Assignment 1: Blood Types with Bayesian Networks

AI2SysProj 2022

Topic:	Bayesian networks
Due on:	July 11, 2022
Version from:	June 4, 2022

1 Task summary

Using Bayesian networks, compute the probability that someone has a particular blood type given the blood types of some of their relatives. The assignment repository [A1] contains example problems for you to solve.

Objectives

1. Model a problem as a Bayesian Network,
2. gain experience working with graphs,
3. use complex conditional probability tables,
4. find and use a suitable library for Bayesian inference (many exist and finding good libraries is an important skill),
5. get to know the JSON format (in case you haven't used it before).

Prerequisites and useful methods

1. Basics of representing graphs in code and working with them,
2. Bayesian networks (Section 21 in the lecture notes).

2 Background: The ABO blood group system

The **ABO blood group system** [ABO] is used to describe the presence/absence of A and B antigens on the red blood cells in humans. We distinguish 4 different **(ABO) blood types**: **A** (if only the A antigens are present), **B** (if only the B antigens are present), **AB** (if both are present) and **O** (if neither are present). Knowing the blood type is in particular important for blood transfusions as a transfusion with an incompatible blood type can be lethal. Another important antigen for blood transfusions is Rh, which we will ignore in this assignment.

2.1 Alleles

The ABO blood type is determined by the **ABO gene**, which comes in three variants, called **alleles**: the **A allele**, the **B allele** and the **O allele**. Humans have chromosome pairs and therefore two versions of the **ABO gene** – one taken from each parent. The A allele and B allele are co-dominant: if at least one ABO gene has the A allele, the person has A antigens, and if at least one ABO gene has the B allele, the person has B antigens. Humans with both the A and B allele have both antigens and therefore blood type AB and only humans with two O alleles have neither antigens and therefore blood type O. In the following, we will sometimes write XY to denote that someone has one X allele and one Y allele.

2.2 Inheritance

A child randomly gets one allele from each parent. For example, if the father has an A allele and an O allele (short AO) and the mother has an A allele and a B allele (short AB), then the child could have one of the following allele combinations:

1. AA (blood type A)
2. OA (blood type A)
3. AB (blood type AB)
4. OB (blood type B)

Each of the four combinations is equally likely.

3 Detailed problem description

The assignment repository [A1] contains problem files and example solutions encoded in the JSON format. Listing 1 shows an example problem file. The `query-for` field specifies the person whose blood type we would like to know – in this case Calvin's. The `evidence` field tells us the known blood types of relatives. In the example, Ahmed has blood type O and Beatrice blood type B. At last, the `family-tree` field specifies the relationship between family members: Ahmed (`subject`) is the father (`relation`) of Beatrice (`object`) and Beatrice is the mother of Calvin. Note that every family member is uniquely identified by their name.

```
{
    "query-for": "Calvin",
    "evidence": {
        "Ahmed": "O",
        "Beatrice": "B"
    },
    "family-tree": [
        {
            "subject": "Ahmed",
            "relation": "father-of",
            "object": "Beatrice"
        },
        {
            "subject": "Beatrice",
            "relation": "mother-of",
            "object": "Calvin"
        }
    ]
}
```

Listing 1: Example problem file.

3.1 Details on the family tree specification

As mentioned above, the family tree is described via relationhips between people. The following types of relationship are used:

1. mother-of, father-of, parent-of (the easier problems only use these),
2. sister-of, brother-of, sibling-of,
3. grandmother-of, grandfather-of, grandparent-of (note that in this case it is not clear if the maternal or paternal grand parent is referred to).

Note that some relationships may only be implicit. For example, if A is a parent of B, and C is a sibling of B, then clearly A is also a parent of C – even though it might not be explicitly stated.

For the ambiguous grandparent relationships, we assume uniform priors. For example, the prior probability of a grandparent to be paternal or maternal is 50% respectively. We also assume that everyone has exactly one male and one female parent. To make things easier, problems that involve the grandparent relationships only use gendered versions.

3.2 Distribution of ABO alleles

To solve the assignment, you need to know the distribution of the ABO alleles in the population, which actually varies significantly across the world. For this assignment, we assume that (without any further knowledge) an ABO gene has the O allele with a 64% prior probability, the A allele with a 28% prior probability and the B allele with an 8% prior probability. We also assume that alleles are independent, so e.g. the prior probability of having blood type AB is $28\% \cdot 8\% = 2.24\%$.

3.3 Solution format

The solution to a problem file, i.e. the blood type distribution of a person, should also be stored in a JSON file. For example, the solution to the example problem in Listing 1 should be represented as

```
{"O": 0.32, "A": 0.14, "B": 0.4, "AB": 0.14}
```

meaning that Calvin has e.g. blood type O with a probability of 32%.

The assignment repository contains more example problems and solutions that you can use for comparison.

4 What to submit

Your solution should be submitted to your team's repository. It should contain:

1. all your code,
2. a README file explaining how to run your code to solve other problem files (including how to install dependencies),
3. a brief summary of how you solved the problem either as a PDF file ($\approx$ 1 page) or as part of your README (in particular, please describe what random variables you used in the Bayesian network),
4. a file solution[NUMBER].json for every problem[NUMBER].json that you managed to solve.

5 A few tips

1. The problems increase in difficulty (see Section 6) – it makes sense to start with the easier ones, which give you the most points anyway.
2. Part of the problem is to find a suitable library for Bayesian inference (inference through variable elimination should be efficient enough). It might make sense to try out the library with a minimal example before integrating it with the rest of the code to make sure it actually works as expected.
3. Getting the encoding as a Bayesian network right is somewhat tricky. Make sure that you correctly distinguish between the ABO allele pairs of a person and their observed blood type (state vs evidence variables).
4. For the more difficult problems, you might have to add people to the family tree that were not explicitly mentioned in the problem file.
5. The problems with ambiguous relationships are much more difficult. You do not have to solve them with a single Bayesian network – an alternative approach is to create one network for each family tree and then combine the results.

6 Points

The problem files increase in difficulty. You will get points for each difficulty that was solved correctly:

- 30 points for problems 0 – 19. They contain only a child and two parents.
- 20 points for problems 20 – 39. They also only use the parent relationships, but in a more complex setting as e.g. shown in Listing 1.
- 15 points for problems 40 – 59. They also include the sibling relationships.
- 15 points for problems 60 – 79. They include the ambiguous grandparent relationships and are more difficult.

Assuming you have at least a partial solution, you can additionally get up to 20 points for the quality of the submission (README, explanation, ...). The maximum number of points is therefore 100. If the grading scheme doesn't seem to work well, we might adjust it later on (likely in your favor).

References

[A1] *Assignment 1*. URL: https://gitlab.rrze.fau.de/wrv/AISysProj/ss22/a1-blood-types/assignment (visited on 06/02/2022).

[ABO] *ABO blood group system*. URL: https://en.wikipedia.org/wiki/ABO_blood_group_system (visited on 06/01/2022).