1 Reasoning under uncertainty II (15/11/2018)

1.1 Objectives

At the end of this repetition you should be able to:

- Define, construct a Bayesian network
- Compute probabilities in the context of a simple Bayesian network.
- Do inference by variable elimination
- Explain the following approximate inference algorithm: ancestral sampling and rejection sampling; Likelihood weighting; Gibbs sampling.

1.2 Exercises

a ≈ 15 min

We have a bag of three biased coins a, b, and c with probabilities of coming up heads of 20%, 60%, and 80%, respectively. One coin is drawn randomly from the bag (with equal probability of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes X1, X2, and X3.

1. Draw the Bayesian network corresponding to this setup and define the necessary CPTs.
2. Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

b ≈ 20 min

Let H_x be a random variable denoting the handedness of an individual x, with possible values l or r. A common hypothesis is that left- or right-handedness is inherited by a simple mechanism; that is, perhaps there is a gene G_x, also with values l or r, and perhaps actual handedness turns out mostly the same (with some probability s) as the gene an individual possesses. Furthermore, perhaps the gene itself is equally likely to be inherited from either of an individual's parents, with a small nonzero probability m of a random mutation flipping the handedness.

1. Which of the three networks in Figure 1 claim that
   $P(G_{father}, G_{mother}, G_{child}) = P(G_{father})P(G_{mother})P(G_{child})$?
2. Which of the three networks make independence claims that are consistent with the hypothesis about the inheritance of handedness?
3. Which of the three networks is the best description of the hypothesis?
4. Write down the CPT for the Gchild node in network (a), in terms of s and m.
5. Suppose that $P(G_{father} = l) = P(G_{mother} = l) = q$. In network (a), derive an expression for $P(G_{child} = l)$ in terms of m and q only, by conditioning on its parent nodes.
6. Under conditions of genetic equilibrium, we expect the distribution of genes to be the same across generations. Use this to calculate the value of q, and, given what you know about handedness in humans, explain why the hypothesis described at the beginning of this question must be wrong. hypothesis about the inheritance of handedness?

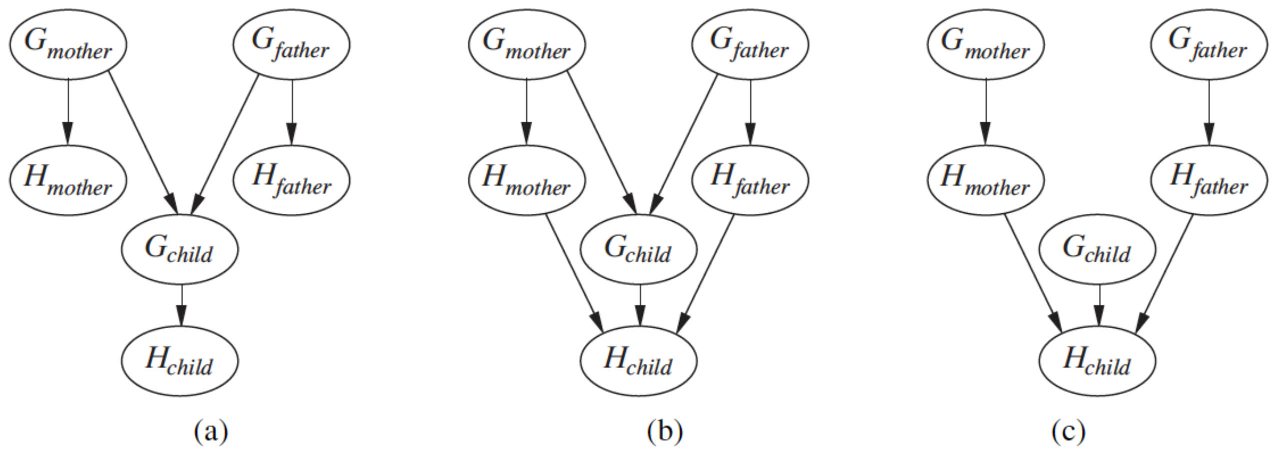

Figure 1: Possible Bayesian Networks of handedness inheritance

c ≈ 15 **min**

You are advised to take a look at d-separation before doing this exercise: `http://web.mit.edu/jmn/www/6.034/d-separation.pdf`.

