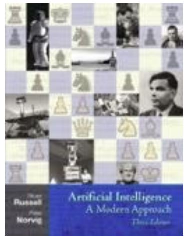
Outline

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



Many slides based on Russell & Norvig's slides
Artificial Intelligence: A Modern Approach

Uncertain Actions

- So far, our agents believe that
  - (logical) statements are true or false (maybe unknown)
  - actions will always do what they think they do
- Unfortunately, the real world is not like that
  - agents almost never have access to the whole truth about the world

→ agents must deal with uncertainty

Uncertain Actions

- So far, our agents believe that
  - (logical) statements are true or false (maybe unknown)
  - actions will always do what they think they do
- Unfortunately, the real world is not like that
  - agents almost never have access to the whole truth about the world

→ agents must deal with uncertainty

- Example:
  - We have different actions for getting to the airport:
    - action A_t = leave for the airport t minutes before departure

Uncertain Actions

- So far, our agents believe that
  - (logical) statements are true or false (maybe unknown)
  - actions will always do what they think they do
- Unfortunately, the real world is not like that
  - agents almost never have access to the whole truth about the world

→ agents must deal with uncertainty

- Example:
  - We have different actions for getting to the airport:
    - action A_t = leave for the airport t minutes before departure
  - Typical problems:
    - Will a given action A_t get me to the airport in time?
    - Which action is the best choice for getting me to the airport?

Atomic Taco, CC BY-SA 4.0, resized, cropped and compressed

Many situations are uncertain.
Agents have to deal with these uncertainties.

Steve Jurvetson, CC BY 2.0, resized, cropped and compressed

Problems with Uncertainty

We leave 90 minutes before departure

- Risks involved in the plan $\boxed{A_{90} \text{ will get me to the airport}}$
  - partial observability (road state, other drivers' plans, etc.)
  - noisy sensors (traffic reports may be wrong)
  - uncertainty in action outcomes (flat tire, accident, etc.)
  - immense complexity of modeling and predicting traffic

Problems with Uncertainty

We leave 90 minutes before departure

- Risks involved in the plan A_{90} will get me to the airport
  - partial observability (road state, other drivers' plans, etc.)
  - noisy sensors (traffic reports may be wrong)
  - uncertainty in action outcomes (flat tire, accident, etc.)
  - immense complexity of modeling and predicting traffic
- A logically correct plan:

A_{90} will get me to the airport as long as my car doesn't break down,
I don't run out of gas, no accident, the bridge doesn't fall down, **etc**.

  - impossible to model all things that can go wrong
    - → recall the **qualification problem**

Problems with Uncertainty

We leave 90 minutes before departure

- Risks involved in the plan A_{90} will get me to the airport
  - partial observability (road state, other drivers' plans, etc.)
  - noisy sensors (traffic reports may be wrong)
  - uncertainty in action outcomes (flat tire, accident, etc.)
  - immense complexity of modeling and predicting traffic
- A logically correct plan:

A_{90} will get me to the airport as long as my car doesn't break down,
I don't run out of gas, no accident, the bridge doesn't fall down, **etc.**

  - impossible to model all things that can go wrong
    - → recall the **qualification problem**
- A more cautious plan:

A_{1440} will get me to the airport

  - will (virtually) certainly succeed, but clearly suboptimal
    - e.g., we have to pay for a night in a hotel

Probabilities

- Probabilities are **one way** of handling uncertainty
  - e.g. $\boxed{A_{90} \text{ will get me to the airport with probability } 0.5}$
- The probability **summarizes effects** that are due to
  - Laziness
    - I don't want to list all things that must not go wrong
  - Theoretical Ignorance
    - Some things just can't be known
      - e.g.: We cannot completely model the weather
  - Practical Ignorance
    - Some things might not be known about the particular situation
      - e.g. Is there a traffic jam at A5?

Probabilities and Beliefs

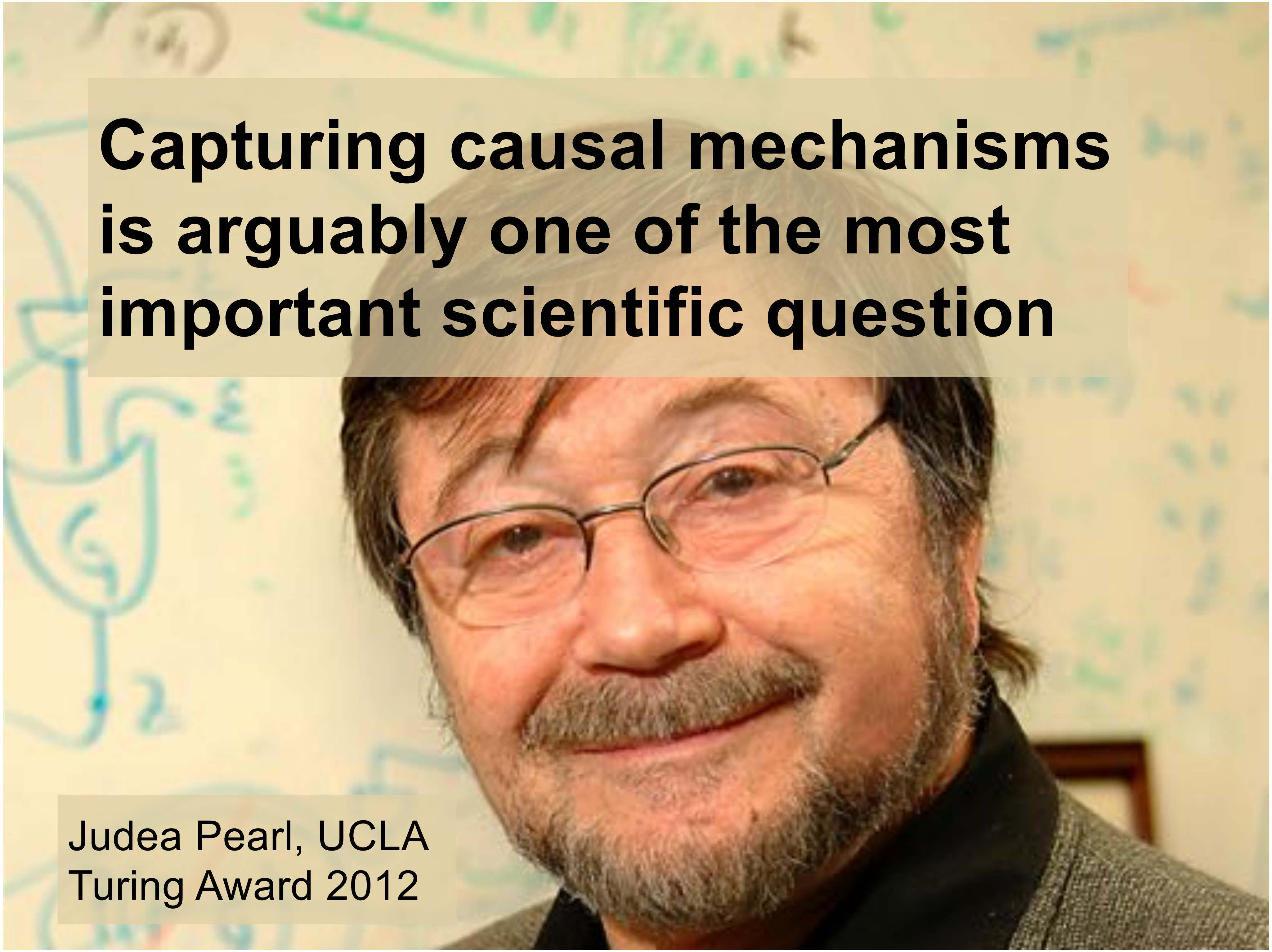
- Probabilities that are related to one's beliefs
    - a probability p attached to a statement means that I believe that the statement will be true in $p{\cdot}100\%$ of the cases
        - there is traffic jam on the A5 in 10% of the cases (meaning: there might be jam, but usually there is none)
    - it does not mean that it is true with $p\%$
        - the traffic on the A5 is jammed with a degree of 10% (meaning: there's a jam, but it could be worse...)

Consider the probability that the sun will still exist tomorrow

- Difficult to observe by an experiment

What is the chance that a patient has a particular disease?

- Doctor wants to consider other patients who are similar. But if you gather too much information to compare patients, there are no similar patients left



Capturing causal mechanisms is arguably one of the most important scientific question

Judea Pearl, UCLA
Turing Award 2012

Probabilities and Beliefs

- Probabilities that are related to one's beliefs
  - a probability p attached to a statement means that I believe that the statement will be true in $p \cdot 100\%$ of the cases
    - there is traffic jam on the A5 in 10% of the cases (meaning: there might be jam, but usually there is none)
  - it does not mean that it is true with $p\%$
    - the traffic on the A5 is jammed with a degree of 10% (meaning: there's a jam, but it could be worse...)

→ Probability theory is about degree of belief

- other techniques (e.g., Fuzzy logic) deal with degree of truth

Probabilities and Beliefs

- Probabilities that are related to one's beliefs
  - a probability p attached to a statement means that I believe that the statement will be true in $p \cdot 100\%$ of the cases
    - there is traffic jam on the A5 in 10% of the cases (meaning: there might be jam, but usually there is none)
  - it does not mean that it is true with $p\%$
    - the traffic on the A5 is jammed with a degree of 10% (meaning: there's a jam, but it could be worse...)

→ Probability theory is about degree of belief

  - other techniques (e.g., Fuzzy logic) deal with degree of truth

- **Probabilities of propositions change with new evidence:**
  - P(A45 gets me there in time | no reported accidents) = 0.06
    - in 6% of the days I get there in time if no accidents reported
  - P(A45 gets me there in time | no reported accidents, 5 a.m.) = 0.15
    - chances are higher at 5 in the morning...

Making Decisions under Uncertainty

- Suppose I believe the following:
  - P(A_{25} gets me there on time | …) = 0.04
  - P(A_{90} gets me there on time | …) = 0.70
  - P(A_{120} gets me there on time | …) = 0.95
  - P(A_{1440} gets me there on time | …) = 0.9999

Which action should I choose?

Making Decisions under Uncertainty

- Suppose I believe the following:
  - P(A_{25} gets me there on time | …) = 0.04
  - P(A_{90} gets me there on time | …) = 0.70
  - P(A_{120} gets me there on time | …) = 0.95
  - P(A_{1440} gets me there on time | …) = 0.9999

Which action should I choose?

- The choice depends on my **preferences**
  - how bad is it to miss the flight?
  - how bad is it to wait for an hour at the airport?

Making Decisions under Uncertainty

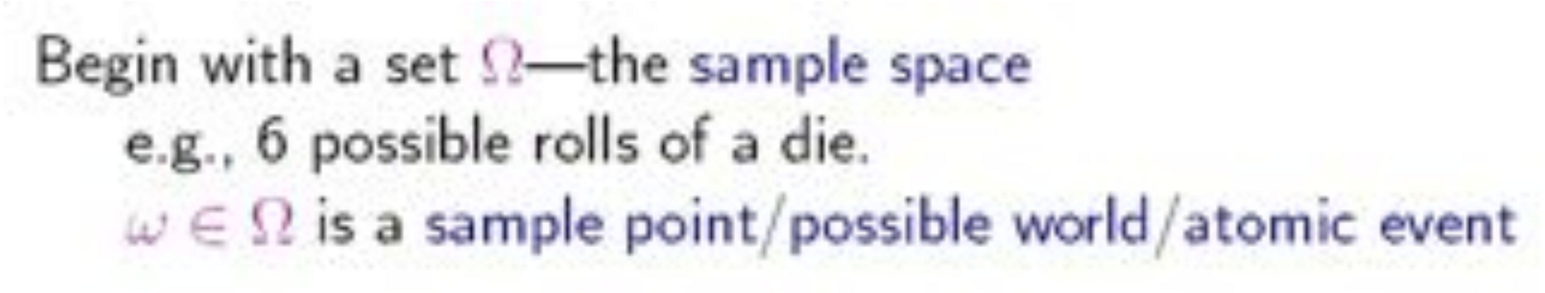
- Suppose I believe the following:
  - $P(A_{25}$ gets me there on time | …) $= 0.04$
  - $P(A_{90}$ gets me there on time | …) $= 0.70$
  - $P(A_{120}$ gets me there on time | …) $= 0.95$
  - $P(A_{1440}$ gets me there on time | …) $= 0.9999$

Which action should I choose?

