DAT405 Introduction to Data Science and AI, 2019 – 2020, Reading period 1

Assignment 5: Approximate inference in Bayesian networks

There will be an overall grade for this assignment. To get a pass grade (grade 3), you need to pass item 1 below. To receive higher grades, finish items 2 and 3 as well.

In this assignment you will implement three different resampling algorithms

I. rejection sampling
II. likelihood weighted sampling
III. Gibbs sampling

to perform approximate inference on a Bayesian network. The Bayesian network to use, and its corresponding prior distribution, is described by the graph below.

Apply each of the three algorithms to the following tasks.

1. Compute the following probabilities, directly, without sampling. Then employ each of the three sampling algorithms to approximate the probabilities. Use 1000 samples for each method and document your results. How do the approximations compare to the true values?
    a. $P(D|B,C) = P(D = true|B = true, C = true)$
    b. $P(X|V) = P(X = true|V = true)$
    c. $P(C|V^c, S) = P(C = true|V = false, S = true)$
2. Now focus on the probability in 1a, $P(D|B,C)$. We know that the accuracy of the sampling approximations depends on the number of samples used. For each of the three sampling methods, plot the probability $P(D|B,C)$ as a function of the number of samples used by the sampling method. Is there any difference between the methods?
3. Choose your own query (i.e. pick a conditional probability over a suitable subset of variables and estimate using the sampling methods) of this Bayes net such that the convergence and effectiveness of rejection sampling is noticeable worse than for the other two algorithms. Report which query you chose and plot the probability as a function of the number of samples used. Why is it that rejection sampling is so much worse for this example?

What to submit

- All Python code written.
- A report stating
    - Your names and results and how many hours each person spent on the assignment.
    - Results, plots and answers to the questions.

Make sure to give the names of all the people in the group on all files you submit!

If you upload a zip file, please also upload any PDF files separately (so that they can be viewed more conveniently in Canvas).

Deadline: Monday 7 October 2019 at 12:00 (noon).

Bayesian network

The network variables are all binary {true,false} variables. Each probability is given for the state "true" of the current variable, e.g. $P(T|V = true) = P(T = true|V = true) = 0.05$.