COMP111: Artificial Intelligence

Section 9a. Reasoning under Uncertainty 2a

Frank Wolter

Content

- Random variables
- (Full) joint probability distribution
- Marginalization
- Probabilistic inference problem

Random variable

Let (S, P) be a probabality space. A random variable F is a function $F : S \to \mathbb{R}$ that assigns to every $s \in S$ a single number $F(s)$.

- Neither a variable nor random
- English translation of variabile casuale

We still assume that the sample space is finite. Thus, given a random variable F from some sample space S, the set of numbers r that are values of F is finite as well.

The event that F takes the value r, that is $\{s \mid F(s) = r\}$, is denoted $(F = r)$. The probability $(F = r)$ of the event $(F = r)$ is then

$$P(F = r) = P(\{s \mid F(s) = r\})$$

Example 1

Let

$$S = \{\text{car}, \text{train}, \text{plane}, \text{ship}\}$$

Then the function $F : S \to \mathbb{R}$ defined by

$$F(\text{car}) = 1, \quad F(\text{train}) = 1, \quad F(\text{plane}) = 2, \quad F(\text{ship}) = 2$$

is a random variable.

$(F = 1)$ denotes the event $\{s \in S \mid F(s) = 1\} = \{\text{car}, \text{train}\}$.

Define a uniform probability space (S, P) by setting

$$P(\text{car}) = P(\text{train}) = P(\text{plane}) = P(\text{ship}) = \frac{1}{4}$$

Then $P(F = 1) = P(\{s \in S \mid F(s) = 1\}) = P(\{\text{car}, \text{train}\}) = \frac{1}{2}$.

Example 2

Suppose that I roll two dice. So the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}^2$$

and $P(ab) = \frac{1}{36}$ for every $ab \in S$.
Let

$$F(ab) = a + b.$$

F is a random variable. The probability that

$$F = r$$

for a number r (say, 12) is given by

$$P(F = r) = P(\{ab \mid F(ab) = r\})$$

For example, $P(F = 12) = P(\{ab \mid F(ab) = 12\}) = P(66) = \frac{1}{36}$.

Random Variable

When defining a probability distribution P for a random variable F, we often do not specify its sample space S but directly assign a probability to the event that F takes a certain value. Thus, we directly define the probability

$$P(F = r)$$

of the event that F has value r. Observe:

- $0 \leq P(F = r) \leq 1$;
- $\sum_{r \in \mathbb{R}} P(F = r) = 1$.

Thus, the events $(F = r)$ behave in the same way as outcomes of a random experiment.

Notation and Rules

- We write $\neg(F = r)$ for the event $\{s \mid F(s) \neq r\}$. For example, assume the random variable *Die* can take values $\{1, 2, 3, 4, 5, 6\}$ and

  $$P(Die = n) = \frac{1}{6}$$

  for all $n \in \{1, 2, 3, 4, 5, 6\}$ (thus we have a fair die). Then $\neg(Die = 1)$ denotes the event

  $(Die = 2)$ or $(Die = 3)$ or $(Die = 4)$ or $(Die = 5)$ or $(Die = 6)$

  We have the following complementation rule:

  $$P(\neg(F = r)) = 1 - P(F = r)$$

- We write $(F_1 = r_1, F_2 = r_2)$ for '$(F_1 = r_1)$ and $(F_2 = r_2)$'.

Notation and Rules

- We write $(F_1 = r_1) \vee (F_2 = r_2)$ for '$(F_1 = r_1)$ or $(F_2 = r_2)$'. Then

$$\begin{aligned} P((F_1 = r_1) \vee (F_2 = r_2)) &= P(F_1 = r_1) + P(F_2 = r_2) \\ &\quad - P(F_1 = r_1, F_2 = r_2) \end{aligned}$$

- Conditional probability: if $P(F_2 = r_2) \neq 0$, then

$$P(F_1 = r_1 \mid F_2 = r_2) = \frac{P(F_1 = r_1, F_2 = r_2)}{P(F_2 = r_2)}$$

- Product rule:

$$P(F_1 = r_1, F_2 = r_2) = P(F_1 = r_1 \mid F_2 = r_2) \times P(F_2 = r_2)$$

Notation

We sometimes use symbols distinct from numbers to denote the values of random variables.
For example, for a random variable *Weather* rather than using values $1, 2, 3, 4$, we use

$$sunny, rain, cloudy, snow$$

Thus,

$$(Weather = sunny)$$

denotes the event that it is sunny.
To model a visit to a dentist, we use random variables *Toothache*, *Cavity*, and *Catch* (the dentist's steel probe catches in the tooth) that all take values 1 and 0 (for true and false).
For example, $(Toothache = 1)$ states that the person has toothache and $(Toothache = 0)$ states that the person does not have toothache.