- The choice depends on my **preferences**
  - how bad is it to miss the flight?
  - how bad is it to wait for an hour at the airport?
- **Utility theory is used to represent and infer preferences**
- **Decision theory = probability theory + utility theory**

SPIEGEL TV
WISSE

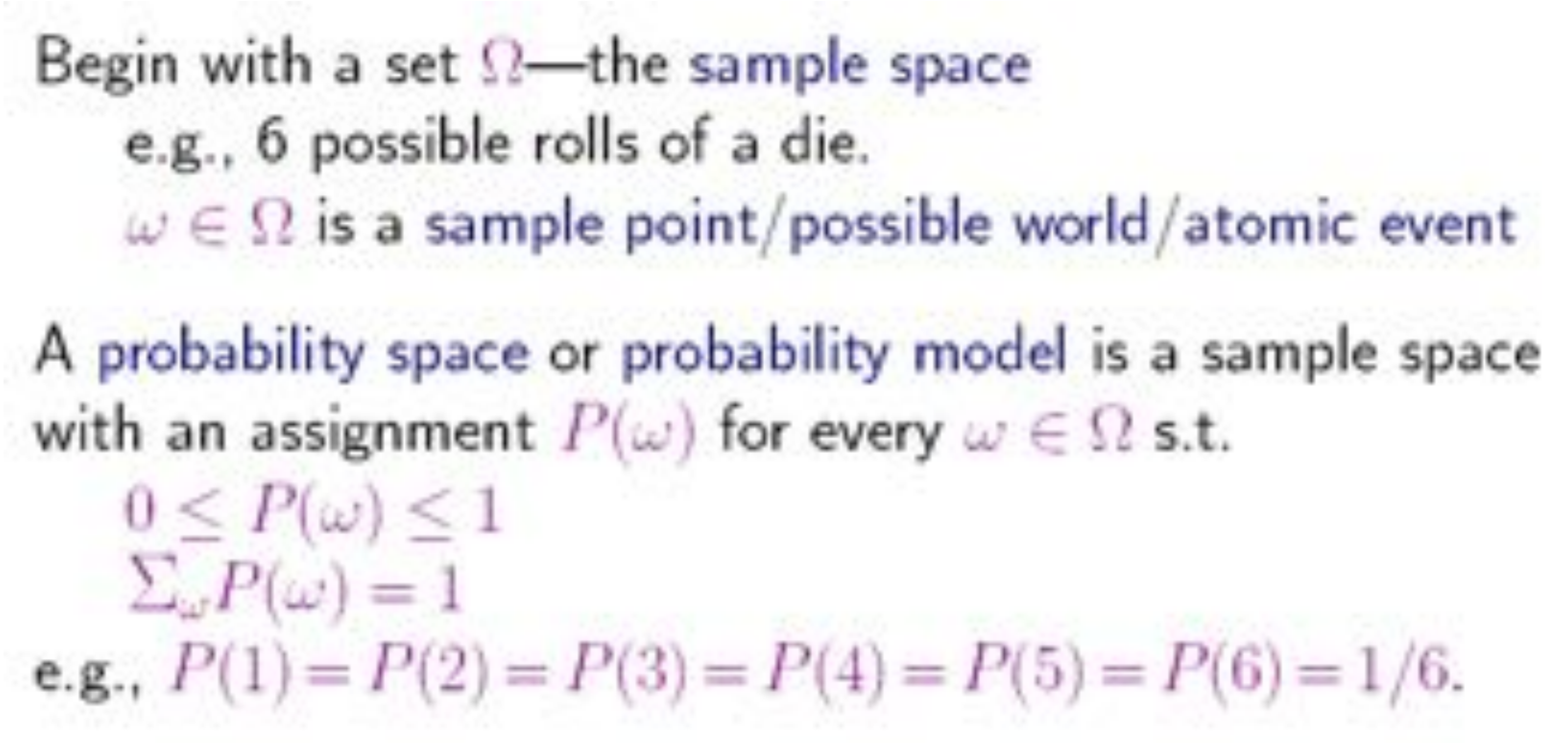
Probability Basics



Begin with a set Ω—the sample space

- e.g., 6 possible rolls of a die.
- $\omega \in \Omega$ is a sample point/possible world/atomic event

Probability Basics



Begin with a set Ω—the sample space

e.g., 6 possible rolls of a die.

$\omega \in \Omega$ is a sample point/possible world/atomic event

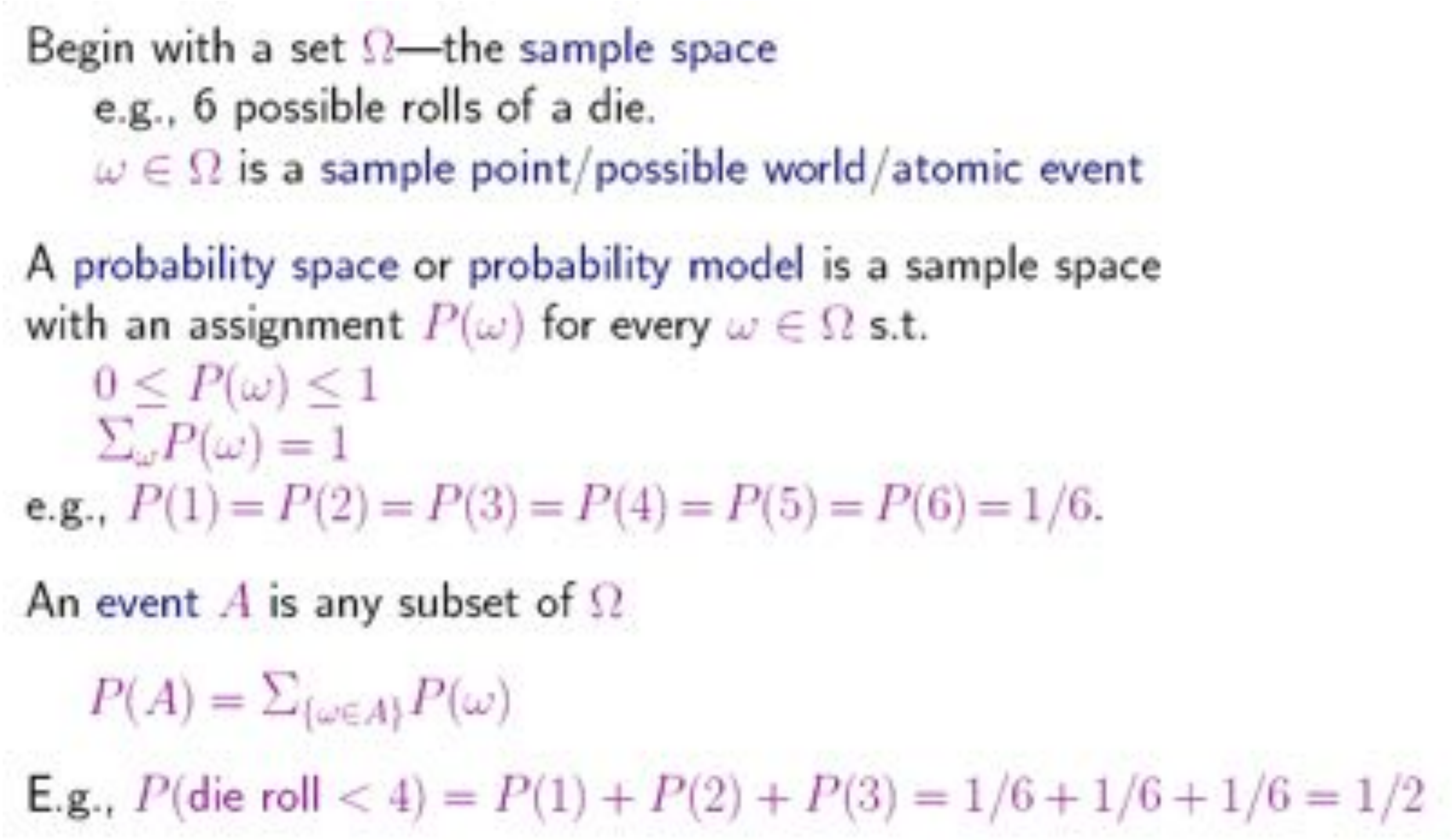
A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

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

$\Sigma_\omega P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6.$

Probability Basics



Begin with a set Ω—the sample space

e.g., 6 possible rolls of a die.

$\omega \in \Omega$ is a sample point/possible world/atomic event

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

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

$\Sigma_\omega P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6.$

An event A is any subset of Ω

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

E.g., $P(\text{die roll} < 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2$

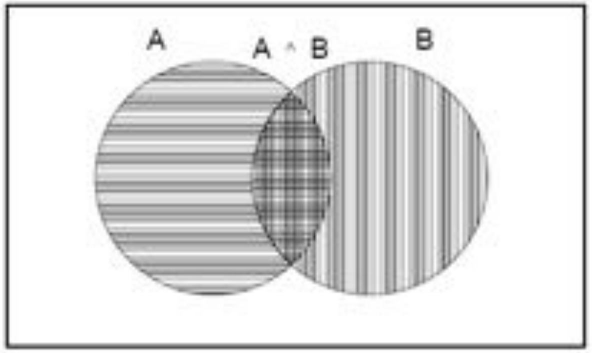
Kolmogorov's Axioms of Probability

1. All probabilities are between 0 and 1

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

2. Necessarily true propositions have probability 1, necessarily false propositions have probability 0

$$P(false) = 0 \qquad P(true) = 1$$



3. The probability of a disjunction is

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

4. These axioms restrict the set of probabilistic beliefs that an agent can (reasonably) hold.
   similar to logical constraints like A and $\neg A$ can't both be true

Violation of Axioms of Probability

„put its money where its probabilities are“

Dutch Book Theorem, Bruno de Finetti (1931)

- an agent who bets according to probabilities that violate the axioms of probability can be forced to bet so as to lose money *regardless of outcome!*

Example:

- suppose Agent 1 believes the following

$$P(a)=0.4 \qquad P(b)=0.3 \qquad P(a \vee b)=0.8$$

axioms of probability are violated because $P(a \vee b) > P(a) + P(b)$

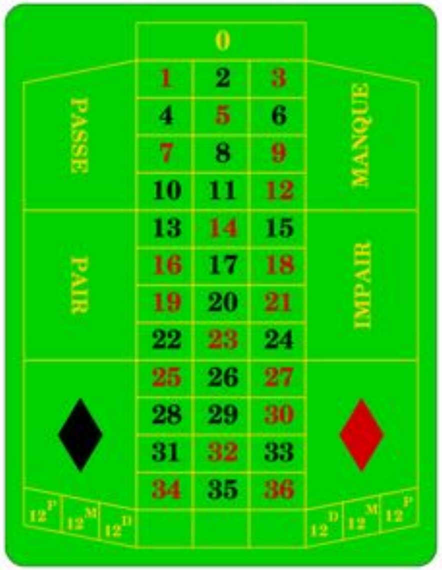
- Agent 2 can now select a set of events and bet on them according to these probabilities so that she cannot loose

Agent 1		Agent 2		Outcome for Agent 1			
proposition	belief	bet	stakes	$a \wedge b$	$a \wedge \neg b$	$\neg a \wedge b$	$\neg a \wedge \neg b$
a	0.4	a	4:6	-6	-6	4	4
b	0.3	b	3:7	-7	3	-7	3
$a \vee b$	0.8	$\neg(a \vee b)$	2:8	2	2	2	-8
				-11	**-1**	**-1**	**-1**

Random Variables

- A random variable is a function from atomic events to some range of values
- Example: Roulette
  - atomic events: numbers 0-36
  - random variables with outcomes true or false
    - rouge / noir, pair / impair, passe / manque
    - transversale, carre, cheval
    - douzaines premier/milieu/dernier
    - etc.

  e.g. rouge(36) = true

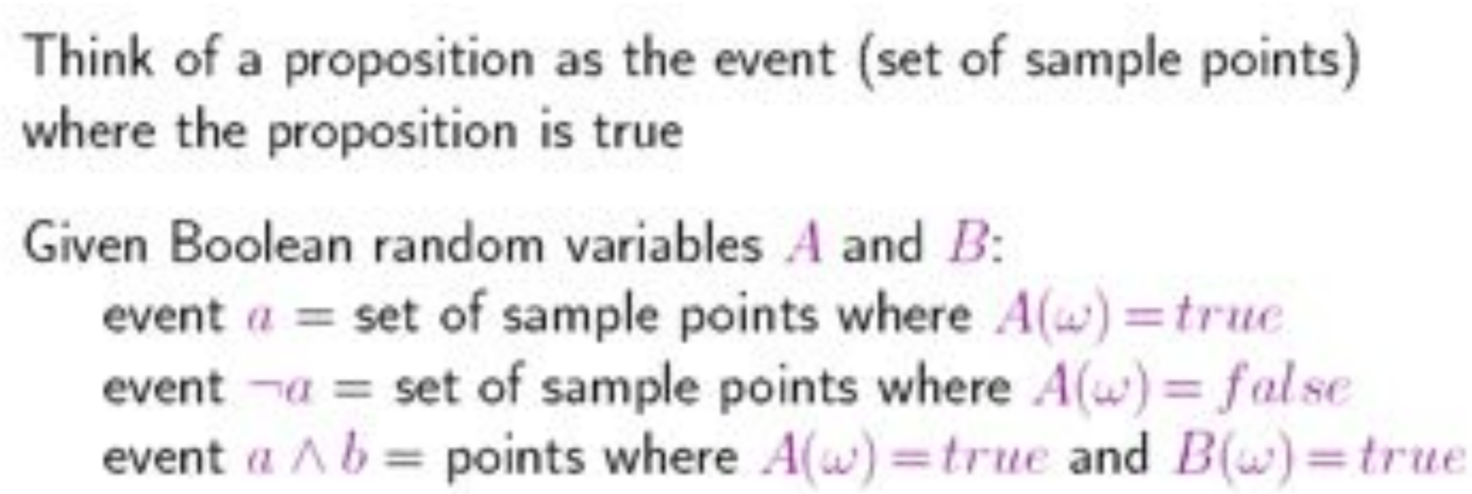


- The probability function P over atomic events induces a probability distribution over all random variables X

$$P(X=x_i)=\sum_{\{\omega:X(\omega)=x_i\}} P(\omega)$$

$$P(Rouge=true)=P(1)+P(3)+\ldots+P(34)+P(36)=\frac{1}{37}+\frac{1}{37}+\ldots+\frac{1}{37}+\frac{1}{37}=\frac{18}{37}$$

Propositions



Think of a proposition as the event (set of sample points) where the proposition is true

