Probabilistic Reasoning

Question and Answer form 2012

Name: František Farka

Student nr: F120127

BN software used: GeNIe, Samlam

PART A

No software is required to answer the following questions.

Consider a set of n discrete variables, with m values per variable. You have no information about the independence relation, and therefore have to assume all variables to be dependent.

A1. Indicate for each of the following combinations of n and m how many parameters[1] would be required to specify a full joint probability distribution on the variables. Explain your answers.

(a) $n = 15$ and $m = 2$:

(b) $n = 15$ and $m = 5$:

(c) $n = 100$ and $m = 20$:

Explanation:
Given the definition of joint probability, the number of parameters c is the cardinality of universe of BA of propositions $\mathcal{V}$ except values required by definition:

$$c = (n+1)^m - 2$$

ie.

(a) $c = 14348905$

(b) $c = 470184984574$

(c) $c \approx 1.6 * 10^{132}$

Suppose the above set of n variables corresponds to the set of nodes of a probabilistic network. You now know that the independence relation among the variables is such that the network has one root node, one node with one parent, one node with two parents, and the remaining nodes all have 3 parents.

A2. Indicate for each of the following combinations of n and m how many parameters would be required to define this probabilistic network. Explain your answers.

(a) $n = 15$ and $m = 2$:

(b) $n = 15$ and $m = 5$:

(c) $n = 100$ and $m = 20$:

[1] The term "parameters" is used to refer to the numbers that are required to completely specify a probability distribution.

Explanation:

Given properties of discrete probability distribution function and disjunctive partitioning of probability space to d sets we have the number of parameters c_n necessary to define probability distribution of node with n parents:

$$c_n = (m-1)m^n$$

Thus we conclude number c of parameters of the probability distribution of an entire network

$$c = (m-1) + (m-1)m + 2(m-1)m^2 + (n-4)(m-1)m^3$$

$$c = (m-1)\left(2m^2 + m + 1 + (n-4)m^3\right)$$

and for given values of m and n

(a) $c = 99$

(b) $c = 5724$

(c) $c = 14607599$

A3. *Causal independence* is an assumption that allows you to define a complete assessment function for a node V_i, by specifying only part of the function. As such, the number of parameters that require elicitation is reduced. More specifically, for a node V_i with $k > 0$ parents V_j, $j = 1, \ldots, k$, a causal independence model typically only requires the specification of k distributions $\Pr(V_i \mid V_j)$, $j = 1, \ldots, k$, that each capture the independent effect of the single parent V_j. These k local distributions are subsequently combined using some (for now irrelevant) deterministic rule to obtain a complete assessment function for the child node V_i.

Assuming causal independence, indicate for each of the following combinations of n and m how many parameters would be required to define the probabilistic network with again one root node, one node with one parent, one node with two parents, and the remaining nodes all having 3 parents. Explain your answers.

(a) $n = 15$ and $m = 2$:

(b) $n = 15$ and $m = 5$:

(c) $n = 100$ and $m = 20$:

Explanation:

Given the causal independence number c_n of parameters defining probability distribution of node with $n0$ parents is:

$$c_n = \begin{cases} (m-1)mn & : n > 0 \\ (m-1) & : n = 0 \end{cases}$$

Thus we conclude number c of parameters of probability distribution of entire network given causal independence:

$$c = (m-1) + (m-1)m + 4(m-1)m + (n-4)(m-1)m$$

$$c = (m-1)\left(1 + m\left(n+1\right)\right)$$

and for given values of m and n

(a) $c = 33$

(b) $c = 324$

(c) $c = 38399$

Conclusion from PART A:

What is your conclusion from the above regarding the relation between independence assumptions and the space complexity of defining a discrete joint probability distribution?

Explanation:
Independence assumptions significantly reduces space complexity of representation of probability distribution function. This complexity is generally exponential in number of variables and domain size. Introducing graph representation reduces this complexity to polynomial (assuming fixed maximum degree).

PART B

B1. Clearly explain the difference between the following concepts:

I. probabilistic independence   II. d-separation

Explanation:
I. is a property of a probability distribution function whereas II. is a property of graph representing the distribution.

B2. Clearly explain the difference between the following concepts:

I. blocking   II. d-separation

Explanation:
I. considers only certain path or chain in graph representing a distribution whereas II. represents cut between two sets of nodes considering every path/simple chain

The answers to the following questions follow from Figure 1 in the description; you can try and verify your answers once you have constructed/imported the car diagnosis network in a Bayesian network software package.

B3. *Spark functioning* is d-separated from *Main fuse* given *Voltage at plug*:

> **True; explanation:**
> Only chain which is not block by *Voltage at plug*contains *Spark functioning* $\rightarrow$ *Car starts* $\leftarrow$ *Car cranks* thus is blocked.

