University of Cagliari

MSc programs in Computer Engineering, Cybersecurity and Artificial Intelligence, and in Electronic Engineering

Artificial Intelligence

Academic Year: 2023/2024
Instructor: Giorgio Fumera

Exercises on Bayesian networks

1. Consider the following Boolean random variables related to the state of a car: *Battery* (it equals `f` if the battery is dead), *Fuel* (it equals `f` if the fuel tank is empty), *Ignition* (it equals `t` if the ignition system works), *Moves* (it equals `t` if the car moves after one tries to start the engine), *Radio* (it equals `t` if the radio works when one tries to switch it on), *Starts* (it equals `t` if the engine fires when one tries to start it).

    (a) Define a proper set of causal relationships between the above random variables, and write the corresponding expression of their joint probability density function using the chain rule.

    (b) Make suitable conditional independence assumptions, *clearly motivating them*, to simplify the expression of the joint density.

    (c) Graphically represent the joint density using a Bayesian network.

    (d) How many probability values need to be estimated to define the conditional and prior densities associated to the nodes of your Bayesian network? How could such density funcitons be estimated, in practice?

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 they see. Their counts may be inaccurate for several reasons, including the fact that their telescopes can occasionally (with a small probability) be out of focus. Define a set of random variables to describe the above domain, define a proper set of causal relationships between them, make suitable conditional independence assumptions, and draw the corresponding Bayesian network.

3. Mary's car has an alarm that sounds when a motion sensor detects someone entering the car. The alarm and the sensor are powered by two distinct batteries, which can be occasionaly dead. Draw a Bayesian network for this problem, clearly motivating the underlying conditional independence assumptions.

4. Headache and fever are among the symptoms of several health problems, including influenza and food poisoning.

    (a) Represent the above knowledge as a Bayesian network, and clearly explain the conditional independence assumptions you make.

    (b) Write down the expression of the joint density function of the chosen random variables, according to your Bayesian network.

    (c) Derive an expression of the probability that a person who is suffering from headache has caught influenza, in terms of the density functions associated to your Bayesian network.

5. In a nuclear power station an alarm sounds and warning lights flash in the control room, when a sensor detects that the temperature of the core exceeds a given threshold. The sensor measurement may be incorrect (very rarely), resulting in false positive or false negative detections; this is less unlikely to happen, when the external temperature gets too high. Occasionally, also the alarm and the warning lights can fail; to limit joint failures, they are implemented as physically separated systems.

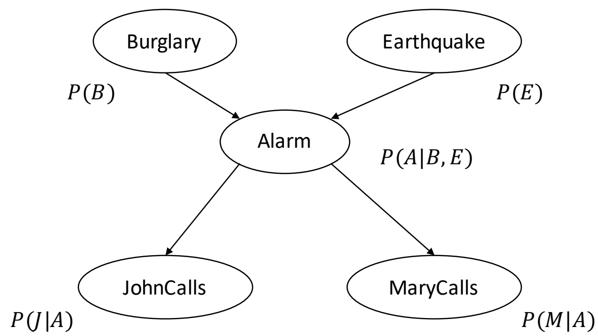
(a) Represent the above domain as a Bayesian network, and clearly explain the conditional independence assumptions you make.

(b) Write down the expression of the joint density function of the chosen random variables using the chain rule, according to your Bayesian network.

(c) Derive an expression of the probability that a core overheating occurred, if warning lights are flashing in the control room, as a function of the density functions associated to your Bayesian network.

6. Consider four random variables A, B, C and D. Write their joint probability density function using the chain rule, by considering the following causal ordering between them: $D \to C \to B \to A$ (i.e., D is assumed to be the "root cause"). Assuming that *no* conditional independence assumption can be made, draw the corresponding Bayesian network. What observations can be made about the resulting network topology? Assuming all four random variables are Boolean, how many probability values need to be estimated to define the density functions associated to the above Bayesian network?

7. Consider the following Bayesian network (see the textbook for its meaning):



Assume you are interested in computing the following conditional probabilities of the form $P(\text{query}|\text{evidence})$ (for the sake of brevity only the initial letter of each random variable is used, and the symbols `t` and `f` are used to denote the values `t` and `f`):

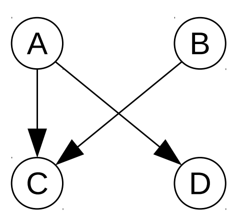
(a) $P(B = \mathtt{t}|J = \mathtt{t})$ (what is the probability that there was a burglary, given that John called?)

