# UNCERTAINTY - PROBABILITY

# UNCERTAINTY

Let action *At* = leave for airport *t* minutes before flight
Will *At* get me there on time?

Problems:

1. partial observability (road state, etc.)
2. multi-agent problem (other drivers' plans)
3. noisy sensors (uncertain traffic reports)
4. uncertainty in action outcomes (flat tire, etc.)
5. immense complexity of modeling and predicting traffic

# UNCERTAINTY

Let action $A_t$ = leave for airport $t$ minutes before flight  Will $A_t$ get me there on time?

Problems:

1. partial observability (road state, etc.)
2. multi-agent problem (other drivers' plans)
3. noisy sensors (uncertain traffic reports)
4. uncertainty in action outcomes (flat tire, etc.)
5. immense complexity of modeling and predicting traffic

Hence a purely logical approach either

1. risks falsehood: "$A_{25}$ will get me there on time", or
2. leads to conclusions that are too weak for decision making:

"$A_{25}$ will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact, etc., etc."

"$A_{1440}$ should get me there on time but I'd have to stay overnight in the airport."

# PROPOSITIONAL LOGIC AND PROBABILITY

Their ontological commitments are the same

The world is a set of facts that do or do not hold

**Ontology** is the philosophical study of the nature of being, becoming, existence, or reality; <u>what exists in the world?</u>

Their epistemological commitments differ

- **Logic agent** believes true, false, or no opinion
- **Probabilistic agent** has a numerical degree of belief between 0 (false) and 1 (true)
  - **Epistemology** is the philosophical study of the nature and scope of knowledge; <u>how, and in what way, do we know about the world?</u>

# UNCERTAINTY IN THE WORLD

- Uncertainty due to
  - Randomness
  - Overwhelming complexity
  - Lack of knowledge
  - …

# UNCERTAINTY IN THE WORLD

- Uncertainty due to
  - Randomness
  - Overwhelming complexity
  - Lack of knowledge
  - …
- Probability gives
  - natural way to describe our assumptions
  - rules for how to combine information

# UNCERTAINTY IN THE WORLD

- Uncertainty due to
  - Randomness
  - Overwhelming complexity
  - Lack of knowledge
  - …
- Probability gives
  - natural way to describe our assumptions
  - rules for how to combine information
- Subjective probability
  - Relate to agent's own state of knowledge: $P(A25|\text{no accidents}) = 0.05$
  - Not assertions about the world; indicate degrees of belief
  - Change with new evidence: $P(A25 \mid \text{no accidents, 5am}) = 0.20$

# MAKING DECISIONS UNDER UNCERTAINTY

- Suppose I believe the following:
  - P(A25  gets me there on time | …) = 0.04
  - P(A90  gets me there on time | …) = 0.70
  - P(A120 gets me there on time | …) = 0.95
  - P(A1440 gets me there on time | …) = 0.9999

- Which action to choose?

# MAKING DECISIONS UNDER UNCERTAINTY

- Which action to choose?
- Depends on my preferences for missing flight vs. time spent  waiting, etc.
  - Utility theory is used to represent and infer preferences
  - Decision theory= probability theory + utility theory

- Expected utility of action a in state s

  $$= \sum_{\text{outcome in Results}(s,a)} P(\text{outcome}) * \text{Utility}(\text{outcome})$$

- A rational agent acts to maximize expected utility

# EX: AIRPORT

- Suppose I believe the following:
  - P(A25 gets me there on time | …) = 0.04
  - P(A90 gets me there on time | …) = 0.70
  - P(A120 gets me there on time | …) = 0.95
  - P(A1440 gets me there on time | …) = 0.9999
  - Utility(on time) = $1,000
  - Utility(not on time) = −$10,000
- Expected utility of action a in state s

  $= \sum_{\text{outcome in Results}(s,a)} \text{P(outcome)} * \text{Utility(outcome)}$

E(Utility(A25)) = 0.04*$1,000 + 0.96*(−$10,000) = −$9,560
E(Utility(A90)) = 0.7*$1,000 + 0.3*(−$10,000) = −$2,300 E(Utility(A120))
= 0.95*$1,000 + 0.05*(−$10,000) = $450 E(Utility(A1440)) =
0.9999*$1,000 + 0.0001*(−$10,000) = $998.90

  - Have not yet accounted for disutility of staying overnight at the airport, etc.

