University of Pittsburgh
CS 2740 Knowledge Representation — Handout 15
Professor Milos Hauskrecht — December 5, 2008

Problem set 6 solutions

Uncertainty

Problem 1

Let A,B,C be random variables. A and B have two possible values T,F and C has three values: high, medium, low. The full joint distribution is defined by the following contingency table.

C / (A,B)	(T T)	(T F)	(F T)	(F F)
high	0.02	0.08	0.03	0.12
medium	0.1	0.1	0.05	0.15
low	0.05	0.05	0.15	0.1

Calculate the following **probability distributions**:

- Part a. $P(A, B)$.
- Part b. $P(A, C)$.
- Part c. $P(B)$.
- Part d. $P(A, B|C = medium)$.
- Part e. $P(A|B = T, C = high)$.

Solutions.

$P(A, B)$:

A / B	T	F
T	0.17	0.23
F	0.23	0.37

$P(A, C)$:

C / A	T	F
high	0.1	0.15
medium	0.2	0.2
low	0.1	0.25

$P(B)$:

T	F
0.4	0.6

$P(A, B|C = medium)$:

A / B	T	F
T	0.25	0.25
F	0.125	0.375

$P(A|B = T, C = high)$:

T	F
0.4	0.6

Problem 2

Random variables A,B are conditionally independent given C when:

$$P(A, B|C) = P(A|C)P(B|C).$$

Prove that this implies:

$$P(A|B, C) = P(A|C); \text{ and } P(B|A, C) = P(B|C).$$

Solution:

$$\begin{aligned}
P(A|B, C) &= \frac{P(A, B, C)}{P(B, C)} \quad \text{by definition of the conditional probability;} \\
&= \frac{P(A, B|C)P(C)}{P(B, C)} \quad \text{by product rule;} \\
&= \frac{P(A|C)P(B|C)P(C)}{P(B, C)} \quad \text{from conditional independence;} \\
&= \frac{P(A|C)P(B, C)}{P(B, C)} \quad \text{by definition of the conditional probability;} \\
&= P(A|C)
\end{aligned}$$

To prove that $P(B|A, C) = P(B|C)$ we use the same sequence of steps, with A and B interchanged.

Problem 3

A pharmaceutical company has developed a nearly accurate test for the disease A. The accuracy of the test is 99%, that is, with probability 0.99 it gives the correct result (the same probability for disease-positive-test and no-disease-negative-test combinations) and only in 1% of tested cases (probability 0.01) the result is wrong. The incidence of the disease in the population is 0.01% (probability 0.0001). Compute the probability that somebody from wide population who has tested positive indeed suffers from the disease. Would you recommend the test to be widely adopted?

Answer:
$P(disease = T) = 0.0001$
$P(disease = F) = 0.9999$
$P(Test = Positive|disease = T) = 0.99$
$P(Test = Negative|disease = T) = 0.01$
$P(Test = Negative|disease = F) = 0.99$
$P(Test = Positive|disease = F) = 0.01$
$P(disease = T|Test = Positive) = \frac{P(Test=Positive|Disease=T).P(disease=T)}{P(Test=Positive)} = \frac{0.99 \times 0.0001}{P(Test=Positive)}$
$P(disease = F|Test = Positive) = \frac{P(Test=Positive|Disease=F).P(disease=F)}{P(Test=Positive}) = \frac{0.01 \times 0.9999}{P(Test=Positive)}$

where

$$\begin{aligned} & P(Test = Positive) \\ = \; & P(Test = Positive|Disease = T).P(disease = T) \\ & +P(Test = Positive|Disease = F).P(disease = F) \\ = \; & 0.99 \times 0.0001 + 0.01 \times 0.9999 \\ = \; & 0.010098 \end{aligned}$$

Therefore

$$P(disease = T|Test = Positive) = \frac{0.99 \times 0.0001}{0.010098} = 0.0098$$

and

$$P(disease = F|Test = Positive) = \frac{0.01 \times 0.9999}{0.010098} = 0.990196.$$

As is evident from the above probabilities, there is a very high chance that the recipient does not have the disease even if he/she tests positive, so it is not advisable to conduct the test for the wide population. However, in practice such tests are widely used if they are inexpensive compared to the cost of the disease and a more specific (and expensive) test is applied to those who test positive (HIV testing).