(b) $P(B = \mathtt{t}|A = \mathtt{t})$ (what is the probability that there was a burglary, given that the alarm sounded?)

(c) $P(B = \mathtt{t}|A = \mathtt{t}, E = \mathtt{f})$ (what is the probability that there was a burglary, given that that the alarm sounded and there was no earthquake?)

Compute the above probabilities by rewriting them in terms of the density functions associated with the nodes of the above Bayesian network.

8. Assume that the Bayesian network below has been built by considering the following ordering between its random variables: A, B, C, D:



(a) Rewrite the joint density function of the four random variables using the chain rule, correspoding to the above Bayesian network.

(b) What conditional independence assumptions does the above Bayesian network encode?

(c) Assuming the above random variables are all Boolean, compute $P(A = \mathtt{t}|C = \mathtt{t}, D = \mathtt{f})$ in terms of the conditional densities associated with the nodes of the Bayesian network.

Solution

1. (a) The state of the battery and of the fuel tank can be seen as the "root causes". The battery state directly affects the working of the radio and of the ignition system. In turn, the state of the fuel tank and of the ignition system directly determine whether the engine fires or not. Finally, the state of the engine determines whether the car moves.

   Accordingly, the considered random variables can be sorted from the "root causes" to the "end effects" as follows (note that the relative ordering between *Fuel* and *Battery* and between *Radio* and *Ignition* is arbitrary): $Fuel \rightarrow Battery \rightarrow Radio \rightarrow Ignition \rightarrow Starts \rightarrow Moves$. The corresponding expression of their joint probability density function through the chain rule is given by (only the initial letters of the random variables are used, for the sake of brevity):

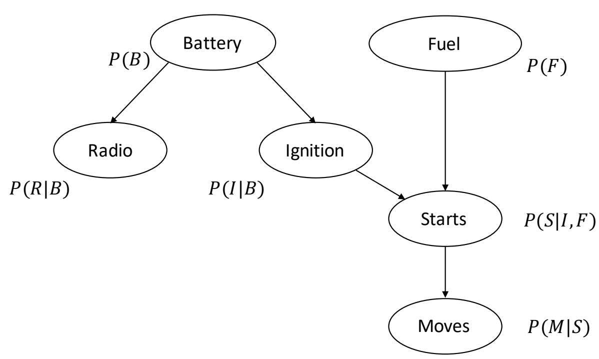
$$P(M,S,I,R,B,F) = P(M|S,I,R,B,F)P(S|I,R,B,F)P(I|R,B,F)P(R|B,F)P(B|F)P(F) \ .$$

   (b) The state of the battery and of the fuel tank (B and F, the "root causes") can be considered independent on each other, i.e., $P(B|F) = P(B)$. Whether the radio works or not, given the battery state, is independent on the state of the fuel tank: $P(R|B,F) = P(R|B)$. The working of the ignition system can be considered independent on the state of the radio and of the fuel tank, given the state of the battery: $P(I|R,B,F) = P(I|B)$. Given the state of the fuel tank and of the ignition system, the working of the engine is independent on the state of the radio and of the battery (note that the battery affects the engine only *indirectly*, through the ignition system): $P(S|I,R,B,F) = P(S|I,F)$. Finally, given the state of the engine, whether the car moves or not can be considered independent on all the other considered factors: $P(M|S,I,R,B,F) = P(M|S)$.

   The above conditional independence assumptions allow the joint density to be simplified as follows:

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

   (c) The corresponding Bayesian network is shown below.



   (d) Remember that, to define the *unconditional* density $P(X)$ of any Boolean random variable X, only one value needs to be estimated: either $P(X = \mathtt{t})$ or $P(X = \mathtt{f})$, since the other value is determined by the constraint $P(X = \mathtt{t}) + P(X = \mathtt{f}) = 1$. To define the *conditional* density of X given the values of n other Boolean random variables $Y_1, \ldots, Y_n$, $P(X|Y_1, \ldots, Y_n)$, either the probability that $X = \mathtt{t}$ or the probability that $X = \mathtt{f}$ need to be estimated, for *each* of the 2^n possible combinations of values of $Y_1, \ldots, Y_n$.