# PROBABILITY

- P(a) is the probability of proposition “a”
  - E.g., P(it will rain in London tomorrow)
  - The proposition “a” is actually true or false in the real world
  - P(a) is our degree of belief that proposition “a” is true in the real world
  - P(a) = “prior” or marginal or unconditional probability
  - Assumes no other information is available

- Axioms of probability:
  - 0 <= P(a) <= 1
  - P(NOT(a)) = 1 – P(a)
  - P(true) = 1
  - P(false) = 0
  - P(a OR b) = P(a) + P(b) – P(a AND b)

# INTERPRETATIONS OF PROBABILITY

- **Relative Frequency**: *Usually taught in school*
  - P(*a*) represents the frequency that event *a* will happen in repeated trials.
  - Requires event *a* to have happened enough times for data to be collected.
- **Degree of Belief**: *A more general view of probability*
  - P(*a*) represents an agent's degree of belief that event *a* is true.
  - Can predict probabilities of events that occur rarely or have not yet occurred.
  - Does not require new or different rules, just a different interpretation.
- <u>Examples</u>:
  - a = "life exists on another planet"
    - What is P(a)? We all will assign different probabilities
  - a = "California will secede from the US"
    - What is P(a)?
  - a = "over 50% of the students in this class will get A's"
    - What is P(a)?

# CONCEPTS OF PROBABILITY

- **Unconditional Probability**
  - **P(a)**, the probability of “a” being true, or **P(a=True)**
  - Does not depend on anything else to be true (**unconditional**)
  - Represents the probability prior to further information that may adjust it (**prior**)
  - Also sometimes “**marginal**” probability (vs. joint probability)

# CONCEPTS OF PROBABILITY

- **Conditional Probability**
  - **P(a|b)**, the probability of “a” being true, given that “b” is true
  - Relies on “b” = true (**conditional**)
  - Represents the prior probability adjusted based upon new information “b” (**posterior**)
  - Can be generalized to more than 2 random variables:
    - e.g. P(a|b, c, d)

# CONCEPTS OF PROBABILITY

- <u>**Conditional Probability**</u>
  - **P(a|b)**, the probability of "a" being true, given that "b" is true
  - Relies on "b" = true (**conditional**)
  - Represents the prior probability adjusted based upon new information "b" (**posterior**)
  - Can be generalized to more than 2 random variables:
    - e.g. P(a|b, c, d)

**We often use comma to abbreviate AND.**

- <u>**Joint Probability**</u>
  - **P(a, b) = P(a ^b)**, the probability of "a" and "b" both being true
  - Can be generalized to more than 2 random variables:
    - e.g. P(a, b, c, d)

# RANDOM VARIABLES

- **Random Variable**:
  - Basic element of probability assertions
  - Similar to CSP variable, but values reflect probabilities not constraints.
    - Variable: A
    - Domain: $\{a_1, a_2, a_3\}$ <-- events / outcomes

- Types of Random Variables:
  - Boolean random variables : *{ true, false }*
    - e.g., *Cavity (= do I have a cavity?)*
  - Discrete random variables : one value from a set of values
    - e.g., *Weather is one of {sunny, rainy, cloudy ,snow}*
  - Continuous random variables : a value from within constraints
    - *e.g., Current temperature is bounded by (10°, 200°)*

- Domain values must be exhaustive and mutually exclusive:
  - One of the values must always be the case (**Exhaustive**)
  - Two of the values cannot both be the case (**Mutually Exclusive**)

