TP2: Maximum Entropy

Francesco Saverio Pezzicoli[*], Guillaume Charpiat

March 23rd, 2023

Credits: Vincenzo Schimmenti

The Maximum Entropy Principle (MEP) states that the best probability distribution that represents the current state of knowledge of the data is the probability distribution with the maximum possible entropy. In other words, the principle provides a model to interpret our data given some constraints which represent our knowledge. We will see that the MEP provides a justification for the wide use in the data modelling of distributions such as the Exponential $p(x) = \lambda e^{-\lambda x}$ or the Gaussian one $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Our data is labelled by x. It can be either discrete or continuous and can have any domain of interest (e.g. $(-\infty, \infty)$ or $[0, \infty)$ for continuous and $\{-1, 1\}$, $\{0, 1\}$ or $\{0, 1, \dots, N\}$ for discrete).

We label our Maximum Entropy Probability distribution by $p(x)$. The constraints representing out knowledge are a collection of functions $f_j(x)$ $j = 1, \dots, n$ for which we fix the expected value (imagine that experimentally you can measure the average of some function of the data):

$$\mathbb{E}[f_j(x)] = F_j \tag{1}$$

where the expected value is done w.r.t. $p(x)$. It can be shown that $p(x)$ has the form:

$$p(x) = \frac{1}{Z(\lambda_1, \dots, \lambda_n)} \exp\left(-\sum_{i=1}^{n} \lambda_i f_i(x)\right) \tag{2}$$

The quantities $\lambda_1, \dots, \lambda_n$ are free parameters and the quantity $Z(\lambda_1, \dots, \lambda_n)$ is called partition function and must be fixed to ensure normalization:

$$Z(\lambda_1, \dots, \lambda_n) = \int \mathrm{d}x \exp\left(-\sum_{i=1}^{n} \lambda_i f_i(x)\right) \tag{3}$$

The parameters λ_j must be fixed by imposing:

$$\mathbb{E}[f_j(x)] = \frac{1}{Z(\lambda_1, \dots, \lambda_n)} \int dx f_j(x) \exp\left(-\sum_{i=1}^{n} \lambda_i f_i(x)\right) = F_j \tag{4}$$

[*]francesco.pezzicoli@universite-paris-saclay.fr

An important property of $Z(\lambda_1, \ldots, \lambda_n)$ is that:

$$F_j = -\frac{\partial \log Z(\lambda_1, \ldots, \lambda_n)}{\partial \lambda_j} \tag{5}$$

Example: if we want to constraint the mean to have some value μ we need to choose a constraint $f_1(x) = x$ and impose: $\mathbb{E}[f_1(x)] = \mathbb{E}[x] = F_j = \mu$. How to proceed practically? For example we take our data to be on our positive real axis $[0, \infty)$ and we want to find the probability distribution fixing the average $\mathbb{E}[f(x)] = \mu$. Using the equation above:

$$p(x) = \frac{1}{Z} \exp\left(-\lambda_1 x\right) \tag{6}$$

We have to find Z:

$$1 = \int_0^\infty p(x)dx = \frac{1}{Z} \int_0^\infty \exp\left(-\lambda_1 x\right) dx = \tag{7}$$

$$= \frac{1}{Z\lambda_1} = 1 \tag{8}$$

This implies:

$$Z = 1/\lambda_1 \tag{9}$$

Since we want to fix $\mathbb{E}[x] = \mu$:

$$-\frac{\partial \log Z}{\partial \lambda_1} = \mu \tag{10}$$

$$\frac{\partial \log \lambda_1}{\partial \lambda_1} = \mu \tag{11}$$

$$\lambda_1 = 1/\mu \tag{12}$$

Thus we obtained:

$$p(x) = \frac{1}{\mu} e^{-x/\mu} \tag{13}$$

which is exactly the exponential distribution.

Problem 1. *Proceed in the same way to find the distribution which maximizes the entropy, has domain $(-\infty, \infty)$ and has the following constraints:*

$$f_1(x) = x \tag{14}$$

$$f_2(x) = x^2 \tag{15}$$

with values:

$$F_1 = \mu \tag{16}$$

$$F_2 = \mu^2 + \sigma^2 \tag{17}$$

To solve this, one needs to do a Gaussian integral:

$$\int_{-\infty}^{\infty} e^{-ax^2 + bx} dx = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{4a}} \tag{18}$$

Hint: at the end you will get a Gaussian distribution with mean μ and variance σ^2!

Problem 2. *Find the discrete distribution over with values in* $-1, 1$ *with maximum entropy and constraining the mean:*

$$f_1(x) = x \tag{19}$$

with value:

$$F_1 = m \tag{20}$$

Instead of the integral here one needs to do a sum!

Comments