Sapienza University of Rome

Master in Engineering in Computer Science

Artificial Intelligence & Machine Learning

A.Y. 2022/2023

Prof. Fabio Patrizi

4. Probability

Fabio Patrizi

Outline

- Uncertainty
- Probability
- Syntax and Semantics
- Inference
- Independence and Bayes' Rule

References

- [AIMA] 12 until 12.6 (included)

Uncertainty

Consider action $A_t =$ leave for airport t minutes before flight.

Will A_t get me there on time?

Problems:

- partial observability (road state, other drivers' plans, etc.)
- noisy sensors (traffic reports)
- uncertainty in action outcomes (flat tire, etc.)
- complexity of modelling and predicting traffic

Probability

- Degree of belief about a fact or an event, given available evidence

Given the available evidence, A_{25} will get me at the airport on time with probability 0.04

Given the available evidence, A_{60} will get me at the airport on time with probability 0.85

Given the available evidence, A_{1440} will get me at the airport on time with probability 0.999

Probability

Sample Space:

- Ω: *sample space* (possible worlds)
- $\omega \in \Omega$: *sample point* (possible world, atomic event, ...)

Probability space (or *Probability Model*)

- Function $P : \Omega \mapsto \mathbb{R}$, such that:
  - $0 \leq P(\omega) \leq 1$
  - $\sum_{\omega \in \Omega} P(\omega) = 1$

Example (rolling a die):

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $P(1) = 1/6, P(2) = 1/6, \ldots, P(6) = 1/6$

Event

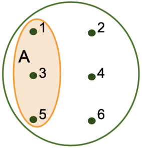
- *Event*: subset $A \subseteq \Omega$
- Probability of event A (sum of probability of its points):

$$P(A) = \sum_{\{\omega \in A\}} P(\omega)$$

- Observe: $0 \leq P(A) \leq 1$

Example:

- $A = \{1, 3, 5\} \subset \Omega$



- $P(A) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$

Random variables

Random variable:

- function from Ω to some range (e.g., reals or Booleans)

$$X : \Omega \mapsto B$$

Value of random variable depends on outcome of random phenomenon (point in Ω)

Examples:

- $Odd : \Omega \mapsto Boolean$
  - $Odd(1) = true, Odd(2) = false, Odd(5) = false, \ldots$ (odd outcome)
- $Leq4 : \Omega \mapsto Boolean$
  - $Leq4(1) = true, Leq4(5) = false, \ldots$ (outcome is less or equal than 4)
- $Total : \Omega \times \Omega \mapsto \{2, \ldots, 12\}$
  - $Total(1, 1) = 2, \ldots, Total(2, 3) = 5, \ldots$
    (outcome of rolling two dice)

Propositions

- Expressions over random variables can represent events, e.g.,:
  - $Odd = true \equiv \{\omega \in \Omega | Odd(\omega) = true\} = \{1, 3, 5\}$
  - $Total < 4 \equiv \{\omega \in \Omega \times \Omega | Total(\omega) < 4\} = \{\langle 1, 1\rangle, \langle 1, 2\rangle, \langle 2, 1\rangle\}$
  - Expressions: $Var = val$ (e.g., $Odd = true$), $X = 10$, $X < 10$, etc.
- *Proposition*: Boolean expression over assignments
- Represent events, e.g.:
  - $\neg Odd = true \equiv \{2, 4, 6\}$
  - $Odd = true \wedge Leq(4) \equiv \{1, 3\}$
  - $\neg(Odd = true) \vee Leq(4) \equiv \{1, 2, 3, 4, 6\}$
- Also for *continuous* random variables:
  - $Temp = 21.6$, $Temp \geq 18.0 \wedge Temp \leq 22.0$

We represent events as propositions

Prior Probability

Prior probabilities model beliefs prior to any (new) evidence

Examples:

- $P(Odd = true) = 0.5$
- $P(Cavity = true) = 0.1$
- $P(Weather = sunny) = 0.72$

Probability Distribution

Probability Distribution (for discrete random variables only):

- function assigning probability to each assignment of a random variable
- (must sum up to 1)

Examples:

- $P(Odd) = \langle 0.5, 0.5 \rangle$:
  - $P(Odd = true) = 0.5, P(Odd = false) = 0.5$
- $P(Cavity) = \langle 0.1, 0.9 \rangle$:
  - $P(Cavity = true) = 0.1, P(Cavity = false) = 0.9$
- $P(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$:
  - $P(Weather = sunny) = 0.72$
  - $P(Weather = rain) = 0.1$
  - $P(Weather = cloudy) = 0.08$
  - $P(Weather = snow) = 0.1$

(Real random variables: probabilities computed starting by continuous *probability density function (pdf)*)

Joint probability distribution

Joint probability distribution for a set of random variables gives the probability of every atomic joint event on those random variables (i.e., every sample point in the joint sample space).