# RANDOM VARIABLES

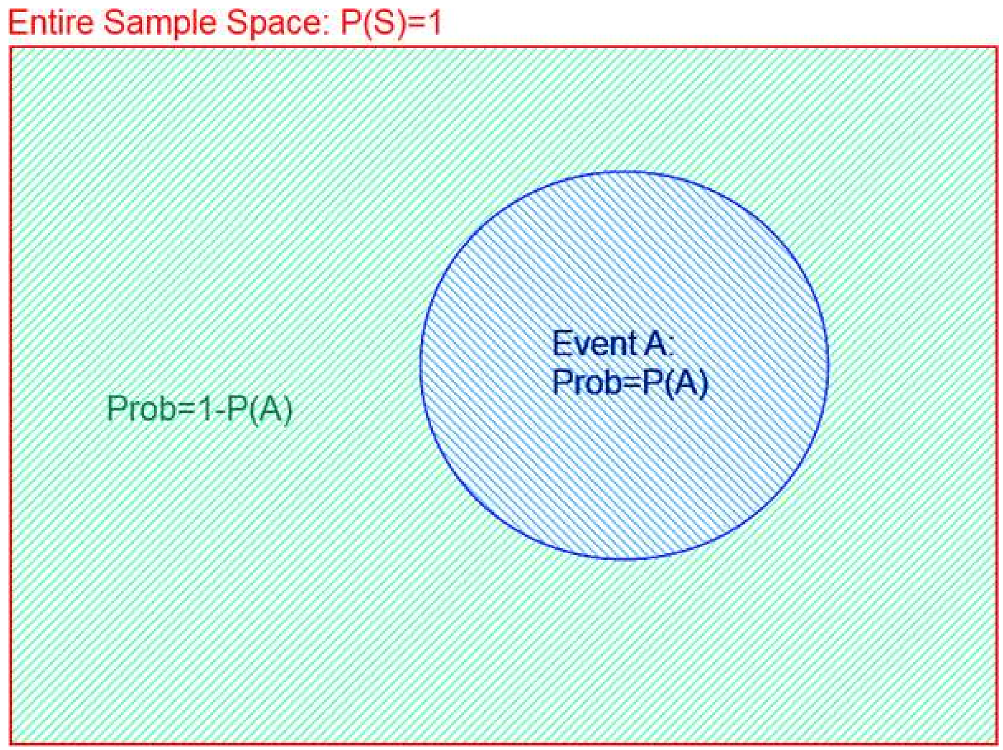
- **Example**: Coin flip
  - Variable = R, the result of the coin flip
  - Domain = {heads, tails, edge}          <-- must be exhaustive
  - P(R = heads) = 0.4999          }
  - P(R = tails) = 0.4999          } <-- must be exclusive
  - P(R = edge) = 0.0002          }

- Shorthand is often used for simplicity:
  - <u>Upper-case letters for variables, lower-case letters for values.</u>
  - E.g., P(A) $\equiv \langle P(A{=}a1), P(A{=}a2), \ldots, P(A{=}an)\rangle$ for all n values in Domain(A)
  - E.g., P(a) $\equiv$ P(A = a)<br>P(a|b) $\equiv$ P(A = a | B = b)<br>P(a, b) $\equiv$ P(A = a $\wedge$ B = b)

- Two kinds of probability propositions:
  - **Elementary propositions** are an assignment of a value to a random variable:
    - e.g., *Weather = sunny;* e.g., *Cavity = false* (abbreviated as *¬cavity*)
  - **Complex propositions** are formed from elementary propositions and standard logical connectives :
    - e.g., *Cavity = false* $\vee$ *Weather = sunny*

