Probability

Uncertainty in the World

- An person can often be uncertain about the state of the world/domain since there is often ambiguity and uncertainty
- Plausible/**probabilistic inference**
  - I've got this evidence; what's the chance that this conclusion is true?
    - I've got a sore neck; how likely am I to have meningitis?
    - A mammogram test is positive; what's the probability that the patient has breast cancer?

Uncertainty

- Say we have a rule:

  if *toothache* ***then*** *problem is cavity*
- But not all patients have toothaches due to cavities, so we could set up rules like:

  if *toothache and ¬gum-disease and ¬filling and ...*
  then *problem = cavity*
- This gets complicated; better method:

  if *toothache* ***then*** *problem is cavity with 0.8 probability*

  or $P(cavity \mid toothache) = 0.8$

  the probability of cavity is 0.8 given toothache is observed

Example of Uncertainty

- Assume a camera and vision system is used to estimate the curvature of the road ahead
- There's uncertainty about which way it curves
  - limited pixel resolution, noise in image
  - algorithm for "road detection" is not perfect
- This uncertainty can be represented with a simple probability model:

  $P(\textit{road curves to left} \mid E) = 0.6$
  $P(\textit{road goes straight} \mid E) = 0.3$
  $P(\textit{road curves to right} \mid E) = 0.1$
  - where the probability of an event is a measure of observer's belief in the event given the evidence E

Logics

Logics are characterized by what they commit to as "primitives"

Logic	What Exists in World	Knowledge States
Propositional	facts	true/false/unknown
First-Order	facts, objects, relations	true/false/unknown
Temporal	facts, objects, relations, times	true/false/unknown
Probability Theory	facts	degree of belief 0..1
Fuzzy	degree of truth	degree of belief 0..1

Probability Theory

- Probability theory serves as a formal means for
  - Representing and reasoning with uncertain knowledge
  - Modeling **degrees of belief** in a proposition (event, conclusion, diagnosis, etc.)
- *Probability is the "language" of uncertainty*
  - A key modeling method in modern AI

Source of Probabilities

- Frequentists
  - probabilities come from experiments
  - if 10 of 100 people tested have a cavity, $P(cavity) = 0.1$
  - probability means the fraction that would be observed in the limit of infinitely many samples
- Objectivists
  - probabilities are real aspects of the world
  - objects have a propensity to behave in certain ways
  - coin has propensity to come up heads with probability 0.5
- Subjectivists
  - probabilities characterize an agent's belief
  - have no external physical significance

Sample Space/Outcome Space

- S is a outcome space: collection of all possible outcome
- Let, A be a part of the collection of outcomes in S; that is, $A \subset S$. Then A is called an event.
- Events can be binary, multi-valued, or continuous

Outcome and Event

- Outcome and event are not synonymous.
- **Outcome** is the result of a random experiment. Example: rolling a die has **six** possible outcomes.
- **Event** is a set of outcomes to which a probability is assigned. Example: One possible event is "rolling a number less than 3".

Mutually Exclusive Event

- **Mutually exclusive** events are events that **cannot occur together (simultaneously).**
- $A_1, A_2, \dots, A_k$ are mutually exclusive events means that $A_i \cap A_j = \emptyset, i \neq j$; that is, $A_1, A_2, \dots, A_k$ are disjoint sets.
- **Example:**
  - A = queen of diamonds; B = queen of clubs
  - Events A and B are mutually exclusive if only one card is selected

Mutually Exhaustive Event

- $A_1, A_2, \dots, A_k$ are mutually exhaustive events means that $A_i \cup A_j \cup \cdots \cup A_k = S$

Example:

Consider the experiment of throwing a die.