B4. *Distributor* is independent of *Starter motor* given *Main fuse*:

> **True; explanation:**
> It is d-separated given *Main fuse*.

B5. *Starter system* is d-separated from *Fuel system* given *Car starts*, *Car cranks*, *Distributor* and *Voltage at plug*:

> **True; explanation:**
> Every chain is block by either *Voltage at plug* or *Car cranks*.

B6. *Headlights* may be independent of *Distributor* given *Car starts*:

> **True; explanation:**
> We give *Battery voltage* besides *Car starts*.

B7. The Markov blanket[2] of *Voltage at plug* is $\{$*Distributor*, *Main fuse*, *Battery voltage*, *Spark functioning*$\}$:

> **False; explanation:**
> It contains *Spark plugs*.

B8. *Voltage at plug* is independent of all other variables given its Markov blanket:

> **True; explanation:**
> It blocks all the chains.

B9. The chain *Main fuse* — *Starter system* — *Battery voltage* — *Charging system* is *active*, i.e. not blocked, given *Spark functioning* and *Headlights*:

> **True; explanation:**
> None of the blocking criteria is satisfied.

[2]The Markov blanket $M(A)$ for a node A is defined as $M(A) = \rho(A) \cup \sigma(A) \cup \rho(\sigma(A)) \setminus \{A\}$.

PART C

The following questions can be answered only after you have completely constructed/imported the car-diagnosis network using Bayesian network software.

Make sure that you verify that your network computes the correct prior (marginal) probability distributions $\Pr(V)$ for each variable V in the car-diagnosis network before evidence is entered (see page 5 of the description). From here on it is assumed that your network is correctly constructed. **Therefore, if incorrect probabilities are returned for the questions below, it is assumed that the source for these errors is the approach you used, rather than an incorrect network specification!**

C1. What is the probability of the car starting given fouled spark plugs?

$$\Pr(\textit{Car starts} = \text{true} \mid \textit{Spark plugs} = \text{fouled}) = 0.38977987$$

C2. What is the probability of the car cranking *or* starting given a weak battery?

$$\Pr(\textit{Car cranks} = \text{true} \vee \textit{Car starts} = \text{true} \mid \textit{Battery voltage} = \text{weak}) = 0.498033021447005$$

Explanation:
Result of MAP computation

C3. For the variable *Battery voltage*, first compute the following:

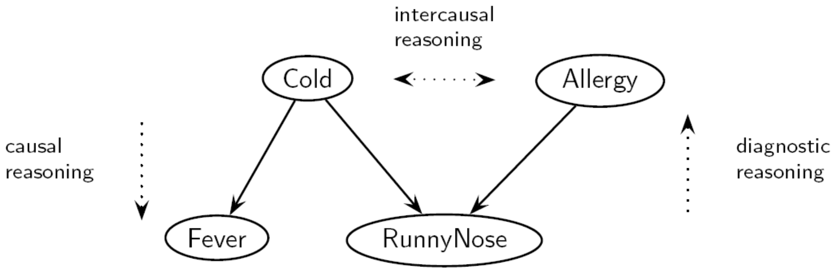
- I. its prior (marginal) probability distribution;
- II. its posterior (marginal) distribution given that the main fuse has *not* blown;
- III. its posterior (marginal) distribution given that the car does not start;
- IV. its posterior (marginal) distribution given that the car does not start and the main fuse has not blown:

	strong	weak	dead
$\Pr(\textit{Battery voltage})$	0.583	0.193	0.224
$\Pr(\textit{Battery voltage} \mid \textit{Main fuse} = \text{okay})$	0.583	0.193	0.224
$\Pr(\textit{Battery voltage} \mid \textit{Car starts} = \text{false})$	0.765	0.203	0.032
$\Pr(\textit{Battery voltage} \mid \textit{Car starts} = \text{false} \wedge \textit{Main fuse} = \text{okay})$	0.765	0.203	0.032

C4. Consider your answers for C3. Explain the differences between I vs II, I vs III, and IV vs II and III:

Explanation:
Battery voltage is independent on *Main fuse* thus evidence of *Main fuse* does not change probability distribution of *Battery voltage*. Evidence of *Car starts* propagates upwards and changes distribution of *Battery voltage*. Due to independency on *Main fuse* is probability distribution the same regardless *Main fuse* evidence.

In probabilistic networks we roughly distinguish between three types of reasoning: causal (in the direction of the arcs), diagnostic (against the direction of the arcs), and inter-causal. See the figure below.



