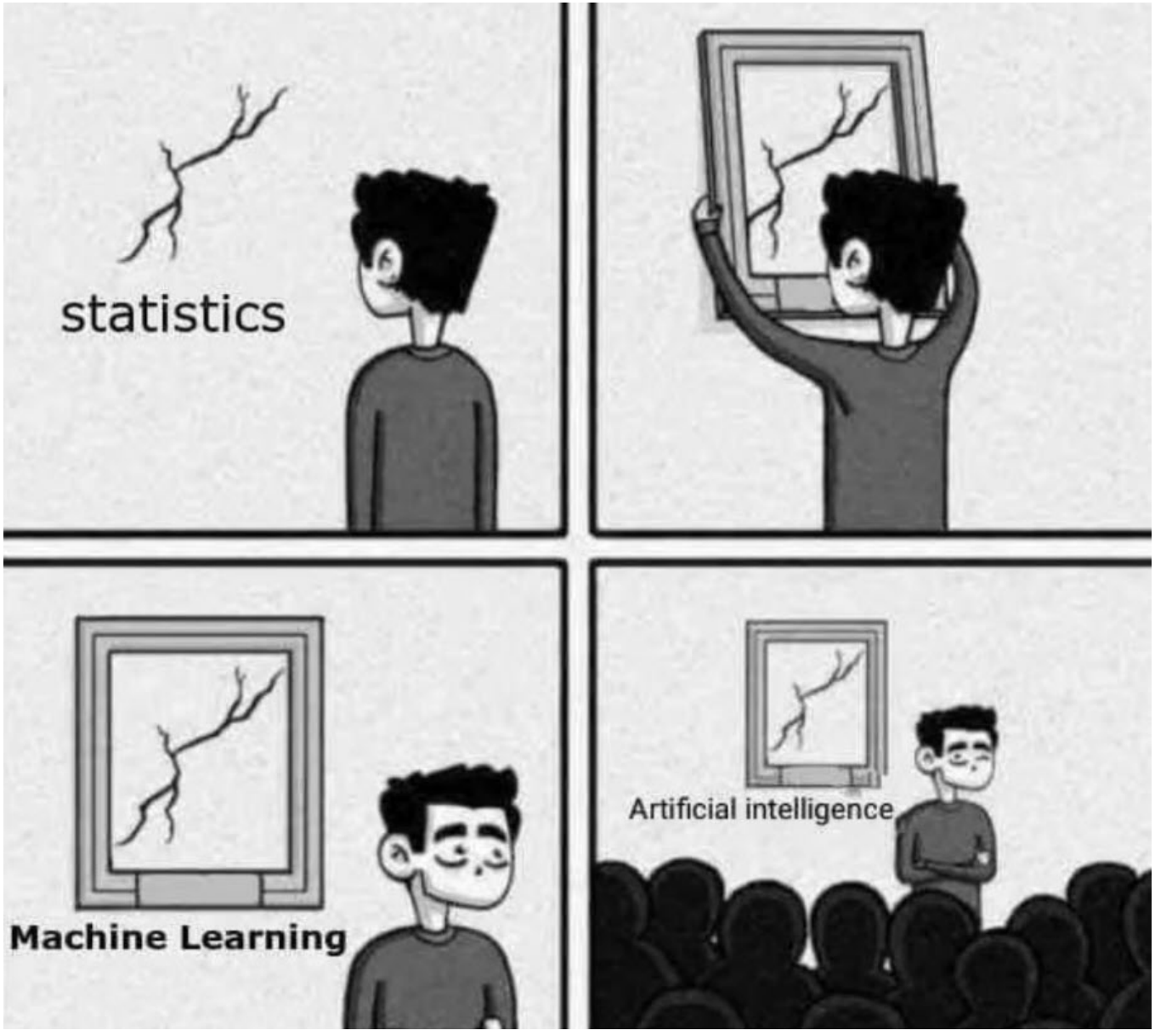
MOD500 Decision Analysis with Artificial Intelligence Support

Enrico Riccardi[1]

Department of Energy Resources, University of Stavanger (UiS).[1]