Sample space S = {1, 2, 3, 4, 5, 6}

Assume that A, B and C are the events associated with this experiment.
Define: A be the event of getting a number greater than 3

B be the event of getting a number greater than 2 but less than 5

C be the event of getting a number less than 3

We can write these events as:

A = {4, 5, 6}

B = {3, 4}

and C = {1, 2}

We observe that

A ∪ B ∪ C = {4, 5, 6} ∪ {3, 4} ∪ {1, 2} = {1, 2, 3, 4, 5, 6} = S

The Axioms of Probability

1. 0 ≤ *P*(*A*) ≤ 1
2. *P*(true) = 1, *P*(false) = 0
3. For any two disjoint events A and B, we have
$$P(A \cup B) = P(A) + P(B)$$
4. For any infinite sequence of mutually disjoint events $A_1, A_2, A_3, \dots$, we have
$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

Empirical Probablity

- Refers to a probability that is based on historical data.

$$P(A) = \frac{\text{\# of times event A occurs}}{\text{total \# of observed occurences}}$$

Empirical Probablity

Find the probability of selecting a male taking statistics from the population described in the following table:

	Taking Stats	Not Taking Stats	Total
Male	84	145	229
Female	76	134	210
Total	160	279	439

$$\text{Probability of Male Taking Stats} = \frac{\text{number of males taking stats}}{\text{total number of people}} = \frac{84}{439} = 0.191$$

Equiprobable Probability Space

- All outcomes equally likely (fair coin, fair die...)
- Laplace's definition of probability (only in finite equiprobable space)

$$P(A) = \frac{|A|}{|S|}$$

Theoritical Probablity

- Theoretical probability is finding the probability of events that come from an equiprobable sample space.

$$P(A) = \frac{\#\ of\ outcomes\ in\ A}{number\ of\ outcomes\ in\ S} = \frac{|A|}{|S|}$$

Theoritical Probablity

Find the probability of selecting a face card (Jack, Queen, or King) from a standard deck of 52 cards.

$$P(Face\ Card) = \frac{|A|}{|S|} = \frac{12}{52} = \frac{3}{13}$$

Simple vs Joint Probability

- Simple (Marginal) Probability refers to the probability of a simple event.
  - Example: P(King)
- Joint Probability refers to the probability of an occurrence of two or more events.
  - Example: P(King and Spade)

Simple vs Joint Probability

Computing Joint and Marginal Probabilities:

- The probability of a **joint** event, A and B:

$$P(A\ and\ B) = \frac{\text{number of outcomes satisfying A and B}}{\text{total number of elementary outcomes}}$$

- Computing a **marginal (or simple)** probability:

$$P(A) = P(A\ and\ B_1) + P(A\ and\ B_2) + \cdots + P(A\ and\ B_k)$$

Where $B_1, B_2, \ldots, Bk$ are k mutually exclusive and collectively exhaustive events

Example of Joint Probability

	Ace	Not Ace	Total
Black	2	24	26
Red	2	24	26
Total	4	48	52

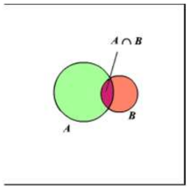
$$P(\text{Red and Ace}) = \frac{\text{number of cards that are red and ace}}{\text{total number of cards}} = \frac{2}{52}$$

Example of Marginal Probability

	Ace	Not Ace	Total
Black	2	24	26
Red	2	24	26
Total	4	48	52

$$\text{P(Ace)} = \text{P(Ace and Red)} + \text{P(Ace and Black)} = \frac{2}{52} + \frac{2}{52} = \frac{4}{52}$$

Laws of Probability: Additive Rule



- If A and B are two events in a probability experiment, then the probability that either one of the events will occur is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Or

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

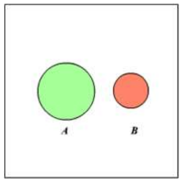
Laws of Probability: Additive Rule (Example)

- Example: If I roll a number cube and flip a coin, What is the probability I will get a tails or a 3?