   Taking into account that all the considered random variables are Boolean:

   - $P(F)$ and $P(B)$ require the estimation of one probability value each, e.g., $P(F = \mathtt{t})$ and $P(B = \mathtt{t})$, therefore 2 values in total;
   - $P(R|I)$ requires one value for $I = \mathtt{t}$ and one for $I = \mathtt{f}$; similarly for $P(R|B)$ and $P(M|S)$; therefore, these densities require 6 values in total;
   - $P(S|I,F)$ requires one value for each of the four combinations of values of I and F, therefore 4 values in total.

The joint density can therefore be estimated by $2+6+4=12$ probability values, thanks to the above conditional independence assumptions, instead of $2^6-1=63$ values.

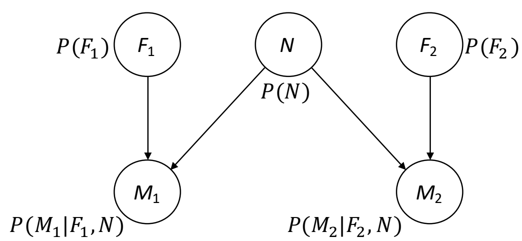
Some values of the above probability densities could be estimated *a priori*, by considering physical constraints. In particular, the ignition system and the radio cannot work, if the battery is dead, i.e., $P(I=\texttt{t}|B=\texttt{f})=P(R=\texttt{t}|B=\texttt{f})=0$. Similarly, if either the ignition system does not work or the fuel tank is empty, the engine cannot fire: $P(S=\texttt{t}|I=\texttt{f},F=f)=0$, where $f\in\{\texttt{t},\texttt{f}\}$, and $P(S=\texttt{t}|I=i,F=\texttt{f})=0$, where $i\in\{\texttt{t},\texttt{f}\}$. On the other hand, for instance, even if the battery is not dead, the ignition system may not work, due for instance to a broken fuel pump, or to a fault in the ignition system itself: therefore, $P(I=\texttt{f}|B=\texttt{t})$ should not be estimated as zero, to account for other possible causes of the event $I=\texttt{f}$ that one may not know, or may not be willing to consider *explicitly*. In general, the remaining probabilities may be estimated either using a frequentist approach (i.e., by carrying out several experiments involving several cars over some period of time), or subjectively, possibly based on expert knowledge (depending also on the real application scenario).

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 (`t`) or not (`f`);
   - 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 N of stars and the states of the two telescopes (F_1 and F_2) can be considered as independent root causes that determine the number of stars estimated by two astronomers. Such a number 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).

   The above considerations lead to the following Bayesian network:



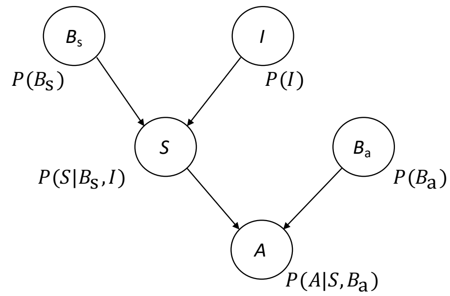
   The corresponding expression of the joint density is therefore:

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

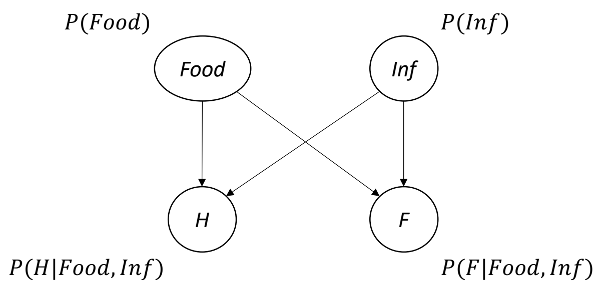
   Note that $P(M_1|N,F_1=\texttt{f})$ and $P(M_2|N,F_2=\texttt{f})$ (i.e., the probability density funciton of the estimated number of stars, for a certain portion of sky containing N stars, if the telescope is not out of focus) should be greater than zero, even if $M_1\neq N$ or $M_2\neq N$; the reason is that the estimated number of stars may differ in several ways from the actual value N due to other possible causes that are not explicitly taken into account in this problem formulation, beside the telescope being out of focus, e.g., the sky may be not perfectly clear when the observation is made.