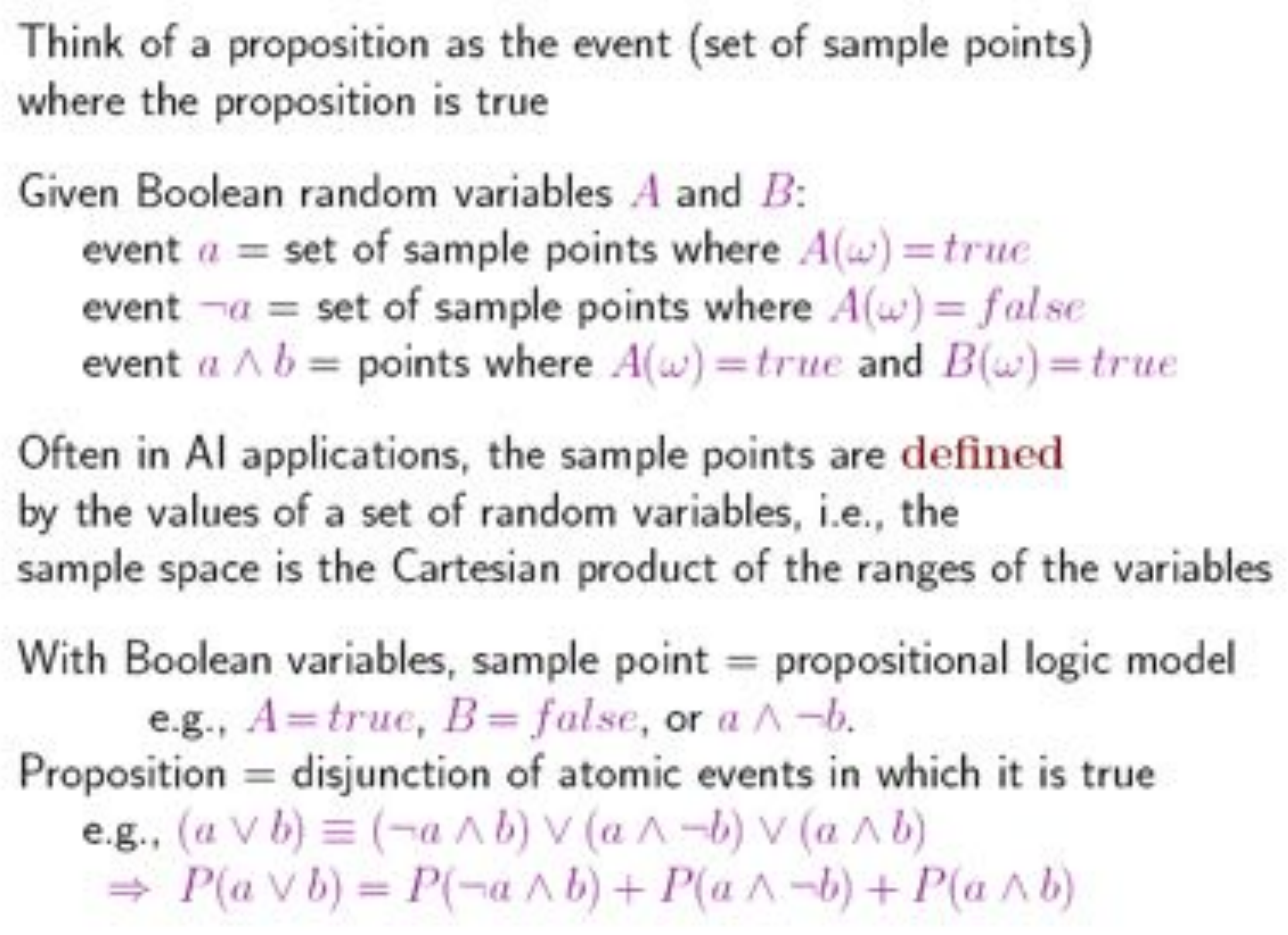
Given Boolean random variables A and B:

event a = set of sample points where $A(\omega) = true$

event $\neg a$ = set of sample points where $A(\omega) = false$

event $a \wedge b$ = points where $A(\omega) = true$ and $B(\omega) = true$

Propositions



Think of a proposition as the event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

event a = set of sample points where $A(\omega) = true$
event $\neg a$ = set of sample points where $A(\omega) = false$
event $a \land b$ = points where $A(\omega) = true$ and $B(\omega) = true$

Often in AI applications, the sample points are defined by the values of a set of random variables, i.e., the sample space is the Cartesian product of the ranges of the variables

With Boolean variables, sample point = propositional logic model

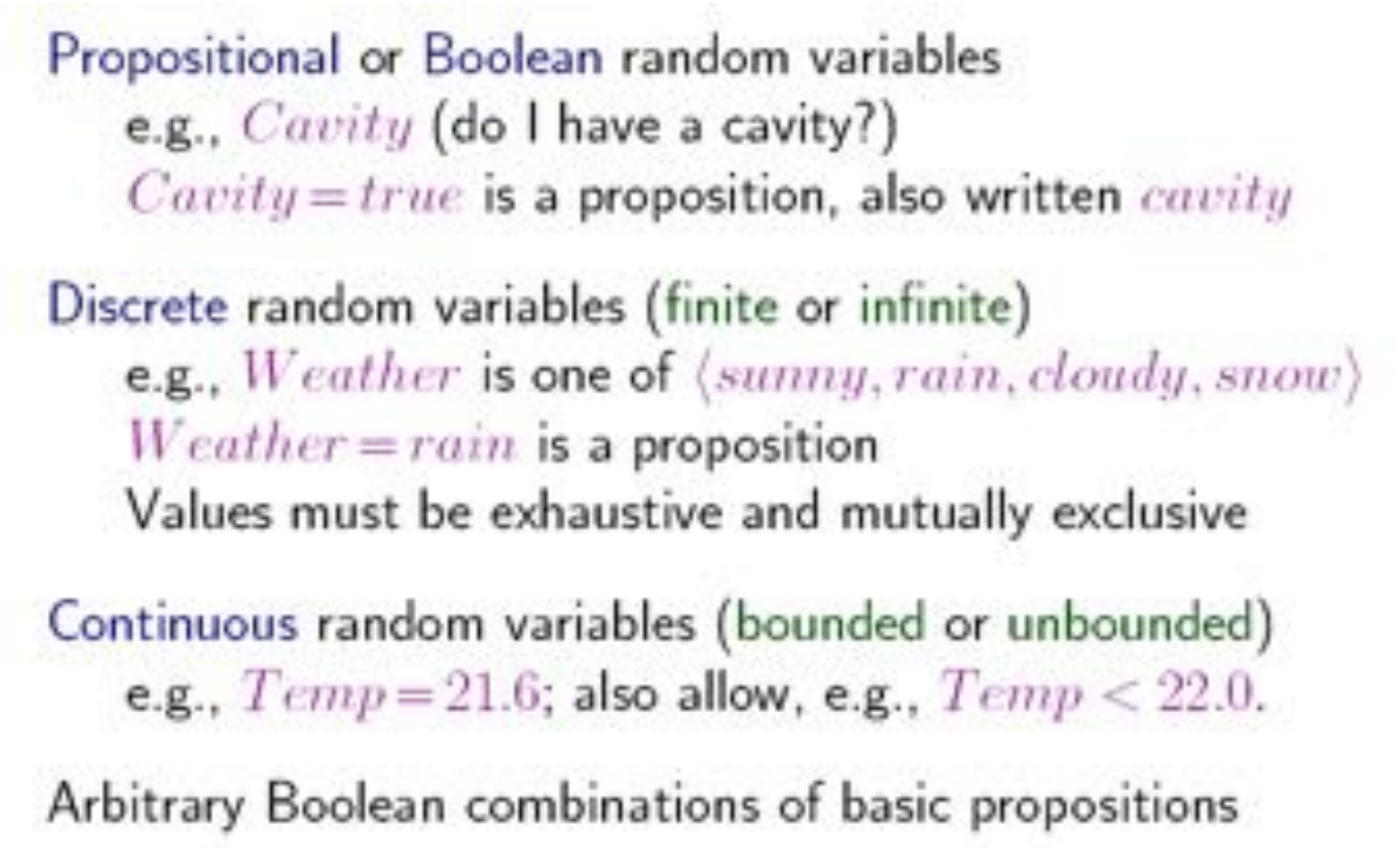
e.g., $A = true$, $B = false$, or $a \land \neg b$.

Proposition = disjunction of atomic events in which it is true

e.g., $(a \lor b) \equiv (\neg a \land b) \lor (a \land \neg b) \lor (a \land b)$

$\Rightarrow P(a \lor b) = P(\neg a \land b) + P(a \land \neg b) + P(a \land b)$

Syntax for Propositions



Propositional or Boolean random variables

e.g., $Cavity$ (do I have a cavity?)

$Cavity = true$ is a proposition, also written $cavity$

Discrete random variables (finite or infinite)

e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow\rangle$

$Weather = rain$ is a proposition

Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)

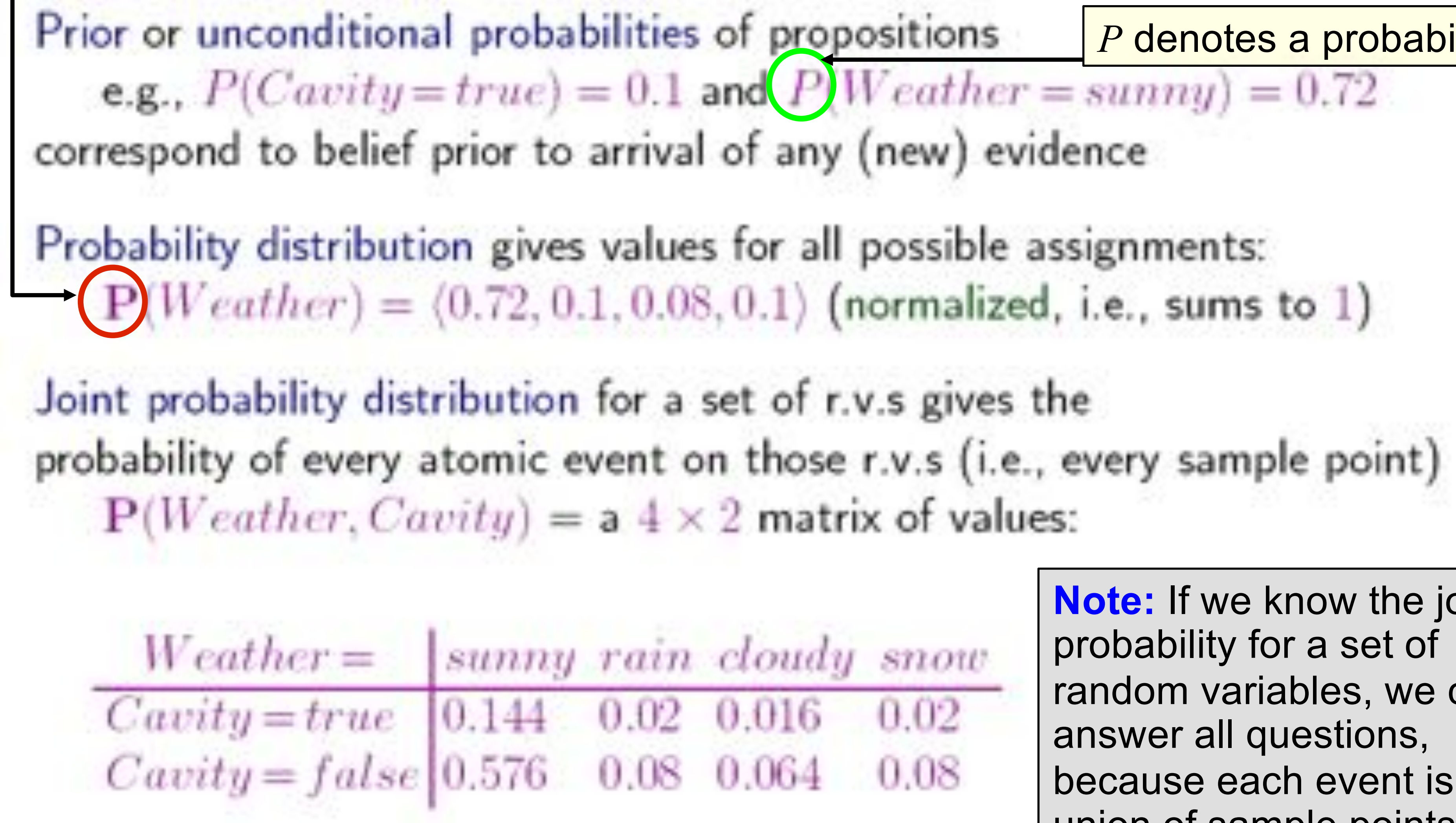
e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Arbitrary Boolean combinations of basic propositions

(Joint) Probability Distribution

P denotes a probability distribution

P denotes a probability



Prior or unconditional probabilities of propositions

e.g., $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

$\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

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

Note: If we know the joint probability for a set of random variables, we can answer all questions, because each event is a union of sample points

Marginalization (Summing Out)

Marginalization (aka *Summing Out*)

- For any set of variables $\mathbf{Y}$ and $\mathbf{Z}$

$$\mathbf{P}(\mathbf{Y}) = \sum_z \mathbf{P}(\mathbf{Y}, z)$$

- In particular, this means that given the joint probability distribution, the probability distribution of any random variable can be computed by summing out
  - the resulting distribution is then also called marginal distribution and its probabilities the marginal probabilities

Marginalization (Summing Out)

Marginalization (aka *Summing Out*)