Answer:

$$P(\text{tails or a 3}) = \frac{1}{2} + \frac{1}{6} = \frac{8}{12} = \frac{2}{3}$$

Laws of Probability: Additive Rule



- If A and B are two mutually exclusive events then $P(A \cap B) = 0$.

$$P(A \text{ or } B) = P(A) + P(B)$$

Or

$$P(A \cup B) = P(A) + P(B)$$

Laws of Probability: Additive Rule (Example)

If you take out a single card from a regular pack of cards, what is probability that the card is either an ace or spade?

Answer

Let X be the event of picking an ace and Y be the event of picking a spade.

$$P(X) = \frac{4}{52}$$

$$P(Y) = \frac{13}{52}$$

The two events are not mutually exclusive, as there is one favorable outcome in which the card can be both an ace and spade.

$$P(X \cap Y) = \frac{1}{52}$$

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{4}{13}$$

Complement Rule

- For any event A, we have

$$P(A^c) = 1 - P(A)$$

Complement Rule

- Suppose that we flip eight fair coins. What is the probability that we have at least one head showing?

Answer:

The complement of the event "we flip at least one head" is the event "there are no heads."

$$P(\text{At least one head}) = 1 - P(\text{No head})$$
$$= 1 - \frac{1}{256} = 0.99609375$$

Complement Rule

- Suppose that we flip eight fair coins. What is the probability that we have at least one head showing?

Answer:

The complement of the event "we flip at least one head" is the event "there are no heads."

$$P(\text{At least one head}) = 1 - P(\text{No head})$$
$$= 1 - \frac{1}{256} = 0.99609375$$

Conditional Probability

- Conditional probabilities
  - formalizes the process of accumulating evidence and updating probabilities based on new evidence
  - specifies the belief in a proposition (event, conclusion, diagnosis, etc.) that is *conditioned on* a proposition (evidence, feature, symptom, etc.) being true
- $P(a \mid e)$: conditional probability of $A=a$ *given* $E=e$ evidence is *all that is known true*
  - $P(a \mid e) = P(a \wedge e) / P(e) = P(a, e) / P(e)$
  - conditional probability can viewed as the joint probability $P(a, e)$ normalized by the prior probability, $P(e)$

Conditional Probability

Conditional probabilities behave exactly like standard probabilities; for example:

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

conditional probabilities are between 0 and 1 inclusive

$P(a_1 \mid e) + P(a_2 \mid e) + \ldots + P(a_k \mid e) = 1$

conditional probabilities sum to 1 where $a_1, \ldots, a_k$ are all values in the domain of random variable A

$P(\neg a \mid e) = 1 - P(a \mid e)$

negation for conditional probabilities

Conditional Probability

$P(\text{conjunction of events} \mid e)$

$P(a \wedge b \wedge c \mid e)$ or as $P(a, b, c \mid e)$ is the agent's belief in the sentence $a \wedge b \wedge c$ conditioned on e being true

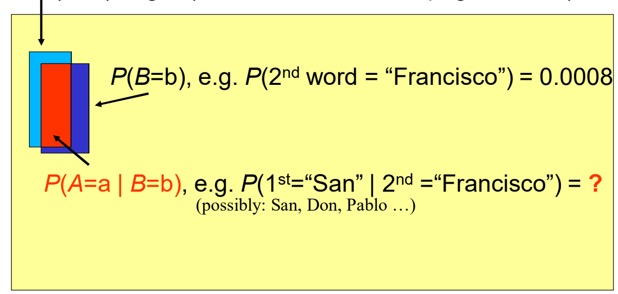
$P(a \mid \text{conjunction of evidence})$

$P(a \mid e \wedge f \wedge g)$ or as $P(a \mid e, f, g)$ is the agent's belief in the sentence a conditioned on $e \wedge f \wedge g$ being true

Conditional Probability

The conditional probability $P(A=\text{a} \mid B=\text{b})$ is the fraction of time A=a, within the region where B=b

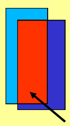
$P(A=\text{a})$, e.g. $P(1^{st}$ word on a random page = "San") = 0.001