3. This domain can be described by five Boolean random variables denoting the presence of someone inside the car (I), the state (dead or not dead) of the batteries powering the sensor (B_s) and the alarm (B_a), the state (detection or no detection) of the sensor (S) and the state (sounding or not sounding) of the alarm (A). It is reasonable to assume that: the "root causes" correspond to the random variables I, B_s and

B_a; such random variables are independent on each other; the state of the sensor is directly influenced by the state of its battery, and on the presence of someone inside the car; the state of the alarm is directly influenced only by the state of its battery and of the sensor. The corresponding Bayesian network is:



4. (a) Four Boolean random variables can be used to represent the occurrence of the two symptoms (H for headache and F for fever) and of the two health problems (Inf for influenza and $Food$ for food poisoning). In general, symptoms are caused by health problems. In particular, the two health problems considered here can be assumed to be independent on each other, and to be direct causes of both symptoms. Finally, if the presence or absence of influenza and food poisoning is known, then headache can be considered to be independent on fever. The corresponding Bayesian network is shown below.



(b) The joint density function corresponding to the above Bayesian network is:

$$P(H, F, Inf, Food) = P(H|Inf, Food)P(F|Inf, Food)P(Inf)P(Food) \ .$$

(c) The probability to compute is $P(Inf = \mathtt{t}|H = \mathtt{t})$. Since the Bayesian network encodes the "reverse" dependency between the two random variables involved, i.e., $Inf \to H$, it is convenient to start by rewriting the above probability using Bayes' rule:

$$P(Inf = \mathtt{t}|H = \mathtt{t}) = \frac{P(H = \mathtt{t}|Inf = \mathtt{t})P(Inf = \mathtt{t})}{P(H = \mathtt{t})} \ . \tag{1}$$

The term $P(Inf = \mathtt{t})$ is known (the corresponding density function is associated with the node Inf of the Bayesian network). The term $P(H = \mathtt{t}|Inf = \mathtt{t})$ can be computed through the conditional density $P(H|Inf, Food)$ corresponding to the node H; to this aim, the sum rule can be first applied to introduce the random variable $Food$, then the product rule, leading to:

$$\begin{aligned} P(H = \mathtt{t}|Inf = \mathtt{t}) &= \sum_{food \in \{\mathtt{t},\mathtt{f}\}} P(H = \mathtt{t}, Food = food|Inf = \mathtt{t}) \\ &= \sum_{food \in \{\mathtt{t},\mathtt{f}\}} P(H = \mathtt{t}|Food = food, Inf = \mathtt{t})P(Food = food|Inf = \mathtt{t}) \ . \end{aligned}$$

The term $P(H = \mathtt{t}|Food = food, Inf = \mathtt{t})$ is known form the Bayesian network. Moreover, since $Food$ and Inf are assumed to be independent, it follows that $P(Food = food|Inf = \mathtt{t}) = P(Food = food)$, which is known as well from the Bayesian network.

Finally, the term $P(H = \mathtt{t})$ in expression (1) can be computed similarly, by applying first the sum rule and then the product rule:

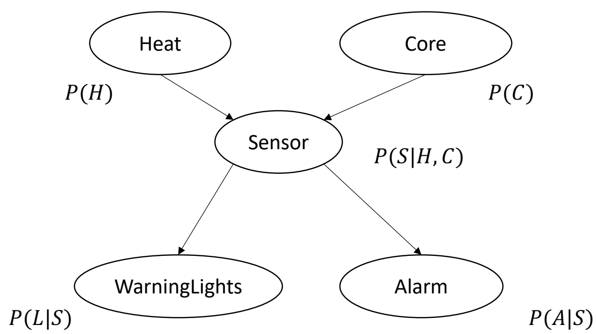
$$\begin{aligned} P(H = \mathtt{t}) &= \sum_{inf, food \in \{\mathtt{t},\mathtt{f}\}} P(H = \mathtt{t}, Inf = inf, Food = food) \\ &= \sum_{inf, food \in \{\mathtt{t},\mathtt{f}\}} P(H = \mathtt{t} | Inf = \mathtt{t}, Food = food) P(Inf = inf, Food = food) \ . \end{aligned}$$

The first term is the last expression above is known from the density function $P(H|Inf, Food)$ associated with node H. The second term can be rewritten as $P(Inf = inf)P(Food = food)$, taking into account the independence assumption between Inf and $Food$, and teh values of both probabilities can be directly obtained from the unconditional density functions associated to nodes Inf and $Food$.

