Statistical Methods
(summer term 2024)

Maximum Entropy

(based on original lectures by Prof. Dr. N. Christlieb and Dr. Hans-G. Ludwig)

Dr Yiannis Tsapras

ZAH – Heidelberg

Overview

- One can define a quantity analogous to the thermodynamic entropy, which provides a measure of the information content of statistical distributions
- Principle of Maximum Entropy: The belief/model/distribution which best represents the current state of knowledge about a system is the one which maximizes the entropy
- The Maximum Entropy (MaxEnt) approach is rooted in information theory and is widely used across various fields, including physics and natural language processing
- MaxEnt creates a model that best accounts for the available data while ensuring that, without any additional information, the model should maximize entropy

Entropy and statistical independence

- A toy problem: "The kangaroo problem" (Gull & Skilling 1984)
  - Information: $1/3$ of all kangaroos have blue eyes
    $1/4$ of all kangaroos are left-handed
  - Question: On the basis of this information alone, what proportion of kangaroos are both blue-eyed **and** left-handed?

→ blackboard → kangaroo.ipynb

The Normal distribution as Maximum Entropy distribution

- The previous example was a special case of Shannon-Jaynes entropy (also know as the Kullback number, or cross-entropy)

$$S = -\sum p_i \log(p_i)$$

- This can be directly obtained from considering the number of possible combinations of elementary events making up a "macrocopic" event (e.g., left-handed & blue-eyed kangaroos)

- When only the expectation value μ of a distribution is known, then the distribution that maximizes entropy is the exponential (Boltzmann) distribution $\rightarrow$ blackboard

- When both the expectation value μ and the variance σ^2 of a distribution are known, then the distribution that maximizes entropy is the Normal distribution!

Entropy of a distribution – intuitive example

- Suppose we have elections. A particular politician has probability p of being elected, or probability $1-p$ of not being elected. These events are mutually exclusive and exhaustive

- Somebody who is not well informed about politics guesses that the chances of the politician are 50-50. This statement does not contain much information (high S):

$$S = -\left[0.5\log(0.5) + (1-0.5)\log(1-0.5)\right] = 0.693...$$

- Another person has looked at poll results and states that the chance of that politician to win is actually closer to 20%. This additional information implies lower S:

$$S = -\left[0.2\log(0.2) + (1-0.2)\log(1-0.2)\right] = 0.500...$$

- The same holds for the case for winning by 80%.

- S approaches zero for small or large p.