Fundamentals of Probability

Alexander G. Ororbia II

Introduction to Machine Learning

CSCI-635

9/1/2023

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, other drivers' plans, etc.)
2. noisy sensors (traffic reports)
3. uncertain (non-deterministic) action outcomes (flat tire, etc.)
4. immense complexity of modeling and predicting traffic

Set of actions:
{A_1, A_2,…A_t,…,A_T}

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} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …

Probability in Context

Probability theory

- Branch of mathematics concerned with analysis of random phenomena
  - ***Randomness***: a non-order or non-coherence in a sequence of symbols or steps, such that there is no intelligible pattern or combination
- Central objects of probability theory are:
  random variables, stochastic processes, and **events**
  - Mathematical abstractions of non-deterministic events or measured quantities that may either be single occurrences or evolve over time in an apparently random fashion

Uncertainty

- A lack of knowledge about an event
- Can be represented by a probability
  - Ex: role a die, draw a card
- Can be represented as an error

A statistic (a measure in **statistics**)

- Can use probability in determining that measure

Founders of Probability Theory

Blaise Pascal

(1623-1662, France)

Pierre Fermat

(1601-1665, France)

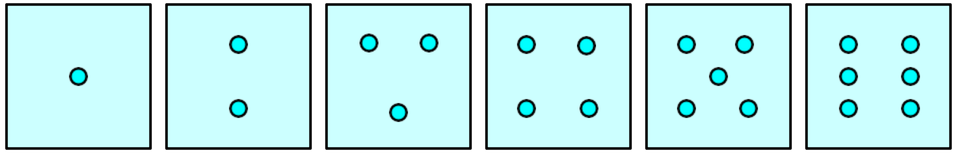
Laid the foundations of the probability theory in a correspondence on a dice game posed by a French nobleman

Sample Spaces – Measures of Events

Collection (list) of all possible outcomes

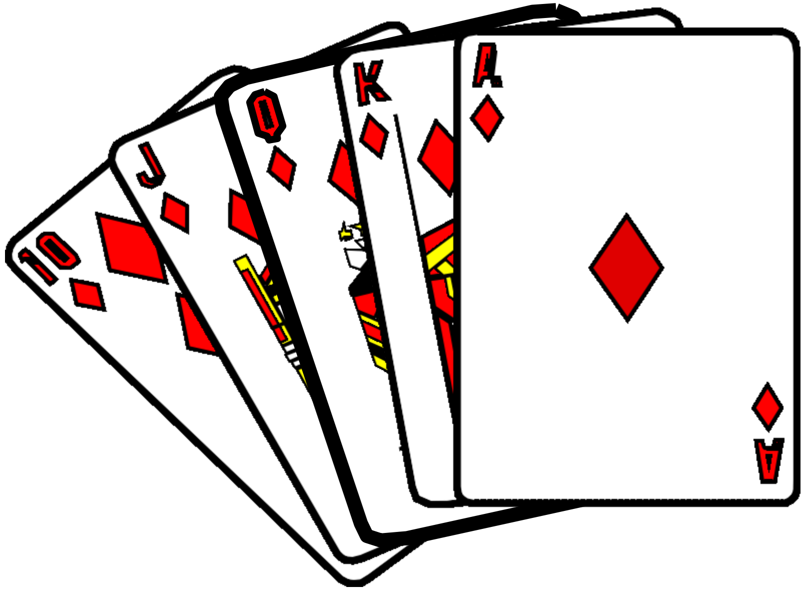
Experiment: Roll a die!

– e.g.: All six faces of a die:



Experiment: Draw a card!

e.g.: All 52 cards in a deck:



Types of Events

Event

– Subset of sample space (set of outcomes of experiment)

Random event

– Different likelihoods of occurrence

Simple event

– Outcome from a sample space with one characteristic in simplest form

– e.g.: King of clubs from a deck of cards

Joint event

– Conjunction (AND, □, “,”); disjunction (OR,∨)

– Contains **several** simple **events**

– e.g.: A red ace from a deck of cards,
(ace on hearts OR ace of diamonds)

