CSC242: Intro to AI

Lecture 14 Probability


优酷
CANADA BILLIARD
CANADA BILLIARD

Causal rule: Cavity $\Rightarrow$ Toothache

Causal rule: Cavity ⇒ Toothache

Diagnostic rule: Toothache ⇒ Cavity

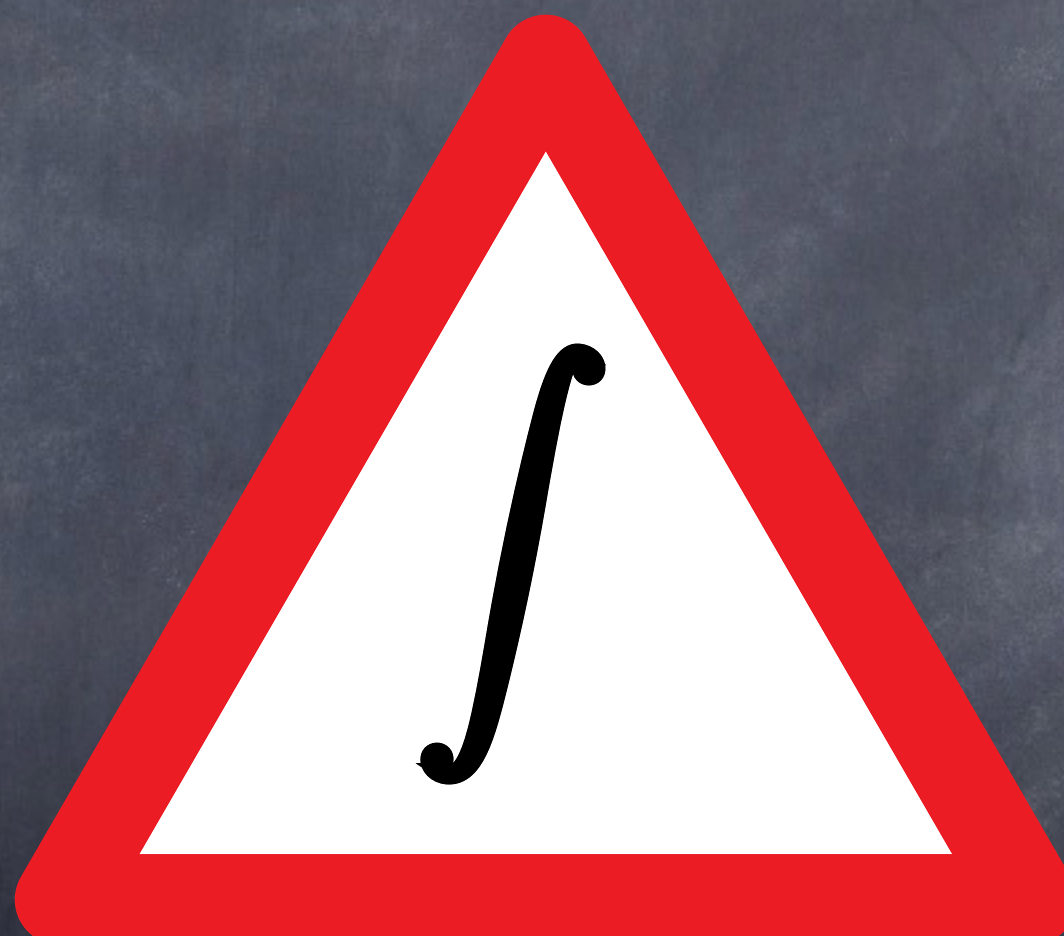
Causal rule: Cavity ⇒ Toothache

Diagnostic rule: ~~Toothache ⇒ Cavity~~

Toothache ⇒ Cavity ∨ GumProblem ∨ Abcess ∨ ...

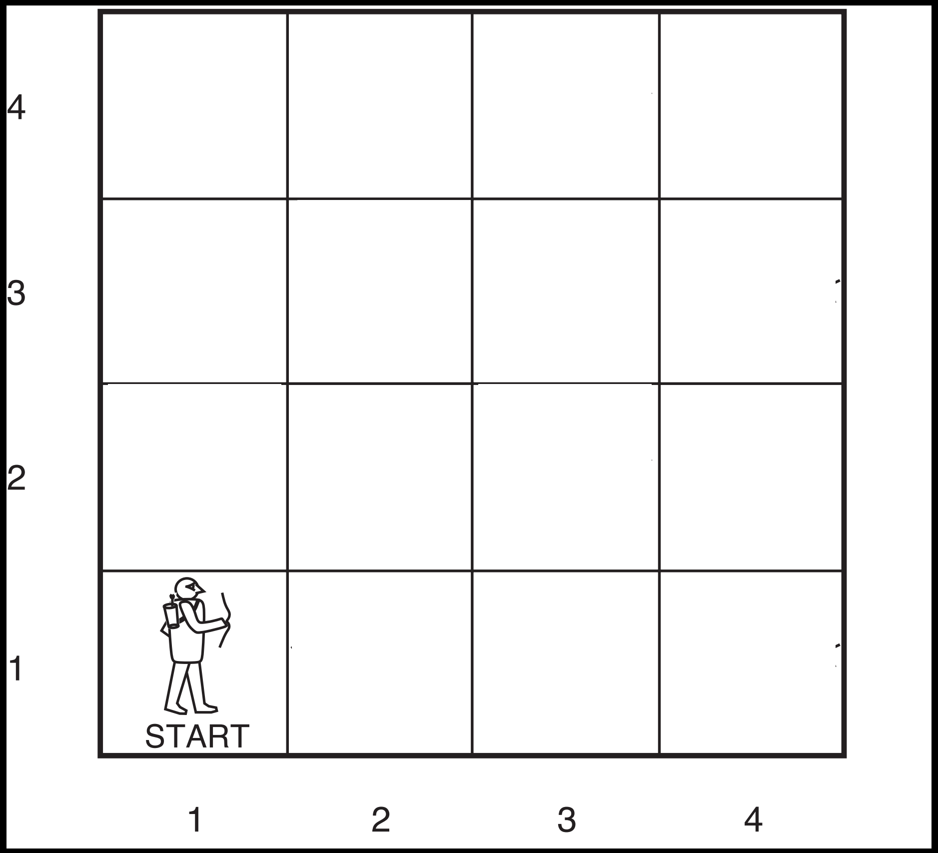
Probability

WARNING!