- For any set of variables $\mathbf{Y}$ and $\mathbf{Z}$

$$\mathbf{P}(\mathbf{Y}) = \sum_z \mathbf{P}(\mathbf{Y}, z)$$

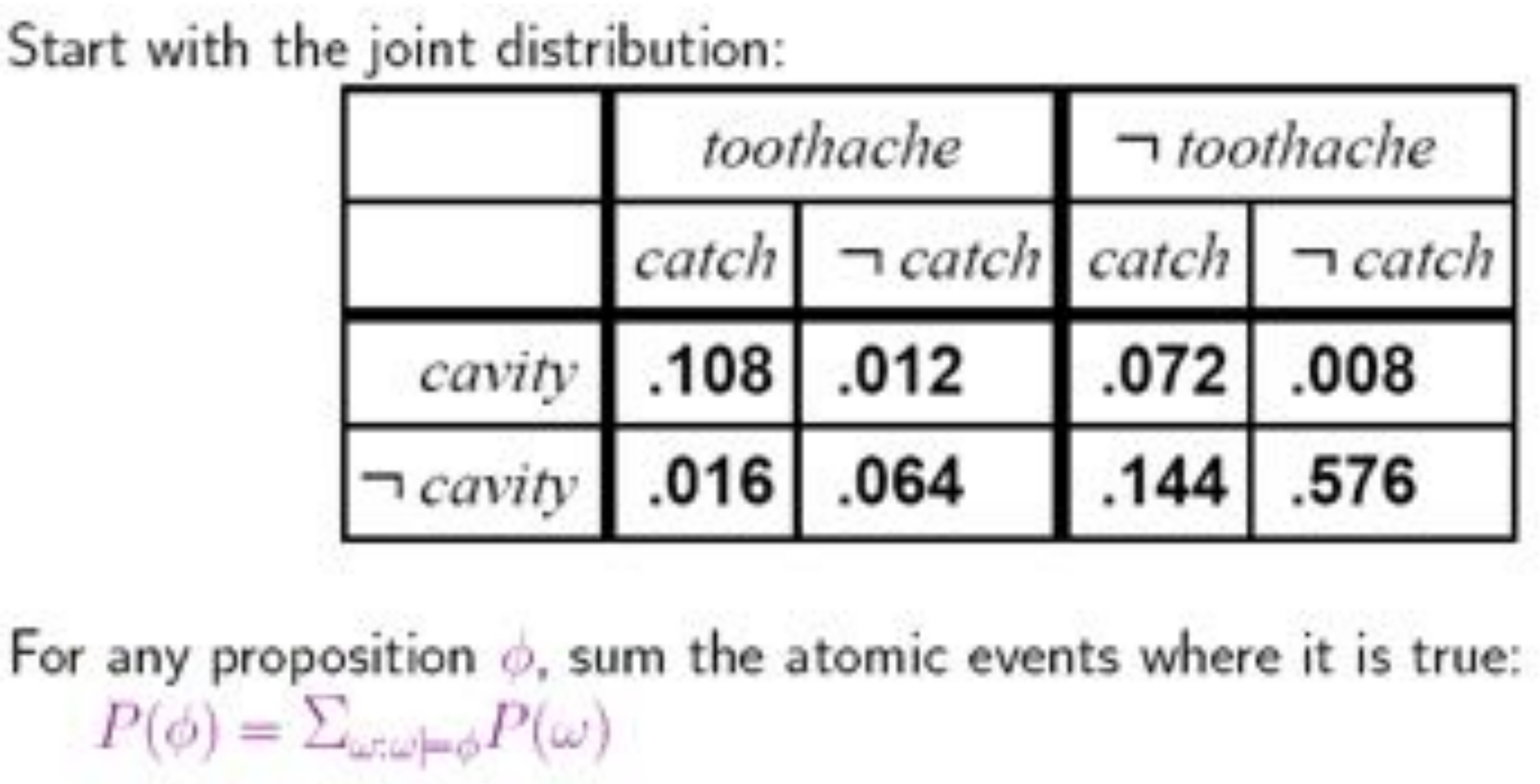
- In particular, this means that given the joint probability distribution, the probability distribution of any random variable can be computed by summing out
  - the resulting distribution is then also called marginal distribution and its probabilities the marginal probabilities

Conditioning

- A variant of the above rule that uses conditional probabilities

$$\mathbf{P}(\mathbf{Y}) = \sum_z \mathbf{P}(\mathbf{Y}|z) \cdot P(z)$$

Marginalization



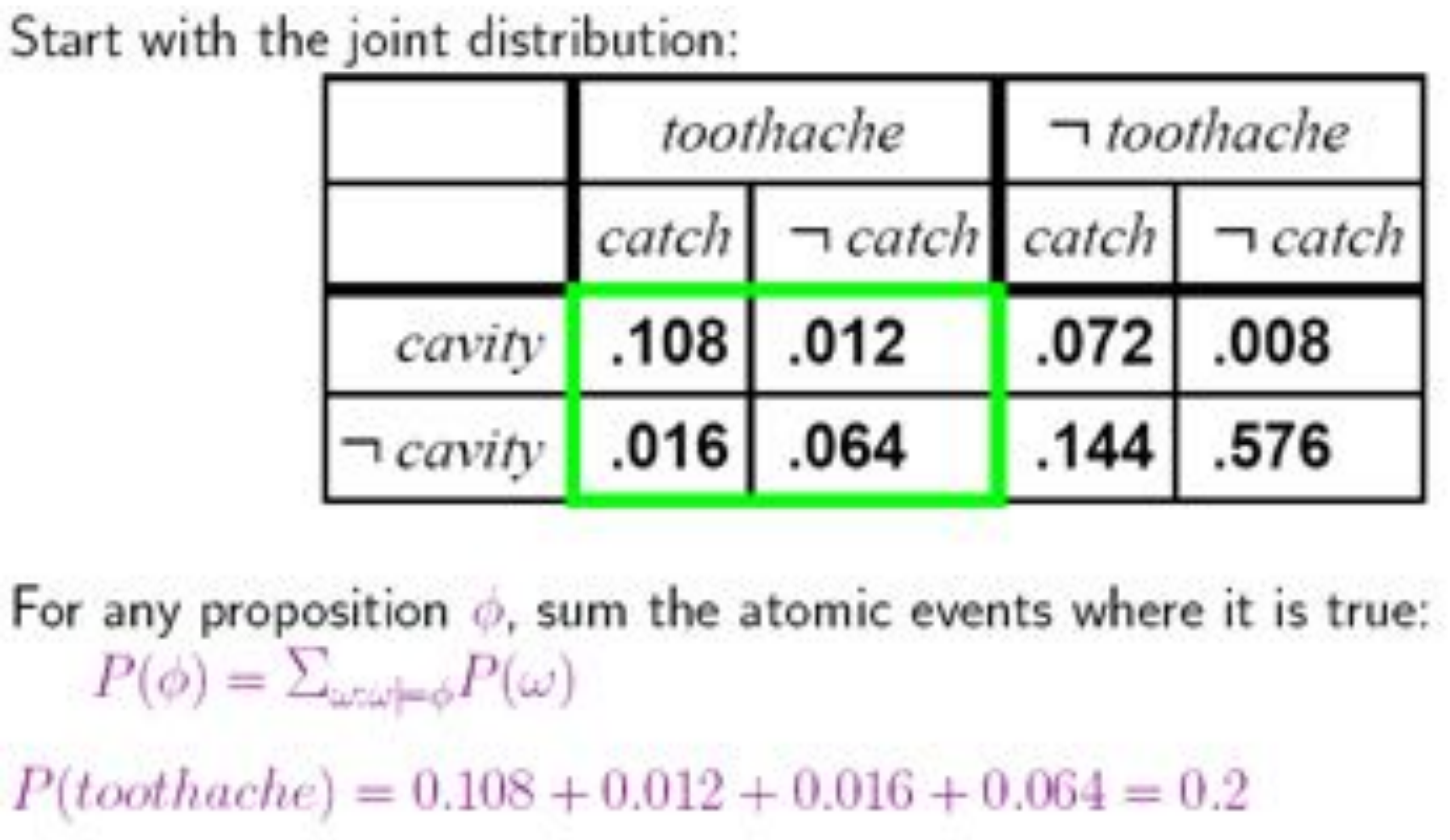
Start with the joint distribution:

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

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

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

Marginalization



Start with the joint distribution:

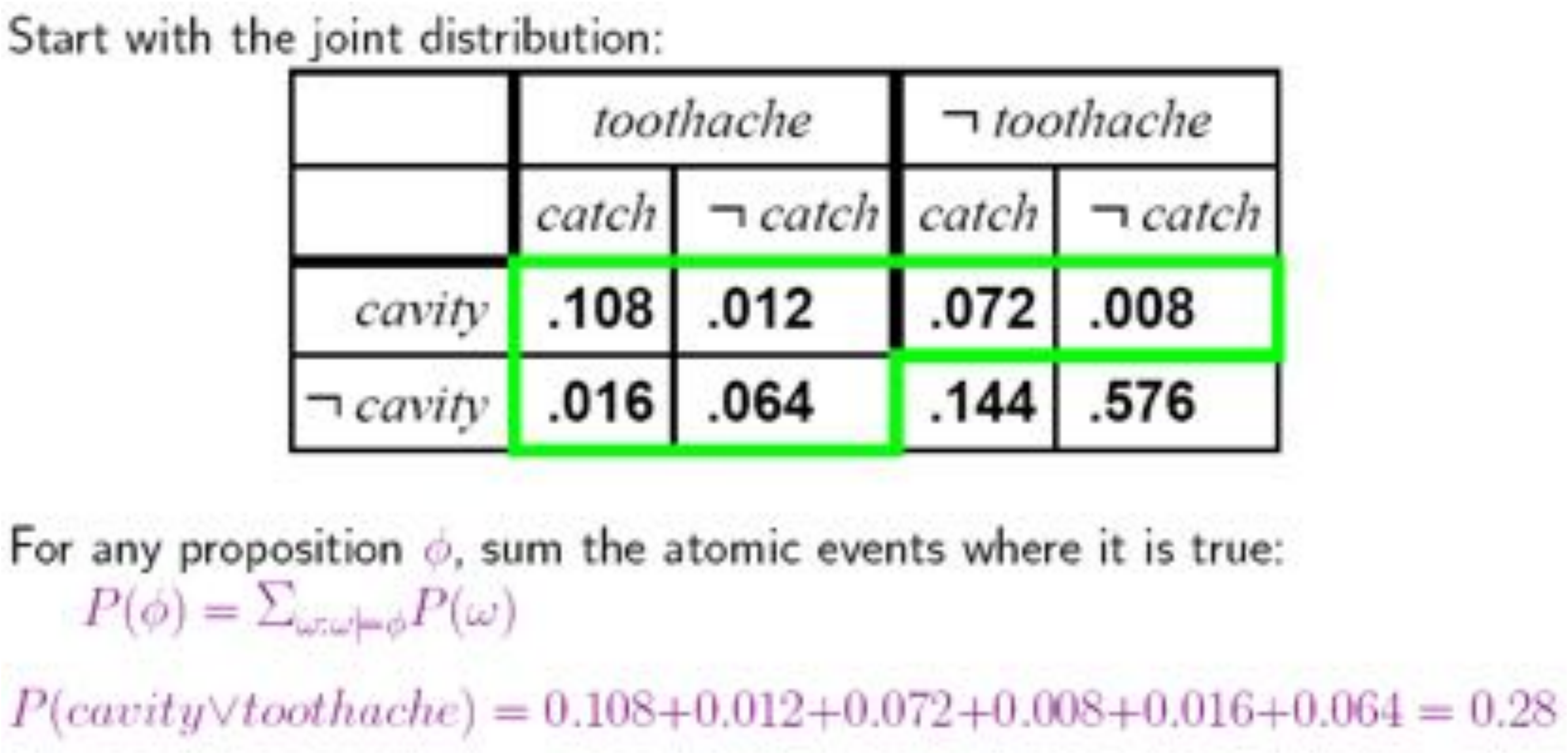
	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

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

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

$$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inference by Enumeration



Start with the joint distribution:

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

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

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

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

Conditional Probabilities

$\mathbf{P}(Cavity, Toothache)$	$toothache$	$\neg\, toothache$
$cavity$	0.12	0.08
$\neg\, cavity$	0.08	0.72

Conditional or posterior probabilities

e.g., $P(cavity\,|toothache) = 0.6$

i.e., **given that** *toothache* **is all I know**

NOT "if *toothache* then 60% chance of *cavity*"

(Notation for conditional distributions:

$\mathbf{P}(Cavity|Toothache)$ = 2-element vector of 2-element vectors)

$$\mathbf{P}\ \ Cavity|Toothache = \langle\langle 0.6, 0.4\rangle, \langle 0.1, 0.9\rangle\rangle$$