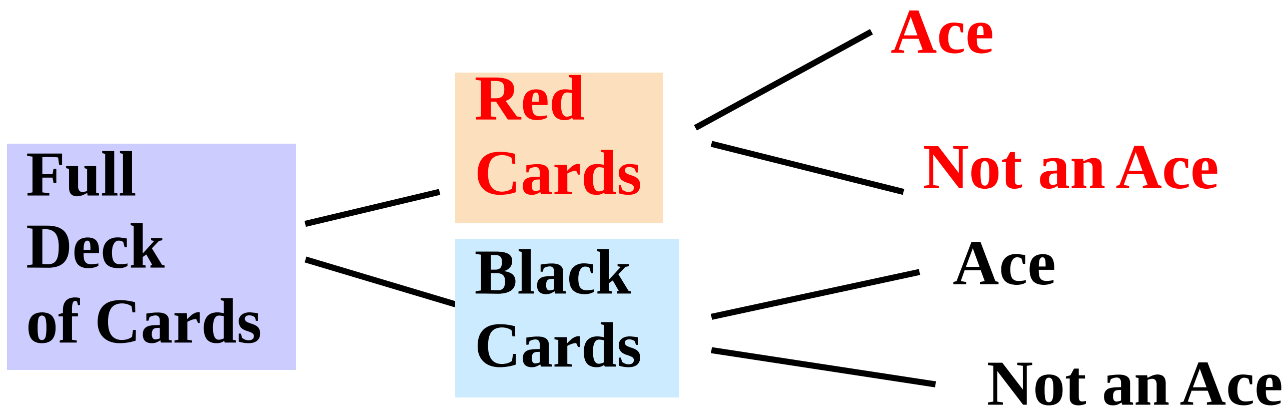
Visualizing Events

Excellent ways of determining probabilities, can be built from data

Contingency tables (nice way to look at probability):

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

Tree diagrams:



Maximum Likelihood Estimation (MLE)

- Uses relative frequencies as estimates
- Maximizes likelihood of training data D under a simple model M, or P(D|M)

- With discrete data, we can employ a *counting function* **c(**A=a**)**, that returns frequency of a particular value taken on by attribute A
  - ***Note***: **c(**A=a**)** is actually **c(**A=a, *D***)**, where *D* is a dataset

- **Issue:** What happens with sparse data?

You're thinking like a frequentist now!

An Example: Unigram Model

w_i is particular word in W, where W is set of unique words (or vocabulary)

- Do not use history:

Probability of a word given a word sequence/history

$$P(w_i|w_1 \ldots w_{i-1}) \approx P(w_i) = \frac{c(w_i)}{\sum_{\tilde{w}} c(\tilde{w})}$$

i live in osaka . </s>
i am a graduate student . </s>
my school is in nara . </s>

P(nara) = 1/20 = 0.05
P(i) = 2/20 = 0.1
P(</s>) = 3/20 = 0.15

P(W=i live in nara . </s>) =
$0.1 * 0.05 * 0.1 * 0.05 * 0.15 * 0.15 = 5.625 * 10^{-7}$

Axioms of Probability

<u>Given 2 events:</u> x, y

1) P(x OR y) = P(x) + P(y) - P(x AND y);
   for **mutually exclusive events**, P(x AND y) = 0
2) P(x and y) = P(x) * P(y | x), also written as P(y | x) = P(x and y)/P(x)
3) If x and y are independent, P(y | x) = P(y), thus P(x AND y) = P(x) * P(y)
4) P(x) > P(x) * P(y); P(y) > P(x) * P(y)

Probability Mass Function (PFM)

- The domain of P must be the set of all possible states of x.
- $\forall x \in \mathrm{x}, 0 \leq P(x) \leq 1$. An impossible event has probability 0 and no state can be less probable than that. Likewise, an event that is guaranteed to happen has probability 1, and no state can have a greater chance of occurring.
- $\sum_{x \in \mathrm{x}} P(x) = 1$. We refer to this property as being **normalized**. Without this property, we could obtain probabilities greater than one by computing the probability of one of many events occurring.