5. (a) This domain can be represented using five Boolean random variables: C (the core temperature exceeding the safety threshold), S (the sensor detecting a core overheating), H (the external temperature being higher than some specific value), L (warning lights flashing) and A (alarm sounding).

   The "root" causes are the core temperature and the external temperature, which can also be considered independent on each other. They both directly influence the sensor measurement. In turn, the sensor measurement directly affects both the warning lights and the alarm; on the other hand, given the state of the sensor, both the warning lights and the alarm can be assumed to be independent on the core and external temperatures, as well as on each other (the latter assumption is motivated by the fact that the warning lights and the alarm are implemented as physically separated systems).

   The corresponding Bayesian network is shown below.



   (b) The joint density function of the five random variables, corresponding to the above Bayesian network, is given by:

$$P(A, L, S, C, H) = P(A|S)P(L|S)P(S|C, H)P(C)P(H) \ .$$

   (c) The probability to compute is $P(C = \mathtt{t}|L = \mathtt{t})$. The dependency relationship encoded by the Bayesian network is $C \to L$, thus the above probability can be conveniently computed by first applying Bayes' formula:

$$P(C = \mathtt{t}|L = \mathtt{t}) = \frac{P(L = \mathtt{t}|C = \mathtt{t})P(C = \mathtt{t})}{P(L = \mathtt{t})} \ .$$

   The term $P(C = \mathtt{t})$ is known from the density function $P(C)$ associated with the corresponding node of the Bayesian network. The term $P(L = \mathtt{t}|C = \mathtt{t})$ can be computed as a function of the conditional densities $P(L|S)$ and $P(S|H, C)$ associated with the corresponding nodes. To this aim, the sum rule can be applied to introduce the random variables S and H, followed by the product rule:

$$\begin{aligned} P(L = \mathtt{t}|C = \mathtt{t}) &= \sum_{s,h \in \{\mathtt{t},\mathtt{f}\}} P(L = \mathtt{t}, S = s, H = h|C = \mathtt{t}) \\ &= \sum_{s,h \in \{\mathtt{t},\mathtt{f}\}} P(L = \mathtt{t}|S = s, H = h, C = \mathtt{t})P(S = s, H = h|C = \mathtt{t}) \qquad (2) \end{aligned}$$

Due to the conditional independence assumption between L and its non-descendants H and C, given S, the first term of expression (2) equals $P(L = \texttt{t}|S = s)$, which is known from the conditional density associated to node L. The second term of expression (2) can be rewritten, using the product rule, as:

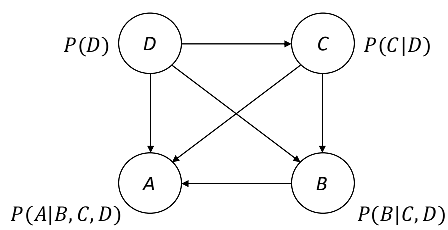
$$P(S = s|H = h, C = \texttt{t})P(H = h|C = \texttt{t}) \ .$$

The first term above is known from the conditional density of node S. The second term is equal to $P(H = h)$, due to the independence assumption between H and C, and is known as well, from the unconditional density associated to node H.

6. By applying the chain rule, the joint density 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 nodes. This is a general characteristic of any Bayesian network when no conditional independence assumption can be made on the corresponding conditional densities.

The number of probability values that have to be estimated to specify the density functions associated to the nodes of the above Bayesian network is:

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

for a total of 15 values. Note that this is the same number of probability values that have to be estimated for the full joint density of the 4 random variables, $P(A, B, C, D)$, that is, $2^4 - 1 = 15$. In general, for n Boolean random variables this number is $\sum_{k=1}^{n} 2^{k-1} = 2^n - 1$. This means that the number of required probability values increases *exponentially* with the number of random variables. Therefore, conditional independence assumptions are useful in practical applications to reduce the effort required to estimate the joint density through the chain rule.