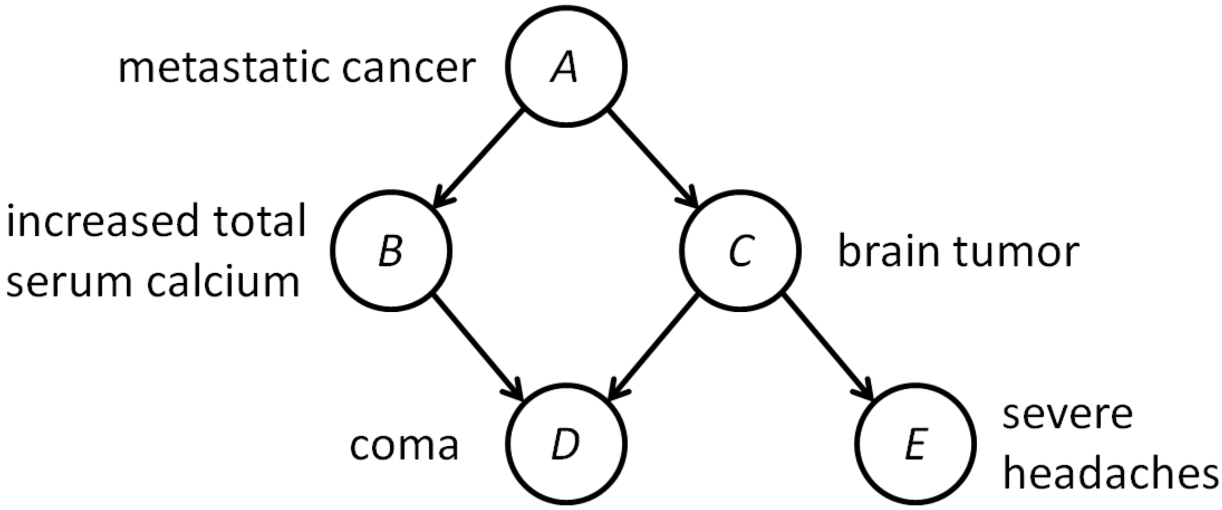

Figure 2: Bayesian Network of metastatic cancer

Consider the Bayesian network of Figure 2, which, if any, of the following are asserted by the network structure ?

- $P(b, c) = P(b)P(c)$
- $P(b, c|a) = P(b|a)P(c|a)$
- $P(b, c|a, d) = P(b|a, d)P(c|a, d)$
- $P(c|a, d, e) = P(c|a, b, d, e)$
- $P(b, e|a) = P(b|a)P(e|a)$
- $P(b, e) = \sum_{a \in A, c \in C, d \in D} P(a)P(b|a)P(c|a)P(e|c)P(d|b, c)$

Use inference by variable elimination to compute $P(E|a, b)$.

function PRIOR-SAMPLE(bn) **returns** an event sampled from the prior specified by bn
  inputs: bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $\mathbf{x} \leftarrow$ an event with n elements
  foreach variable X_i **in** $X_1, \ldots, X_n$ **do**
    $\mathbf{x}[i] \leftarrow$ a random sample from **???**
  return x

Figure 3: Incomplete Prior Sampling Algorithm

function GIBBS-ASK(X, **e**, bn, N) **returns** an estimate of $\mathbf{P}(X|\mathbf{e})$
  local variables: **N**, a vector of counts for each value of X, initially zero
    Z, the nonevidence variables in bn
    x, the current state of the network, initially copied from **e**

  initialize **x** with random values for the variables in **Z**
  for j = 1 to N **do**
    for each Z_i in **Z do**
      set the value of Z_i in **x** by sampling from **???**
      $\mathbf{N}[x] \leftarrow \mathbf{N}[x] + 1$ where x is the value of X in **x**
  return NORMALIZE(**N**)

Figure 4: Incomplete Gibbs Sampling Algorithm

d Quizz (≈ 10 min)

1. What is the Markov Blanket of a node ?
2. Complete the Prior Sampling algorithm of Figure 3.
3. Why is Rejection sampling inefficient ?
4. Complete the Gibbs Sampling algorithm of Figure 4.

1.3 Supplementary material

http://bayes.cs.ucla.edu/WHY/why-intro.pdf