UNIVERSITY OF WARSAW

Faculty of Mathematics, Informatics and Mechanics
University of Warsaw

Statistical Data Analysis 2

Exercises # 1

Ewa Szczurek
Lukasz Kozlowski

September 20, 2018

Problem 1: Conditional independence
Let X, Y, Z be random variables. X and Y are said to be *conditionally independent* given Z (in symbols $X \perp Y \mid Z$) if

$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z).$$

This condition is equivalent to

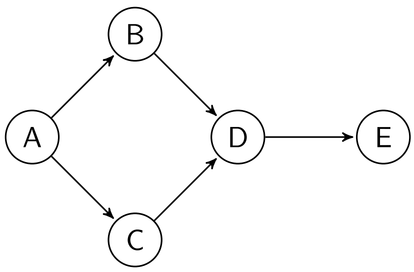
$$P(X \mid Y, Z) = P(X \mid Z).$$

Using the laws of probability, show that this equivalence holds (show both directions of the proof).

Solution:

- If the defining property holds, then we have by the laws of probability:
  $P(X \mid Y, Z) = \frac{P(X,Y,Z)}{P(Y,Z)} = \frac{P(X,Y|Z)P(Z)}{P(Y,Z)} = \frac{P(X|Z)P(Y|Z)P(Z)}{P(Y,Z)} = \frac{P(X|Z)P(Y,Z)}{P(Y,Z)} = P(X \mid Z)$.
- The converse is shown as follows:
  $P(X, Y \mid Z) = P(X \mid Y, Z)P(Y \mid Z) = P(X \mid Z)P(Y \mid Z)$.

Problem 2: Markov blanket
Consider the following graphical structure of a Bayesian network:



Determine the Markov blanket $\mathrm{MB}(C)$ of the node C and show that the conditional probability $P(C \mid A, B, D, E)$ can be expressed as

$$P(C \mid A, B, D, E) = P(C \mid \mathrm{MB}(C)).$$

Solution: By the law of total probability we have

$$P(C \mid A,B,D,E) = \frac{P(A)P(B \mid A)P(C \mid A)P(D \mid B,C)P(E \mid D)}{\sum_C P(A)P(B \mid A)P(C \mid A)P(D \mid B,C)P(E \mid D)}$$
$$= \frac{P(A)P(B \mid A)P(C \mid A)P(D \mid B,C)P(E \mid D)}{P(A)P(B \mid A)P(E \mid D)\sum_C P(C \mid A)P(D \mid B,C)}$$
$$= \frac{P(C \mid A)P(D \mid B,C)}{\sum_C P(C \mid A)P(D \mid B,C)}$$

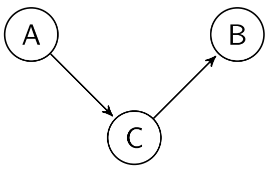
Since this expression does not depend on E we have

$$P(C \mid A,B,D,E) = P(C \mid A,B,D)$$

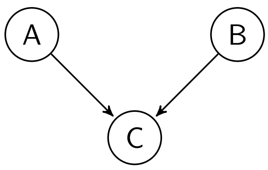
Problem 3: Conditional independence and BNs

Consider the following graphical structures, corresponding to (different) Bayesian networks. For which network does the statement $A \perp B \mid C$ hold? For which does the statement $A \perp B$ hold? Prove your answers by the laws of probability.

a)



b)



Solution:

	a)	b)
$A \perp B \mid C$	Yes	No
$A \perp B$	No	Yes

Explanation for the conditional independences:

- For a) we have $P(A,B,C) = P(B \mid C)P(C \mid A)P(A)$.
  Inserting this and using Bayes' rule we get the conditional probability:
  $P(A,B \mid C) = \frac{P(A,B,C)}{P(C)} = P(B \mid C)\frac{P(C|A)P(A)}{P(C)} = P(B \mid C)P(A \mid C) \Leftrightarrow A \perp B \mid C$
- For b) we have $P(A,B,C) = P(C \mid A,B)P(A)P(B)$.
  Consequently, we have
  $$P(A,B \mid C) = \frac{P(A,B,C)}{P(C)} = \frac{P(C \mid A,B)P(A)P(B)}{P(C)}$$
  $$= \frac{P(C \mid A,B)P(A)P(B)P(C \mid A)}{P(C)P(C \mid A)} = \frac{P(A \mid C)P(C \mid A,B)P(B)}{P(C \mid A)}$$
  $$= \frac{P(A \mid C)P(C \mid A,B)P(B)P(C \mid B)P(C)}{P(C \mid A)P(C \mid B)P(C)} = P(A \mid C)P(B \mid C)\frac{P(C \mid A,B)P(C)}{P(C \mid A)P(C \mid B)}$$
  Since the last term will only equal 1 in special cases, in general $A \perp B \mid C$ does not hold.

Explanation for the marginal independences:

- For a) applying Bayes' rule, we get
  $$P(A,B) = \sum_C P(A,B,C) = \sum_C P(B \mid C)P(C \mid A)P(A)$$
  $$= \sum_C \frac{P(C \mid B)P(B)P(C \mid A)P(A)}{P(C)} = P(A)P(B)\sum_C \frac{P(C \mid B)P(C \mid A)}{P(C)}.$$
  Therefore, $A \perp B$ does not generally hold.

- For b) we have $P(A,B,C) = P(C \mid A,B)P(A)P(B)$. Marginalizing over C: $\sum_C P(A,B,C) = \sum_C P(C \mid A,B)P(A)P(B)$ yields $P(A,B) = P(A)P(B) \Leftrightarrow A \perp B$

Problem 4: Conjugate distributions
Let X be a binomially distributed random variable with parameters $N \in \mathbb{N}$ and $\theta \in [0,1]$. Further, assume that a prior $P(\theta)$ of the parameter θ is *beta distributed* with parameters α, β, i.e. its probability density function is given by

$$\rho(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt}$$

Show that the prior $P(\theta)$ is conjugate to the binomial likelihood $L(\theta) := P(X \mid \theta)$. In other words, show that the posterior distribution $P(\theta \mid X)$, which is defined as

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)}$$

also obeys a beta distribution with suitable parameters.

Solution: Given that

$$P(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta),} \quad \text{where } B(\cdot,\cdot) \text{ is the beta function, and}$$

$$P(X \mid \theta) = \binom{N}{x}\theta^x(1-\theta)^{N-x},$$

we obtain

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)} = \frac{P(X \mid \theta)P(\theta)}{\int\limits_0^1 P(X \mid \theta)P(\theta)d\theta}$$

$$= \frac{\binom{N}{x}\theta^{\alpha+x-1}(1-\theta)^{\beta+N-(x+1)}(B(\alpha,\beta))^{-1}}{\int\limits_0^1 \binom{N}{x}\theta^{\alpha+x-1}(1-\theta)^{\beta+N-(x+1)}(B(\alpha,\beta))^{-1}d\theta}$$

$$= \frac{\theta^{\alpha+x-1}(1-\theta)^{\beta+N-(x+1)}}{\int\limits_0^1 \theta^{\alpha+x-1}(1-\theta)^{\beta+N-x-1}d\theta} = \frac{\theta^{\alpha+x-1}(1-\theta)^{\beta+N-x-1}}{B(\alpha+x,\beta+N-x)}$$

Therefore $P(\theta \mid X)$ is beta distributed with parameters $\alpha + x$ and $\beta + N - x$.