MATH AHEAD

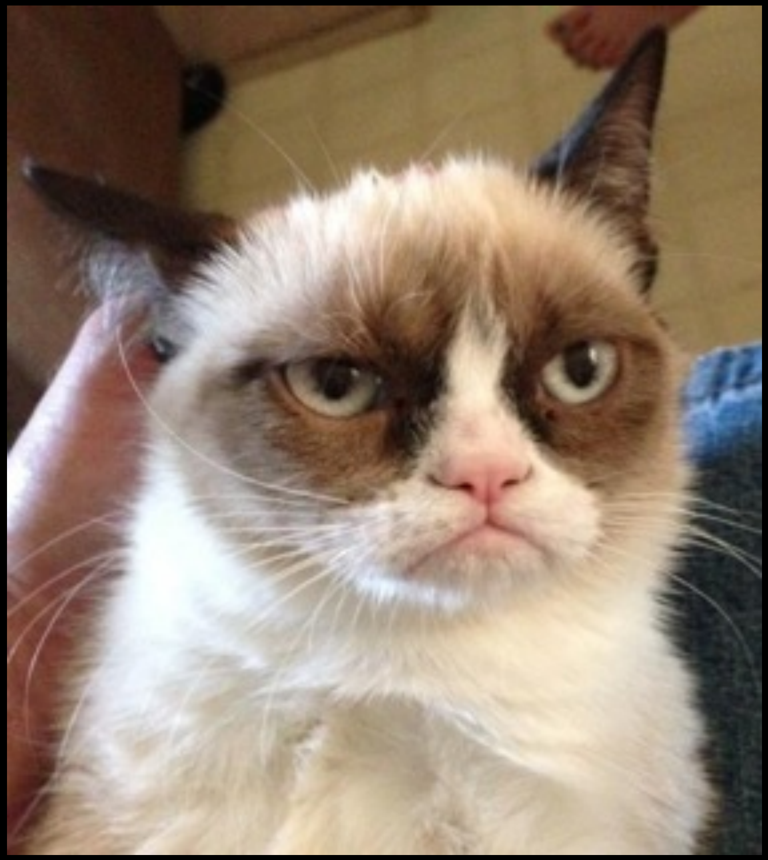
Sources of Uncertainty



Partial Observability

Nondeterminism

Possible Worlds



Hungry=true,
Cranky=false

Hungry=false,
Cranky=true

Hungry=true,
Cranky=true

Hungry=false,
Cranky=false

Possible Worlds

Hungry ∨ Cranky

Hungry=true, Cranky=false

Hungry=false, Cranky=true

Hungry=true, Cranky=true

Hungry=false, Cranky=false

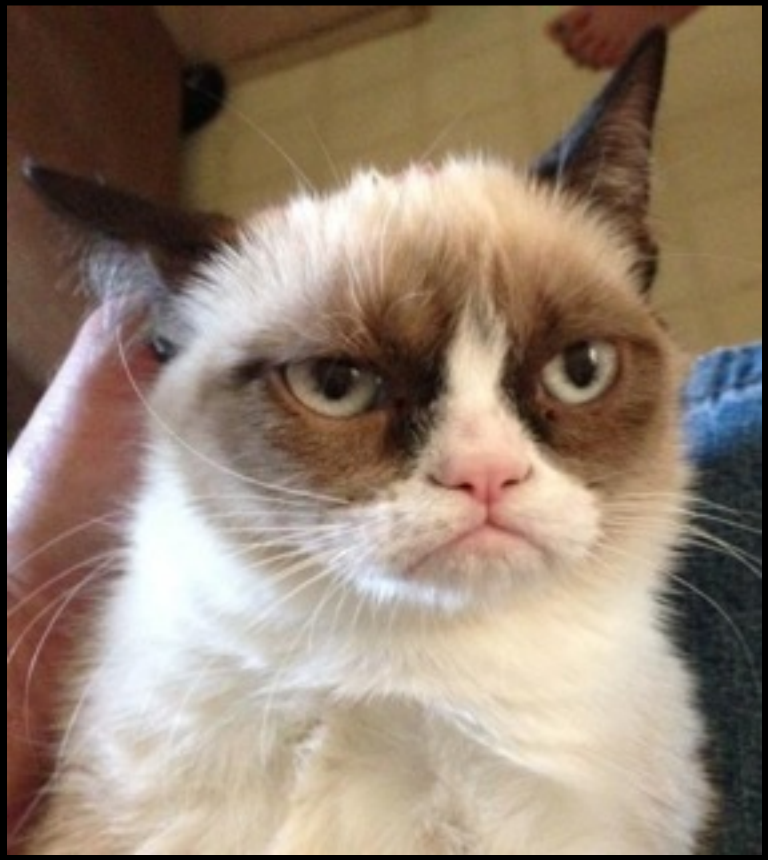
Sample Space

- In probability theory, the set of all possible worlds is called the sample space:

