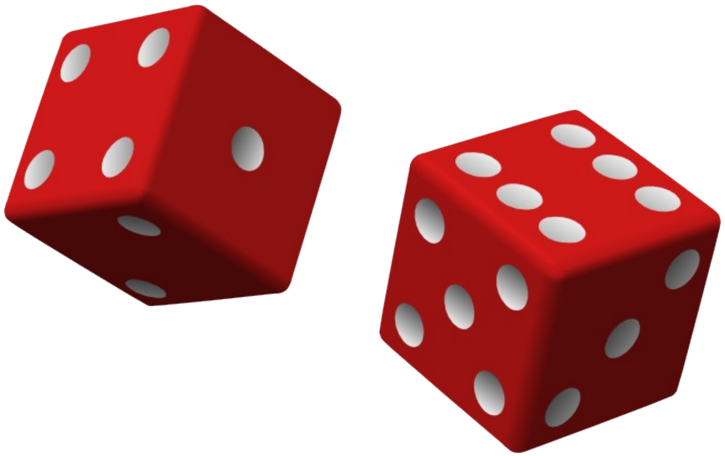
Artificial Intelligence

Lecture 1 supplement:
What you should know about probability



IN4010 Artificial Intelligence Techniques

Prof. dr. C.M. Jonker, dr. F. Oliehoek, ir. Rolf Starre

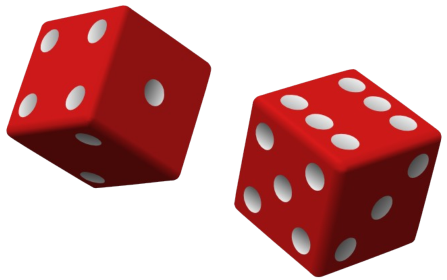
4-9-2019

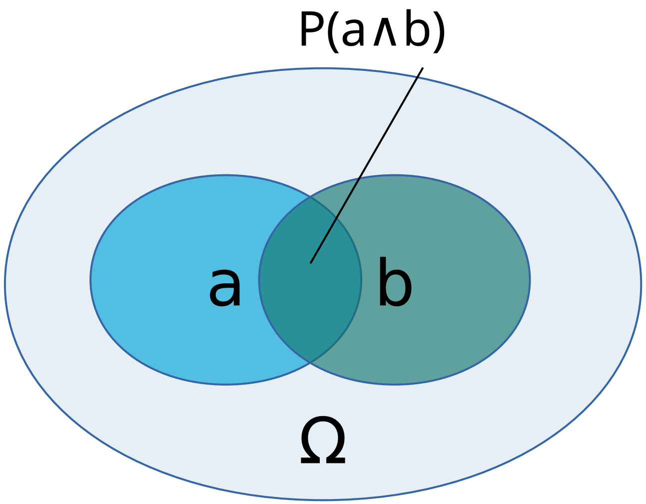
Interactive Intelligence group
Department of Electrical Engineering, Mathematics and Computer Science

Delft University of Technology

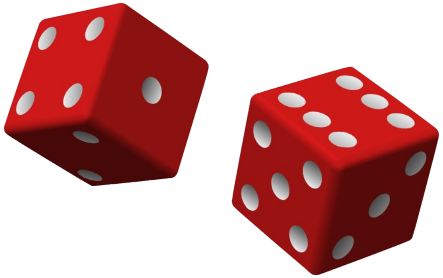
Short answer:
All of Russel&Norvig(v3) Chap 13

- events, random variables
- joint probability
- inference by enumeration
- independence
- conditional independence
- Bayes Rule



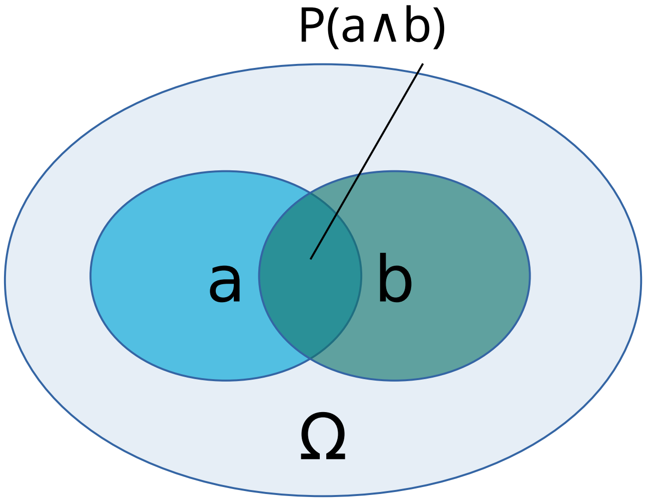


Short answer:
All of Russel&Norvig(v3) Chap 13



- So, for example:
  - sample space Ω={(1,1), (1,2), … }
  - Σ_ω P(ω)=1
  - events: **subsets** of sample space
    - "dice-1-odd", "at least one odd", or "both even"
    - P("dice-1-odd" ∧ "dice-2-even")
  - inclusion-exclusion:

  $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



- RN notation:
  - (boolean) random variable: *Even*
  - values: {*true,false*} also written {*even, ¬even*}

More notation

- Prior probability
  - Weather can be {*sunny,rain,cloudy,snow*}
  - probability: $P(Weather{=}sunny) = 0.2$
  - prob. distribution: $\mathbf{P}(Weather) = [0.2, 0.3, 0.45, 0.05]$

- Joint probability

'product rule'

scalar
  - $P(toothache, cavity) = P(toothache \mid cavity)\ P(cavity)$

table
  - $\mathbf{P}(Toothache, Cavity) = \mathbf{P}(Toothache \mid Cavity)\ \mathbf{P}(Cavity)$

TU Delft

More notation

- Prior probability
  - Weather can be {*sunny,rain,cloudy,snow*}
  - probability: *P(Weather=sunny)* = 0.2
  - prob. distribution: ***P****(Weather)* = [0.2, 0.3, 0.45, 0.05]

- Joint probability
  - scalar — *P(toothache, cavity) = P(toothache | cavity) P(cavity)*
  - table — ***P****(Toothache, Cavity)* = ***P****(Toothache | Cavity)* ***P****(Cavity)*

'product rule'

Conditional Probability

$$P(a|b)=\frac{P(a\wedge b)}{P(b)}$$

Marginal Probability

$$P(b)=\sum_a P(a\wedge b)$$

Joint Probability Tables

- E.g.:

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

- Can be use to compute all kinds of probabilities

Inference by enumeration

Start with the joint distribution:

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

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$$

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

TU Delft

Normalization

- Compute: $P(Cavity|toothache) = \dfrac{P(Cavity, toothache)}{P(toothache)}$

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

Denominator can be viewed as a normalization constant α

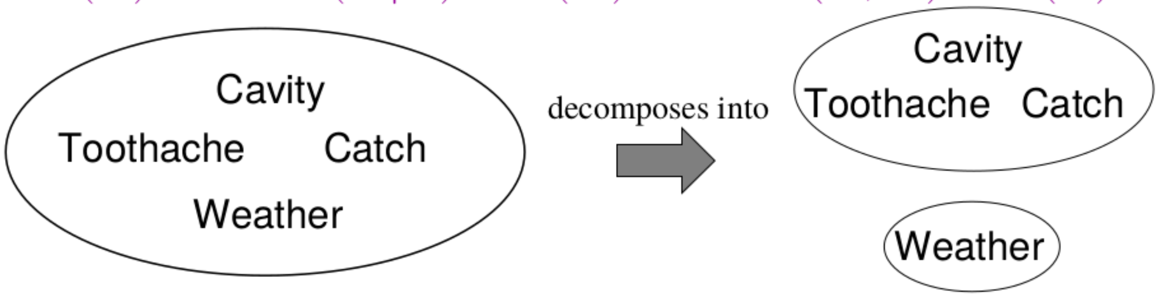
$$\begin{aligned}\mathbf{P}(Cavity|toothache) &= \alpha\,\mathbf{P}(Cavity, toothache)\\ &= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]\\ &= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]\\ &= \alpha\,\langle 0.12, 0.08\rangle = \langle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

General idea: compute distribution on query variable
by fixing evidence variables and summing over hidden variables

TU Delft

Independence

A and B are independent iff
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$

Absolute independence powerful but rare

Dentistry is a large field with hundreds of variables,
none of which are independent. What to do?

Conditional Independence

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

The same independence holds if I haven't got a cavity:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

$Catch$ is conditionally independent of $Toothache$ given $Cavity$:

$\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$

Again: why probability?

De Finetti's argument

1) (Non-negativity) $P(A) \geq 0$, for all $A \in F$.
2) (Normalization) $P(\Omega) = 1$.
3) (Finite additivity) $P(A \vee B) = P(A) + P(B)$ for all A, $B \in F$ such that $A \cap B = \emptyset$.

Bruno de Finetti:

If agent's beliefs violate the axioms of probability, then there exists a combination of bets against it which it is willing to accept that guarantees it will lose money, every time.