Joint probability distribution of the random variables *Weather* and *Cavity*:

$Weather =$	*sunny*	*rain*	*cloudy*	*snow*
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution because every event is a set of sample points

Conditional/Posterior Probability

Conditional (or *Posterior*) probabilities model beliefs after new evidence

I know the outcome of a random variable, how does this affect probability of other random variables?

Example:
I know that today $Weather = sunny$, how this information affects the random variable $Cavity$?

Notation:
$P(Cavity = true|Weather = sunny)$: conditional/posterior probability

Conditional/Posterior Probability

Posterior $\neq$ Joint $\neq$ Prior!!!

$$P(Cavity = true | Weather = sunny) \neq$$
$$P(Cavity = true, Weather = sunny) \neq$$
$$P(Cavity = true)$$

Conditional/Posterior Probability

Consider Boolean random variable Toothache (I have toothache)

How does having a toothache affects the probability of having a cavity?

Example:

- Have a cavity: $cavity \equiv Cavity = true$
- Have a toothache: $tootache \equiv Toothache = true$
- Prior: $P(cavity) = 0.1$
- Posterior: $P(cavity|toothache) = 0.6$

(Having a toothache increases the probability of having a cavity)

Conditional Probability Distributions

Conditional Probability Distribution:

- Probability Distribution of a variable, given the value of another variable

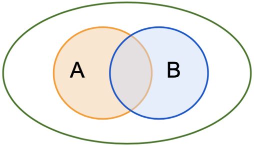
Example:

$$P(Cavity|Toothache)$$

	$Toothache = true$	$Toothache = false$
$Cavity = true$	0.6	0.1
$Cavity = false$	0.4	0.9

Conditional probability

- Definition of conditional probability: $P(a|b) \equiv \frac{P(a \wedge b)}{P(b)}$ if $P(b) \neq 0$



- Product rule: $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$
- In general, for distributions:
  $P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X)$
- If Die_1 is 4, what is the probability of obtaining < 10?

$$P(Total < 10 | Die_1 = 4) = \frac{P(Total < 10 \wedge Die_1 = 4)}{P(Die_1 = 4)} = \frac{5/36}{6/36} = 5/6$$

Total probabilities

- For propositions a and b:

$$P(a) = P(a \wedge b) + P(a \wedge \neg b) = P(a|b)P(b) + P(a|\neg b)P(\neg b)$$

- In general, for random variable Y taking mutually exclusive values y_i:

$$P(X) = \sum_{y_i \in \mathcal{Y}} P(X|Y = y_i)P(Y = y_i)$$

  - $\mathcal{Y}$: set of values for variable Y

Chain Rule

- *Chain rule* is derived by successive application of product rule:

$$P(X_1, X_2) = P(X_1|X_2)P(X_2)$$

- In general:

$$\begin{aligned}
&P(X_1, \ldots, X_n) = \\
&\qquad P(X_1|X_2, \ldots, X_n)P(X_2, \ldots, X_n) = \\
&\qquad P(X_1|X_2, \ldots, X_n)P(X_2|X_3, \ldots, X_n)P(X_3, \ldots, X_n) = \\
&\qquad \cdots = \\
&\qquad P(X_1|X_2, \ldots, X_n)P(X_2|X_3, \ldots, X_n) \ldots P(X_n) = \\
&\qquad = \prod_{i=1}^{n} P(X_i|X_{i+1}, \ldots, X_n)
\end{aligned}$$

Inference by Enumeration

- From joint distribution can compute probability of any event

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

- Probability of any proposition p:
  - sum of probabilities of atomic events favorable to p

Example:

- $P(cavity \vee toothache) =$
  $0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

Inference by Enumeration

- Also for conditional probabilities

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

$$
\begin{aligned}
P(\neg cavity | toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \\
&\frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
\end{aligned}
$$

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

- Distribution of *query variable*
  - denominator viewed as *normalization constant* α

$$\begin{aligned}
P(Cavity|toothache) &= \alpha\, P(Cavity, toothache) \\
&= \alpha\, [P(Cavity, toothache, catch) + P(Cavity, toothache, \neg catch)] \\
&= \alpha\, [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] \\
&= \alpha\, \langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle
\end{aligned}$$