$$\Omega = \{ \omega_i \}$$

- Possible worlds ω_i are:
  - Mutually exclusive
  - Exhaustive

Sample Space

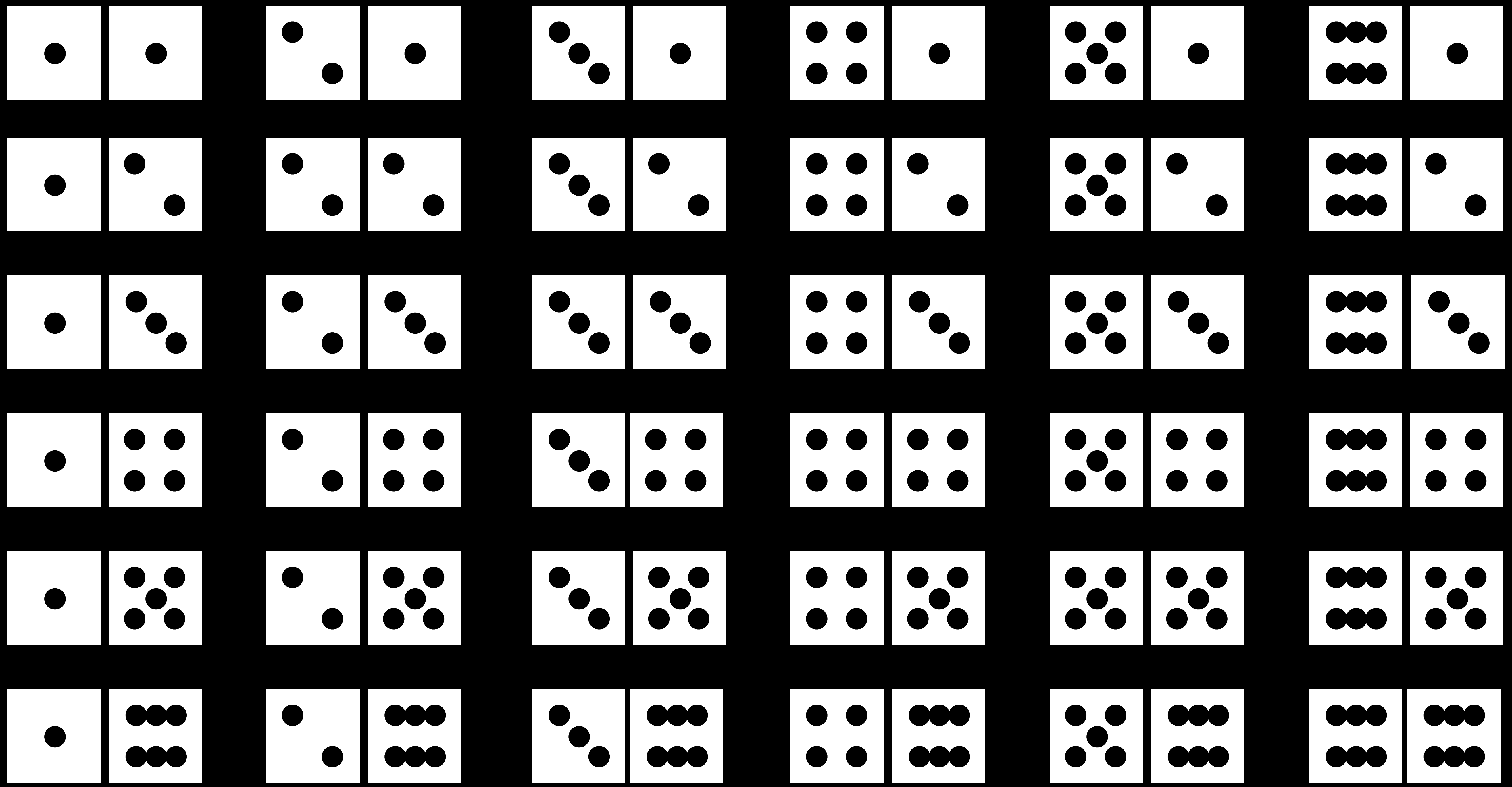


Hungry=true, Cranky=false

Hungry=false, Cranky=true

Hungry=true, Cranky=true

Hungry=false, Cranky=false



$$\Omega = \{\ \omega_i\ \} = \{\ (1,1),\ (1,2),\ (2,1),\ (3,1),\ \dots\ \}$$

Probability Model

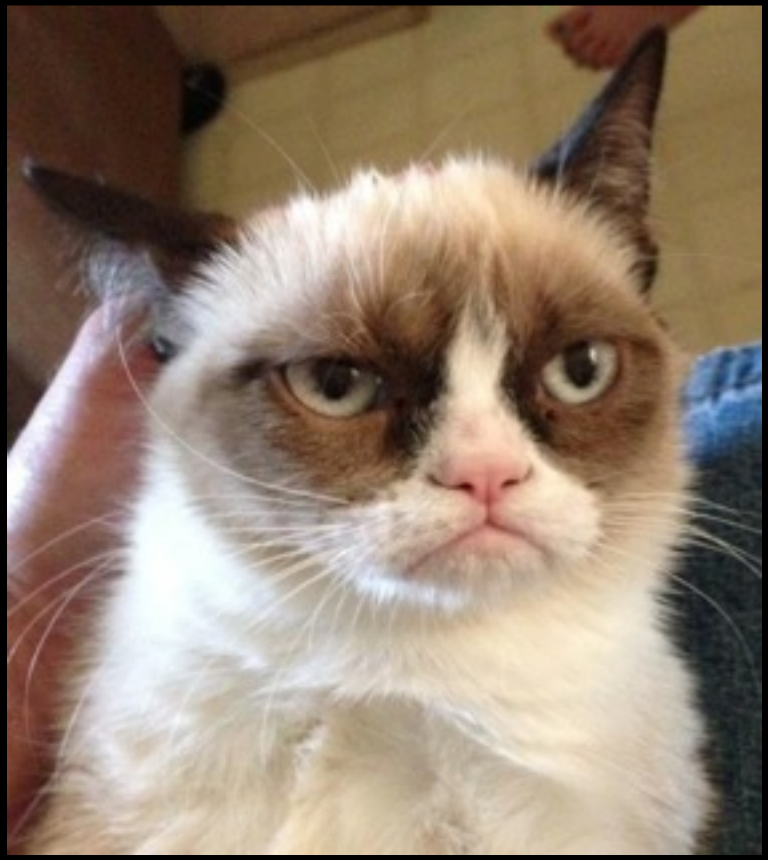
- Assigns a numerical probability $P(\omega)$ to each possible world*, such that:

$$0 \leq P(\omega) \leq 1$$

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

*Discrete, countable set of worlds

Probability Model

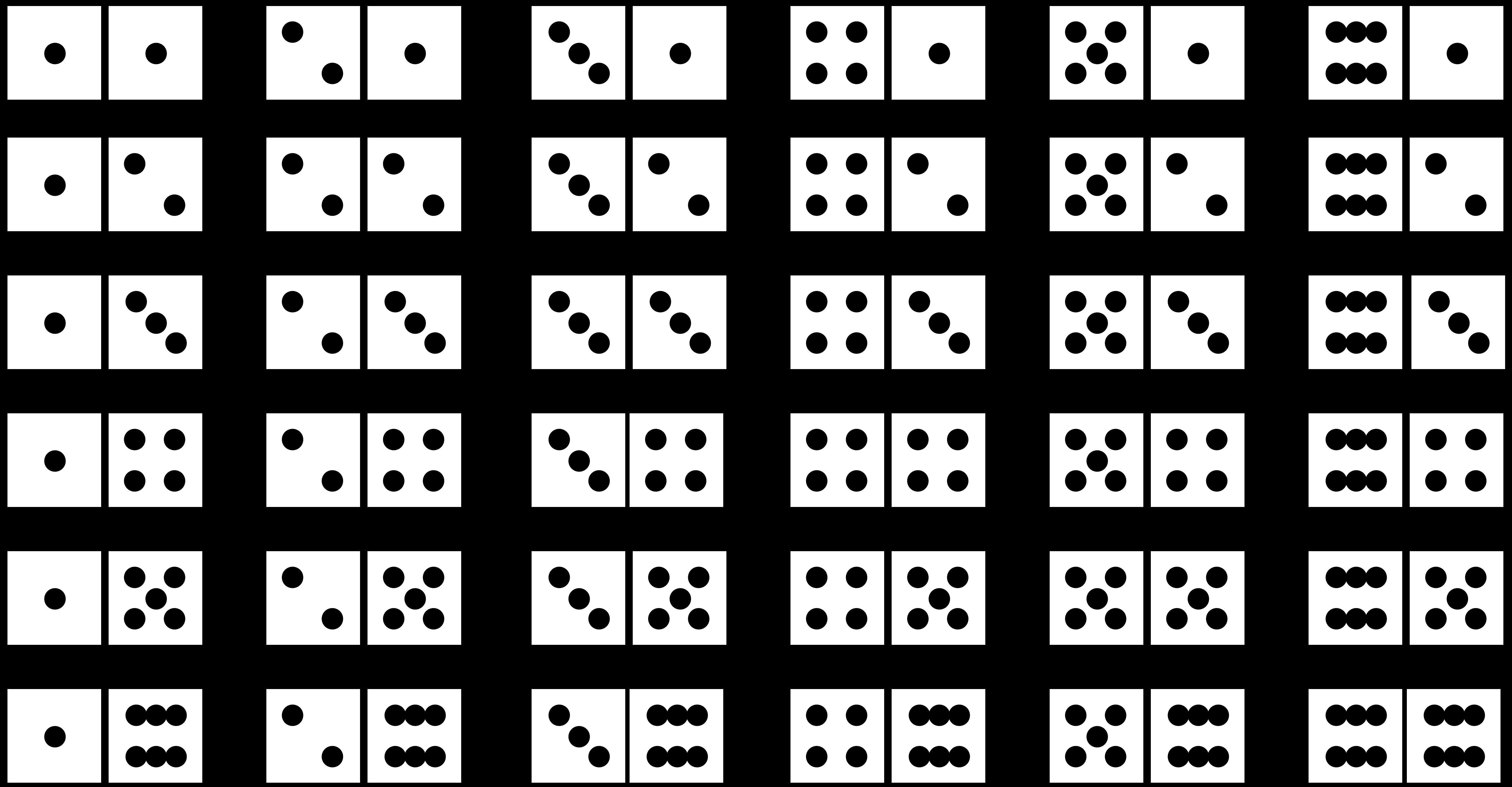


Hungry=true, Cranky=false
0.2

Hungry=false, Cranky=true
0.1

Hungry=true, Cranky=true
0.4

Hungry=false, Cranky=false
0.3



$$P(\omega_i) = 1/36 \text{ for all } \omega_i \in \Omega$$

What *are* Probabilities?

- Our answer: The degree to which an agent believes a possible world is the actual world
    - From 0: certainly not the case (i.e., false)
    - To 1: certainly is the case (i.e., true)
- Could come from statistical data, general principles, combination of evidence, ...

Pierre Simon Laplace

Propositions (Events)

- A proposition (event) corresponds to the set of possible worlds in which the proposition holds

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

Propositions

Cranky

~~Hungry=true, Cranky=false~~ 0.2	Hungry=false, Cranky=true 0.1
Hungry=true, Cranky=true 0.4	~~Hungry=false, Cranky=false~~ 0.3

P(Cranky) = 0.1+0.4 = 0.5

Propositions

Hungry ∨ Cranky

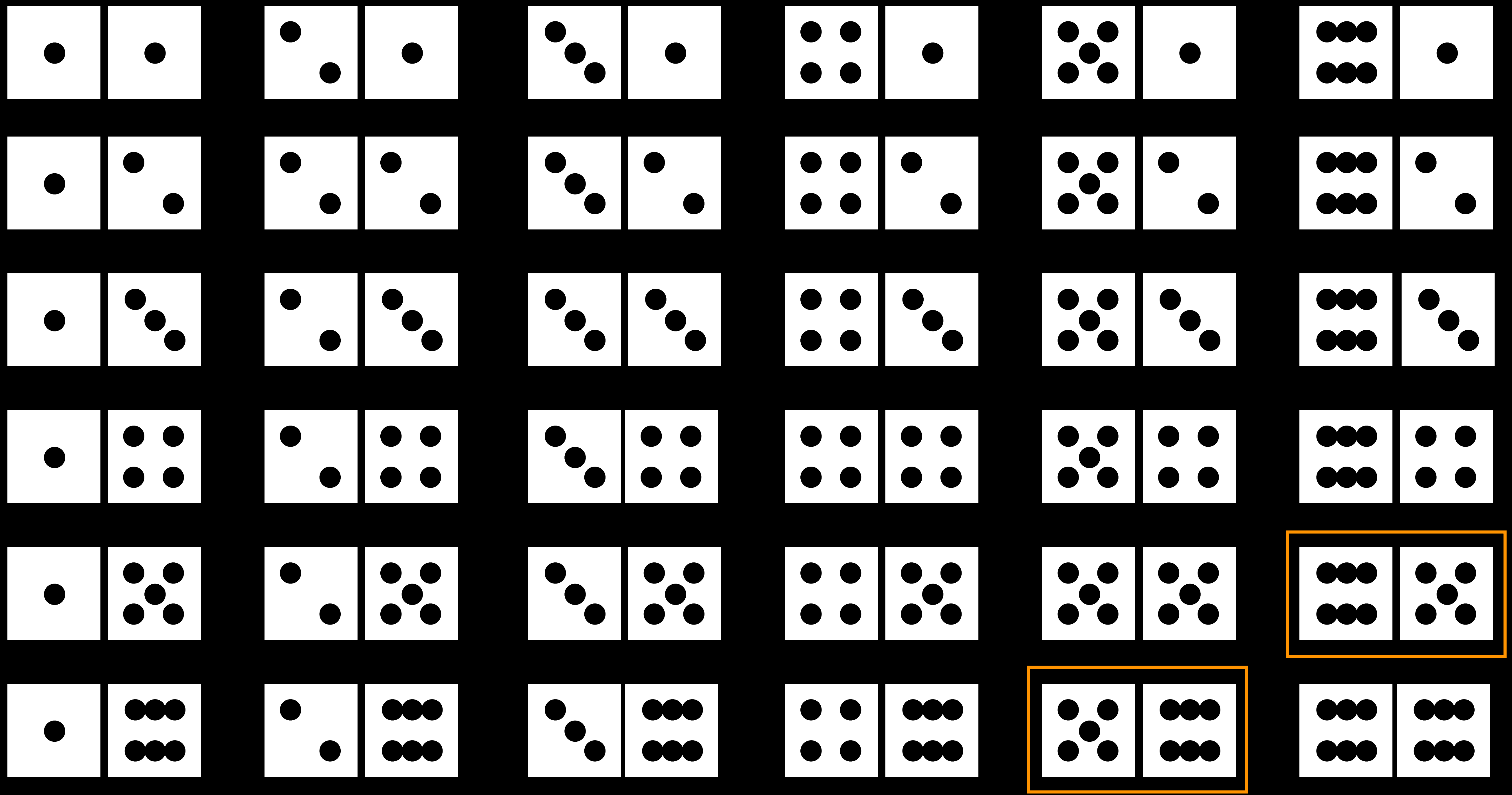
Hungry=true, Cranky=false
0.2

Hungry=false, Cranky=true
0.1

Hungry=true, Cranky=true
0.4

~~Hungry=false, Cranky=false~~
0.3

P(Hungry ∨ Cranky) = 0.2+0.1+0.4 = 0.7



$$P(Total = 11) = P((5,6)) + P((6,5)) = 1/36 + 1/36 = 1/18$$

Probability (so far)

- Sample space (possible worlds)
- Probability model (degrees of belief in worlds)
- Propositions (subsets of worlds in which proposition holds)

