Intelligent Systems

Escuela Técnica Superior de Informática

Universitat Politècnica de València

Block 2 Chapter 1
Probabilistic Reasoning

DSIC – UPV: October, 2016

Index

Index

Uncertainty

Let A_t be the consequent (action) of a rule:

$A_t =$ LEAVING HOME FOR THE AIRPORT t MINUTES BEFORE THE FLIGHT DEPARTURE

Let $t = 25$. Key question: Will I get on time with A_{25}?

Problems:

- partial observability (road conditions, other drivers' plans, etc.)
- imprecise information on the traffic conditions
- other uncertainties (unexpected conditions) such that the car doesn't break or runs out of gas
- huge complexity to model and predict the traffic conditions

Rule-based reasoning (logic) presents two limitations:

- possible falsity: (will I get on time with A_{25} in all cases?)
- conclusions need to consider a lot of factors: (I'll be on time with A_{25} if there is no car accident AND it does not rain AND I don't get a flat tyre AND ...).

Other plans, such as A_{1440}, might increase the belief that that I will get to the airport on time, but also increase the likelihood of a very long wait!

Approaches to uncertainty

Historically, many approaches to uncertainty:

- non-monotonic logic [1]
- certainty factors in RBS (e.g.: MYCIN expert system [2])
- fuzzy logic (*fuzzy sets*) [3]
- methods based on the probability theory [4,5]

In 1931, Finetti proved the following statement [5, p. 489-490]:

> If an agent [broker] A expresses a set of degrees of belief [makes investments] that violate the axioms of probability theory then there is a combination of bets by another agent B that guarantees that A will lose [money] every time.

Currently, *probabilistic methods* prevail as the general framework to represent uncertainty. These methods enable to:

- adequately and consistently model and combine:
  - the *inaccuracy or vagueness of a priori knowledge*
  - the *imprecision of facts, observations or data*
- *Automated learning* of the representation models

Index

Sample space and probability space

Let Ω be a set named *sample space* (set of all possible worlds)

Example: the 6 possible outcomes (worlds) when we roll a dice, $\Omega = \{t \in \mathbb{N} : 1 \leq t \leq 6\}$

The possible worlds of Ω are **mutually exclusive** and **exhaustive**.

One item $\omega \in \Omega$ is called *simple event*, *world*, or simply *sample*.

Probability model or probabilistic space is a sample space along with a function $P : \Omega \to \mathbb{R}$ that assigns a real number to each $\omega \in \Omega$ such that:

$$0 \leq P(\omega) \leq 1; \qquad \sum_{\omega} P(\omega) = 1$$

Example: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$,

$$\sum_{t=1}^{6} P(t) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

Events, random variables and probability distribution

An *event* $\mathcal{A}$ is a subset of possible worlds of Ω; its probability is:

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

Example: $P(1 < t < 4) \; = \; P(2) + P(3) \; = \; 1/6 + 1/6 \; = \; 1/3$

A *random variable* is a function that maps the sample space to a domain; for instance the boolean domain ($\mathbb{B}$) (boolean random variable).

Example: the function *odd* (O). $\;O : \Omega \rightarrow \mathbb{B};\;\; O(5) = \mathbf{true},\;\; O(2) = \mathbf{false}.$

If X is a random variable, "$(X = x)$" denotes the event:

$$(X = x) \;\equiv\; \{\omega \in \Omega : X(\omega) \; = \; x\}$$

Given a random variable X, P induces a *probability distribution*:

$$P(X = x) \;\stackrel{\text{def}}{=}\; \sum_{\omega \in (X = x)} P(\omega)$$

Example: $\;P(O = \mathbf{true}) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$

Events, random variables and propositions

A (logic) proposition is interpreted as an event (subset of possible worlds) in which the proposition is true.