If we know more, e.g., *cavity* is also given, then we have

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

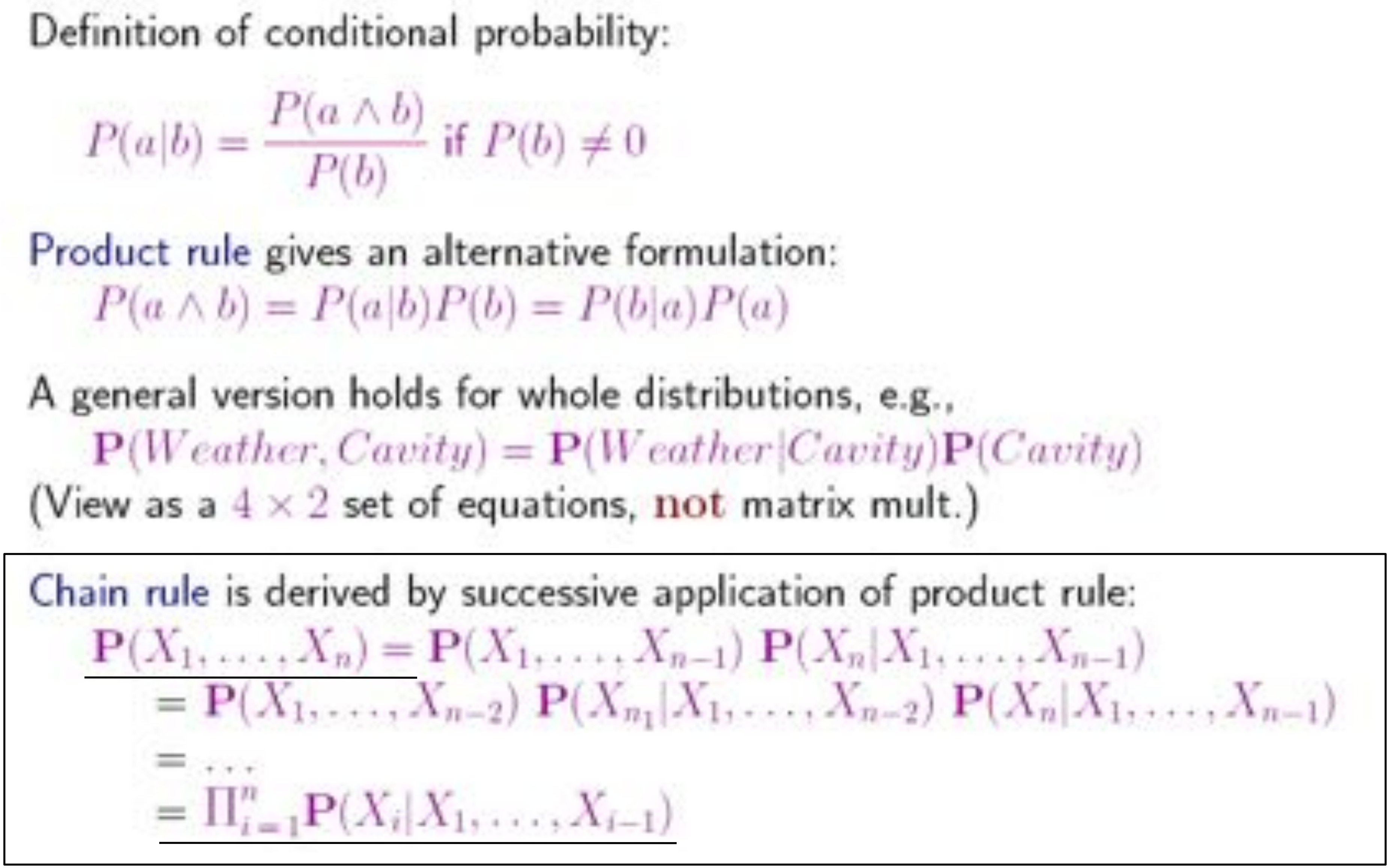
Note: the less specific belief **remains valid** after more evidence arrives, but is not always **useful**

New evidence may be irrelevant, allowing simplification, e.g.,

$P(cavity\,|toothache\,, sunny) = P(cavity\,|toothache) = 0.6$

This kind of inference, sanctioned by domain knowledge, is crucial

Definition of Conditional Probability



Definition of conditional probability:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Product rule gives an alternative formulation:

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

A general version holds for whole distributions, e.g.,

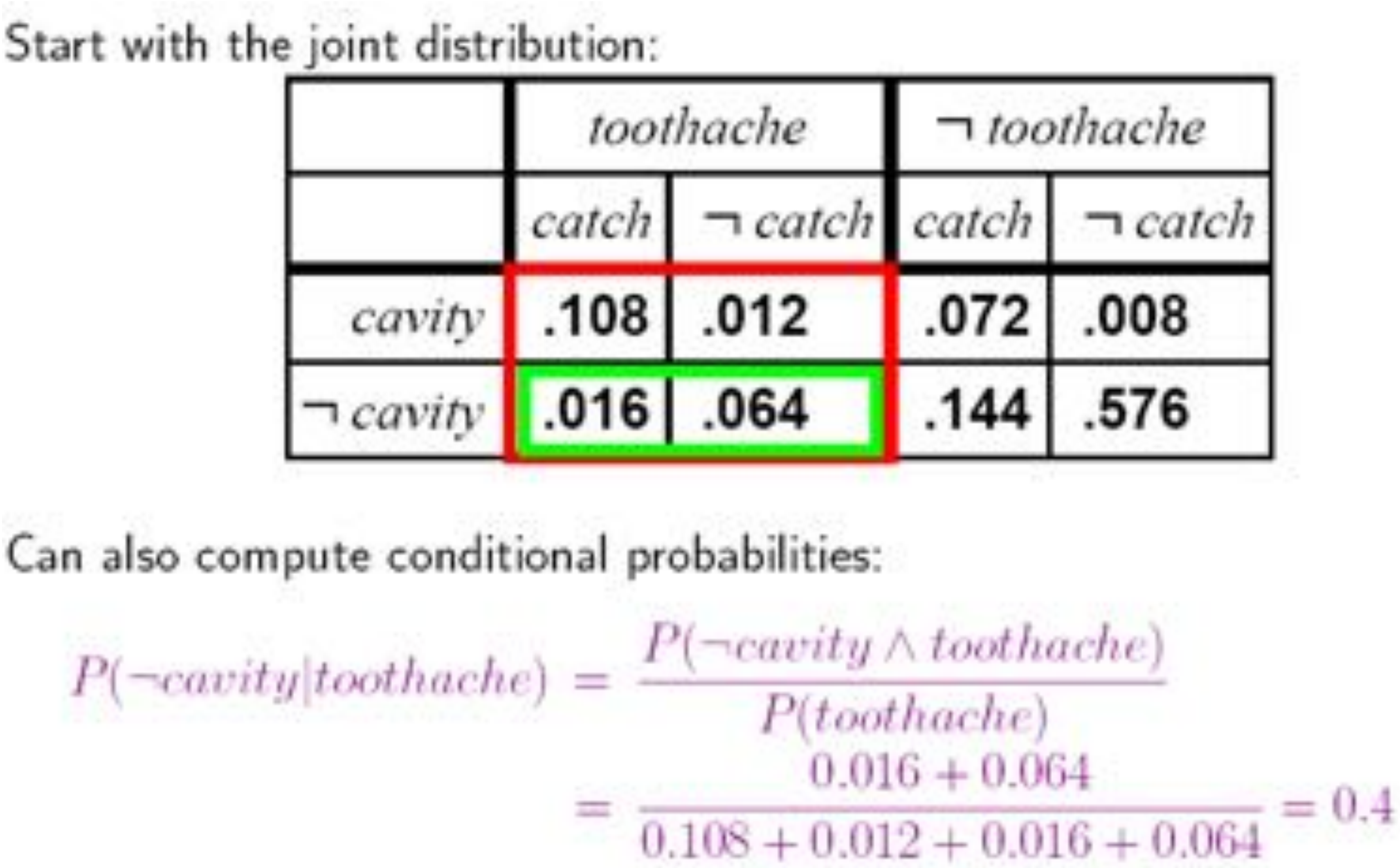
$$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$$

(View as a 4×2 set of equations, **not** matrix mult.)

Chain rule is derived by successive application of product rule:

$$\begin{aligned}\underline{\mathbf{P}(X_1, \ldots, X_n) =} &\ \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n_1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &\underline{= \Pi_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})}\end{aligned}$$

Inference by Enumeration



Start with the joint distribution:

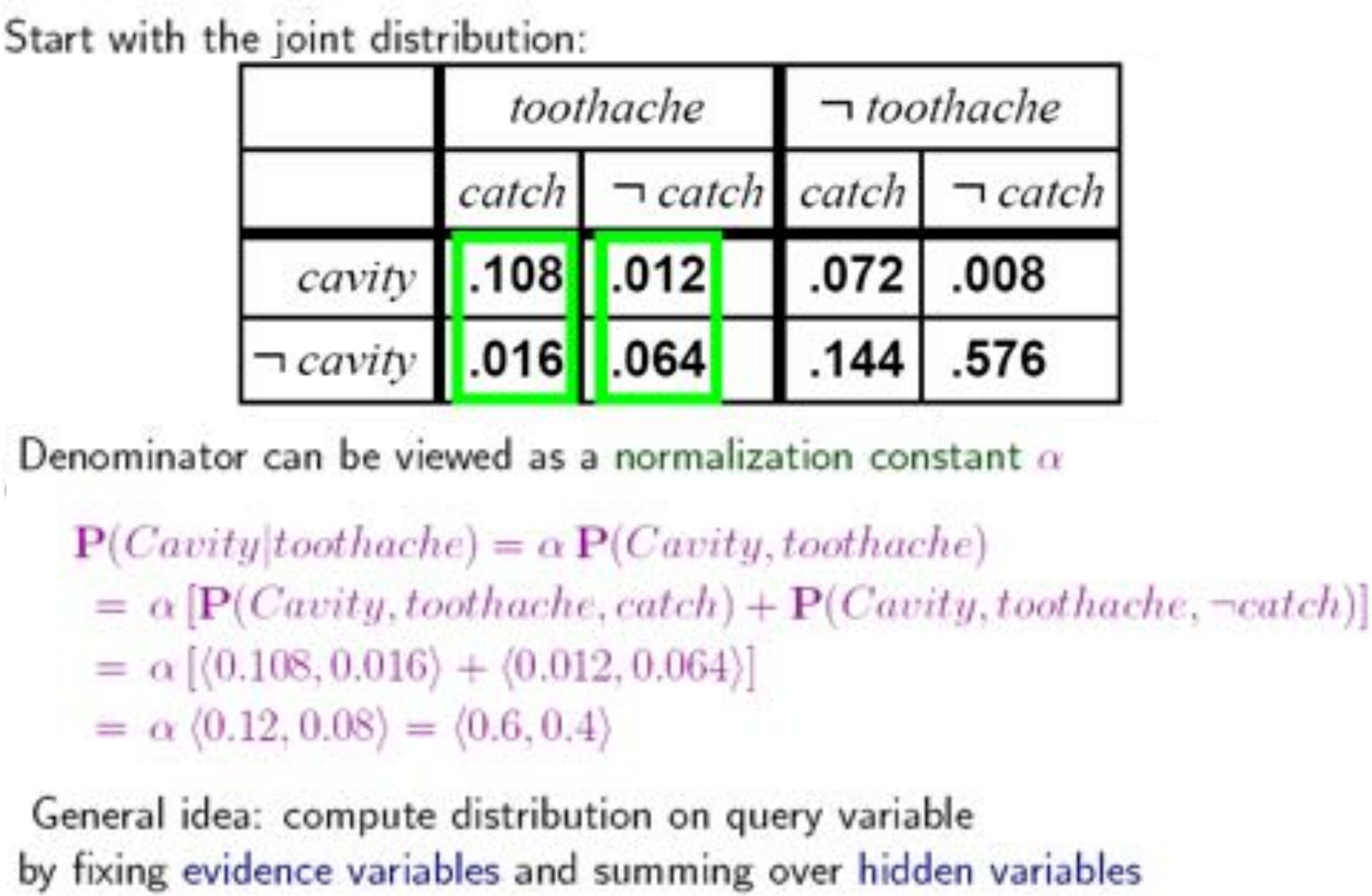
	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Can also compute conditional probabilities:

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

$$= \frac{0.016+0.064}{0.108+0.012+0.016+0.064} = 0.4$$

Normalization



Start with the joint distribution:

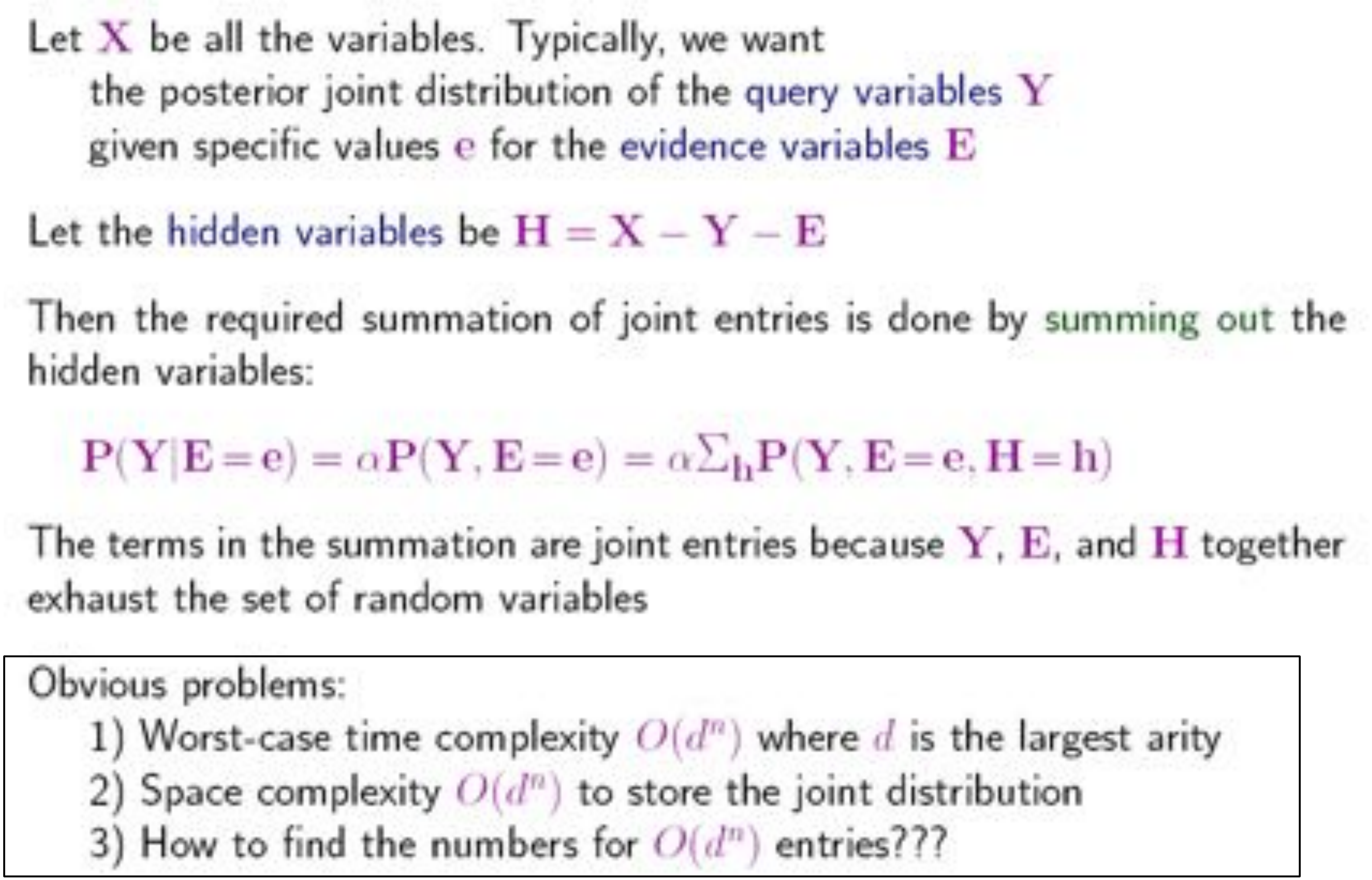
	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

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

Inference by Enumeration (Ctd.)