Conditional Probability

- P(san | francisco)
  = #(1st=s and 2nd=f) / #(2nd=f)
  = P(san ∧ francisco) / P(francisco)
  = 0.0007 / 0.0008
  = 0.875

P(s)=0.001
P(f)=0.0008
P(s,f)=0.0007



$P(B$=b), e.g. P(2nd word = “Francisco”) = 0.0008

$P(A$=a | B=b), e.g. P(1st=“San” | 2nd =“Francisco”) = **0.875**
(possibly: San, Don, Pablo …)

Although “San” is rare and “Francisco” is rare, given “Francisco” then “San” is quite likely!

Conditional Probability

- In general, the conditional probability is

$$P(A=a \mid B)=\frac{P(A=a,B)}{P(B)}=\frac{P(A=a,B)}{\sum_{\text{all } a_i} P(A=a_i,B)}$$

- We can have everything *conditioned* on some other event(s), C, to get a conditionalized version of conditional probability:

$$P(A \mid B,C)=\frac{P(A,B \mid C)}{P(B \mid C)}$$

‘|’ has low precedence.
This should read: P(A | (B,C))

The Chain Rule

- From the definition of conditional probability we have

  $P(A, B) = P(B) * P(A \mid B) = P(A \mid B) * P(B)$
- It also works the other way around:

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

Called “**Product Rule**”

- It works with more than 2 events too:

$P(A_1, A_2, \ldots, A_n) =$
$P(A_1) * P(A_2 \mid A_1) * P(A_3 \mid A_1, A_2) * \ldots$
$\quad * P(A_n \mid A_1, A_2, \ldots, A_{n-1})$

Called “**Chain Rule**”

Probabilistic Reasoning

How do we use probabilities in AI?

- You wake up with a headache
- Do you have the flu?
- H = headache, F = flu

Logical Inference: if H then F
(but the world is usually not this simple)

Statistical Inference: compute the probability of a query/diagnosis/decision given (i.e., conditioned on) evidence/symptom/observation, i.e., $P(F \mid H)$

[Example from Andrew Moore]

Example

Statistical Inference: Compute the probability of a diagnosis, F, given symptom, H, where H = "has a headache" and F = "has flu"

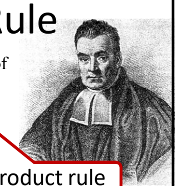
That is, compute $P(F \mid H)$

You know that

- $P(H) = 0.1$ "one in ten people has a headache"
- $P(F) = 0.01$ "one in 100 people has flu"
- $P(H \mid F) = 0.9$ "90% of people who have flu have a headache"

[Example from Andrew Moore]

Inference with Bayes's Rule



Thomas Bayes, "Essay Towards Solving a Problem in the Doctrine of Chances," 1764

$$P(F \mid H) = \frac{P(F,H)}{P(H)} = \frac{P(H \mid F)P(F)}{P(H)}$$

Def of cond. prob.

Product rule

- $P(H) = 0.1$ "one in ten people has a headache"
- $P(F) = 0.01$ "one in 100 people has flu"
- $P(H|F) = 0.9$ "90% of people who have flu have a headache"

- $P(F|H) = 0.9 * 0.01 / 0.1 = 0.09$
- So, there's a 9% chance you have flu – much less than 90%
- But it's higher than $P(F)$ = 1%, since you have a headache

Bayes's Rule

- Bayes's Rule is the basis for probabilistic reasoning given a prior model of the world, P(Q), and a new piece of evidence, E, Bayes's rule says how this piece of evidence decreases our ignorance about the world
- Initially, know P(Q) ("prior")
- Update after knowing E ("posterior"):

$$P(Q|E) = P(Q)\,\frac{P(E|Q)}{P(E)}$$

Inference with Bayes's Rule

***P*(*A*|*B*) = *P*(*B* | *A*)*P*(*A*) / *P*(*B*)** **Bayes's rule**

- Why do we make things this complicated?
  - Often $P(B|A)$, $P(A)$, $P(B)$ are easier to get
  - Some names:
    - **Prior *P*(*A*)**: probability of A *before* any evidence
    - **Likelihood *P*(*B*|*A*)**: assuming A, how likely is the evidence
    - **Posterior *P*(*A*|*B*)**: probability of A after knowing evidence B
    - **(Deductive) Inference**: deriving an unknown probability from known ones
- If we have the full joint probability table, we can simply compute $P(A|B) = P(A, B) / P(B)$