Given two boolean random variables A and B:

$$\begin{aligned}
\text{event} \quad a \quad &\equiv \quad \{\omega \in \Omega : A(\omega) = \mathbf{true}\} \\
\text{event} \quad \neg a \quad &\equiv \quad \{\omega \in \Omega : A(\omega) = \mathbf{false}\} \\
\text{event} \quad \neg a \wedge b \quad &\equiv \quad \{\omega \in \Omega : A(\omega) = \mathbf{false} \wedge B(\omega) = \mathbf{true}\}
\end{aligned}$$

When using boolean variables, the set of possible worlds are just those worlds in which the proposition holds (propositional logic). Example: $A = \mathbf{true}$, $B = \mathbf{false}$, $a \wedge \neg b$, $\ldots$

Simplification of the notation (whenever the semantics is clear):

$$P(A = \mathbf{true}) \;\rightarrow\; P(a), \qquad P(A = \mathbf{false}) \;\rightarrow\; P(\neg a),$$

$$P(X = x) \;\rightarrow\; P(x)$$

Probability axioms

Various axiomatic formulation to the Probability theory have been proposed. For example, the Kolmogorov's axioms:

$$0 \leq P(\omega) \leq 1 \tag{1}$$

$$\sum_{\omega \in \Omega} P(\omega) = 1 \tag{2}$$

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b) \tag{3}$$

We can build up the rest of probability theory from this simple foundation:

Exercise: prove that $P(\neg a) = 1 - P(a)$

As commented above, any agent whose set of degrees of belief (rating system) violates the axioms (1-3) will fall into contradictions obtaining undesirable practical results.

Unconditional, joint and conditional probability

Unconditional or prior probability of a random variable X:

$$P(X = x) \;\equiv\; P(x): \qquad \sum_x P(x) = 1$$

Joint probability of two random variables X, Y:

$$P(X = x; Y = y) \;\equiv\; P(x, y): \qquad \sum_x \sum_y P(x, y) = 1$$

Conditional probability:

$$P(X = x \mid Y = y) \;\equiv\; P(x \mid y): \qquad \sum_x P(x \mid y) = 1 \quad \forall y$$

Random variable and unconditional probability: examples

Sample space: road trips (Ω). Elements to consider:

- *Weather* (W): *clear* (CLE), *cloudy* (CLO), *rainy* (RAI)
- *Daylight* (D): *day* (DAY), *night* (NIG)
- *Safety* (S): *safe trip* (SAF), *accident* (ACC)

Random variable and unconditional probability: examples

Sample space: road trips (Ω). Elements to consider:

- *Weather* (W): *clear* (CLE), *cloudy* (CLO), *rainy* (RAI)
- *Daylight* (D): *day* (DAY), *night* (NIG)
- *Safety* (S): *safe trip* (SAF), *accident* (ACC)

Random variables:

$W : \Omega \to \{\text{CLE,CLO,RAI}\}, \quad D : \Omega \to \{\text{DAY,NIG}\}, \quad S : \Omega \to \{\text{SAF,ACC}\}.$

Example: $(D = \text{DAY}) \equiv \{\omega \in \Omega : D(\omega) = \text{DAY}\} \to$ daytime trips . . .

Random variable and unconditional probability: examples

Sample space: road trips (Ω). Elements to consider:

- *Weather* (W): *clear* (CLE), *cloudy* (CLO), *rainy* (RAI)
- *Daylight* (D): *day* (DAY), *night* (NIG)
- *Safety* (S): *safe trip* (SAF), *accident* (ACC)

Random variables:

$W : \Omega \rightarrow \{\text{CLE,CLO,RAI}\}, \quad D : \Omega \rightarrow \{\text{DAY,NIG}\}, \quad S : \Omega \rightarrow \{\text{SAF,ACC}\}.$

Example: $(D = \text{DAY}) \equiv \{\omega \in \Omega : D(\omega) = \text{DAY}\} \rightarrow$ daytime trips . . .

Unconditional (prior) probabilities.

Examples: $P(D{=}\text{DAY}) = 0.62$, $P(D{=}\text{NIG}) = 0.38$

Random variable and unconditional probability: examples

Sample space: road trips (Ω). Elements to consider:

- *Weather* (W): *clear* (CLE), *cloudy* (CLO), *rainy* (RAI)
- *Daylight* (D): *day* (DAY), *night* (NIG)
- *Safety* (S): *safe trip* (SAF), *accident* (ACC)

Random variables:

$W:\Omega\rightarrow\{\text{CLE,CLO,RAI}\},\quad D:\Omega\rightarrow\{\text{DAY,NIG}\},\quad S:\Omega\rightarrow\{\text{SAF,ACC}\}.$

Example: $(D=\text{DAY})\equiv\{\omega\in\Omega : D(\omega)=\text{DAY}\}\rightarrow$ daytime trips ...

Unconditional (prior) probabilities.

