuals of the population of interest. Let M and S denote the events corresponding respectively to the subset of individuals suffering from meningitis and from stiff neck. The doctor's knowledge can be expressed as $P[S|M] = 0.5$, $P[M] = \frac{1}{50,000} = 0.00002$ and $P[S] = 0.05$. The probability she is interested in is $P[M|S]$. By applying the definition of conditional probability she gets:

$$P[M|S] = \frac{P[M \cap S]}{P[S]}.$$

The value $P[S]$ is known. However $P[M \cap S]$, i.e., the probability that an individual suffers *both* from meningitis *and* from stiff neck, cannot be directly computed from the above statistics. The doctor can nevertheless resort to Bayes' formula:

$$P[M|S] = \frac{P[S|M]P[M]}{P[S]}.$$

In this case all three terms in the above expression are known, which immediately leads to:

$$P[M|S] = \frac{P[S|M]P[M]}{P[S]} = \frac{0.5 \times 0.00002}{0.05} = 0.0002.$$

7. Remember that two events A and B defined on the same sample space are *defined* to be independent if and only if any of the following (equivalent) conditions is satisfied:

- $P[A \cap B] = P[A]P[B]$
- $P[A|B] = P[A]$, if $P[B] > 0$
- $P[B|A] = P[B]$, if $P[A] > 0$

- (a) In this case the easiest condition to be checked is the first one above. First, it is easy to see that $P[A] = \frac{2}{6} = \frac{1}{3}$ and $P[B] = \frac{3}{6} = \frac{1}{2}$. To evaluate $P[A \cap B]$ one can consider that the event “the outcome is \square or \square , and is an even number” corresponds to the elementary event \square , whose probability is $P[A \cap B] = \frac{1}{6}$. It immediately follows that $P[A \cap B] = P[A]P[B]$, and thus A and B are independent.
- (b) The probabilities of the three events can be easily computed as $P[A] = \frac{1}{2}$, $P[B] = \frac{1}{6}$ and $P[C] = \frac{1}{6}$ (there are 6 outcomes where the total is 7). The event $A \cap B$ corresponds to “odd sum *and* \square on the first dice”, which is made up of the three elementary events (\square, \square) , (\square, \square) and (\square, \square) . Therefore $P[A \cap B] = \frac{3}{36} = \frac{1}{12}$, which equals $P[A]P[B]$; it follows that A and B are independent. The event $A \cap C$ corresponds to “odd sum *and* a total of 7”, which is made up of the two elementary events (\square, \square) and (\square, \square) , leading to $P[A \cap C] = \frac{2}{36} = \frac{1}{18}$; this is different from $P[A]P[C]$, and therefore A and C are not independent. Finally, the event $B \cap C$ corresponds to “ \square on the first dice *and* a total of 7”, which corresponds to the elementary event (\square, \square) ; therefore $P[B \cap C] = \frac{1}{36}$, which equals $P[B]P[C]$, and thus B and C are independent.

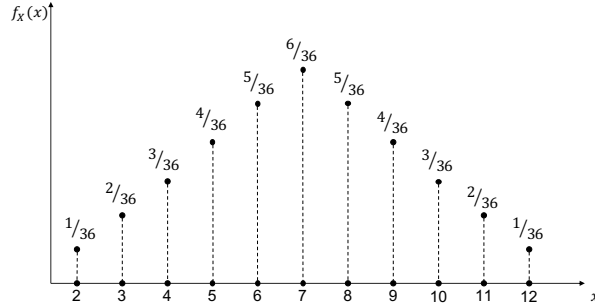
(c) The sample space can be represented as the set of pairs of colours. It is not difficult to see that the possible mutually disjoint and equally likely outcomes are 100, among which:

- (green, green): 9 outcomes (3 different green balls from the first draw \times 3 different green balls from the second draw)
- (green, yellow): $3 \times 7 = 21$ outcomes
- (yellow, green): $7 \times 3 = 21$ outcomes
- (yellow, yellow): $7 \times 7 = 49$ outcomes

The event of drawing two green balls corresponds to 9 outcomes out of 100 (see above), and therefore its probability is $\frac{9}{100}$.

