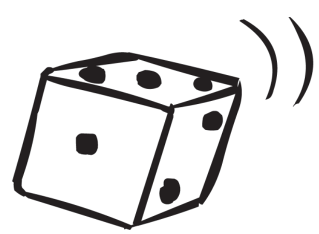
Prof. Alex Rogers
Department of Computer Science
University of Oxford
14th January 2023

Probabilistic Machine Learning
Session 1

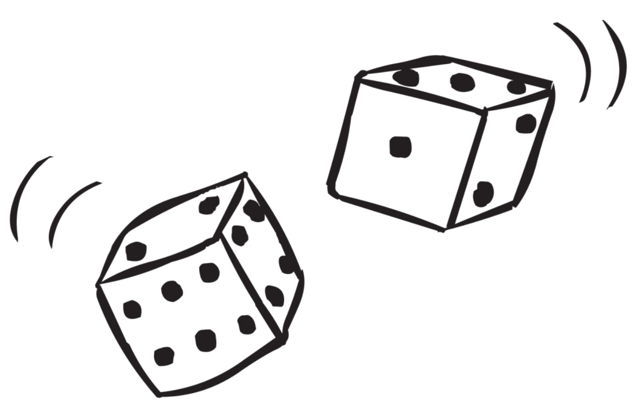
Probability Theory

Single and Joint Probabilities

Probability theory typically describes random variables, $P(X)$, and the probability that these random variables take on actual values.



$$P(D = 6) = 1/6$$

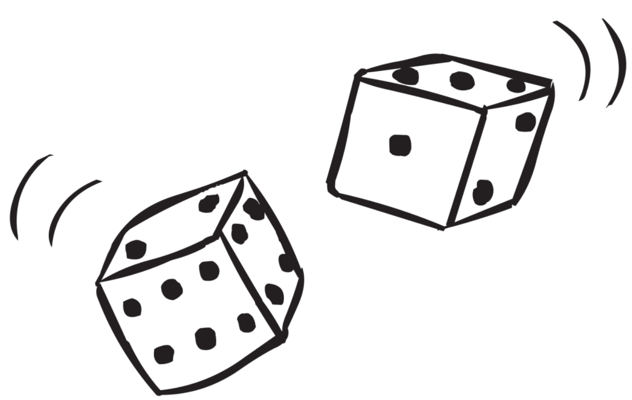


$$P(D_1 = 6, D_2 = 6) = 1/36$$

probability.ipynb

Independent Probabilities

When events are **independent** their **joint** probability is given by the product of their individual probabilities.

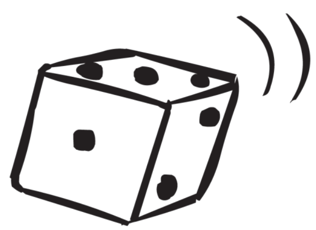


$$P(D_1 = 6, D_2 = 6) = 1/36$$

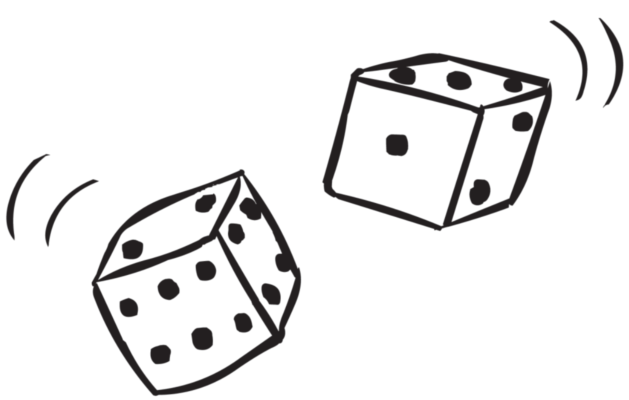
$$P(D_1 = 6, D_2 = 6) = P(D_1 = 6) \times P(D_2 = 6)$$

Conditional Probabilities

Conditional probabilities describe the probability of any event **given** the occurrence of another event.



$$P(D = 6 | D \text{ is even}) = 1/3$$



$$P(D_1 = 6 | D_1 + D_2 > 8) = \ ?$$

probability.ipynb

The Rules of Probabilities

There are just two fundamental rules of probability theory.

Sum rule

$$P(X) = \sum_{Y} P(X, Y)$$

Product rule

$$P(X, Y) = P(Y|X) \times P(X)$$

Plus the observation that probabilities must sum to 1.

Frequentist probability is the interpretation of probability that defines an event's probability as the limit of its relative frequency in a large number of trials.

Bayesian probability is the interpretation of the probability that holds that it can be defined as the degree to which a person (or community) believes that a proposition is true.

Rev. Thomas Bayes (1702 - 1761)

- Framework for plausible reasoning developed over the nineteenth century

- Developed by three physicists:

  - Harold Jeffreys (1891 – 1989)
  - Richard Cox (1898 - 1991)
  - Edwin Jaynes (1922 – 1998)

- Growth of computational power since 2000 has seen this become one of the dominant model for much of machine learning.

- Recent development of probabilistic programming make using Bayesian reasoning ever more accessible.

Cox's axioms describe desirable properties of a framework for reasoning about uncertainty

1. Degrees of plausibility are represented by real numbers.

2. Qualitative correspondence with common sense:
    - Increasing the plausibility of a statement, decreases the plausibility of its negative.

3. Consistency:
    - If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.
    - It must be possible to incorporate all the evidence relevant to a question into the reasoning process.
    - All equivalent states of knowledge are represented by equivalent plausibility assignments.

Bayes Theorem

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

Bayes Theorem

Likelihood Prior

Posterior

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Evidence

Example

- A clinical test, designed to diagnose a specific illness, comes out positive for a certain patient.

- We are told that
  - The test is 79 percent reliable (that is, it misses 21 percent of actual cases)
  - On average, this illness affects 1 percent of the population in the same age group as the patient
  - The test has a false positive rate of 10 percent.

- Taking this into account and assuming you know nothing about the patient's symptoms or signs, what is the probability that this patient actually has the illness?

Example

$$P(\mathrm{D}) = 0.01$$

$$P(\mathrm{T} \mid \mathrm{D}) = 0.79$$

$$P(\mathrm{T} \mid \neg \mathrm{D}) = 0.1$$

Example

$$P(\mathrm{D}|\mathrm{T}) = \frac{P(\mathrm{T}|\mathrm{D})P(\mathrm{D})}{P(\mathrm{T})}$$

$$= \frac{P(\mathrm{T}|\mathrm{D})P(\mathrm{D})}{P(\mathrm{T}|\mathrm{D})P(\mathrm{D}) + P(\mathrm{T}|\neg\mathrm{D})P(\neg\mathrm{D})}$$

$$= \frac{0.79 \times 0.01}{0.79 \times 0.01 + 0.1 \times 0.99}$$

$$\approx 0.074$$

Exercises 1

Next Time

Bayesian Inference and
Markov Chain Monte Carlo