Examples of probabilistic models

To model a domain using probability theory, one first introduces the relevant random variables. We have seen two basic examples:

- The weather domain could be modeled using the single random variable *Weather* with values

$$(sunny, rain, cloudy, snow)$$

- The dentist domain could be modeled using the random variables *Toothache*, *Cavity*, and *Catch* with values 0 and 1 for true and false.
  We might be interested in

$$P(Cavity = 1 \mid Toothache = 1, Catch = 1)$$

Student Exam Domain

A very basic model of the performance of students in an exam could be given by the random variables

- *Grade*: takes as values the possible grades of a student in the exam;
- *Answers*: takes as values the possible answers to exam questions;
- *Background*: takes as value the school visited before going to university;
- *Works_hard*: takes as values the degree to which the student works hard.

We might be interested in

$$P(\mathit{Grade} = A \mid \mathit{Works\_hard} = 1, \mathit{Background} = \mathit{Comprehensive})$$

Fire Alarm Domain

A basic model of a fire alarm system and reporting about it could be given by the following random variables (all take value 0 or 1):

- *Fire*: there is fire;
- *Alarm*: the alarm goes off;
- *Tampering*: there is tampering with the alarm system;
- *Smoke*: there is smoke (no smoke detector used);
- *Leaving*: people leave the building;
- *Report*: it is reported that people leave the building (reporting not always correct).

We might be interested in

$$P(\mathit{Fire} = 1 \mid \mathit{Report} = 1)$$

Probability Distribution

- The probability distribution for a random variable gives the probabilities of all the possible values of the random variable.
- For example, let *Weather* be a random variable with values

$$(sunny, rain, cloudy, snow)$$

  such that its probability distribution is given by
    - $P(Weather = sunny) = 0.7;$
    - $P(Weather = rain) = 0.2$
    - $P(Weather = cloudy) = 0.08;$
    - $P(Weather = snow) = 0.02.$
- Assume the order of the values is fixed. Then we write instead

$$\mathbf{P}(Weather) = (0.7, 0.2, 0.08, 0.02)$$

  where the bold **P** indicates that the result is a vector of numbers representing the individual values of *Weather*.

More Probability Distributions

- Assume the random variable *Die* can take the values $1, 2, 3, 4, 5, 6$ and represents a fair die. Then we can define its probability distribution as

$$\mathbf{P}(Die) = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$$

- Recall the random variable $F(ab) = a + b$ from the sample space $S = \{1, 2, 3, 4, 5, 6\}^2$ with $P(ab) = \frac{1}{36}$ for all $a, b \in \{1, 2, 3, 4, 5, 6\}$. Then $2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ are its possible values. Then

$$\mathbf{P}(F) = (\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \ldots, \frac{1}{36})$$

Joint Probability Distribution

Let $F_1, \ldots, F_k$ be random variables. A joint probability distribution for

$$F_1, \ldots, F_k$$

gives the probabilities

$$P(F_1 = r_1, \ldots, F_k = r_k)$$

for the events

$$(F_1 = r_1) \text{ and } \cdots \text{ and } (F_k = r_k)$$

that F_1 takes value r_1, F_2 takes value r_2, and so on up to k, for all possible values $r_1, \ldots, r_k$.

The joint probability distribution is denoted $\mathbf{P}(F_1, \ldots, F_k)$.

Example

A possible joint probability distribution $\mathbf{P}(Weather, Cavity)$ for the random variables *Weather* and *Cavity* is given by the following table:

$Weather =$	*sunny*	*rain*	*cloudy*	*snow*
$Cavity = 1$	0.144	0.02	0.016	0.02
$Cavity = 0$	0.576	0.08	0.064	0.08

The probabilities of the joint distribution sum to 1!

Full Joint Probability Distribution

A full joint probability distribution

$$\mathbf{P}(F_1, \ldots, F_k)$$

is a joint probability distribution for all relevant random variables $F_1, \ldots, F_k$ for a domain of interest.

Every probability question about a domain can be answered by the full joint distribution because the probability of every event is a sum of probabilities

$$P(F_1 = r_1, \ldots, F_k = r_k)$$

(The $r_1, \ldots, r_k$ are often called data points or sample points.)

Example: Full Joint Probability Distribution for Dentist Domain

Assume the random variables *Toothache*, *Cavity*, *Catch* fully describe a visit to a dentist.
Then a full joint probability distribution is given by the following table:

	$Toothache = 1$		$Toothache = 0$	
	$Catch = 1$	$Catch = 0$	$Catch = 1$	$Catch = 0$
$Cavity = 1$	0.108	0.012	0.072	0.008
$Cavity = 0$	0.016	0.064	0.144	0.576

The probabilities of the joint distribution sum to 1!