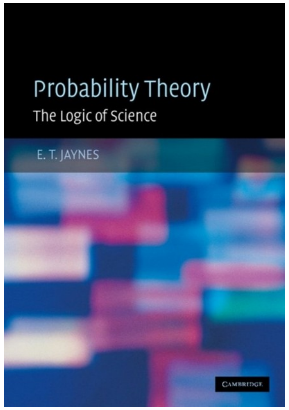
Example

u(6:4 against a) = -6 * bel(a) + 4 * bel(¬a)
= -6 * 0.4 + 4 * 0.6 = 0

A Dutch book:

prop.	belief	taken bet	a,b	a,¬b	¬a,b	¬a,¬b
a	0.4	6:4 against a	-6	-6	4	4
b	0.3	7:3 against b	-7	3	-7	3
a∨b	0.8	8:2 on a∨b	2	2	2	-8
			-11	**-1**	**-1**	**-1**

Cox' Theorem
(See Jaynes 2003)



- Desiderata:
  - Degrees of plausibility: represented by **real numbers**
  - **Qualitative** correspondence how humans reason
  - **Consistency**: If a conclusion can reached in more ways, then every possible way must lead to the same result

→ Need to use probability to represent plausibility

- "Probability theory is nothing but common sense reduced to calculation." — Laplace, 1819

Bayes rule

“One rule to rule them all”

If you are going to remember just one thing…

- …remember Bayes' rule:

  $$P(A|B) = P(B|A)P(A) \,/\, P(B)$$

directly from product rule:
⇔ $P(A|B)\,P(B) = P(B|A)P(A)$
⇔ $P(A,B) = P(B,A)$

Generalized form given background evidence ***e***:

$$P(A|B,\mathbf{e}) = P(B|A,\mathbf{e})P(A|\mathbf{e}) \,/\, P(B|\mathbf{e})$$

- Important why..?

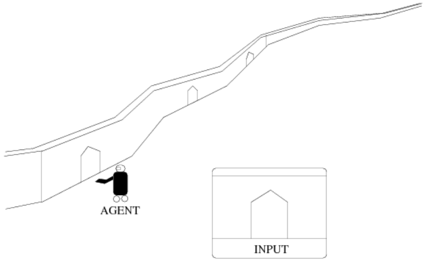
  (What is its importance for an intelligent agent?)

It allows to update a belief

- Rewrite with State and a particular observation *o*:

  $\boldsymbol{P}(State|o) = \boldsymbol{P}(o|State)\boldsymbol{P}(State) / P(o)$

  - ***P***_(State)_ is our prior belief
  - so we can update our belief, based on observations!



Many observations…

- How to deal with **many observations**?

  $\boldsymbol{P}(State|o_1,o_2,o_3,\ldots)=\alpha\, \boldsymbol{P}(o_1,o_2,o_3,\ldots|State)\boldsymbol{P}(State)$

  - we don't know the observation sequence in advance…
  - representing $\boldsymbol{P}(o_1,o_2,o_3,\ldots|State)$ with a table does not scale…

- Solution:
  - conditional independence can help significantly

Many observations…

- How to deal

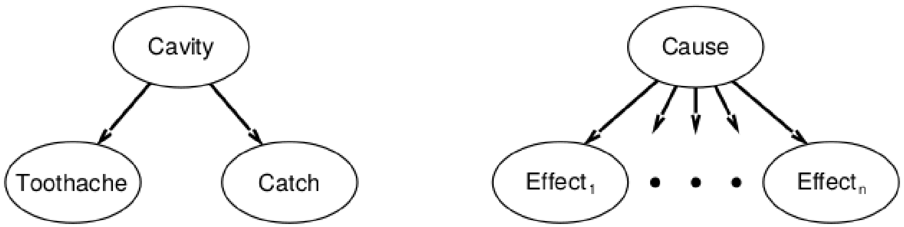
$\boldsymbol{P}(State\,|\,o_1,o_2,o_3,..$

  - we don’t kn
  - representin scale…

$$\mathbf{P}(Cavity|toothache \wedge catch)$$
$$= \alpha\,\mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity)$$
$$= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)$$

This is an example of a naive Bayes model:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i|Cause)$$



Total number of parameters is **linear** in n

- Solution:
  - conditional independence can help significantly

It allows to *maintain* a belief

- Given conditional independent observations
  P$(o_1,o_2|State)$ = ***P***$(o_1|State)$***P***$(o_2|State)$

- ...we can also sequentially update:

  P'*(State) :=* ***P****(o₁|State)*

  $P'(State) := P(o_1|State)P(State) / P(o_1)$
  $P''(State) := P(o_2|State)P'(State) / P(o_2)$

  - then $P''(State) = P(o_1,o_2|State)$
  - (Exercise!)

- We will see later how
  to incorporate robot movement over time