7. The desired result can always be obtained in three steps:

(i) Any conditional probability can first be rewritten as the ratio of two unconditional ones, using the definition of conditional probability. For instance:

$$P(B = \texttt{t}|J = \texttt{t}) = \frac{P(B = \texttt{t}, J = \texttt{t})}{P(J = \texttt{t})} \ .$$

(ii) Then, the joint probability of any subset of random variables can be computed by applying the sum rule to the joint density of *all* random variables. For instance, $P(B = \texttt{t}, J = \texttt{t})$ can be computed by applying the sum rule as follows:

$$P(B = \texttt{t}, J = \texttt{t}) = \sum_{a,e,m \in \{\texttt{t},\texttt{f}\}} P(B = \texttt{t}, J = \texttt{t}, A = a, E = e, M = m) \ .$$

(iii) Finally, the joint density of all the random variables can be rewritten using the chain rule corresponding to the considered Bayesian network, e.g.:

$$\sum_{a,e,m \in \{\texttt{t},\texttt{f}\}} P(B = \texttt{t}, J = \texttt{t}, A = a, E = e, M = m) =$$
$$\sum_{a,e,m \in \{\texttt{t},\texttt{f}\}} P(B = \texttt{t}) P(E = e) P(A = a | B = \texttt{t}, E = e) P(J = \texttt{t} | A = a) P(M = m | A = a) \ .$$

Note however that a sum over the combinations of values of n Boolean random variables contains 2^n terms, which is impractical for large values of n.

An alternative approach to rewrite the probability of interest in terms of the density funcitons associated to the Bayesian network, is to suitably use also the Bayes' formula and the product rule, beside the definition of conditional probability and the sum rule, trying to avoid obtaining expressions involving a large number of summands when the sum rule is used.

(a) To compute $P(B = \texttt{t} | J = \texttt{t})$ one can try to rewrite it in terms of the density functions associated with the Bayesian network that involve the random variables B and J, namely $P(J|A)$, $P(A|B, E)$ and $P(B)$. To this aim, taking into account the ordering of the variables in the Bayesian network (i.e., $B \to A \to J$), one can start from Bayes' formula, followed by the sum rule applied to its denominator (which can conveniently be applied *only* to the random variable B) and by the product rule:

$$\begin{aligned} P(B = \texttt{t} | J = \texttt{t}) &= \frac{P(J = \texttt{t} | B = \texttt{t}) P(B = \texttt{t})}{P(J = \texttt{t})} \text{ (Bayes' rule)} \\ &= \frac{P(J = \texttt{t} | B = \texttt{t}) P(B = \texttt{t})}{P(J = \texttt{t}, B = \texttt{t}) + P(J = \texttt{t}, B = \texttt{f})} \text{ (sum rule)} \\ &= \frac{P(J = \texttt{t} | B = \texttt{t}) P(B = \texttt{t})}{P(J = \texttt{t} | B = \texttt{t}) P(B = \texttt{t}) + P(J = \texttt{t} | B = \texttt{f}) P(B = \texttt{f})} \text{ (product rule).} \quad (3) \end{aligned}$$

The latter expression involves the density function $P(B)$, which is known since it is associated with the corresponding node of the Bayesian network; it also involves the probabilities $P(J = \texttt{t}|B)$, which can be conveniently computed through the know density functions $P(J|A)$ and $P(A|B, E)$ (these density functions are associated with nodes J and A of the Bayesian network); to this aim one can first apply the sum rule over only A and E, and then the product rule:

$$\begin{aligned} P(J = \texttt{t}|B) &= \sum_{a,e \in \{\texttt{t},\texttt{f}\}} P(J = \texttt{t}, A = a, E = e | B) \text{ (sum rule)} \quad (4) \\ &= \sum_{a,e \in \{\texttt{t},\texttt{f}\}} P(J = \texttt{t} | A = a, E = e, B) P(A = a, E = e | B) \text{ (product rule).} \quad (5) \end{aligned}$$