Let $\mathbf{X}$ be all the variables. Typically, we want
the posterior joint distribution of the query variables $\mathbf{Y}$
given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$

Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

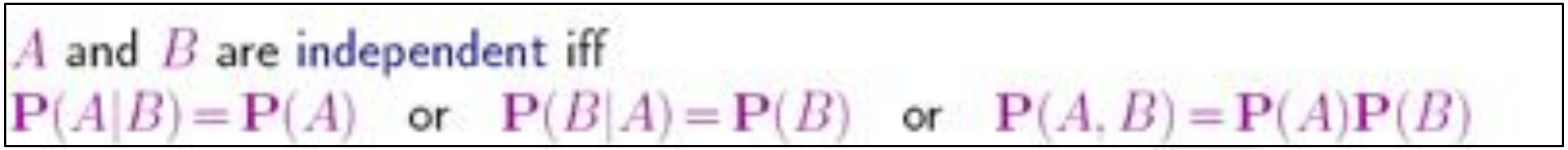
Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \alpha \Sigma_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}, \mathbf{H}=\mathbf{h})$$

The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables

Obvious problems:
1) Worst-case time complexity $O(d^n)$ where d is the largest arity
2) Space complexity $O(d^n)$ to store the joint distribution
3) How to find the numbers for $O(d^n)$ entries???

Independence



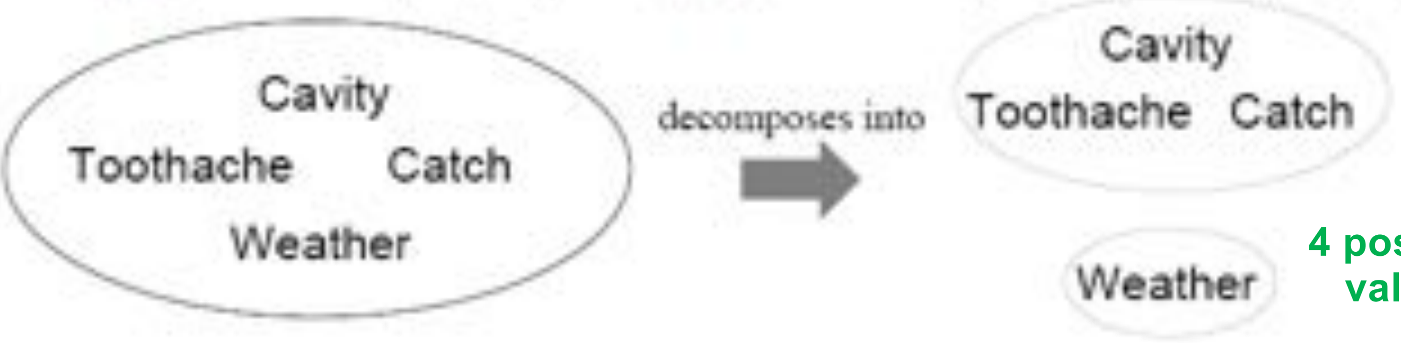
A and B are independent iff

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

Independence

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

2x2x2x4 = 32 possible values



2x2x2 = 8 possible values

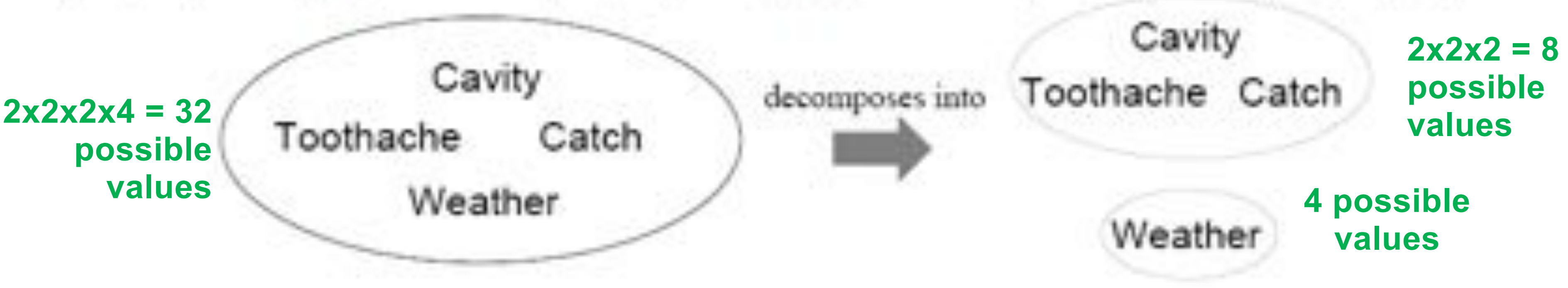
4 possible values

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

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

Independence

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



2x2x2x4 = 32 possible values

Cavity
Toothache Catch
Weather

decomposes into

Cavity
Toothache Catch

2x2x2 = 8 possible values

Weather

4 possible values

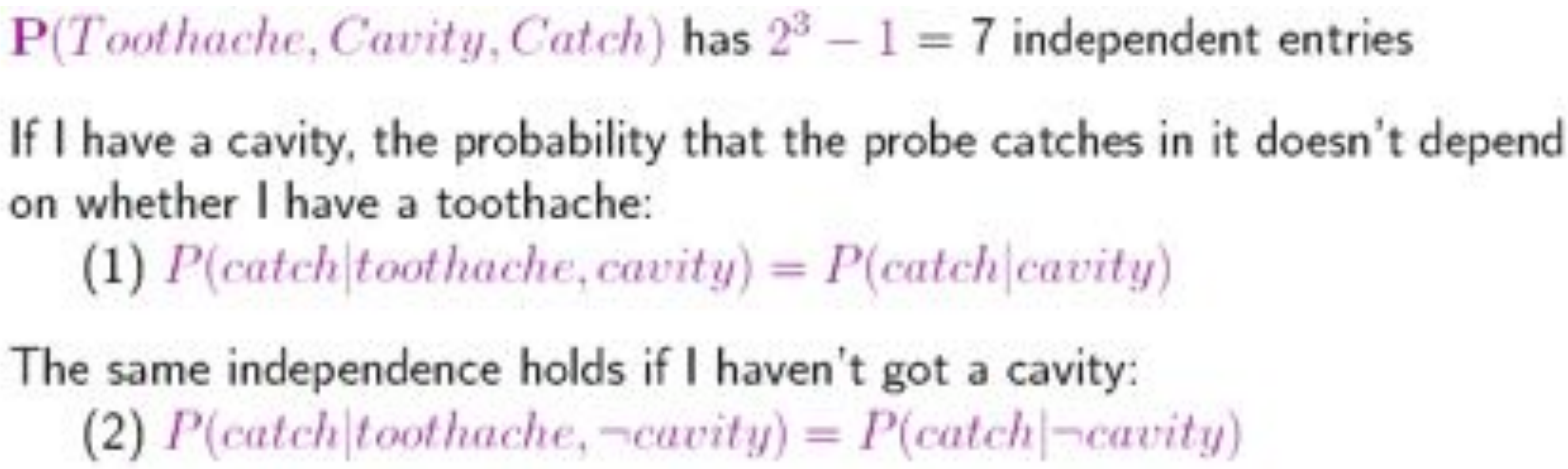
$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

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



$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries

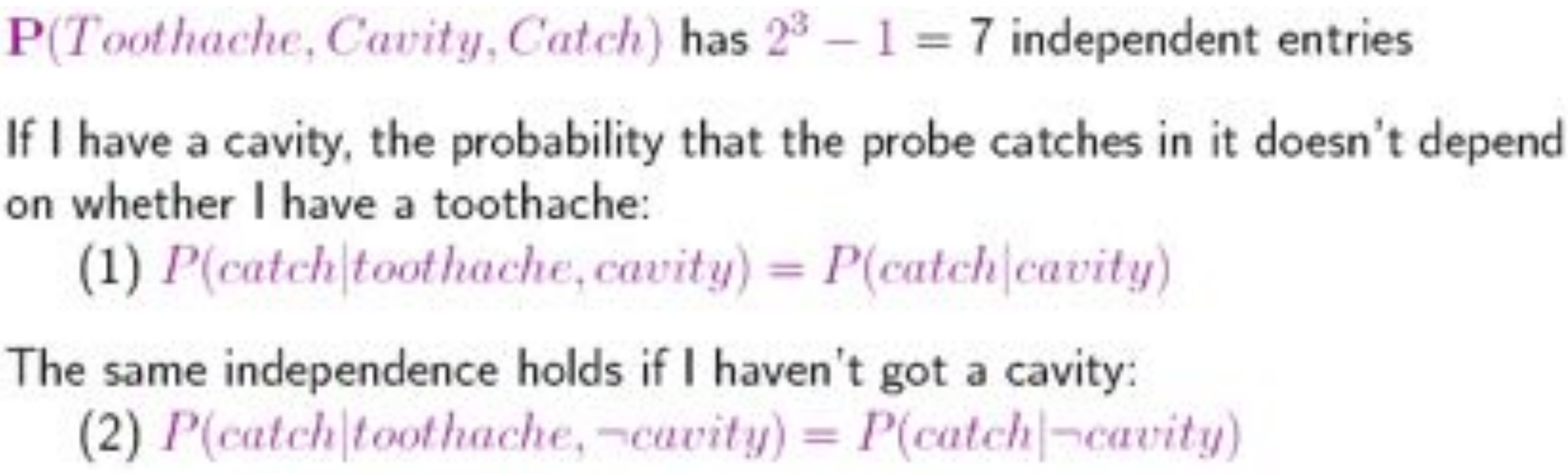
If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

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

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

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

Conditional Independence



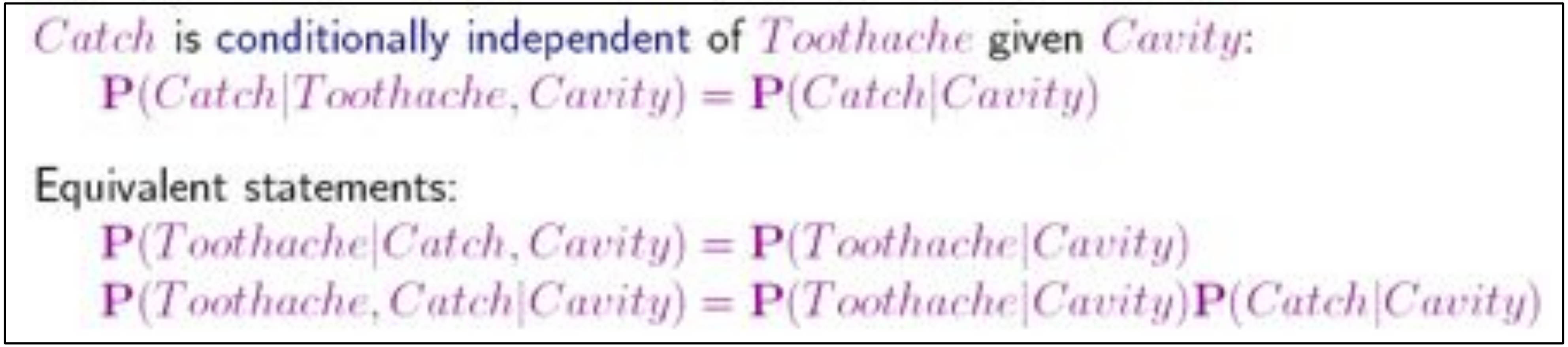
$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries

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 $Cavity$:

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

Equivalent statements:

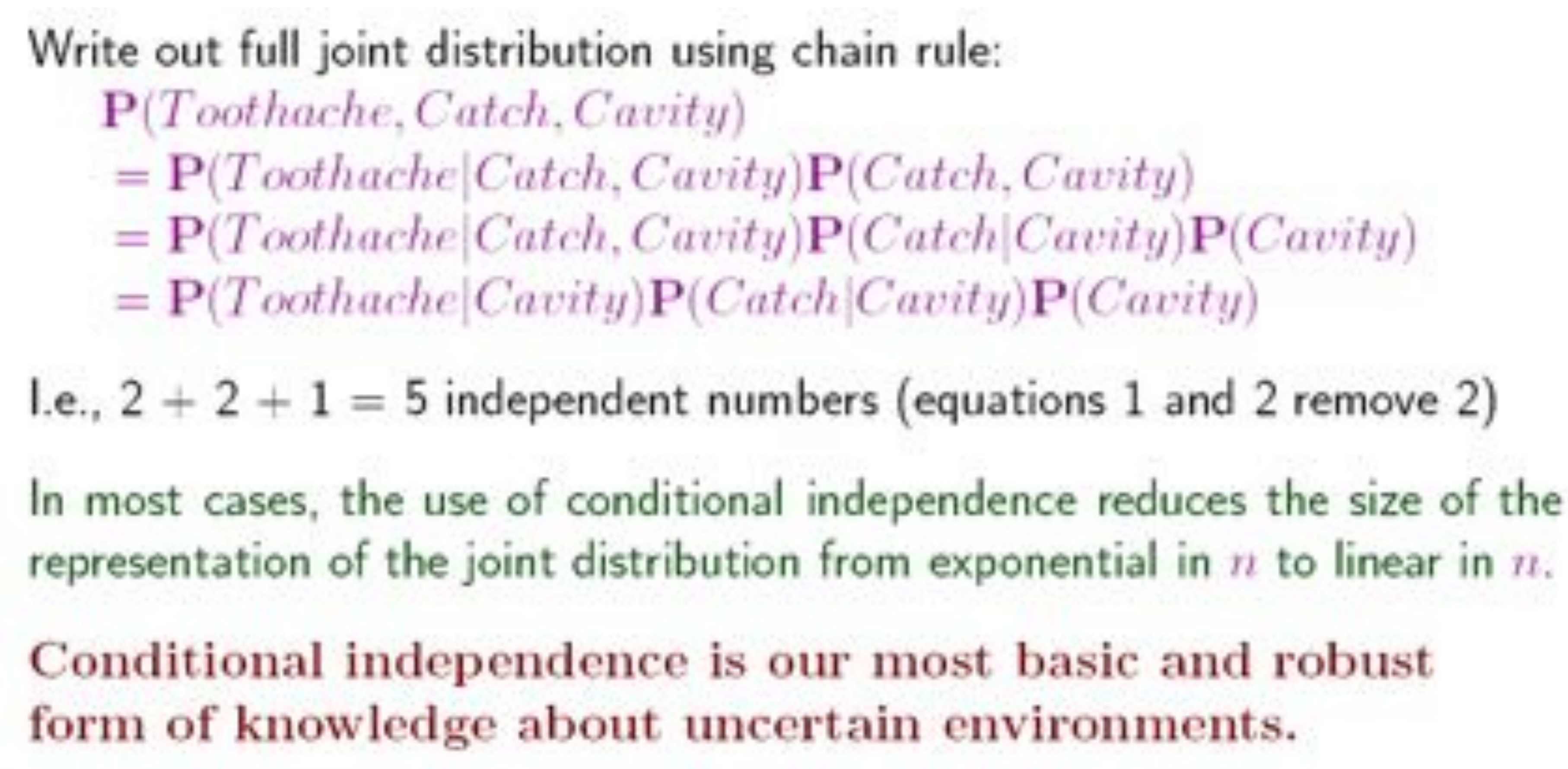
$$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$$

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

Analogous to:

$$\mathbf{P}(A|B) = \mathbf{P}(A) \quad \text{or} \quad \mathbf{P}(B|A) = \mathbf{P}(B) \quad \text{or} \quad \mathbf{P}(A,B) = \mathbf{P}(A)\mathbf{P}(B)$$

Conditional Independence (Ctd.)



Write out full joint distribution using chain rule:

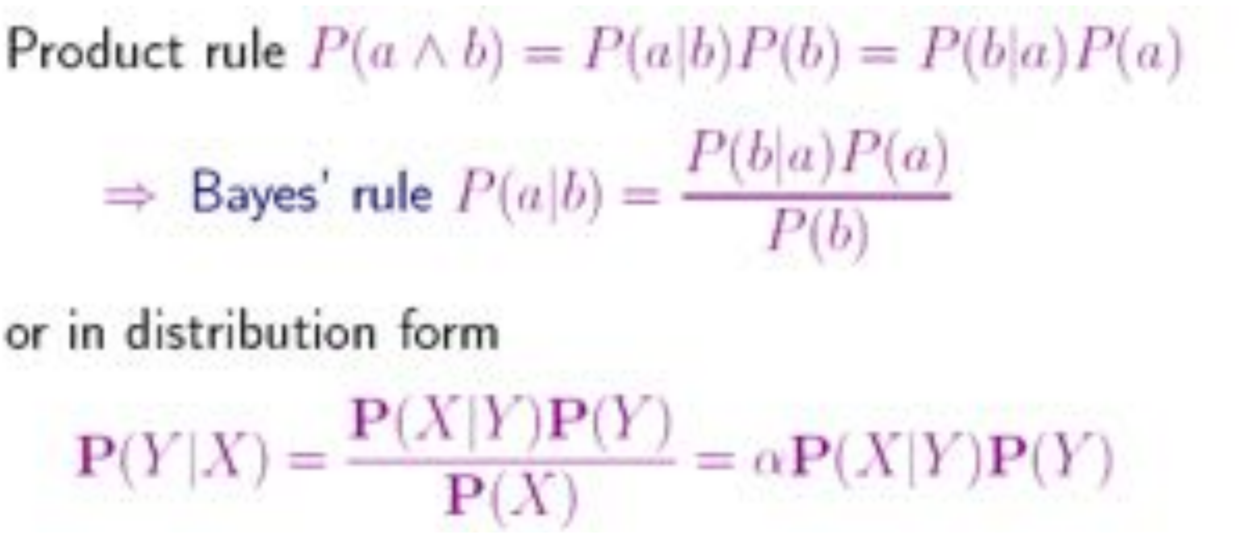
$$\begin{aligned}
&\mathbf{P}(Toothache, Catch, Cavity)\\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity)\\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)\\
&= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

I.e., $2 + 2 + 1 = 5$ independent numbers (equations 1 and 2 remove 2)

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.

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes Rule



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

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

or in distribution form

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

Bayes Rule

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

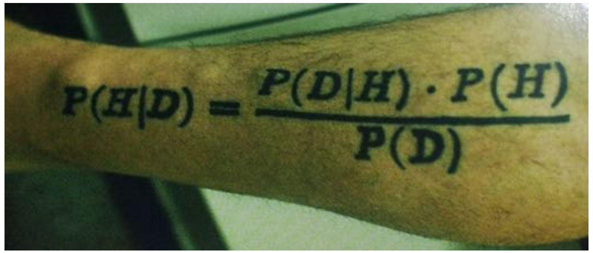
$\Rightarrow$ Bayes' rule $P(a|b) = \dfrac{P(b|a)P(a)}{P(b)}$

or in distribution form

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

Useful for assessing diagnostic probability from causal probability:

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



Tattoo: Gregory von Nessi Foto: Carl Zimmer

E.g., let M be meningitis, S be stiff neck:

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

Note: posterior probability of meningitis still very small!

Example: AIDS-Test

- event $Aids$ = a person has Aids or not
- event $Positive$ = a person has a positive test result

- Assume the test has the following characteristics:

$$P(positive|aids) = 0.99$$
$$P(negative|aids) = 0.01$$

The test makes 1% mistakes
for people that have aids

$$P(positive|\neg aids) = 0.005$$
$$P(negative|\neg aids) = 0.995$$

The test makes 0,5% mistakes
for people that don't have aids

- Looks like a pretty reliable test?

Example: AIDS-Test

- event $Aids$ = a person has Aids or not
- event $Positive$ = a person has a positive test result

- Assume the test has the following characteristics:

$$P\ \ positive|aids\ = 0.99$$
$$P\ \ negative|aids\ = 0.01$$

The test makes 1% mistakes for people that have aids

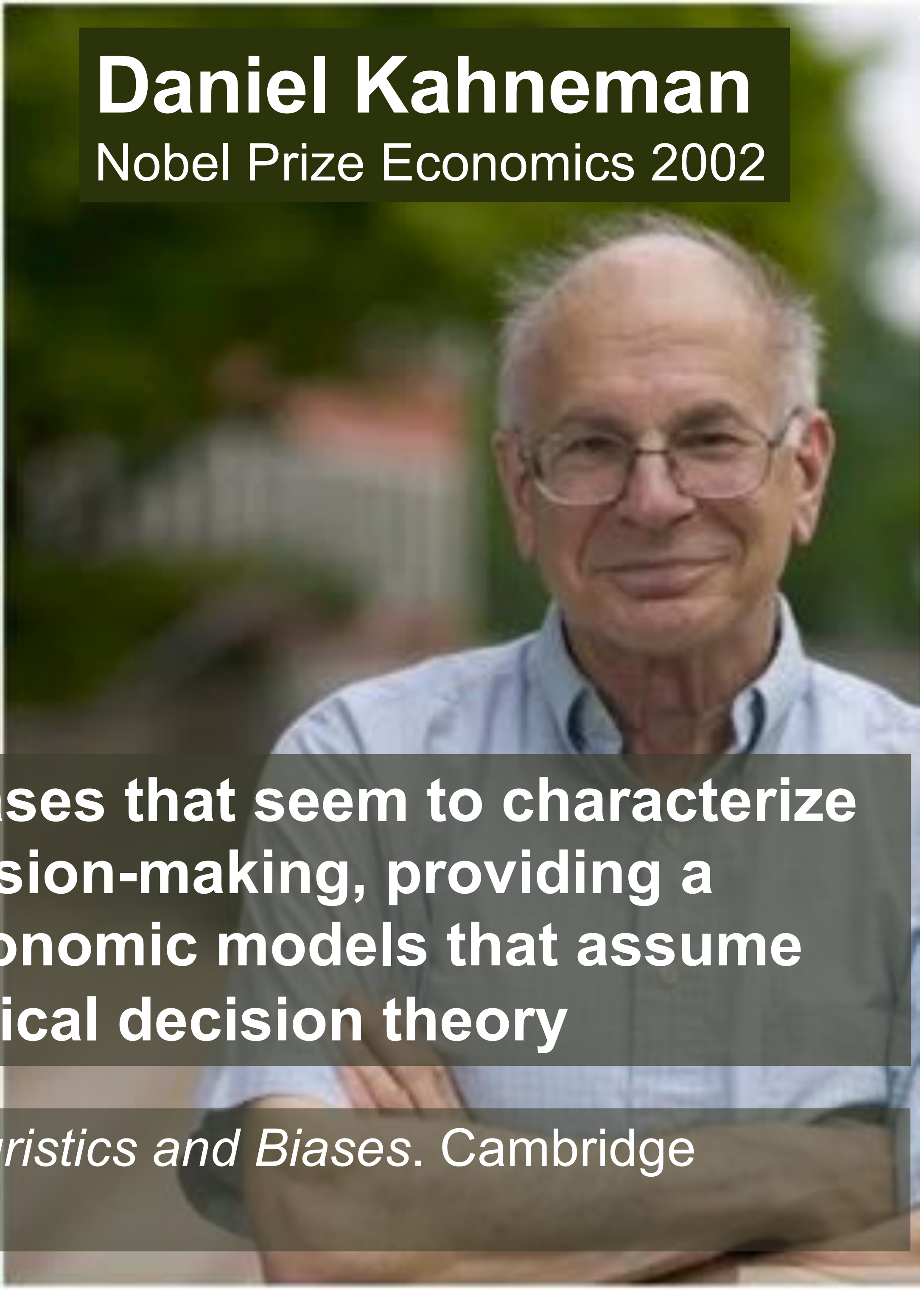
$$P\ \ positive|\neg aids\ = 0.005$$
$$P\ \ negative|\neg aids\ = 0.995$$

The test makes 0,5% mistakes for people that don't have aids

- Now suppose you are in a low-risk group (low a priori probability of having Aids, say $P(aids) = 0.0001$) and have a positive test result. Should you panic?

$$P(a|p) = \frac{P(p|a)\cdot P(a)}{P(p)} = \frac{P(p|a)\cdot P(a)}{P(p|a)\cdot P(a) + P(p|\neg a)\cdot P(\neg a)} = \frac{0.99\cdot 0.0001}{0.99\cdot 0.0001 + 0.005\cdot 0.9999} = 0.0194$$

Amos Tversky



Daniel Kahneman

Nobel Prize Economics 2002

Uncovered a number of biases that seem to characterize human reasoning and decision-making, providing a significant challenge to economic models that assume people simply apply statistical decision theory

Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press 1982

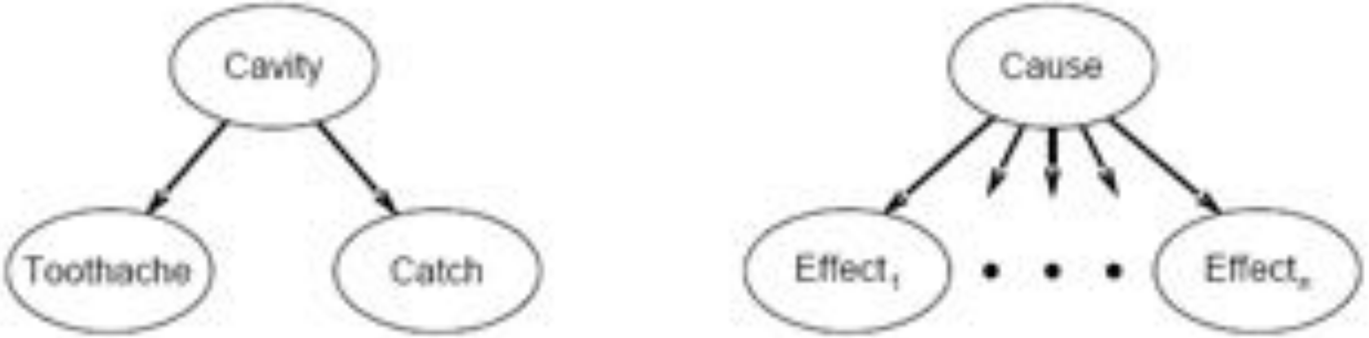
Bayes Rule and Independence

The model is naïve because it assumes that all effects are independent given the cause (which is often not true)