Unconditional (Prior) Probabilities

- Degrees of belief in propositions in the <u>absence</u> of any other information

$$P(Cranky) = \frac{1}{2}$$

$$P(SnakeEyes) = \frac{1}{36}$$

Conditional (Posterior) Probability

- Degree of belief in a proposition given some information (evidence)

$$P(Cranky \mid Hungry) = \frac{2}{3}$$

$$P(SnakeEyes \mid Doubles) = \frac{1}{6}$$

- Whenever evidence is true and we have no further information, conclude probability of proposition

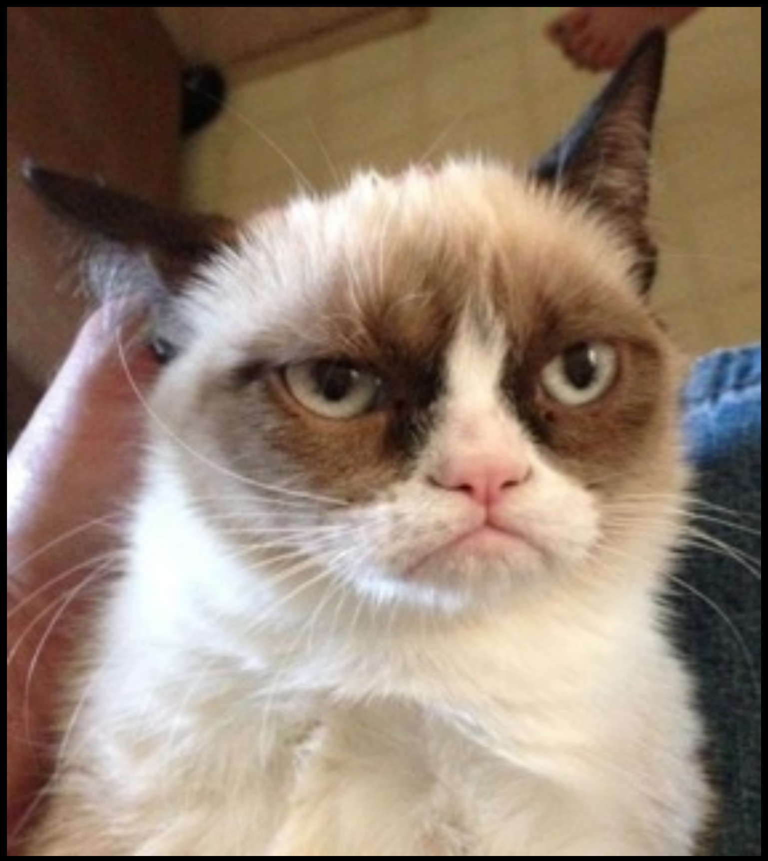
Conditional Probability

- Conditional probability can be defined as:

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \text{ when } P(b) > 0$$

- Why?

