Lecture 3

Cezar Ionescu

25/05 DEPARTMENT FOR CONTINUING EDUCATION

UNIVERSITY OF OXFORD

Administrative

No lecture on the 1st of June!

Homework from 2019-05-18 due **now**!

Questions?

Joint probability distributions

Revisiting the diagnostic problem of the previous lecture:

- `H = {healthy, ill}`
- test results: `R = {⊖, ⊕}`
- `0.008` of the population have the disease.
- patient ill ⇒ test positive in `98%` of the cases
- patient healthy ⇒ test positive in `3%` of the cases

Joint probability distributions

```
Ω = (H, R)
  = {(healthy, ⊖), (healthy, ⊕), (ill, ⊖), (ill, ⊕)
```

Joint probability distributions

$$
\begin{aligned}
\Omega &= (H, R) \\
&= \{(\text{healthy}, \ominus), (\text{healthy}, \oplus), (\text{ill}, \ominus), (\text{ill}, \oplus)
\end{aligned}
$$

$$\text{Healthy} = \{(\text{healthy}, \ominus), (\text{healthy}, \oplus)\}$$

Joint probability distributions

```
Ω = (H, R)
  = {(healthy, ⊖), (healthy, ⊕), (ill, ⊖), (ill, ⊕)}

Healthy = {(healthy, ⊖), (healthy, ⊕)}

pₕ : ℙ(H) → [0, 1]
pₕ (healthy) = p({(healthy, ⊖), (healthy, ⊕)})
             = p (healthy, ⊖) + p (healthy, ⊕)
```

Joint probability distributions

Notation:

- for $x \in \Omega$ $p(x)$ is an abbreviation for $p(\{x\})$.
- similarly $p(\text{healthy}, \ominus)$ denotes $p(\{(\text{healthy}, \ominus)\})$
- $p(\text{healthy})$ denotes $p(\{(\text{healthy}, r) \mid r \in R\})$

Remark: in general, if $\Omega = (\Omega_1, \Omega_2, \ldots, \Omega_n)$ and $\omega_i \in \Omega_i$, we will use p(ω_i) to denote

$$p(\omega_i) = p(\{(x_1, x_2, \ldots, x_n) \mid x_1 \in \Omega_1, \ldots, x_n \in \Omega_n, x_i = \omega_i\})$$

Joint probability distributions

- the "complete" probability measure p is called the **joint probability distribution**
- the probability measures over the individual components (or over tuples of individual components) are called **marginal probability distributions**

Remark: it is trivial to derive the marginal probability distributions, given the joint probability distribution. However, even if we have all the marginal distributions, we cannot derive the joint distribution from them.

Joint probability distributions

If Ω consists of a tuple of n boolean components, then the joint probability distribution can be represented as a table with 2^n rows.

Joint probability distributions

Let $\Omega = (\Omega_1, \Omega_2, \ldots, \Omega_n)$ and $(\omega_1, \omega_2, \ldots, \omega_n) \in \Omega$. Then

$$
\begin{aligned}
p(\omega_1, \omega_2, \ldots, \omega_n) &= p(\omega_1 \cap \omega_2 \cap \ldots \cap \omega_n) \\
&= p(\omega_1 \mid \omega_2 \cap \ldots \cap \omega_n) * p(\omega_2 \cap \ldots \cap \omega_n)
\end{aligned}
$$

It is easy to see that, therefore:

$$
\begin{aligned}
p(\omega_1, \omega_2, \ldots, \omega_n) = \; & p(\omega_1 \mid \omega_2 \cap \ldots \cap \omega_n) * \\
& p(\omega_2 \mid \omega_3 \cap \ldots \cap \omega_n) * \ldots \\
& p(\omega_n)
\end{aligned}
$$

Independence

Definition: Two events $X, Y \subseteq \Omega$ are called **independent** if

$$p(X \cap Y) = p(X) * p(Y)$$

Examples

Example: Consider a fair standard die, and the events `even = {2, 4, 6}`, `big = {5, 6}`. Then `even` and `big` are independent.

Exercise: If `X, Y` are independent, then so are `X, ¬Y`.

Example: The events `big` and `divBy3 = {3, 6}` are **not** independent.

Conditional independence

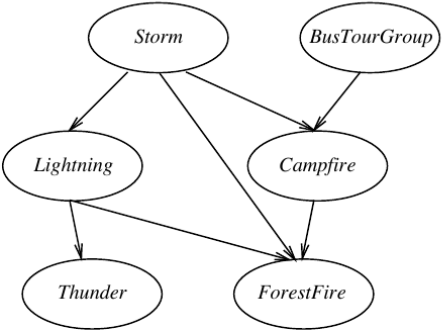
Definition: Consider events `X, Y, Z ⊆ Ω` such that `p(Z) ≠ 0`. Events `X, Y` are called **conditionally independent given** `Z` if

```
p(X ∩ Y | Z) = P(X | Z) * p(Y | Z)
```

Examples

Example:

- The events `even, big` are conditionally dependent given `divBy3`.
- The events `divBy3, big` are conditionally independent given `{2, 3, 5, 6}`.



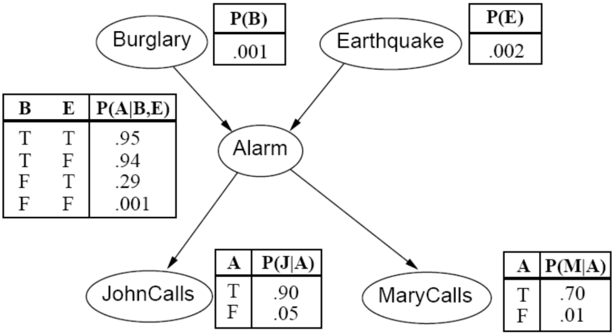
	S,B	$S,\neg B$	$\neg S,B$	$\neg S,\neg B$
C	0.4	0.1	0.8	0.2
$\neg C$	0.6	0.9	0.2	0.8

Bayesian networks

```
  p(thunder, forestfire, campfire, lightning, storm, bustourgroup
=
  p(thunder | forestfire ∩ campfire ∩ lightning ∩ storm ∩ bustourg
  p(forestfire ∩ campfire ∩ lightning ∩ storm ∩ bustourgroup)
=
  p(thunder | lightning) *
  p(forestfire ∩ campfire ∩ lightning ∩ storm ∩ bustourgroup)
=
  p(thunder | lightning) *
  p(forestfire | campfire ∩ lightning) *
  p(campfire | storm ∩ bustourgroup) *
  p(lightning | storm) *
  p(storm) * p(bustourgroup)
```

Homework

(based on Nilsson, 19.4, page 340) Consider the following Bayesian network:



Compute `p(¬J, ¬M, A, B, E)`. This is the probability that there is both an earthquake and a burglary, the alarm rings, but neither John nor Mary call.

Exercise

Exercise: Compute `p(¬J, ¬M, B, E)` (this is exercise 19.4 in Nielsson).