# PROBABILITY SPACE

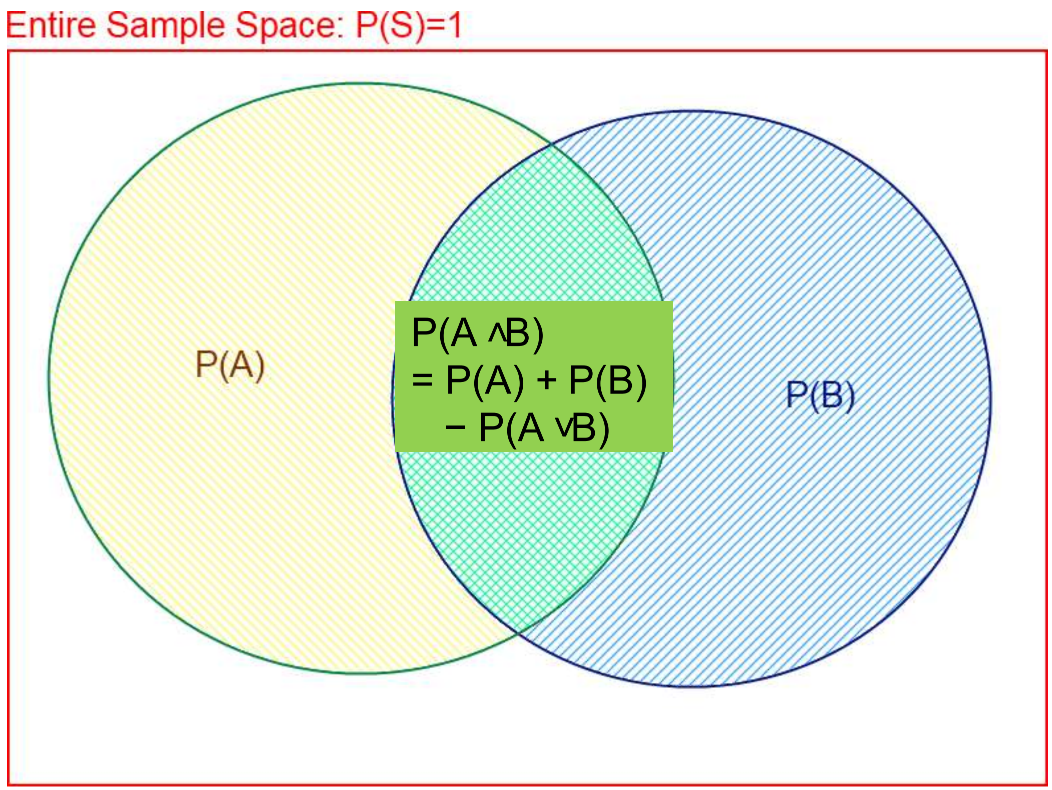
$$P(A) + P(\neg A) = 1$$




**Area = Probability of Event**

# AND PROBABILITY

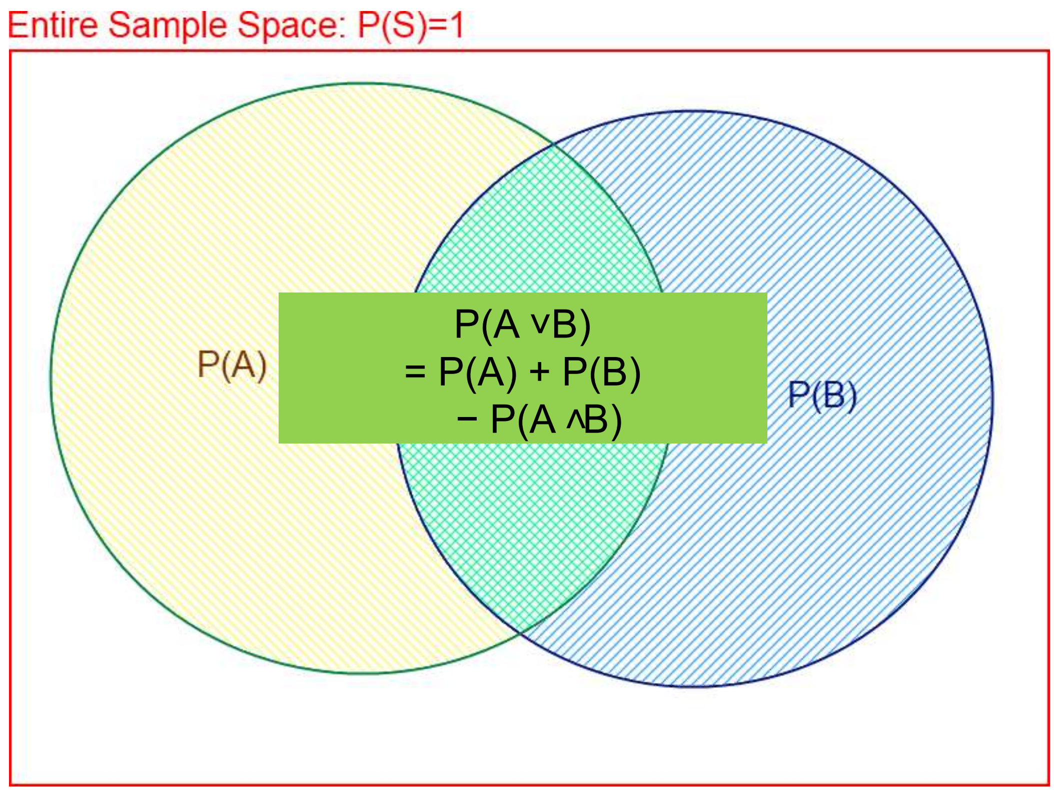
$$P(A, B) = P(A \wedge B) = P(A) + P(B) - P(A \vee B)$$