Taking into account that J is conditionally independent on E and B, given A, the term $P(J = \texttt{t}|A = a, E = e, B)$ equals $P(J = \texttt{t}|A = a)$, which is known from the Bayesian network. Moreover, the term $P(A = a, E = e|B)$ can be rewritten by using again the product rule, as $P(A = a|E = e, B)P(E = e|B)$; the first factor is known from the Bayesian network, whereas the second one equals $P(E = e)$ since B and E are independent; since also the density functions $P(E)$ and $P(B)$ are known from the Bayesian network, one gets:

$$P(J = \texttt{t}|B) = \sum_{a,e \in \{\texttt{t},\texttt{f}\}} P(J = \texttt{t} | A = a) P(A = a | E = e, B) P(E = e) \ . \quad (6)$$

By substituting Eq. (6) into the denominator of Eq. (3), the probability $P(B = \texttt{t}|J = \texttt{t})$ can be finally computed as requested.

(b) To compute $P(B = \mathtt{t}|A = \mathtt{t})$, note first that in the Bayesian network the ordering of the two random variables involved is $B \to A$. Bayes' formula can therefore be used first to get:

$$P(B = \mathtt{t}|A = \mathtt{t}) = \frac{P(A = \mathtt{t}|B = \mathtt{t})P(B = \mathtt{t})}{P(A = \mathtt{t})} .$$

The term $P(B = \mathtt{t})$ is known from the Bayesian network. The probability $P(A = \mathtt{t}|B = \mathtt{t})$ can be computed through the known density function $P(A|B, E)$ using the sum rule over E, then the product rule, and finally considering that B and E are independent:

$$\begin{aligned} P(A = \mathtt{t}|B = \mathtt{t}) &= \sum_{e \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}, E = e|B = \mathtt{t}) \text{ (sum rule)} \\ &= \sum_{e \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}|E = e, B = \mathtt{t})P(E = e|B = \mathtt{t}) \text{ (product rule)} \\ &= \sum_{e \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}|E = e, B = \mathtt{t})P(E = e) . \end{aligned} \tag{7}$$

Finally, $P(A = \mathtt{t})$ can be computed by applying the sum rule over B and E, then the product rule, and finally exploiting again the independence between B and E:

$$\begin{aligned} P(A = \mathtt{t}) &= \sum_{b,e \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}, B = b, E = e) \text{ (sum rule)} \\ &= \sum_{b \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}|B = b, E = e)P(B = b, E = e) \text{ (product rule)} \\ &= \sum_{b \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}|B = b, E = e)P(B = b)P(E = e) . \end{aligned} \tag{8}$$

All the probabilities in the latter expression are known from the Bayesian network.

(c) The probability $P(B = \mathtt{t}|A = \mathtt{t}, E = \mathtt{f})$ can be computed through the distribution $P(A|E, B)$, which is known from the Bayesian network and involves the same variables. To exploit it, Bayes' formula can be used first to obtain:

$$P(B = \mathtt{t}|A = \mathtt{t}, E = \mathtt{f}) = \frac{P(A = \mathtt{t}, E = \mathtt{f}|B = \mathtt{t})P(B = \mathtt{t})}{P(A = \mathtt{t}, E = \mathtt{f})} . \tag{9}$$

The probability $P(B = \mathtt{t})$ is known from the Bayesian network. The probability $P(A = \mathtt{t}, E = \mathtt{f}|B = \mathtt{t})$ can be computed as in Eq. (5).
The term $P(A = \mathtt{t}, E = \mathtt{f})$ in Eq. (9) can be computed using the sum rule over B, then the product rule, and finally exploiting the independence between E and B:

$$\begin{aligned} P(A = \mathtt{t}, E = \mathtt{f}) &= \sum_{b \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}, E = \mathtt{f}, B = b) \text{ (sum rule)} \\ &= \sum_{b \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}|E = \mathtt{f}, B = b)P(E = \mathtt{f}, B = b) \text{ (product rule)} \\ &= \sum_{b \in \{\mathtt{t},\mathtt{f}\}} P(A = \mathtt{t}|E = \mathtt{f}, B = b)P(E = \mathtt{f})P(B = b) . \end{aligned} \tag{10}$$

All the probabilities in the last expression are known from the Bayesian network.