Independence

- Random variables X and Y said *independent* iff:

$$P(X|Y) = P(X)$$

$$(or \quad P(Y|X) = P(Y) \quad or \quad P(X,Y) = P(X)P(Y))$$

Examples:

- $P(Cavity|Weather) = P(Cavity)$
- $P(Wheather|Cavity) = P(Wheather)$
- $P(Toothache, Catch, Cavity, Weather) =$
  $P(Toothache, Catch, Cavity)P(Weather)$

Independence

Independence simplifies representation of probability distributions

- $P(Toothache, Catch, Cavity, Weather)$ has 32 entries
- $P(Toothache, Catch, Cavity)$ and $P(Weather)$: $8 + 4 = 12$ entries

Example:

- n independent biased coins, reduced size from 2^n to n

Remarks:

- Absolute independence desirable but rare
- Complex systems have hundreds of variables, very few or none of which independent

Conditional independence

1. Given one has cavity, prob of its detection independent of toothache:
   $P(catch|cavity, toothache) = P(catch|cavity)$
2. Same independence holds without cavity:
   $P(catch|\neg cavity, toothache) = P(catch|\neg cavity)$
3. *Catch* is *conditionally independent* of *Toothache* **given** *Cavity*:
   $P(Catch|Toothache, Cavity) = P(Catch|Cavity)$
4. Equivalent statements:
   $P(Toothache|Catch, Cavity) = P(Toothache|Cavity)$
   $P(Toothache, Catch|Cavity) = P(Toothache|Cavity)P(Catch|Cavity)$

Conditional independence

General formulation:

- X conditionally independent of Y given Z: $P(X|Y,Z) = P(X|Z)$
- $P(X,Y|Z) = \frac{P(X,Y,Z)}{P(Z)} = \frac{P(X|Y,Z)P(Y,Z)}{P(Z)} = P(X|Z)P(Y|Z)$
- If Y_i conditionally independent of Y_j ($i \neq j$) given Z:
  $P(Y_1,\ldots,Y_n|Z) =$
  $P(Y_1|Y_2,\ldots Y_n,Z)P(Y_2|Y_3\ldots Y_n,Z)\cdots P(Y_n|Z) =$
  $P(Y_1|Z)P(Y_2|Z)\cdots P(Y_n|Z)$

$$P(Y_1,\ldots,Y_n|Z) = P(Y_1|Z)\cdots P(Y_n|Z)$$

Conditional independence

Chain rule + Conditional independence

$$\begin{aligned} P(X, Y, Z) &= P(X|Y, Z)P(Y, Z) = P(X|Y, Z)P(Y|Z)P(Z) \\ &= P(X|Z)\,P(Y|Z)\,P(Z) \end{aligned}$$

$P(Toothache, Catch, Cavity)$
$= P(Toothache|Catch, Cavity)P(Catch, Cavity)$
$= P(Toothache|Catch, Cavity)P(Catch|Cavity)P(Cavity)$
$= P(Toothache|Cavity)P(Catch|Cavity)P(Cavity)$
$2 + 2 + 1 = 5$ independent numbers (instead of $2^3 - 1$)

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n.

Bayes' Rule

- Product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$$\Rightarrow \textit{Bayes' rule } P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

or in distribution form

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} = \alpha P(X|Y)P(Y)$$

Useful for assessing *diagnostic* probability from *causal* probability:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

Bayes' Rule and conditional independence

- Bayes rule: $P(Z|Y_1, \ldots, Y_n) = \alpha P(Y_1, \ldots, Y_n|Z)\, P(Z)$
- if $Y_i, \ldots Y_n$ conditionally independent of each other given Z:

$$P(Z|Y_1, \ldots, Y_n) = \alpha P(Y_1|Z) \cdots P(Y_n|Z)\, P(Z)$$

- Effects conditionally independent of each other given a cause

$$P(Cause|Effect_1, \ldots, Effect_n) = \alpha P(Cause) \prod_i P(Effect_i|Cause)$$

Total number of parameters is *linear* in n

Exercise: Bayes' Rule

An online shop has collected the following data:

- Customers are:
  25% young men (class *ym*); 45% young women (*yw*); 30% other (*o*).
- Young men buy: Shoes 30%; Trousers 50%; Shirts 20%.
- Young women buy: Shoes 50%; Trousers 30%; Shirts 20%.
- Other customers buy: Shoes 30%; Trousers 30%; Shirts 40%.

1. If you receive an order for trousers, which is the most probable class the customer who issued the order belongs to? Why?
2. Which is, and how do you compute, the probability that a random order is for trousers?