**Area = Probability of Event**

# OR PROBABILITY

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

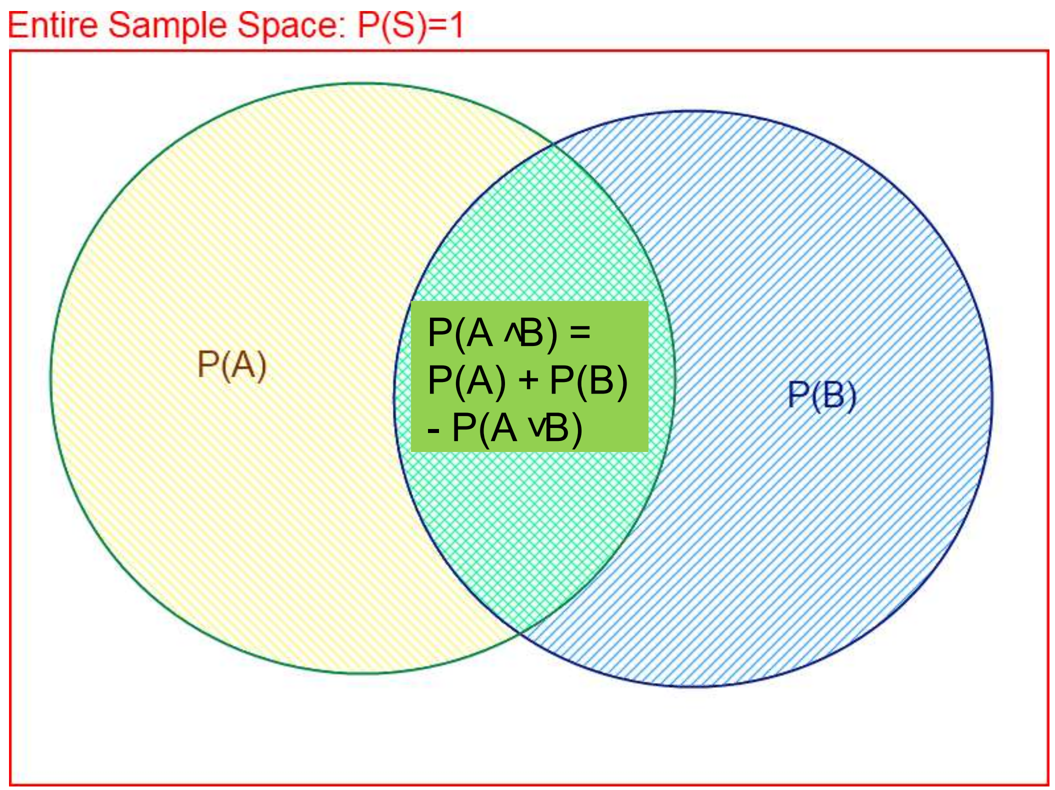



**Area = Probability of Event**

# CONDITIONAL PROBABILITY

We often use comma to abbreviate AND.