Conditional Probability



Hungry=true, Cranky=false
0.2

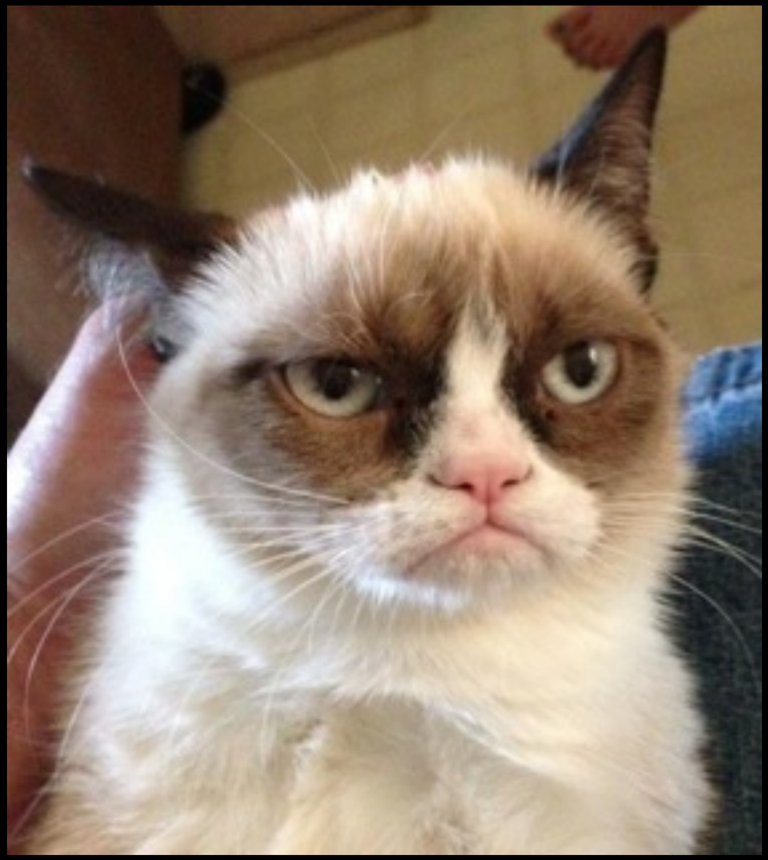
Hungry=false, Cranky=true
0.1

Hungry=true, Cranky=true
0.4

Hungry=false, Cranky=false
0.3

$$P(Cranky \mid Hungry) = ?$$

Conditional Probability



Hungry=true, Cranky=false
0.2

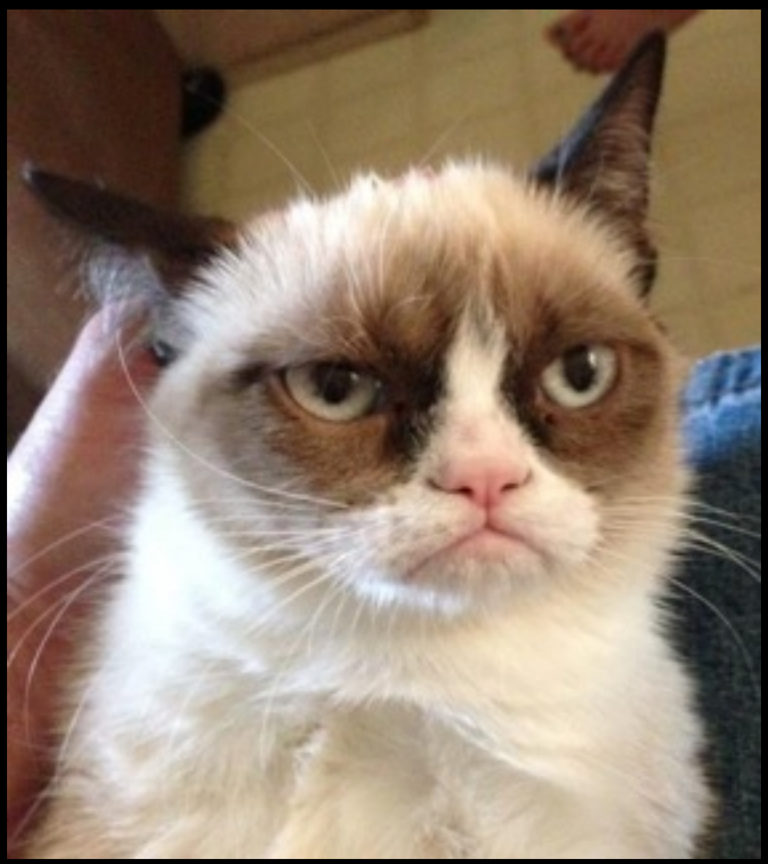
Hungry=false, Cranky=true
0.1

Hungry=true, Cranky=true
0.4

Hungry=false, Cranky=false
0.3

$$P(Cranky \mid Hungry) = \frac{P(Cranky \wedge Hungry)}{P(Hungry)}$$

Conditional Probability



Hungry=true, Cranky=false
0.2

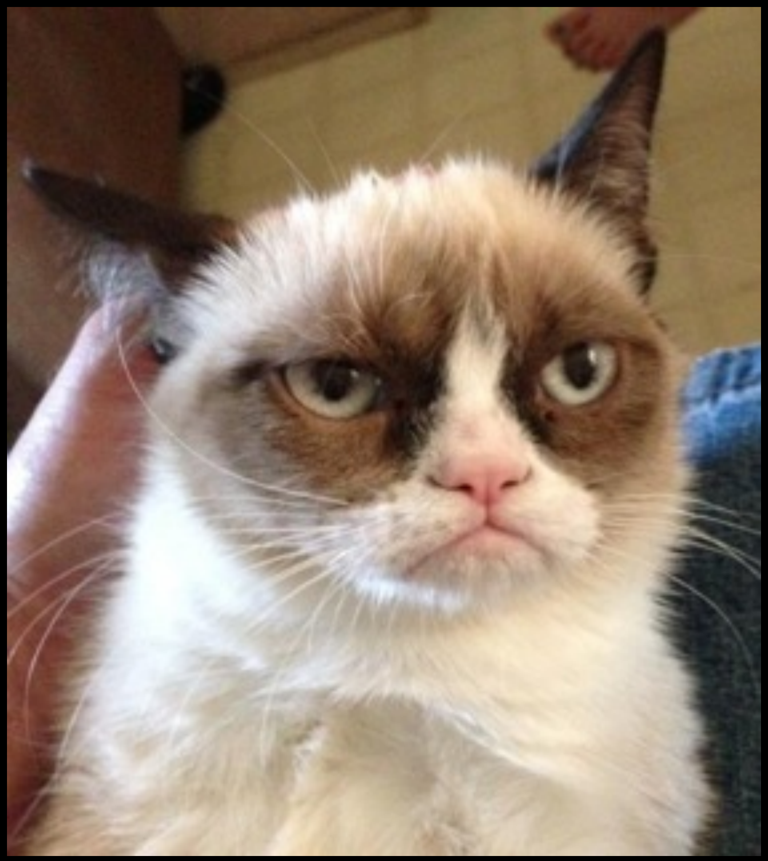
Hungry=false, Cranky=true
0.1

Hungry=true, Cranky=true
0.4

Hungry=false, Cranky=false
0.3

$$P(Cranky \mid Hungry) = \frac{P(Cranky \wedge Hungry)}{P(Hungry)} = \frac{0.4}{}$$

Conditional Probability



Hungry=true, Cranky=false 0.2	Hungry=false, Cranky=true 0.1
Hungry=true, Cranky=true 0.4	Hungry=false, Cranky=false 0.3

$$P(Cranky \mid Hungry) = \frac{P(Cranky \wedge Hungry)}{P(Hungry)} = \frac{0.4}{0.4 + 0.2}$$

Conditional Probability

Hungry=true, Cranky=false
0.2

Hungry=false, Cranky=true
0.1

Hungry=true, Cranky=true
0.4

Hungry=false, Cranky=false
0.3

$$P(Cranky \mid Hungry) = \frac{P(Cranky \wedge Hungry)}{P(Hungry)} = \frac{0.4}{0.4 + 0.2} = \frac{2}{3}$$

Conditional ≠ Implication!

Hungry=true, Cranky=false
0.2

Hungry=false, Cranky=true
0.1

Hungry=true, Cranky=true
0.4

Hungry=false, Cranky=false
0.3

$$P(Hungry \Rightarrow Cranky) = 0.4 + 0.1. + 0.3 = \frac{4}{5} \neq \frac{2}{3}$$

Incomplete Information

↓

Factored Representation

↓

Variables & Values (domains)

Random Variables

- Take on values from a <u>domain</u>
- Boolean random variables have domain { true, false }
- Domains can be finite or infinite
  - Discrete, infinite: integers
  - Continuous, infinite: reals

Random Variables

$$Die_1 : \{1, 2, 3, 4, 5, 6\}$$

$$Total : \{2, \ldots, 12\}$$

$$Doubles : \{true, false\}$$

$$Weather : \{sunny, rain, cloudy, snow\}$$

Atomic Propositions

- Restriction on possible values of a random variable (a.k.a. constraint)
  - Including statement that a random variable takes on a particular value (i.e., domain restricted to a single value)

Atomic Propositions

Boolean random variables:

$$Doubles = true \rightarrow doubles$$

$$Doubles = false \rightarrow \neg doubles$$

Symbolic (unambiguous) value:

$$Weather = sunny \rightarrow sunny$$

Ordered domains:

$$50 \leq Weight < 100$$

Connectives

Same connectives as propositional logic:

$\wedge, \vee, \Rightarrow, \Leftrightarrow$

$$Cavity = \{true, false\}$$

$$Toothache = \{true, false\}$$

$$Age = \{baby, child, teen, adult, senior\}$$

$$P(cavity \mid \neg toothache \wedge teen) = 0.1$$

$$P(Cavity = true \mid Toothache = false \wedge Age = teen) = 0.1$$

The Language of Probability Assertions

- Random variables (and domains)
- Atomic propositions:
  - $Var = value$ (or <, ≤, >, ≥)
- Connectives

Example

Random variable:

$$Cavity : \{True, False\}$$

Probability Assertions:

$$P(Cavity = True) = 0.2$$

$$P(Cavity = False) = 0.8$$

Probability Distribution:

$$\mathbf{P}(Cavity) = \langle 0.2, 0.8 \rangle$$

Example

Random variable:

$$Weather : \{sunny, rain, cloudy, snow\}$$

Probability Assertions:

$$P(Weather = sunny) = 0.6$$

$$P(Weather = rain) = 0.1$$

$$P(Weather = cloudy) = 0.29$$

$$P(Weather = snow) = 0.01$$

Probability Distribution:

$$\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$$

Joint Distributions

- Distributions over multiple variables
- Describe probabilities of all combinations of the values of the variables

Joint Distributions

$$\mathbf{P}(Weather, Cavity)$$

		Cavity	
		true	false
Weather	sunny	0.12	0.48
	rain	0.02	0.08
	cloudy	0.058	0.232
	snow	0.002	0.008