Oct 18, 2024

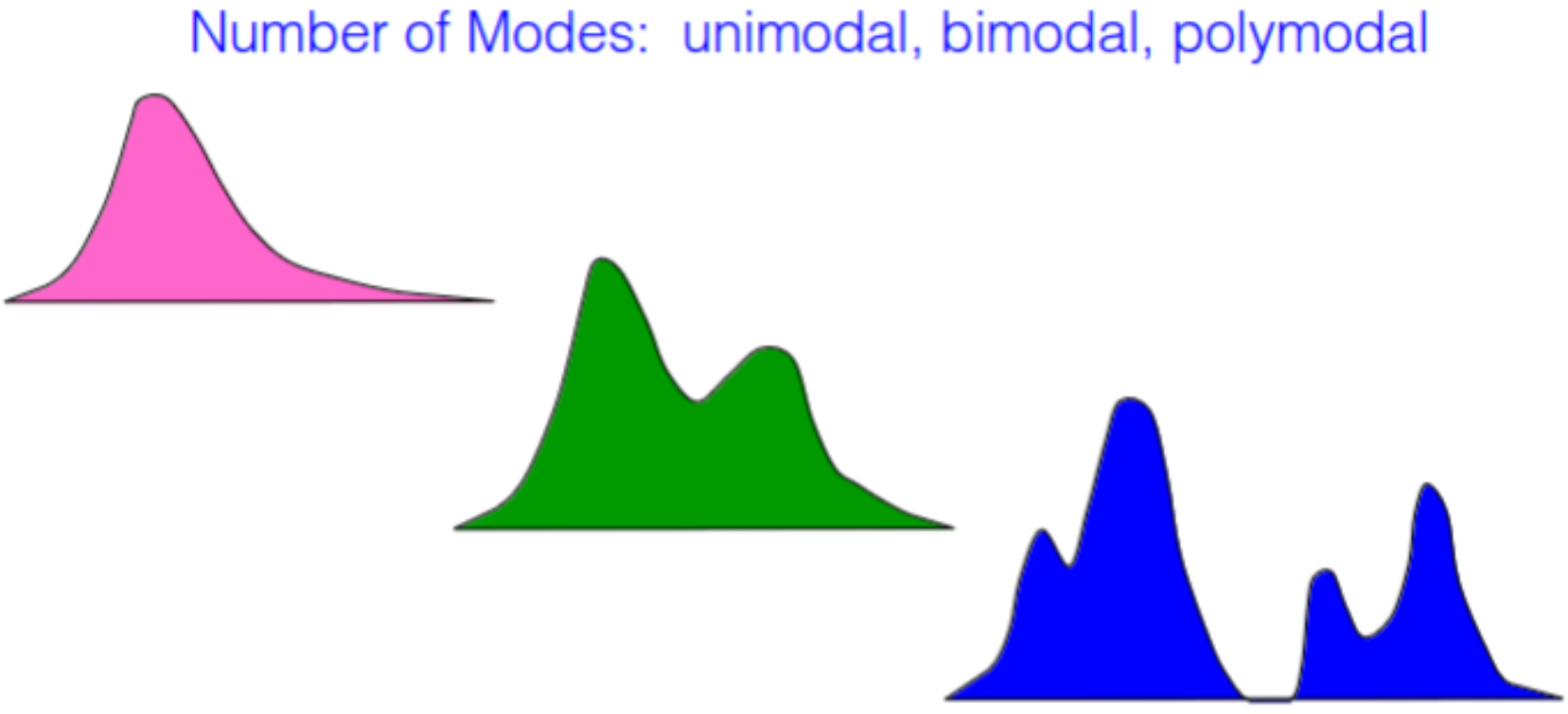
A bias recap from last classes



Probability distribution

A metric is needed

Amount of uncertainty



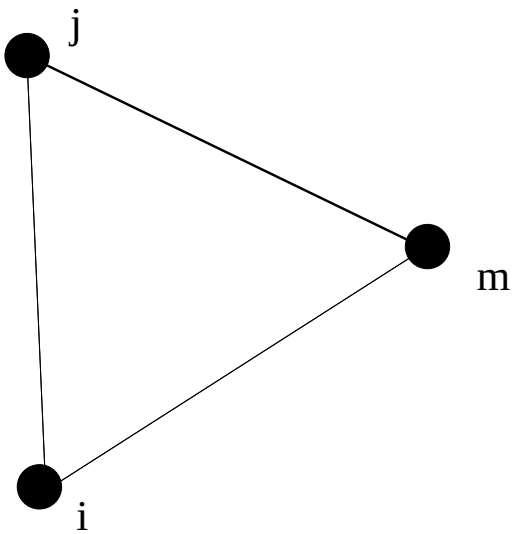
Metrics

Triangle inequality

Considering a vectors in an N-dimensional space, to be a distance it must satisfy the triangle inequality:

$$d_{ij} + d_{im} \geq d_{jm}$$

If also $d_{jj} = 0$, if $d_{ij} - d_{ji} = 0$, then we call it a *metric*.