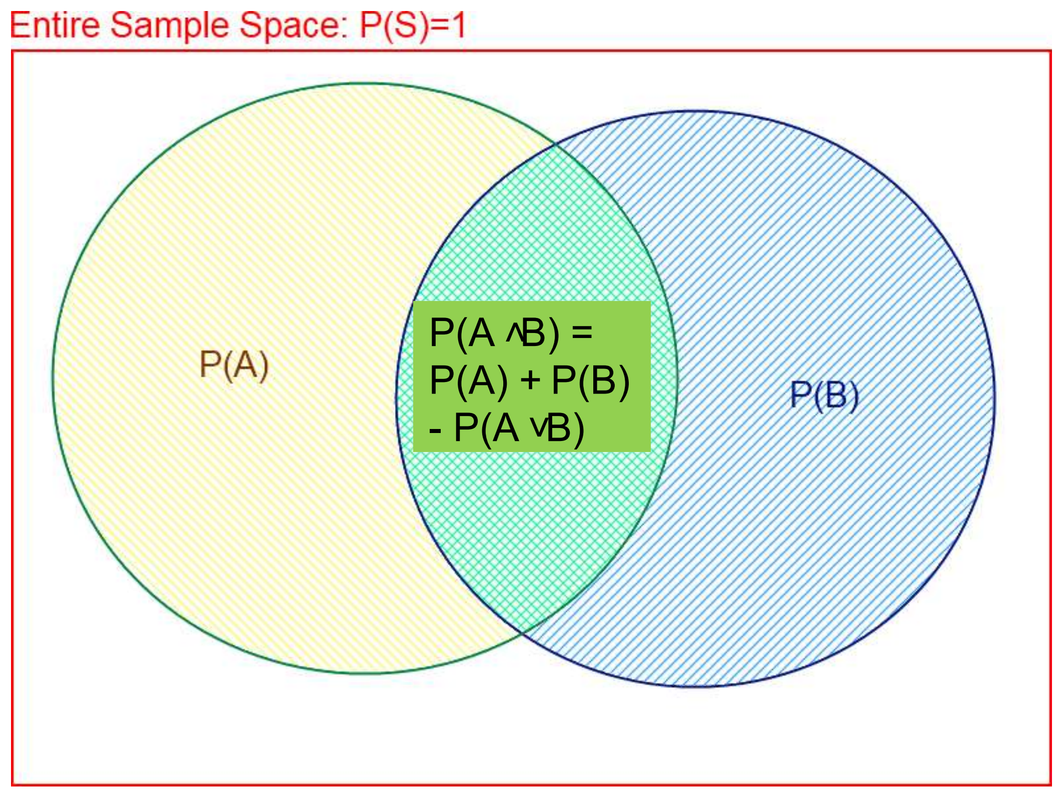
$$P(A \mid B) = P(A, B) / P(B)$$




**Area = Probability of Event**

# PRODUCT RULE

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




**Area = Probability of Event**

# USING THE PRODUCT RULE

- **Applies to any number of variables:**
  - $P(a, b, c) = P(a, b|c)\ P(c) = P(a|b, c)\ P(b, c)$
  - $P(a, b, c|d, e) = P(a|b, c, d, e)\ P(b, c|d, e)$

- **Factoring**: (AKA **Chain Rule** for probabilities)
  - By the product rule, we can always write:

$$P(a, b, c, \ldots z) = P(a \mid b, c, \ldots. z)\ P(b, c, \ldots z)$$

**We often use comma to abbreviate AND.**

  - Repeatedly applying this idea, we can write:

$$P(a, b, c, \ldots z) = P(a \mid b, c, \ldots. z)\ P(b \mid c, .. z)\ P(c \mid .. z) .. P(z)$$