An alternative way to compute the same probability is to observe that, since the draws are made with replacement, the outcome of the second draw is not “affected” by the outcome of the first. More precisely, denoting the event of a green ball from the first draw by A and the event of a green ball from the second draw by B , this means that $P[B|A] = P[B]$, i.e., the two events are independent. It is easy to see that the event of interest is $A \cap B$. Since A and B are independent it follows that $P[A \cap B] = P[A]P[B]$. It is also easy to see that A corresponds to $21 + 9 = 30$ elementary events (favourable outcomes), and B to $9 + 21 = 30$ elementary events as well. Therefore $P[A] = P[B] = \frac{30}{100} = \frac{3}{10}$, and thus $P[A \cap B] = P[A]P[B] = \frac{9}{100}$.

8. The domain of X is the subset of natural numbers $\mathcal{X} = \{2, 3, 4, \dots, 12\}$. The probability density function $f_X(x)$, with $x \in \mathcal{X}$, is defined as $P[X = x]$, and can be computed using the classical definition of probability. For instance, the event $X = 2$ corresponds to the single elementary event (\square, \square) out of the possible 36 mutually disjoint and equally likely outcomes, and therefore $f_X(2) = P[X = 2] = \frac{1}{36}$. The other values can be computed similarly. A graphical representation of $f_X(x)$ is given below:



9. (a) The sample space of the considered experiment (tossing two dice) has already been defined as the set of 36 elementary events $\Omega = \{(\square, \square), (\square, \blacksquare), \dots, (\blacksquare, \blacksquare)\}$. Note now that both random variables X and Y have the same domain $\mathcal{X} = \mathcal{Y} = \{1, 2, 3, 4, 5, 6\}$, and that they can jointly take on only the following pairs of (X, Y) values:

(6, 6),
 (5, 6), (5, 5),
 (4, 6), (4, 5), (4, 4),
 (3, 6), (3, 5), (3, 4), (3, 3),
 (2, 6), (2, 5), (2, 4), (2, 3), (2, 2),
 (1, 6), (1, 5), (1, 4), (1, 3), (1, 2), (1, 1) .

To compute $f_{X,Y}(x, y)$ for any of the above pairs (x, y) one should determine how many elementary events belong to the corresponding event $X = x \cap Y = y$. For instance, the event $X = 6 \cap Y = 6$ corresponds to the 6 elementary events $(\blacksquare, \square), (\blacksquare, \blacksquare), \dots, (\blacksquare, \blacksquare)$, and therefore $f_{X,Y}(6, 6) = \frac{6}{36} = \frac{1}{6}$. The values of $f_{X,Y}(x, y)$ for the other (x, y) pairs above can be determined similarly. For any other (x, y) pair the probability is zero, e.g., $f_{X,Y}(6, 5) = P[X = 6, Y = 5] = 0$.

(b) The solution can be found in a similar way as in the previous exercise.

10. Let A denote the event that a given student *got* the right answer and B denote the event that the student *knew* the right answer. The probability of interest is $P[B|A]$.

The instructor knows from past (oral) exams that $P[B] = 0.7$. She can also assume that the probability that any given student gets the right answer when he knows it is 1, i.e., $P[A|B] = 1$. Note that a value of 1 is not necessarily correct: for instance a student may misunderstand the question or may be nervous, and thus may give a wrong answer even if he knows the right one, which would lead to $P[A|B] < 1$.

The instructor can also estimate the probability that a student gives a correct answer when he does not know it, $P[A|\bar{B}]$, by assuming that in such a case the student guesses; since there are five choices and only one of them is correct, she can assume $P[A|\bar{B}] = \frac{1}{5}$. This assumption may be not realistic as well: for instance, even when a student does not know the right answer, he may know that some alternatives are wrong, in which case the probability of guessing correctly would be better than $\frac{1}{5}$.