Metris example

Common metrics:

- Counting
- Euclidean

$$d_{ij}^{(E)} = \left[\sum_{k=1}^{N} (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}}$$

Categorical variables do not necessarily satisfy these relations.

Task assignment

Code a discrete probability distribution in Python

Calculate the Mean and Standard Deviation

How to get an experiment out of this distribution?

How to relate a distribution to information?

First, the origins:

- Daddy: Claude Shannon (1940)
- His initial work has been done on signal transmission.
- It uses Entropy as key measurement of information uncertainty.

How Entropy can alter a decisions?

It is an interface between data and decisions.

A question to sum up the idea

Does more data bring value?

It has permitted the advances of several fields:

cryptography, neurobiology, signal processing, linguistics, bioinformatics, statistical physics, black holes, quantum computing, information retrieval, intelligence gathering, plagiarism detection, pattern recognition, anomaly detection, etc

zip files, phones, internet!

Entropy of an information source

$H_X = -\sum_i (p_i) log(p_i)$

H_X of a discrete random variable X is a measure of the amount of uncertainty associated with the value of X when only its distribution is known

What is p_i?

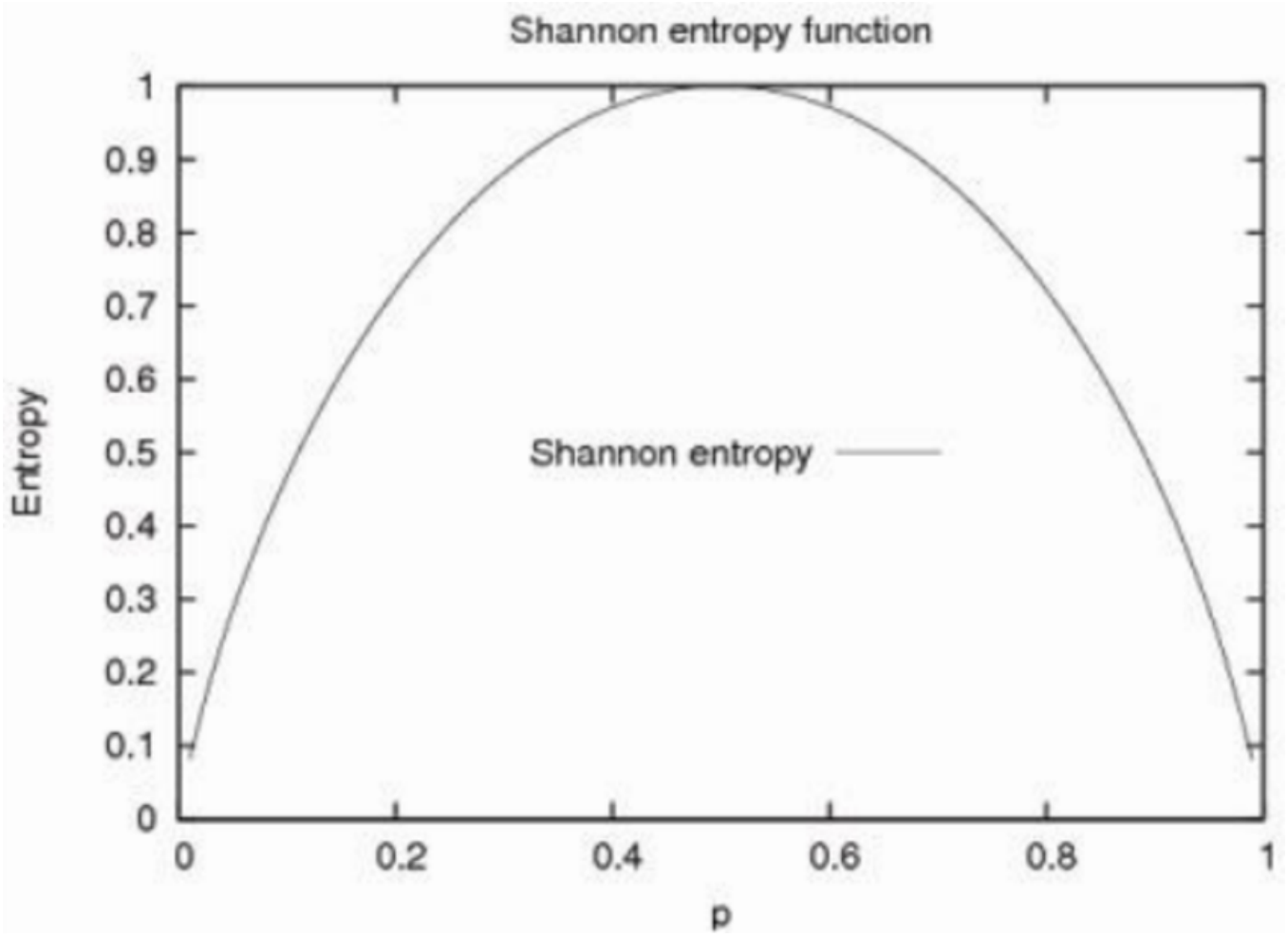
It is a numerical descriptions of how likely an event is to occur