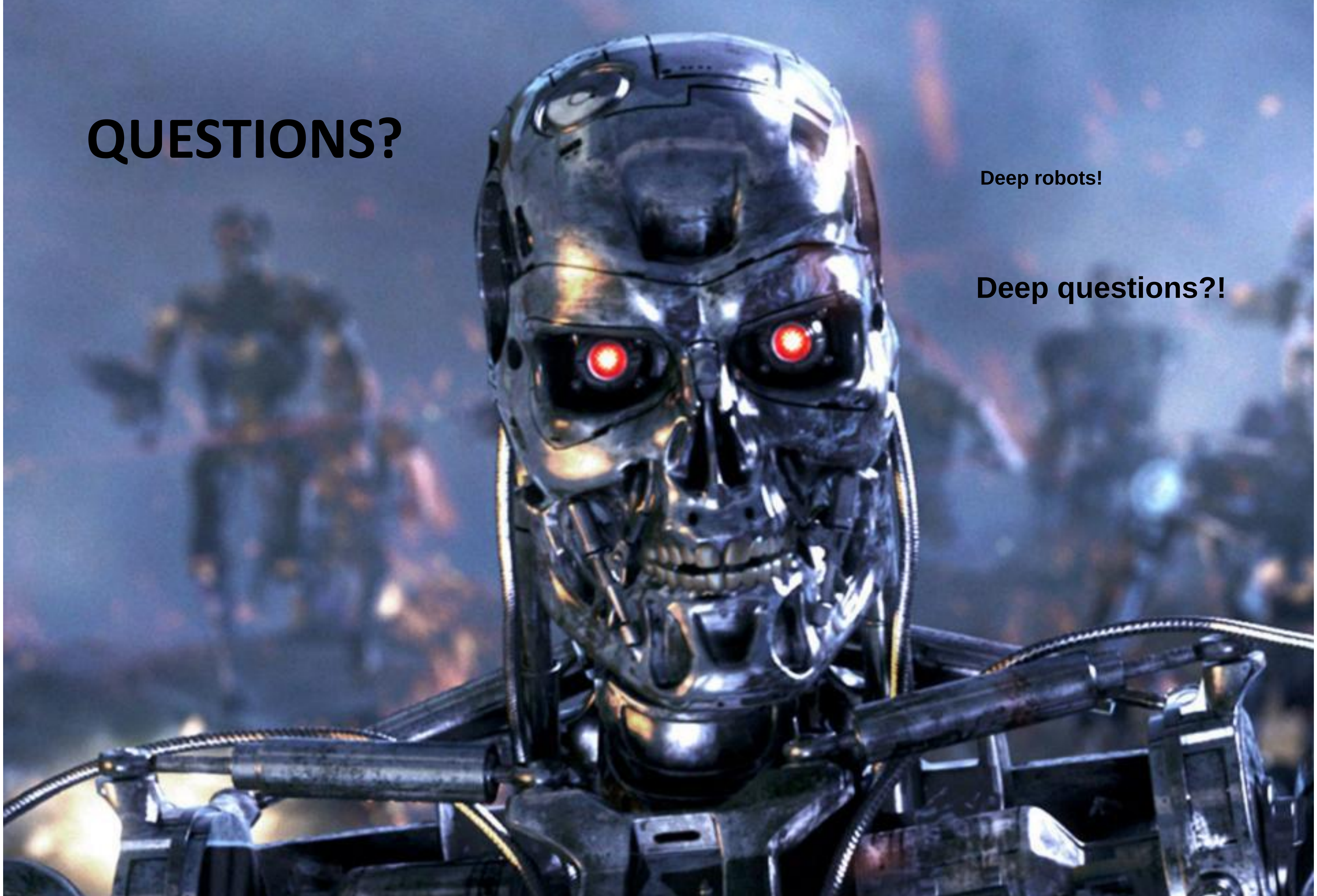
Example: uniform distribution: $$P(\mathrm{x} = x_i) = \frac{1}{k}$$

Probability Density Function (PDF)

- The domain of p must be the set of all possible states of x.
- $\forall x \in \mathrm{x}, p(x) \geq 0$. Note that we do not require $p(x) \leq 1$.
- $\int p(x)dx = 1$.

Example: uniform distribution: $u(x; a, b) = \frac{1}{b-a}$.

QUESTIONS?



Deep robots!

Deep questions?!

Types of Uncertainty

For example, to drive your car in the morning:

- It must not have been stolen during the night
- It must not have flat tires
- There must be gas in the tank
- The battery must not be dead
- The ignition must work
- You must not have lost the car keys
- No truck should obstruct the driveway
- You must not have suddenly become blind or paralytic

Etc…

Not only would it not be possible to list all of them, but would trying to do so be efficient?

Probabilistic Reasoning

Types of probabilistic reasoning

- Reasoning using probabilistic methods
- Reasoning with uncertainty
- Rigorous reasoning vs heuristics or biases