8. (a) Given the considered ordering between the four random variables, the *general* expression of their joint density function obtained from the chain rule (*without* making any conditional independence assumption) is:

$$P(A, B, C, D) = P(D|C, B, A)P(C|B, A)P(B|A)P(A) \tag{11}$$

(b) The expression of the joint density function corresponding to the Bayesian network is:

$$P(A, B, C, D) = P(D|A)P(C|B, A)P(B)P(A) \tag{12}$$

Remind that Eq. (12) is assumed to be obtained under the *same* ordering between the four random variables as in Eq. (11). It follows that, by comparing Eqs. (11) and (12), the Bayesian network encodes two conditional independence assumptions:

- $P(D|C, B, A) = P(D|A)$, i.e., D is *conditionally* independent on A, *given* C and B
- $P(B|A) = P(B)$, i.e., B is independent on A

The same conclusion can be drawn by considering only the structure of the Bayesian network and the specified ordering between its random variables, since, *by definition*, a Bayesian network encodes the assumption that each of its random variables is conditionally independent on the random variables of *all* its *non-descendant* nodes (i.e., nodes that are not its successors, or successors of any of its successors, and so on), *given* the random variables of its *parent* nodes.

(c) The considered probability can be computed using the general approach described at the beginning of the solution of exercise 7, i.e., by first applying the definition of conditional proability, then applying the sum rule to the numerator and denominator, and finally rewriting both of them using the chain rule represented by the Bayesian network.

The alternative approach requires a variable sequence of steps consisting of ad hoc applications of the Bayes' formula, the product rule, the sum rule and the definition of conditional probability, aimed at rewriting the original probability in terms of the density functions associated with the Bayesian network. For instance, since the Bayesian network involves the conditional density functions of both C and D given A, whereas the probability to compute involves the conditional density function of A given C and D, it is convenient to apply Bayes' formula first:

$$P(A = \mathtt{t}|C = \mathtt{t}, D = \mathtt{f}) = \frac{P(C = \mathtt{t}, D = \mathtt{f}|A = \mathtt{t})P(A = \mathtt{t})}{P(C = \mathtt{t}, D = \mathtt{f})} \ . \tag{13}$$

Note that the value $P(A = \mathtt{t})$ is known from the Bayesian network.

The value of $P(C = \mathtt{t}, D = \mathtt{f}|A = \mathtt{t})$ can be conveniently rewritten by first applying the sum rule to include the random variable B:

$$\begin{aligned} P(C = \mathtt{T}, D = \mathtt{F}|A = \mathtt{T}) &= \sum_{b \in \{\mathtt{T},\mathtt{F}\}} P(B = b, C = \mathtt{T}, D = \mathtt{F}|A = \mathtt{T}) \\ &= \sum_{b \in \{\mathtt{T},\mathtt{F}\}} P(D = \mathtt{F}|C = \mathtt{T}, B = b, A = \mathtt{T})P(C = \mathtt{T}, B = b|A = \mathtt{T}) \ . \end{aligned} \tag{14}$$

From the conditional independence assumptions encoded by the Bayesian network, the first term in the above summands equals $P(D = \mathtt{F}|A = \mathtt{T})$. The second term can instead be rewritten using the product rule as:

$$P(C = \mathtt{T}, B = b|A = \mathtt{T}) = P(C = \mathtt{T}|B = b, A = \mathtt{T})P(B = b|A = \mathtt{T}) \ . \tag{15}$$

The first term in the right-hand side of Eq. (15) is known from the Bayesian network, whereas the second one can be rewritten using again the conditional independence assumptions, as $P(B = b|A = \mathtt{T}) = P(B = b)$, which is known as well from the Bayesian network. This completes the computation of the numerator of Eq. (13).

The denominator of Eq. (13) can be computed by directly applying the sum rule to include the remaining random variables A and B, and then using the chain rule corresponding to the Bayesian network:

$$\begin{aligned} P(C = \mathtt{t}, D = \mathtt{f}) &= \sum_{a,b \in \{\mathtt{T},\mathtt{F}\}} P(A = a, B = b, C = \mathtt{t}, D = \mathtt{f}) \\ &= \sum_{a,b \in \{\mathtt{T},\mathtt{F}\}} P(D = \mathtt{f}|A = a)P(C = \mathtt{t}|A = a, B = b)P(B = b)P(A = a) \ . \end{aligned}$$