Problem 2

Thomas Gauthier
Nicolas Gauthier
Raul da Silva Souza

November 2017, 6

Exercice 1

We consider the Bayesian Network given in the exercise. A, B, C and D are random Boolean variables.

The probability of having D true if we know that A is true is P(D|A).

By definition, we know that

$P(D|A) = \frac{P(D,A)}{P(A)}$

P(D,A) can be decomposed the following way :

$P(D,A) = P(D,B,C,A) + P(D,\bar{B},C,A) + P(D,B,\bar{C},A) + P(D,\bar{B},\bar{C},A)$

Then, we can compute the joint probability distribution over all the variables A,B,C,D in the Bayesian net using the formula :

$$P(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i|Parents(X_i))$$

That give us :

$P(D,B,C,A) = P(A) * P(B|A) * P(C|A) * P(D|B,C) = 0,75 * 0,2 * 0,7 * 0,3 = 0,0315$

$P(D,\bar{B},C,A) = P(A) * P(\bar{B}|A) * P(C|A) * P(D|\bar{B},C) = 0,75 * 0,8 * 0,7 * 0,1 = 0,042$

$P(D,B,\bar{C},A) = P(A) * P(B|A) * P(\bar{C}|A) * P(D|B,\bar{C}) = 0,75 * 0,2 * 0,3 * 0,25 = 0,01125$

$P(D,\bar{B},\bar{C},A) = P(A) * P(\bar{B}|A) * P(\bar{C}|A) * P(D|\bar{B},\bar{C}) = 0,75 * 0,8 * 0,3 * 0,35 = 0,063$

Hence, $P(D,A) = 0,14775$ and $\boxed{P(D|A) = 19,7\%}$.

Exercice 2

Question a)

Let's call S the event "Sunny weather", HT the event "High Temperature" and W the event "Weekend". These three events influence the number of visitors, ie. the event "Lots of visitors" (called V). Assuming S, HT and W are independent, we can build the following Bayesian network :

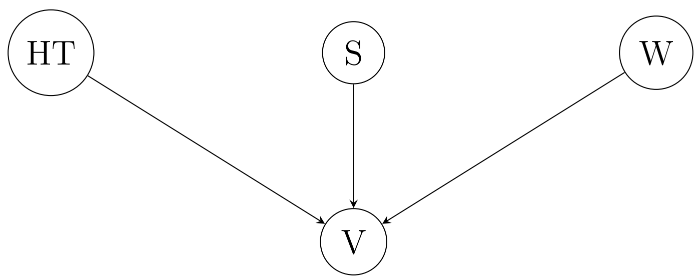

FIGURE 0.1 – Structure of Bayesian Network

From the examples, we can infer the following empirical probabilities : P(S) = 5/7, P(HT) = 4/7, P(W) = 3/7.

The conditional probabilities defining the network are then $P(V|S, W, HT)$, $P(V|\bar{S}, W, HT)$, $P(V|S, \bar{W}, HT)$, $P(V|S, W, \bar{HT})$, $P(V|\bar{S}, \bar{W}, HT)$, $P(V|\bar{S}, W, \bar{HT})$, $P(V|S, \bar{W}, \bar{HT})$ and $P(V|\bar{S}, \bar{W}, \bar{HT})$. One of them corresponds to the probability asked in question 2 : $P(V|\bar{S}, W, HT)$. It could be computed, using the Bayes formula :

$P(V|\bar{S}, W, HT) = \frac{P(\bar{S},W,HT|V)*P(V)}{P(\bar{S},W,HT)}$

Given that S, W and HT are independent, $P(\bar{S}, W, HT) = P(\bar{S}) * P(W) * P(HT)$ and $P(\bar{S}, W, HT|V) = P(\bar{S}|V) * P(W|V) * P(HT|V)$.

Those new conditional probabilities can be computed empirically from the examples : $P(\bar{S}|V) = \frac{P(\bar{S} \cap V)}{P(V)} = 1/5$. The same way, we get P(W|V) = 2/5, P(HT|V) = 4/5.

Thus, using P(V) = 5/7, we get $P(V|\bar{S}, W, H) = 98/375$.

But the order of questions suggests this is not the right modelisation to adopt. Let's rather consider that V is the parent of the other nodes, taking inspiration from the flu example in the third lecture. In this case, the network would be the following :

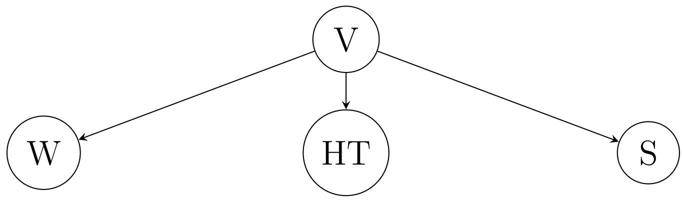

FIGURE 0.2 – Structure of Bayesian Network 2nd version

From the examples, we can infer P(V) = 5/7.

The conditional probabilities defining the network are $P(S|V) = 4/5$, $P(S|\bar{V}) = 1/2$, $P(HT|V) = 4/5$, $P(HT|\bar{V}) = 0$, $P(W|V) = 2/5$, $P(W|\bar{V}) = 1/2$.

Question b)

Now, let's compute the probability of receiving many visitors at Disney Park on a cloudy and hot weekend day

By definition, $P(V|\bar{S}, W, HT) = \frac{P(V,\bar{S},W,HT)}{P(\bar{S},W,HT)}$.

Using the Bayesian network, we can compute $P(V, \bar{S}, W, HT) = P(V)P(\bar{S}|V)P(HT|V)P(W|V) = 8/175$.

The same way, we can compute $P(\bar{S}, W, HT) = P(\bar{S}, W, HT, \bar{V}) + P(\bar{S}, W, HT, V) = 0 + 8/175$ because $P(HT|\bar{V}) = 0$.

So we infer $\boxed{P(V|\bar{S}, W, HT) = 1}$.

The naïve Bayesian model predicts there will be lots of visitors with probability 1, which cannot be true. It shows the limitation of this model. This is due to the few number of examples and the fact that the Bayesian learner considers the event $HT|\bar{V}$ impossible because it is not represented in the training data.