Lecture 11: Selecting prior information

Professor Ilias Bilionis

The principle of maximum entropy for continuous random variables

PREDICTIVE
SCIENCE LABORATORY

The naïve extension of information entropy to continuous distributions and why it doesn't always work

$$\mathbb{H}[p(X)] = -\int p(x)\log p(x)\,dx$$

this is not correct

$$Y = T(X)$$

$$\mathbb{H}[p(Y)] \neq \mathbb{H}[p(X)]$$

PREDICTIVE SCIENCE LABORATORY

The correct information entropy for continuous distributions

$$\mathbb{H}[p(X)] := -\int p(x) \log \frac{p(x)}{q(x)} dx$$

this is the required key

density of maximal uncertainty

PREDICTIVE SCIENCE LABORATORY

Mathematical statement of the principle of maximum entropy for continuous distributions

$$\max \left[\mathbb{H}[p(X)] \right] = \max \left[- \int p(x) \log \frac{p(x)}{q(x)} dx \right]$$

subject to

$$\int p(x) dx = 1$$

and

$$\mathbb{E}[f_k(X)] = F_k, \quad k = 1, \dots, K$$

PREDICTIVE
SCIENCE LABORATORY

The general solution to the maximum entropy problem for continuous distributions

$$p(x) = \frac{q(x)}{Z} \exp\left\{ \sum_{k=1}^{K} \lambda_k f_k(x) \right\}$$

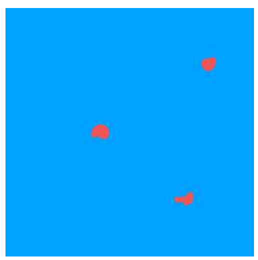
$$Z = \int q(x) \exp\left\{ \sum_{k=1}^{K} \lambda_k f_k(x) \right\} dx$$

$$\boxed{F_k = \frac{\partial Z}{\partial \lambda_k}, \quad k = 1, \dots, K}$$

find

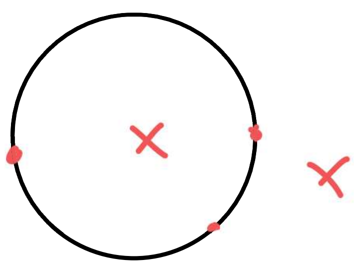
A few comments on the maximal uncertainty density $q(x)$

Example: Particle in a box:



$$q(x) = \begin{cases} 1, & \text{in box} \\ 0, & \text{other.} \end{cases} \propto 1_{Box}(x)$$

Example: Particle restricted on circular guide:



$$q(x) \propto 1_{guide}(x)$$

q(x) depends on the problem

Mathematical theory for finding the maximal uncertainty density $q(x)$

- Principle of transformation groups.
- Theory of Haar measures.

Example 1

- X takes values in $[a, b]$

- $q(x) = 1$

$X \sim U([a,b])$

Example 2

- X takes values in $\mathbb{R}$
- $q(x) = 1$
- $\mathbb{E}[X] = \mu$
- $\mathbb{V}[X] = \sigma^2$

$$X \sim N(\mu, \sigma^2)$$

Example 3

- X takes values in $[0,\infty)$
- $q(x) = 1$
- $\mathbb{E}[X] = \mu$

$$X \sim \mathrm{Exp}\left(\frac{1}{\mu}\right)$$

PREDICTIVE
SCIENCE LABORATORY

A final note on the use of maximum entropy for finding priors

- The principle of maximum entropy is a great tool for assigning “objective” priors.

- However:

    - The cost of “theorizing” and “computing” for finding the ideal distributions should be taken into account. This was called “type-2 reasoning” by I. J. Good.

    - Sometimes, you have subjective information. You should not be afraid to use it.