Using the available probabilities, the probability $P[B|A]$ can be computed using Bayes' formula. First, $P[B|A]$ can be written as:

$$P[B|A] = \frac{P[A|B]P[B]}{P[A]}.$$

The two terms in the numerator are already known. The denominator can be computed using the sum and product rules, taking into account that for any pair of events A and B we have $A = (A \cap B) \cup (A \cap \bar{B})$:

$$P[A] = P[A \cap B] + P[A \cap \bar{B}] = P[A|B]P[B] + P[A|\bar{B}]P[\bar{B}].$$

All terms in the above expression are known as well (taking into account that $P[\bar{B}] = 1 - P[B]$), which leads to:

$$P[B|A] = \frac{P[A|B]P[B]}{P[A|B]P[B] + P[A|\bar{B}]P[\bar{B}]} = \frac{1 \cdot 0.7}{1 \cdot 0.7 + \frac{1}{5} \cdot 0.3} \approx 0.92.$$

11. The experiment consists in randomly drawing a card from a deck which can be stacked or not. The elementary events can be described as a set of $52 + 52 = 104$ deck-card pairs:

$\{(\text{ordinary deck}, 2\heartsuit), (\text{ordinary deck}, 3\heartsuit), \dots, (\text{ordinary deck}, \text{A}\spadesuit),$
 $(\text{stacked deck}, \text{A}\heartsuit), (\text{stacked deck}, 3\heartsuit), \dots, (\text{stacked deck}, \text{A}\heartsuit), \dots, (\text{stacked deck}, \text{A}\spadesuit)\}$

Note that the elementary events corresponding to the stacked deck do not contain any card with value 2, and contain two instances of the ace of each suit. Note also that the above elementary events *cannot* be considered equally likely, since an ordinary deck is not used half of the times by the cheater.

Consider the following random variables:

- X is a Boolean random variable denoting whether the deck is stacked ($X = \text{True}$) or ordinary ($X = \text{False}$);
- Y is a random variable corresponding to the card value, whose domain is therefore $\{2, 3, \dots, 9, \text{J}, \text{Q}, \text{K}, \text{A}\}$.

The probability of interest is $P[X = \text{True}|Y = \text{K}]$. The player's knowledge about the cheater can be expressed by the following probability values:

- $P[X = \text{True}] = 0.3$;
- $P[Y = y|X = \text{False}] = \frac{4}{52} = \frac{1}{13}$ for any value of y (an ordinary deck contains 4 cards of each value);
- $P[Y = 2|X = \text{True}] = 0$ (a stacked deck contains no card of value 2);
- $P[Y = \text{A}|X = \text{True}] = \frac{8}{52} = \frac{2}{13}$ (a stacked deck contains 8 aces);
- $P[Y = y|X = \text{True}] = \frac{4}{52} = \frac{1}{13}$ for any $y \neq 2, \text{A}$ (a stacked deck contains 4 cards of each of the other values, including K, as an ordinary deck).

The probability of interest can now be computed using Bayes' formula:

$$\begin{aligned}
P[X = \text{True}|Y = K] &= \frac{P[Y = K|X = \text{True}]P[X = \text{True}]}{P[Y = K]} \\
&= \frac{P[Y = K|X = \text{True}]P[X = \text{True}]}{P[Y = K|X = \text{True}]P[X = \text{True}] + P[Y = K|X = \text{False}]P[X = \text{False}]} \\
&= \frac{\frac{1}{13} \cdot 0.3}{\frac{1}{13} \cdot 0.3 + \frac{1}{13} \cdot 0.7} = 0.3 .
\end{aligned}$$

This means that $P[X = \text{True}|Y = K] = P[X = \text{True}]$, that is drawing a king from the deck at hand does not provide any *additional* information about the deck being stacked or not (in other words, the events $X = \text{True}$ and $Y = K$ are independent).

If an ace is drawn, instead:

$$\begin{aligned}
P[X = \text{True}|Y = A] &= \frac{P[Y = A|X = \text{True}]P[X = \text{True}]}{P[Y = A]} \\
&= \frac{P[Y = A|X = \text{True}]P[X = \text{True}]}{P[Y = A|X = \text{True}]P[X = \text{True}] + P[Y = A|X = \text{False}]P[X = \text{False}]} \\
&= \frac{\frac{2}{13} \cdot 0.3}{\frac{2}{13} \cdot 0.3 + \frac{1}{13} \cdot 0.7} \approx 0.46 .
\end{aligned}$$

In other words, it is more likely that the deck is stacked if an ace is drawn, than if a king is drawn.

It is also straightforward to see that drawing a 2 gives the *certainty* that the deck is ordinary (i.e., $P[X = \text{True}|Y = 2] = 0$), whereas $P[X = \text{True}|Y = y] = 0.3$ for $y = 3, 4, \dots, Q$ (beside $y = K$ as shown above).