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

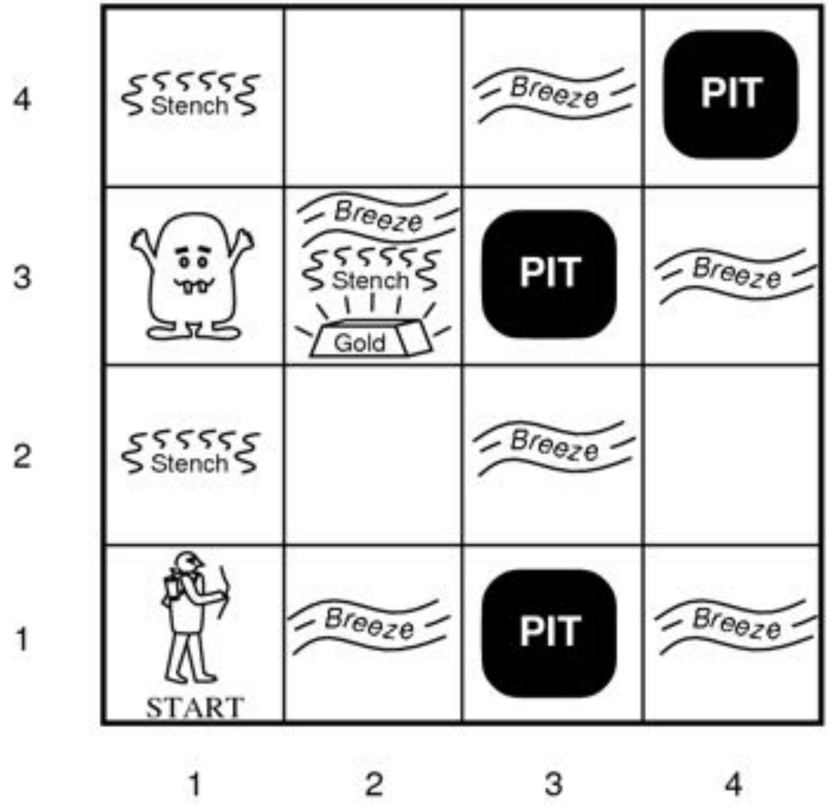
This is an example of a naive Bayes model:

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



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

Example: Wumpus World



Performance measure

gold +1000, death -1000

-1 per step, -10 for using the arrow

Environment

Squares adjacent to wumpus are smelly

Squares adjacent to pit are breezy

Glitter iff gold is in the same square

Shooting kills wumpus if you are facing it

Shooting uses up the only arrow

Grabbing picks up gold if in same square

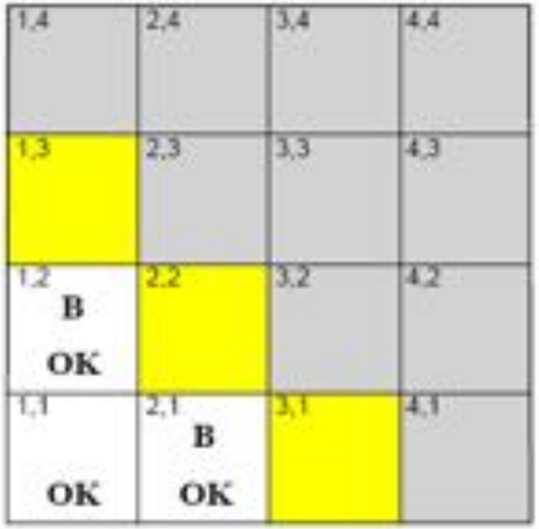
Releasing drops the gold in same square

Actuators Left turn, Right turn,
Forward, Grab, Release, Shoot

Sensors Breeze, Glitter, Smell

Example: Wumpus World

Current knowledge of the agent about the world



- the agent has visited the squares [1,1], [1,2], [2,1]. They are **OK**
- it found a **B**reeze in [1,2] and one in [2,1].
- therefore, no safe explorative step is possible
    - all yellow squares might contain a pit

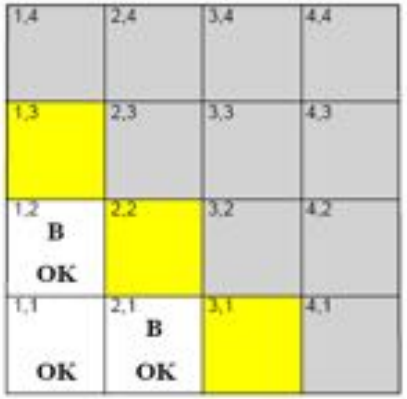
→ **Which of the yellow squares is the safest?**

Example: Wumpus World
Specifying the Probability Model

$P_{ij} = true$ iff $[i, j]$ contains a pit

$B_{ij} = true$ iff $[i, j]$ is breezy

Include only $B_{1,1}, B_{1,2}, B_{2,1}$ in the probability model



The full joint distribution is $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Apply product rule: $\boxed{\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})}$

(Do it this way to get $P(Effect|Cause)$.)

First term: 1 if pits are adjacent to breezes, 0 otherwise

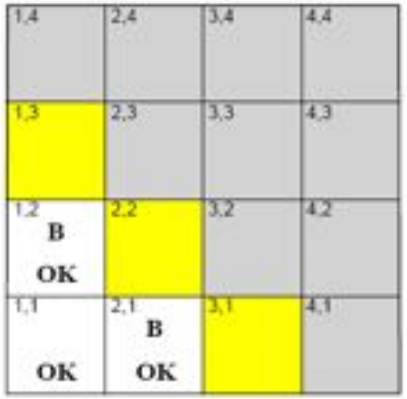
Second term: pits are placed randomly, probability 0.2 per square:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \Pi_{i,j=1,1}^{4,4} \mathbf{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

for n pits.

Example: Wumpus World
Observations and Queries

$P_{ij} = true$ iff $[i, j]$ contains a pit



$B_{ij} = true$ iff $[i, j]$ is breezy

Include only $B_{1,1}, B_{1,2}, B_{2,1}$ in the probability model

We know the following facts:

$$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$$
$$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

Query is $\mathbf{P}(P_{1,3} | known, b)$ ⟵ What is the probability distribution for a pit on [1,3]?

Define $Unknown = P_{ij}$s other than $P_{1,3}$ and $Known$

For inference by enumeration, we have

$$\mathbf{P}(P_{1,3} | known, b) = \alpha \Sigma_{unknown} \underline{\mathbf{P}(P_{1,3}, unknown, known, b)}$$

- There are 12 unknown squares
- The summation contains 212 = 4096 terms

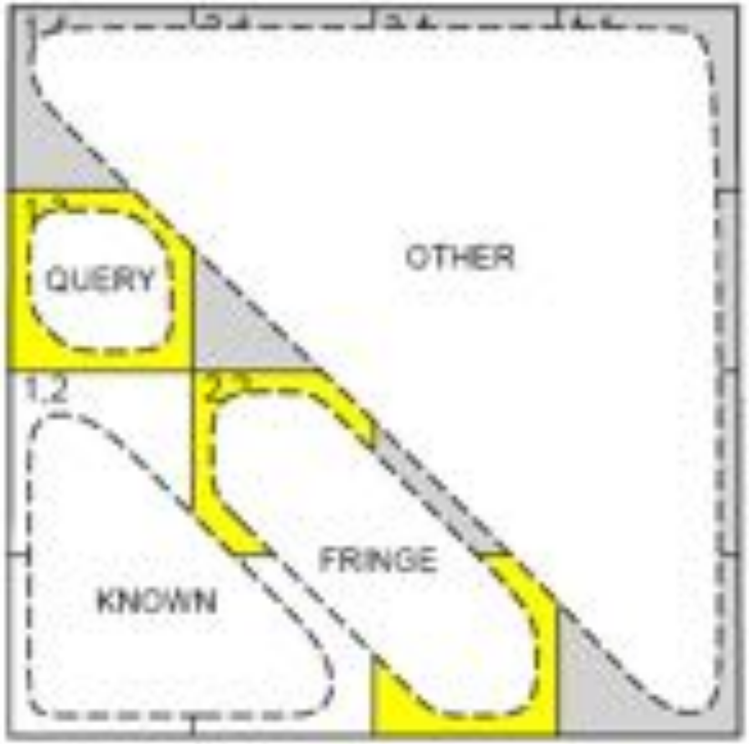
In general the summation grows expoentially with the number of squares!

Example: Wumpus World
Using Conditional Independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares

The square [4,4] will not have an influence on whether the agent has noticed a breeze on [1,2] or not.

In fact, none of the squares in the *Other* region may have influenced the observations in [1,1], [1,2] and [2,1].

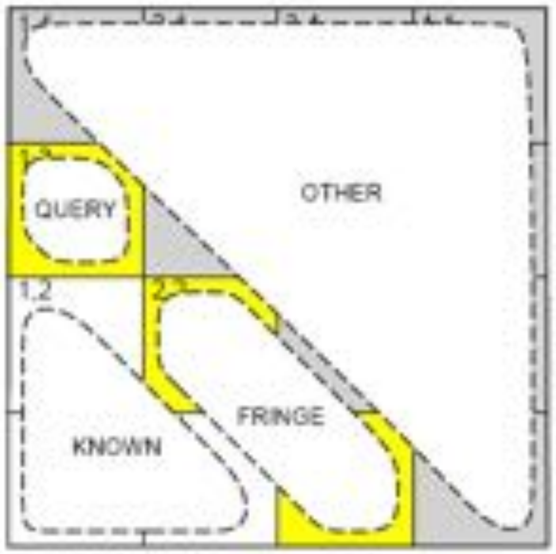


Define $Unknown = Fringe \cup Other$

$\mathbf{P}(b|P_{1,3}, Known, Unknown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$

Manipulate query into a form where we can use this!

Example: Wumpus World
Computation



The query $\mathbf{P}(P_{1,3}|known,b)$ is now transformed in a way so that we can use the equation from the previous slide

Inference by enumration

$$\mathbf{P}(P_{1,3}|known,b) = \alpha \sum_{unknown} \underline{\mathbf{P}(P_{1,3}, unknown, known, b)}$$

$$= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown)\mathbf{P}(P_{1,3}, known, unknown)$$

product rule

$$= \alpha \sum_{fringe}\sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other)$$

conditioning

$$= \alpha \sum_{fringe}\sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other)$$

conditional independence

$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other)$$

pushing sums inwards

$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other)$$

independence

$$= \alpha\, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other)$$

$$= \alpha'\, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe)$$

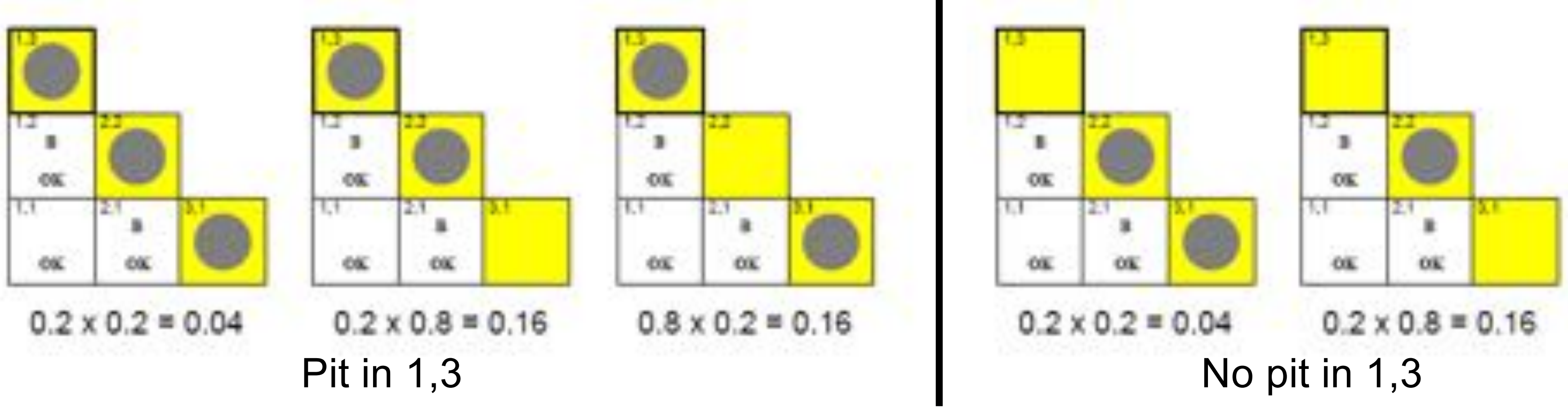
reordering, pushing sums inwards, simplifying

Example: Wumpus World
Computation

We go through all possibilities (filled circle denotes a pit).

is 1 if the breeze observations b are consistent with the fringe, 0 otherwise

$$\mathbf{P}(P_{1,3}|known,b)=\alpha'\mathbf{P}(P_{1,3})\sum_{fringe}\mathbf{P}(b|known,P_{1,3},fringe)P(fringe)$$



Pit in 1,3

No pit in 1,3

$$\begin{aligned}\mathbf{P}(P_{1,3}|known,b) &= \alpha'\langle 0.2(0.04+0.16+0.16),\ 0.8(0.04+0.16)\rangle\\ &\approx \langle 0.31, 0.69\rangle\end{aligned}$$

$$\mathbf{P}(P_{2,2}|known,b) \approx \langle 0.86, 0.14\rangle$$

(by analogous computation)

Summary

- Probability is a rigorous formalism for uncertain knowledge
- Joint probability distribution specifies probability of every atomic event
- Queries can be answered by summing over atomic events
- For nontrivial domains, we must find a way to reduce the joint size
- Independence and conditional independence provide the tools