  - This holds for any ordering of the variables

# 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 $\phi$, sum the atomic events where it is true:

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

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

For any proposition $\phi$, sum the atomic events where it is true:

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

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

# INFERENCE BY ENUMERATION

Start with the joint distribution:

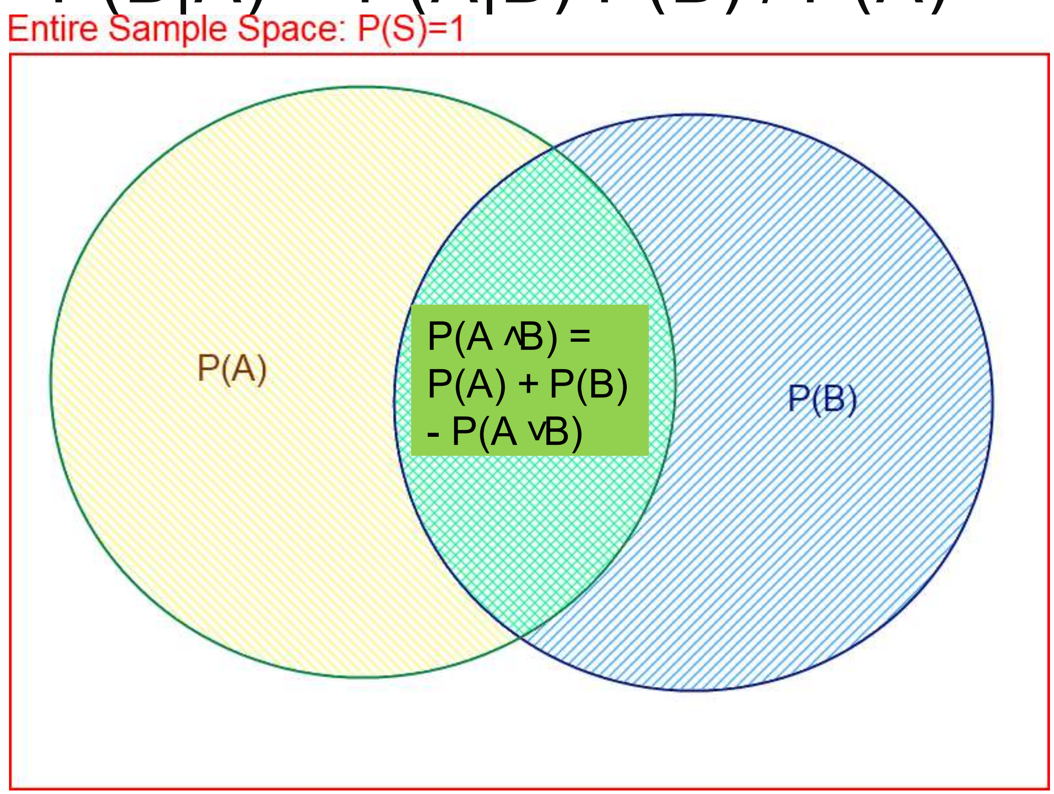
| | *toothache* | | *¬ toothache* | |
|---|---|---|---|---|
| | *catch* | *¬ catch* | *catch* | *¬ catch* |
| *cavity* | .108 | .012 | .072 | .008 |
| *¬ cavity* | .016 | .064 | .144 | .576 |

Can also compute conditional probabilities:

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

# BAYES' RULE

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




**Area = Probability of Event**

# DERIVATION OF BAYES' RULE

- **Start from Product Rule:**
  - P(a, b) = P(a|b) P(b) = P(b|a) P(a)

- **Isolate Equality on Right Side:**
  - P(a|b) P(b) = P(b|a) P(a)

- **Divide through by P(b):**
  - P(a|b) = P(b|a) P(a) / P(b) <-- **Bayes' Rule**

# 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:

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

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!