Bayes's Rule in Practice

Summary of Important Rules

- **Conditional Probability**: $P(A|B) = P(A,B)/P(B)$
- **Product rule**: $P(A,B) = P(A|B)P(B)$
- **Chain rule**: $P(A,B,C,D) = P(A|B,C,D)P(B|C,D)P(C|D)P(D)$
- **Conditionalized version of Chain rule**:
  $P(A,B|C) = P(A|B,C)P(B|C)$
- **Bayes's rule**: $P(A|B) = P(B|A)P(A)/P(B)$
- **Conditionalized version of Bayes's rule**:
  $P(A|B,C) = P(B|A,C)P(A|C)/P(B|C)$
- **Addition / Conditioning rule**: $P(A) = P(A,B) + P(A,\neg B)$
  $P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B)$

Common Mistake

- $P(A) = 0.3$ so $P(\neg A) = 1 - P(A) = 0.7$
- $P(A|B) = 0.4$ so $P(\neg A|B) = 1 - P(A|B) = 0.6$
  because $P(A|B) + P(\neg A|B) = 1$

  but $P(A|\neg B) \neq 0.6$ (in general)
  because $P(A|B) + P(A|\neg B) \neq 1$ in general

Quiz

- A doctor performs a test that has 99% reliability, i.e., 99% of people who are sick test positive, and 99% of people who are healthy test negative. The doctor estimates that 1% of the population is sick.
- Question: A patient tests positive. What is the chance that the patient is sick?
- 0-25%, 25-75%, 75-95%, or 95-100%?

Quiz

- A doctor performs a test that has 99% reliability, i.e., 99% of people who are sick test positive, and 99% of people who are healthy test negative. The doctor estimates that 1% of the population is sick.
- Question: A patient tests positive. What is the chance that the patient is sick?
- 0-25%, 25-75%, 75-95%, or 95-100%?
- Common answer: 99%; Correct answer: 50%

Given:

$P(TP \mid S) = 0.99$

$P(\neg TP \mid \neg S) = 0.99$

$P(S) = 0.01$

TP = "tests positive"
S = "is sick"

Query:

$P(S \mid TP) = ?$

$P(TP \mid S) = 0.99$

$P(\neg TP \mid \neg S) = 0.99$

$P(S) = 0.01$

$P(S \mid TP) =$

$P(TP \mid S)\, P(S) / P(TP)$

$= (0.99)(0.01) / P(TP) = 0.0099/P(TP)$

$P(\neg S \mid TP) = P(TP \mid \neg S)P(\neg S) / P(TP)$

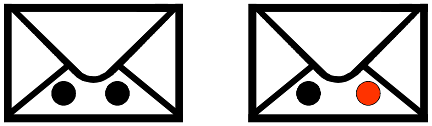
$= (1 - 0.99)(1 - 0.01) / P(TP) = 0.0099/P(TP)$

$0.0099/P(TP) + 0.0099/P(TP) = 1$, so $P(TP) = 0.0198$

So, $P(S \mid TP) = 0.0099 / 0.0198 = 0.5$

Inference with Bayes's Rule

- In a bag there are two envelopes
  - one has a red ball (worth $100) and a black ball
  - one has two black balls. Black balls are worth nothing



- You randomly grab an envelope, and randomly take out one ball – it's **black**
- At this point you're given the option to switch envelopes. Should you switch or not?

Similar to the "Monty Hall Problem"