Do not mix the concepts!

Assigned probability and computed probability are different

Shannon's entropy shape

$$H_X = -\sum_i (p_i) log(p_i)$$



Task assignment [2]

Code a discrete probability distribution in Python

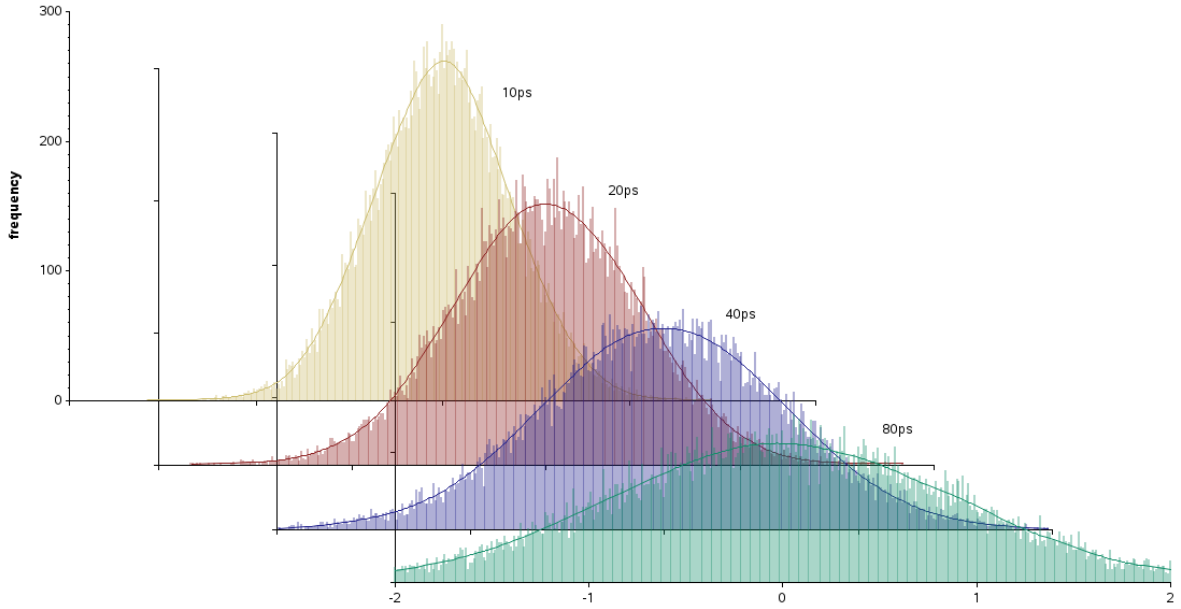
Calculate the Mean and Standard Deviation

How to get an experiment out of this distribution?

Calculate Shannon's Entropy as a function of the number of experiments

Comparing information

A distribution can be the sum of multiple distributions



Kullback–Leibler divergence (information gain)

How to compare multiple information sources?

$D_{KL} = \sum_i (p_i) log(\frac{p_i}{q_i})$

- It is a divergence, as it is asymmetric
- It is NOT a metric

Task assignment [3]

Compare the initial discrete probability distribution with the appriximate ones computed as function of the number of experiments.

Calculate Kullback–Leibler divergence

Find the distribution that minimise K-L divergence