12. (a) Since the events corresponding to the three random variables are sequential, it makes sense to sort the variables in the chain rule in chronological order:

$$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_2, X_1) .$$

- (b) It is straightforward to see that $P(X_1 = A\heartsuit) = \frac{1}{52}$. To evaluate the two conditional probabilities one should first consider that after the first card is drawn without replacement, the probability of drawing any of the *remaining* cards is $\frac{1}{51}$, and therefore: $P(X_2 = J\heartsuit|X_1 = A\heartsuit) = \frac{1}{51}$. Similarly, after the first two cards are drawn without replacement, the probability of drawing any of the remaining ones is $\frac{1}{50}$, and therefore: $P(X_3 = 2\spadesuit|X_2 = J\heartsuit, X_1 = A\heartsuit) = \frac{1}{50}$.
- (c) From the above discussion it is clear that no conditional independence assumption can be made, since after drawing without replacement $n \geq 1$ cards the probability that any of the remaining card is drawn changes from $\frac{1}{52-n+1}$ to $\frac{1}{52-n}$, whereas the probability that any of the cards already drawn is drawn again becomes zero.

2 Bayesian networks

1. The state of the battery and of the fuel tank can be seen as the “root causes”. They can also be considered independent on each other. Whether the radio works or not directly depends only on the battery state. The same for the ignition system, whose functioning can be considered independent on that of the radio, given the state of the battery. The state of the fuel tank, together with the ignition system, directly influences whether the engine fires. Finally, it can be assumed that whether the car moves or not directly depends only on the engine state.

Accordingly, the considered random variables can be sorted from the “root causes” to the “end effects” as follows: *Fuel, Battery; Radio, Ignition; Starts; Moves*.

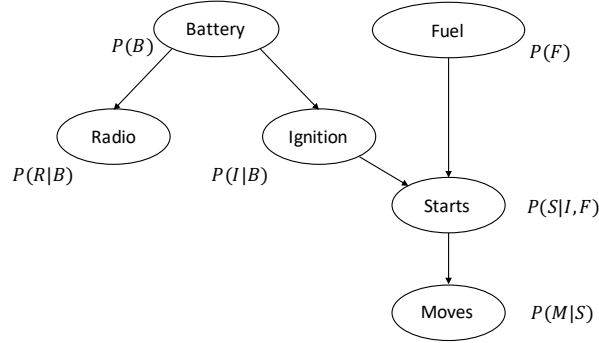
Their joint probability can then be factorized through the chain rule as follows (only the initial letters of the random variables are used, for the sake of brevity):

$$P(F, B, R, I, S, M) = P(F)P(B|F)P(R|B, F)P(I|R, B, F)P(S|I, R, B, F)P(M|S, I, R, B, F) .$$

The conditional independence assumptions described above finally lead to simplify the joint probability as follows:

$$P(F, B, R, I, S, M) = P(F)P(B)P(R|B)P(I|B)P(S|I, F)P(M|S) .$$

The corresponding Bayesian network is shown below.

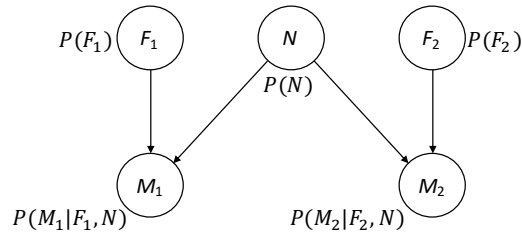


2. Five random variables can be used to describe the relevant information:

- M_1 and M_2 : the number of stars counted by the two astronomers (their domain is the set of natural numbers);
- F_1 and F_2 : Boolean random variables representing whether the two telescopes are out of focus (*True*) or not (*False*);
- N , the actual (unknown) number of stars in the region of the sky under observation (its domain is the set of natural numbers).

The actual number of stars and the states of the two telescopes can be considered as independent root causes. The number of stars estimated by each of two astronomers is directly influenced by the actual number of stars and by the state of the corresponding telescope, but not by the state of the other telescope, nor by the number of stars estimated by the other astronomer (unless they communicate with each other).