C5. Suppose you measure the battery voltage and the read-out of your voltmeter indicates that the battery is sufficiently strong. Compute the probabilities for the variables *Battery age*, *Headlights* and *Alternator* given this observation:

	0-2 yrs	3-5 yrs	> 5 yrs
Pr(*Battery age* | *Battery voltage* = strong)	0.472	0.383	0.144
	bright	dim	no-light
Pr(*Headlights* | *Battery voltage* = strong)	0.94	0.01	0.05
	okay	faulty	
Pr(*Alternator* | *Battery voltage* = strong)	0.9995	0.0005	

C6. Compare the probabilities found for C5 with their priors Pr(*Battery age*), Pr(*Headlights*) and Pr(*Alternator*). Explain the changes in terms of your knowledge from the domain; indicate the type of reasoning — diagnostic, causal and/or intercausal — involved in your explanation.

Explanation:
Regarding *Battery age* probability is higher for 0-2 years and lover for other values given strong battery. This is diagnostic reasoning and represents the property of aging of the *Battery age* with loosing maximal voltage.
Probability of bright lights is much higher given strong battery. This is causal reasoning and represents the fact that lights are powered via battery.
Probability of *Alternator* is biased towards okay. This is diagnostic reasoning and represents that battery is charged by *Alternator*.

Consider a simplified network involving just the variable *Car starts* and the 7 cause variables *Alternator*, *Battery age*, *Distributor*, *Fuel system*, *Main fuse*, *Spark plugs*, *Starter motor*, as direct parents of *Car starts*. The cause variables have the same assessment functions as in the original network; *Car starts* gets a new assessment function, conditioned on the cause variables. The joint distribution over the 8 variables is the same for both the original and the simplified network.

C7. Clearly explain *how* you would determine the values of the new assessment function for *Car starts* to ensure that the distribution over the 8 variables is the same for both the original and the simplified network. (**NB you are not required to give the exact function here: that would mean specifying a large number of values!**)

Explanation:
We set evidence for every combination of variables *Alternator*, *Battery age*, *Distributor*, *Fuel system*, *Main fuse*, *Spark plugs*, *Starter motor*, and compute probabilities for *Car starts*. We use this probabilities as definition of *Car starts* given the evidence.

For the next two questions you are required to construct the simplified car diagnosis network in your Bayesian network software package. However, because of the size of the assessment function for variable *Car starts* you only have to include the 2 cause variables *Battery age* and *Spark plugs*.

C8. Construct the 3-variable network described above. Take the assessment functions for *Battery age* and *Spark plugs* to be equivalent to those in the original network. Define the assessment function for variable *Car starts* such that Pr(*Car starts* ∧ *Battery age* ∧ *Spark plugs*) is the same in both networks. Give the entire assessment function for *Car starts*:

Explanation:

Spark plugs	okay	okay	okay	faulty	faulty	faulty
Battery age	0-2	3 - 5	> 5	0 - 2	3 - 5	> 5
true	0.548	0.505	0.448	0.414	0.386	0.348
false	0.452	0.495	0.552	0.586	0.614	0.652

C9. Compute the probability of the car starting given fouled spark plugs from the 3-variable network, and compare it to the same probability from the original network (see C1). Explain the difference/equivalence.

Pr(*Car starts* = true | *Spark plugs* = fouled) = 0.38977987

Explanation:
It's equal, we have constructed simplified network with equal distribution over variables *Spark plugs*, *Battery age*.

C10. Compare the simplified 8-variable network to the original network and indicate when you would prefer one over the other.

Explanation:
Simplified network is easier to construct and less demanding regarding space and computation time. I would use it in case we have no direct access to values of inner nodes i. e. we cannot get evidence on values of inner nodes.

For the next two questions you are required to change the specification of some of the parameters in your original *car diagnosis network and subsequently select the (single!) correct answer among four possibilities.*

Apply the following change to the network specification: adapt γ(*Starter motor*) to represent that the starter motor is okay only 50% of the time.

C11. Clearly explain for which of the following variables the prior (marginal) probability distributions are affected as a result of the change in γ(*Starter motor*):

I. *Starter system*

II. *Starter system*, *Car cranks*, and *Car starts*

III. *Starter motor*, *Starter system*, *Car cranks*, and *Car starts*

IV. none of the above answers

III; explanation:
We have changed prior distribution of *Starter motor* thus all dependent prior distributions change.

C12. You observe that the car does not start. Clearly explain for which of the following variables the posterior (marginal) probability distributions are affected as a result of the change in γ(*Starter motor*):

I. *Spark functioning*, *Fuel system*, and *Car cranks*

II. all 15 variables

III. all 15 variables, except *Car starts*

IV. none of the above answers

II; explanation:
Evidence of *Car starts* propagates to all reachable nodes and changes posterior probabilities. *Car starts*probability distribution collapses to observed evidence as necessity thus changes.