Examples: $P(D{=}\text{DAY})=0.62$, $P(D{=}\text{NIG})=0.38$

More examples:

w	CLE	CLO	RAI	$\sum$
$P(W=w)$	0.46	0.33	0.21	1.00

D	DAY	NIG	$\sum$
$P(D=d)$	0.62	0.38	1.00

s	SAF	ACC	$\sum$
$P(S=s)$	0.86	0.14	1.00

Joint probability: examples

Joint probabilities. Example:

Probability of a trip under the rain and at night $\rightarrow P(W = \textsf{RAI},\ D = \textsf{NIG}) = 0.11$

Probability of a trip on a clear day and with no accident $\rightarrow P(W = \textsf{CLE}, S = \textsf{SAF}) = 0.43$

Joint probability: examples

Joint probabilities. Example:

Probability of a trip under the rain and at night $\rightarrow P(W = \text{RAI},\ D = \text{NIG}) = 0.11$

Probability of a trip on a clear day and with no accident $\rightarrow P(W = \text{CLE}, S = \text{SAF}) = 0.43$

More examples*:

$P(w,s)$	CLE	CLO	RAI
SAF	0.43	0.30	0.13
ACC	0.03	0.03	0.08

$\sum = 1$

$P(d,s)$	DAY	NIG
SAF	0.57	0.29
ACC	0.05	0.09

$\sum = 1$

$P(w,d)$	CLE	CLO	RAI
DAY	0.31	0.21	0.10
NIG	0.15	0.12	0.11

$\sum = 1$

	DAY			NIG		
$P(s,w,d)$	CLE	CLO	RAI	CLE	CLO	RAI
SAF	0.30	0.20	0.07	0.13	0.10	0.06
ACC	0.01	0.01	0.03	0.02	0.02	0.05

$\sum = 1$

* Probabilities are invented but they are not arbitrary. The values on these tables and the ones in the previous and next page are related. See exercise in page 18.

Conditional probability: examples

Probability of accident *given* a night trip: $P(S = \text{ACC} \mid D = \text{NIG}) = 0.24$
Probability of *no* accident *given* a night trip: $P(S = \text{SAF} \mid D = \text{NIG}) = 0.76$

Probability of daytime trip *given* a rainy day: $P(D = \text{DAY} \mid W = \text{RAI}) = 0.48$
Probability of a night trip *given* a rainy day: $P(D = \text{NIG} \mid W = \text{RAI}) = 0.52$

Conditional probability: examples

Probability of accident *given* a night trip: $P(S = \text{ACC} \mid D = \text{NIG}) = 0.24$
Probability of *no* accident *given* a night trip: $P(S = \text{SAF} \mid D = \text{NIG}) = 0.76$

Probability of daytime trip *given* a rainy day: $P(D = \text{DAY} \mid W = \text{RAI}) = 0.48$
Probability of a night trip *given* a rainy day: $P(D = \text{NIG} \mid W = \text{RAI}) = 0.52$

More examples:

$P(s \mid w)$	CLE	CLO	RAI
SAF	0.93	0.91	0.62
ACC	0.07	0.09	0.38
$\sum$	1.00	1.00	1.00

$P(s \mid d)$	DAY	NIG
SAF	0.92	0.76
ACC	0.08	0.24
$\sum$	1.00	1.00

$P(d \mid w)$	CLE	CLO	RAI
DAY	0.67	0.64	0.48
NIG	0.33	0.36	0.52
$\sum$	1.00	1.00	1.00

Conditional probability: examples

Probability of accident *given* a night trip: $P(S = \text{ACC} \mid D = \text{NIG}) = 0.24$
Probability of *no* accident *given* a night trip: $P(S = \text{SAF} \mid D = \text{NIG}) = 0.76$

Probability of daytime trip *given* a rainy day: $P(D = \text{DAY} \mid W = \text{RAI}) = 0.48$
Probability of a night trip *given* a rainy day: $P(D = \text{NIG} \mid W = \text{RAI}) = 0.52$

More examples:

$P(s \mid w)$	CLE	CLO	RAI
SAF	0.93	0.91	0.62
ACC	0.07	0.09	0.38
$\sum$	1.00	1.00	1.00

$P(s \mid d)$	DAY	NIG
SAF	0.92	0.76
ACC	0.08	0.24
$\sum$	1.00	1.00

$P(d \mid w)$	CLE	CLO	RAI
DAY	0.67	0.64	0.48
NIG	0.33	0.36	0.52
$\sum$	1.00	1.00	1.00

More examples:

$P(w \mid s)$	CLE	CLO	RAI	$\sum$
SAF	0.50	0.35	0.15	1.0
ACC	0.21	0.21	0.58	1.0

$P(d \mid s)$	DAY	NIG	$\sum$
SAF	0.66	0.34	1.0
ACC	0.36	0.64	1.0

$P(w \mid d)$	CLE	CLO	RAI	$\sum$
DAY	0.50	0.34	0.16	1.0
NIG	0.39	0.32	0.29	1.0

Continuous random variables: probability density function

If X is a random variable in $\mathbb{R}$, then $P(X = x) \equiv 0 \ \forall x \in \mathbb{R}$

Probability density function: $\quad p(x) \stackrel{\text{def}}{=} \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x}$

In general, $p(x) \in [0, \infty[$, although: $\quad \int_{-\infty}^{+\infty} p(x)\, dx \; = \; 1$

All of the previous formulation for discrete variables applies to continuous variables replacing $\sum$ by $\int$.

Example: for the *joint probability* $P(x, y)$:

if Y is continuous: $\sum_x \int_y p(x, y)\, dy = 1; \quad$ if X, Y continuous: $\int_x \int_y p(x, y)\, dx\, dy = 1$

Example: For the *conditional probability* $P(x \mid y)$, if X is continuous:

$$\int_x p(x \mid y)\, dx \; = \; 1 \quad \forall y$$

Index

Marginal, chain rule and Bayes' rule

The unconditional (or marginal) probability $P(x)$ is the *marginalization* of the joint probability $P(x, y)$:

$$P(x) \;=\; \sum_y P(x, y), \quad \text{equivalently} \;\; P(y) \;=\; \sum_x P(x, y)$$

The joint probability is related to the conditional and unconditional probabilities (*product rule*):

$$P(x, y) \;=\; P(x)\, P(y \mid x) \;=\; P(y)\, P(x \mid y)$$

Chain rule:

$$P(x_1, x_2, \ldots, x_n) \;=\; P(x_1) \prod_{i=2}^{n} P(x_i \mid x_1, \ldots, x_{i-1})$$

Bayes' rule:

$$P(x \mid y) \;=\; \frac{P(y \mid x) P(x)}{P(y)} \;=\; \frac{P(x, y)}{P(y)} \;=\; \frac{P(x)\, P(y \mid x)}{\sum_{x'} P(x')\, P(y \mid x')}$$

Inference: examples

In the example of page 11, we can infer the unconditional probabilities through marginalization of the joint probabilities in page 12. For example:

$$P(W = \text{CLE}) = \sum_{s \in \{\text{SAF,ACC}\}} P(W = \text{CLE},\, S = s) = \sum_{d \in \{\text{DAY,NIG}\}} P(W = \text{CLE},\, D = d) = 0.46$$

$$P(S = \text{ACC}) = \sum_{w \in \{\text{CLE,CLO,RAI}\}} P(W = w,\, S = \text{ACC}) = \sum_{d \in \{\text{DAY,NIG}\}} P(D = d,\, S = \text{ACC}) = 0.14$$

We can infer the conditional probabilities of page 13 through the application of the Bayes' rule. For example:

$$P(S = \text{ACC} \mid D = \text{NIG}) = \frac{P(S = \text{ACC}, D = \text{NIG})}{P(D = \text{NIG})} = 0.24$$

$$P(D = \text{DAY} \mid W = \text{RAI}) = \frac{P(W = \text{RAI}, D = \text{DAY})}{P(W = \text{RAI})} = 0.48$$

Exercise

The last table in page 12 contains the values of the joint probability $P(s, w, d)$ of the example in page 11.

Calculate:

- $P(S = \text{ACC})$, $P(w)\ \forall w \in \{\text{CLE,CLO,RAI}\}$
- $P(w, s)\ \forall w \in \{\text{CLE,CLO,RAI}\},\ s \in \{\text{SAF,ACC}\}$;
- $P(s \mid w)\ \forall s \in \{\text{SAF,ACC}\},\ w = \text{CLO}$

Check that the results are the same as the ones shown in pages 11, 12 and 13.