This is the resulting Bayesian network:



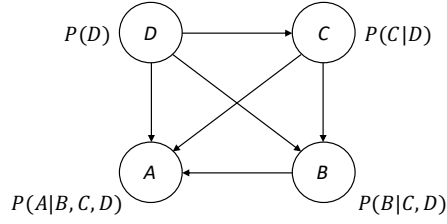
The corresponding factorization of the joint probability is therefore:

$$P(N, F_1, F_2, M_1, M_2) = P(N)P(F_1)P(F_2)P(M_1|N, F_1)P(M_2|N, F_2) .$$

3. By applying the chain rule, the joint probability can be rewritten as:

$$P(A, B, C, D) = P(D)P(C|D)P(B|C, D)P(A|B, C, D) .$$

The corresponding Bayesian network is:



Note that the above graph is *fully connected*, i.e., there is an (oriented) arc between every pair of node. This is a general characteristic of any Bayesian network when no conditional independence assumption can be made on the corresponding conditional probabilities.

To define the unconditional probability distribution of a Boolean random variable X , only one value need to be provided: either $P(X = \text{True})$ or $P(X = \text{False})$, since the other value is determined by the constraint $P(X = \text{True}) + P(X = \text{False}) = 1$.

To define the probability distribution of X conditioned on n other Boolean random variables Y_1, \dots, Y_n , $P(X|Y_1, \dots, Y_n)$, one has to provide either the probability that $X = \text{True}$ or the probability that $X = \text{False}$ for *every* of the 2^n possible combinations of values of Y_1, \dots, Y_n .

Accordingly, the number of probability value that have to be provided (e.g., estimated from available statistics) for the considered Bayesian network is:

- 1 for $P(D)$,
- 2 for $P(C|D)$,
- 4 for $P(B|C, D)$, and
- 8 for $P(A|B, C, D)$,

for a total of 15 values.

In general, it is easy to see that to completely define a fully connected Bayesian network involving n Boolean random variables (equivalently, to define their joint probability distribution) the number of probability values to be provided is $\sum_{k=1}^n 2^{k-1} = 2^n - 1$. This value increases exponentially with the number of random variables; in practice, this means that conditional independence assumptions are useful to reduce the effort required to define the joint probability distribution for a problem at hand.

4. Only the derivation of probability (a) is reported; the other ones are left as exercises. In the following only the initial letter of each random variable is used, for the sake of brevity.

$P(B|J) = \frac{P(J|B)P(B)}{P(B)}$ (Bayes' formula). The term $P(B)$ can be obtained from node B .

The term $P(J|B)$ can be computed as follows:

$$\begin{aligned}
 P(J|B) &= P(J, A = \text{True}|B) + P(J, A = \text{False}|B) \quad (\text{sum rule, or marginalization}) \\
 &= P(J|A = \text{True}, B)P(A = \text{True}) + P(J|A = \text{False}, B)P(A = \text{False}) \quad (\text{product rule}) \\
 &= P(J|A = \text{True})P(A = \text{True}) + P(J|A = \text{False})P(A = \text{False}) \quad (\text{conditional independence encoded in the Bayesian network}).
 \end{aligned}$$

Now the distribution $P(A)$ needs to be computed:

$$\begin{aligned}
 P(A) &= \sum_{B, E=\{\text{True}, \text{False}\}} P(A, B, E) \quad (\text{sum rule}) \\
 &= \sum_{B, E=\{\text{True}, \text{False}\}} P(A|B, E)P(B, E) \quad (\text{product rule}) \\
 &= \sum_{B, E=\{\text{True}, \text{False}\}} P(A|B, E)P(B)P(E) \quad (\text{conditional independence encoded in the Bayesian network}).
 \end{aligned}$$

Finally, the distribution $P(J)$ remains to be computed. Proceeding as in the first derivation above:

$$\begin{aligned}
 P(J) &= P(J, A = \text{True}) + P(J, A = \text{False}) \quad (\text{sum rule}) \\
 &= P(J|A = \text{True})P(A = \text{True}) + P(J|A = \text{False})P(A = \text{False}) \quad (\text{product rule}).
 \end{aligned}$$

The distribution $P(B|J)$ is now expressed only in terms of distributions associated to nodes of the Bayesian network.