Joint Distributions

$$\mathbf{P}(sunny, Cavity)$$

		Cavity	
		true	false
Weather	sunny	0.12	0.48

Joint Distributions

$$\mathbf{P}(sunny, cavity)$$

		Cavity
		true
Weather	sunny	0.12

$$\mathbf{P}(sunny, cavity) = P(sunny, cavity) = P(sunny \land cavity)$$

Marginal Distribution

		Cavity	
		true	false
Weather	sunny	0.12	0.48
	rain	0.02	0.08
	cloudy	0.058	0.232
	snow	0.002	0.008

Marginal Distribution

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	
	rain	0.02	0.08	
	cloudy	0.058	0.232	
	snow	0.002	0.008	

Marginal Distribution

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	
	cloudy	0.058	0.232	
	snow	0.002	0.008	

Marginal Distribution

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	
	snow	0.002	0.008	

Marginal Distribution

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	

Marginal Distribution

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01

Marginal Distribution

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01
		0.2		

Marginal Distribution

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01
		0.2	0.8	

What's Special?

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01
		0.2	0.8	

What's Special?

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01
		0.2	0.8	

$$P(Weather = sunny, Cavity = true) = P(Weather = sunny)P(Cavity = true)$$

What's Special?

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01
		0.2	0.8	

What's Special?

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01
		0.2	0.8	

$$P(Weather = cloudy, Cavity = true) = P(Weather = cloudy)P(Cavity = true)$$

What's Special?

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01
		0.2	0.8	

What's Special?

		Cavity		
		true	false	
Weather	sunny	0.12	0.48	0.6
	rain	0.02	0.08	0.1
	cloudy	0.058	0.232	0.29
	snow	0.002	0.008	0.01
		0.2	0.8	

$$P(Weather = snow, Cavity = false) = P(Weather = snow)P(Cavity = false)$$

Independence

- In this example, we see that

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

- When this holds, we say that the variables *Weather* and *Cavity* are independent
- Leveraging independence - when it holds -will be a key to efficient probabilistic inference!

Probabilistic Inference

- If all variables that describe the world were independent, probabilistic inference would be trivial
- You either know the probability of a variable, or you don't
- Nothing you observed (other than that variable) could affect its probability

Probabilistic Inference

- Computing posterior probabilities for propositions given prior probabilities and observed evidence
- Let's see an example of this…

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

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditional Probability

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Conditional Probability

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \underline{0.12}$$

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditional Probability

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = \frac{0.12}{0.2}$$

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Conditional Probability

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Conditional Probability

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)}$$

Conditional Probability

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.08}{0.2} = 0.4$$

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Conditional Probability
Probabilistic Inference

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)} = 0.4$$

$$\mathbf{P}(Cavity \mid toothache) = \langle 0.6, 0.4 \rangle$$

Simplifying Calculation

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

Normalization

$$P(cavity \mid toothache) = \alpha \, P(cavity \wedge toothache) = \alpha \, 0.12$$

$$P(\neg cavity \mid toothache) = \alpha \, P(\neg cavity \wedge toothache) = \alpha \, 0.08$$

Normalization

$$P(cavity \mid toothache) = \alpha\, P(cavity \wedge toothache) = \alpha\, 0.12$$

$$P(\neg cavity \mid toothache) = \alpha\, P(\neg cavity \wedge toothache) = \alpha\, 0.08$$

$$\mathbf{P}(Cavity \mid toothache) = \alpha\, \langle 0.12, 0.08 \rangle$$

Normalization

$$P(cavity \mid toothache) = \alpha\, P(cavity \wedge toothache) = \alpha\, 0.12$$

$$P(\neg cavity \mid toothache) = \alpha\, P(\neg cavity \wedge toothache) = \alpha\, 0.08$$

$$\mathbf{P}(Cavity \mid toothache) = \frac{1}{0.12 + 0.08}\, \langle 0.12, 0.08 \rangle$$

Normalization

$$P(cavity \mid toothache) = \alpha\, P(cavity \wedge toothache) = \alpha\, 0.12$$

$$P(\neg cavity \mid toothache) = \alpha\, P(\neg cavity \wedge toothache) = \alpha\, 0.08$$

$$\mathbf{P}(Cavity \mid toothache) = \langle 0.6, 0.4 \rangle$$

We didn't need to compute $P(toothache)$!

Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Query variable $X : Domain(X) = \{x_1, \ldots, x_m\}$

Evidence variables $\mathbf{E} : \{E_1, \ldots, E_k\}$

Observations $\mathbf{e} : \{e_1, \ldots, e_k\}$ s.t. $E_i = e_i$

Unobserved variables $\mathbf{Y} : \{Y_1, \ldots, Y_l\}$

$Domain(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$

Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

For each possible value x_i for X

For each possible combination of values $\mathbf{y}$ for Y

Add $\mathbf{P}(x_i, \mathbf{e}, \mathbf{y})$

Result: vector $\mathbf{P}(X|\mathbf{e}) = \langle\, \mathbf{P}(x_i|\mathbf{e})\, \rangle = \langle\, \mathbf{P}(X = x_i|\mathbf{e})\, \rangle$

Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

For each possible value x_i for X

For each possible combination of values $\mathbf{y}$ for Y

Add $\mathbf{P}(x_i, \mathbf{e}, \mathbf{y})$

$$\mathbf{P}(X = x_i, E_1 = e_1, \ldots, E_k = e_k, \ldots, Y_1 = y_{1,i_1}, \ldots, Y_l = y_{l,i_l})$$

Result: vector $\mathbf{P}(X|\mathbf{e}) = \langle\, \mathbf{P}(x_i|\mathbf{e})\, \rangle = \langle\, \mathbf{P}(X = x_i|\mathbf{e})\, \rangle$

Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

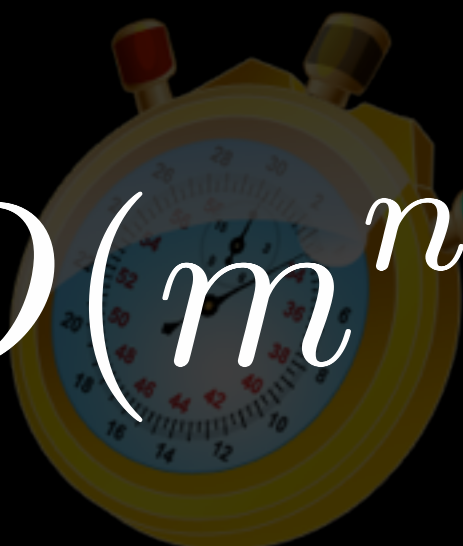
```
int m;    // Number of values in domain of query variable X
int k;    // Number of evidence variables E
int[k] e; // e[i]: index of i'th evidence value in domain of Ei
int l;    // Number of unobserved variables Y
int[l] n; // n[i]: number of values in domain of Yi
double[l][n[l]] D; // D[i][j]: j'th value of domain for Yi

for i from 1 to m
  PXe[i] = 0
  for i1 from 1 to n[1]
    for i2 from 1 to n[2]
      ...
        for il from 1 to n[l]
          PXe[i] += JPDF[i][e[1]]...[e[k]][D[1][i1]]...[D[l][il]]
return Pxe // vector of length m
```