Calculate $\forall s \in \{\text{SAF,ACC}\}$:

- $P(S = s \mid W = \text{RAI}, L = \text{NIG})$
- $P(S = s \mid W = \text{CLE}, L = \text{DAY})$

Index

Uncertainty and optimal decisions: Decision theory

In the 'airport' example of page 3, let's assume that:

$$
\begin{aligned}
P(A_{25} \text{ WILL ALLOW ME TO GET ON TIME} \mid \ldots) &= 0.04 \\
P(A_{90} \text{ WILL ALLOW ME TO GET ON TIME} \mid \ldots) &= 0.70 \\
P(A_{120} \text{ WILL ALLOW ME TO GET ON TIME} \mid \ldots) &= 0.95 \\
P(A_{1440} \text{ WILL ALLOW ME TO GET ON TIME} \mid \ldots) &= 0.999
\end{aligned}
$$

Which actions do we choose?

It depends on our *preferences* on the possibility of missing the flight over enjoying the airport shops or a nice restaurant at the airport, etc.

The *Utility theory* can be used to represent and infer preferences or the cost of the undesirable effects of the decisions

Probability theory + Utility theory

=

Statistical decision theory

Decision theory: minimizing the error risk

Simplification: decisions can only be *"right"* o *"wrong"* and the costs are 0 and 1, respectively.

Let $x \in \mathcal{X}$ be a *fact* or *data* and let $d \in \mathcal{D}$ be a *decision* for x.

Probability of error if we take decision d:

$$P_d(\text{error} \mid x) = 1 - P(d \mid x)$$

Minimum probability of error:

$$\forall x \in \mathcal{X} : \quad P_\star(\text{error} \mid x) = \min_{d \in \mathcal{D}} P_d(\text{error} \mid x) \; = \; 1 - \max_{d \in \mathcal{D}} P(d \mid x)$$

That is, for each x, the minimum probability of error is obtained if we take the decision with the highest (maximum) conditional (posterior) probability.

Minimum average probability of error:

$$P_\star(\text{error}) = \sum_{x \in \mathcal{X}} P_\star(\text{error} \mid x)\, P(x)$$

Bayes decision rule for minimizing the probability of error (error risk):

$$\forall x \in \mathcal{X} : \quad d^\star(x) = \operatorname*{argmax}_{d \in \mathcal{D}} P(d \mid x)$$

Exercise (to do in class)

A classical decision problem is to classify *Iris* flowers into three classes: *setosa, versicolor* and *virgínica*, on the basis of their petal and sepal sizes (x).

Using the histograms of petal surface of a sample of 50 flowers of each class, and normalizing the values, we get the following estimate of distribution of the petal size for each class (c):

	petal sizes in cm^2											
$P(x \mid c)$	$<$1	1	2	3	4	5	6	7	8	9	10	$>$10
SETO	0.90	0.10	0	0	0	0	0	0	0	0	0	0
VERS	0	0	0	0.20	0.30	0.32	0.12	0.06	0	0	0	0
VIRG	0	0	0	0	0	0	0.08	0.12	0.24	0.14	0.20	0.22

Assuming that the three classes have the same probability, calculate:

a) The conditional (posterior) probabilities $P(c \mid x)$, $c \in \{$SETO,VERS,VIRG$\}$, for a flower whose petal size is $x =$7 cm^2

b) The decision of optimal classification for this flower and the probability of taking a wrong decision.

c) The best decision and probability of error for petals $1, 2, \ldots, 10$ cm^2

d) The minimum probability of error for any iris flower; that is, $P_\star(\text{error})$

e) Repeat the same calculations assuming that the prior probabilities are:
$P(\text{SETO}) = 0.3,\ P(\text{VERS}) = 0.5,\ P(\text{VIRG}) = 0.2$

Algunas soluciones: a) 0.0, 0.33, 0.67; b) VIRG, 0.33; d) 0.05 (5 %) e.a) 0.0, 0.55, 0.44; e.b) VERS, 0.44; e.d) 0.04 (4 %)

Index

Bibliography

[1] A.N. Abdallah. The Logic of Partial Information. Springer Verlag, 1995.

[2] B.G. Buchanan, E.H. Shortliffe (editires): Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison Wesley, 1984. (También en http://aitopics.net/RuleBasedExpertSystems).

[3] J.F. Baldwin. Fuzzy sets and expert systems. Wiley, 1985.

[4] R.O. Duda, D.G. Stork, P.E. Hart. Pattern Classification. Wiley, 2001.

[5] S. Russell, P. Norvig. Artificial Intelligence: A Modern Approach. Pearson, third edition, 2010.

The material of this chapter is basically taken from [5].