Full Joint Probability Distributions

- The full joint probability distribution for the student exam domain, denoted

$$\mathbf{P}(Grade, Answers, Background, Works\_hard)$$

  gives the probability for every possible combination of values of the random variables *Grade*, *Answers*, *Background*, and *Works_hard*.
- The full joint probability distribution for the fire alarm domain gives the probability for every possible combination of values of the random variables *Fire*, *Alarm*, *Tampering*, *Smoke*, *Leaving*, and *Report*.

Marginalization

Given a joint distribution $\mathbf{P}(F_1, \ldots, F_k)$, one can compute the marginal probabilities of the random variables F_i by summing out the remaining variables.

For example,

$$P(Cavity = 1) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

is the sum of the entries in the first row of:

	$Toothache = 1$		$Toothache = 0$	
	$Catch = 1$	$Catch = 0$	$Catch = 1$	$Catch = 0$
$Cavity = 1$	0.108	0.012	0.072	0.008
$Cavity = 0$	0.016	0.064	0.144	0.576

Conditional Distributions

- We can also compute conditional distributions from the full joint distribution.
- We use the **P** notation for conditional distributions.
- $\mathbf{P}(F \mid G)$ gives the conditional distribution of F given G given by the probabilities $P(F = r \mid G = s)$ for all values r and s.
- Using **P** notation, the general version of the product rule is as follows:

$$\mathbf{P}(F, G) = \mathbf{P}(F \mid G)\mathbf{P}(G)$$

stands for the list of equations:

$$\begin{aligned} P(F = r_1, G = s_1) &= P(F = r_1 \mid G = s_1)P(G = s_1) \\ P(F = r_1, G = s_2) &= P(F = r_1 \mid G = s_2)P(G = s_2) \\ \cdots &= \cdots \end{aligned}$$

Probabilistic Inference

Probabilistic inference can be characterized as the computation of posterior probabilities

$$\mathbf{P}(Q \mid E_1 = e_1, \ldots, E_n = e_n)$$

for query variables Q given observed evidence $e_1, \ldots, e_n$.
In principle, we can use the full joint distribution to do this:

	$Toothache = 1$		$Toothache = 0$	
	$Catch = 1$	$Catch = 0$	$Catch = 1$	$Catch = 0$
$Cavity = 1$	0.108	0.012	0.072	0.008
$Cavity = 0$	0.016	0.064	0.144	0.576

Example: $\mathbf{P}(Cavity \mid Toothache = 1)$

We want to compute the conditional probability distribution for $Cavity$ given the observation/evidence $Toothache = 1$.

Thus we want to compute:

- $P(Cavity = 1 \mid Toothache = 1)$ and
- $P(Cavity = 0 \mid Toothache = 1)$

We can easily obtain this using the table:

$$P(Cavity = 1 \mid Toothache = 1) = \frac{P(Cavity = 1, Toothache = 1)}{P(Toothache = 1)} = \frac{0.12}{0.2} = 0.6$$

$$P(Cavity = 0 \mid Toothache = 1) = \frac{P(Cavity = 0, Toothache = 1)}{P(Toothache = 1)} = \frac{0.08}{0.2} = 0.4$$

The denominator 0.2 can be viewed as a normalization constant $\frac{1}{\alpha} = 5$ for the distribution $\mathbf{P}(Cavity | Toothache = 1)$, ensuring that it adds up to 1.

Example: $\mathbf{P}(Cavity \mid Toothache = 1)$

Instead of

$$P(Cavity = 1 \mid Toothache = 1) = \frac{P(Cavity = 1, Toothache = 1)}{P(Toothache = 1)} = \frac{0.12}{0.2} = 0.6$$

$$P(Cavity = 0 \mid Toothache = 1) = \frac{P(Cavity = 0, Toothache = 1)}{P(Toothache = 1)} = \frac{0.08}{0.2} = 0.4$$

consider

$$\begin{aligned}\mathbf{P}(Cavity \mid Toothache = 1) &= \alpha \mathbf{P}(Cavity, Toothache = 1) \\ &= \alpha(0.12, 0.08) \\ &= 5(0.12, 0.08) \\ &= (0.6, 0.4)\end{aligned}$$

Combinatorial Explosion

This approach does not scale well: for a domain described by n random variables taking k distinct values each we face two problems:

- Writing up the full joint distribution requires $k^n - 1$ entries;
- How do we find the numbers (probabilities) for the entries?

For these reasons, the full joint distribution in tabular form is not a practical tool for building reasoning systems.