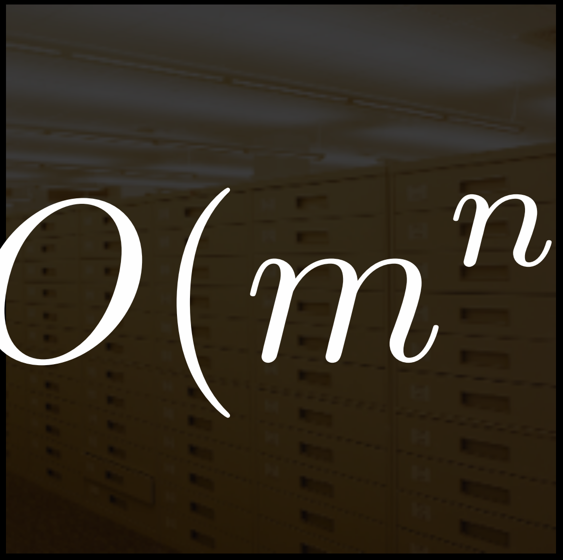
l nested loops

Full joint prob. dist.

Inference
(Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$





$O(m^n)$

Time Complexity

$O(m^n)$

Space Complexity

Probabilistic Inference

- In logic, we can use rules for deduction in place of enumerating all truth assignments
- Similarly: for probabilistic reasoning, we can use rules in place of enumerating all assignments to the random variables

Rules for Probabilistic Inference

Range of probabilities: $0 \leq P(\alpha) \leq 1$

Range of a random variable: $\sum_c P(X = c) = 1$

Conditional probability: $P(\alpha \mid \beta) = \dfrac{P(\alpha \wedge \beta)}{P(\beta)}$

Logical connectives: $P(\neg \alpha) = 1 - P(\alpha)$

$$P(\alpha \vee \beta) = P(\alpha) + P(\beta) - P(\alpha \wedge \beta)$$

Rules for Probablistic Inference

Independence: $\alpha \,||\, \beta$ if and only if:

$$P(\alpha) = P(\alpha \mid \beta)$$

$$P(\alpha \wedge \beta) = P(\alpha)P(\beta)$$

Conditional independence: $\alpha \,||\, \beta \mid \varphi$ iff:

$$P(\alpha \mid \varphi) = P(\alpha \mid \beta, \varphi)$$

$$P(\alpha \wedge \beta \mid \varphi) = P(\alpha \mid \varphi)P(\beta \mid \varphi)$$

A Little Theorem

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

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

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

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

$$P(a \mid b)P(b) = P(b \mid a)P(a)$$

$$P(b \mid a)P(a) = P(a \mid b)P(b)$$

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

Bayes' Rule

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

Thomas Bayes
(c. 1702 – 1761)

Causal and Diagnostic Knowledge

Causal knowledge: $P(\mathit{effect} \mid \mathit{cause})$

Diagnostic knowledge: $P(\mathit{cause} \mid \mathit{effect})$

Causal and Diagnostic Knowledge

Causal knowledge: $P(symptom \mid disease)$

Diagnostic knowledge: $P(disease \mid symptom)$

Bayesian Diagnosis

$$P(disease \mid symptom) = \frac{P(symptom \mid disease)P(disease)}{P(symptom)}$$

Bayesian Diagnosis

Meningitis causes a stiff neck 70% of the time

$$P(stiffneck \mid meningitis) = 0.7$$

Prior probability of meningitis $P(meningitis) = 0.00002$

Prior probability of stiff neck $P(stiffneck) = 0.01$

$$P(meningitis \mid stiffneck) = \frac{P(stiffneck \mid meningitis)P(meningitis)}{P(stiffneck)}$$

$$= \frac{0.7 \times 0.00002}{0.01}$$

$$= 0.0014$$

Combining Evidence

$toothache$

$(Toothache = True)$

$catch$

$(Catch = True)$

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

Combining Evidence

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$
$$= \alpha \langle 0.180, 0.016 \rangle \approx \langle 0.871, 0.129 \rangle$$

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Exponential Growth of Combinations of Evidence

$$P(Cavity \mid toothache \land catch \land \neg bleeding)$$

	$toothache$				$\neg toothache$			
	$catch$		$\neg catch$		$catch$		$\neg catch$	
	$bleeding$	$\neg bleeding$	$bleeding$	$\neg bleeding$	$bleeding$	$\neg bleeding$	$bleeding$	$\neg bleeding$
$cavity$	?	?	?	?	?	?	?	?
$\neg cavity$	?	?	?	?	?	?	?	?

Combining Evidence

- In general, if there are n possible evidence variables, then there are $O(2^n)$ possible combinations of observed values for which we would need to know the conditional probabilities
- How can avoid the need for an exponential number of conditional probabilities?

Conditional Independence

- Both $toothache$ and $catch$ are caused by a cavity, but neither has a direct effect on the other
- The variables are independent given the presence or absence of a cavity
- Notation: $Toothache \perp\!\!\!\perp Catch \mid Cavity$

Conditional Independence

Assuming $Toothache \| Catch \mid Cavity$

$$\mathbf{P}(toothache \wedge catch \mid Cavity) =$$
$$\mathbf{P}(toothache \mid Cavity)\mathbf{P}(catch \mid Cavity)$$

After Applying Bayes Rule

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

Only need these probabilities -
linear in the number of evidence variables!

Combining Evidence

- For n symptoms (e.g., $Toothache$, $Catch$) that are all conditionally independent given a disease (e.g., $Cavity$), we need $O(n)$ probabilities rather than $O(2^n)$
  - Representation scales to much larger problems!

Probabilistic Reasoning

- Full joint distribution: intractable as problem grows
- Independence assumptions reduce number of probabilities required to represent full joint distribution
- Next class: Bayes Nets: an efficient data structure for reasoning with independence assumptions

Summary

- Probability model assigns a degree of belief to possible worlds
- Factored representation: random variables
- Probabilities:
  - Joint vs. marginal distribution
  - Unconditional (Prior) vs. conditional (posterior)
- Inference: Computing posterior (conditional) probabilities for propositions given observed evidence

Summary (continued)

- Rules of probability can be used for inference
- Independence assumptions reduce number of probabilities required to represent and reason about the full joint probability distribution

Project 1-II: Timed Othello

Due Tonight (1 minute to midnight)

Project 2: Planning

Due April 8

Homework 4: Probability & Learning I

Exam April 8