Lecture 11: Selecting prior information

Professor Ilias Bilionis

The principle of maximum entropy for discrete random variables

PREDICTIVE
SCIENCE LABORATORY

Prequel to the principle of maximum entropy

- You have a discrete random variable X.
- You know what values it takes, say $x_1, \ldots, x_N$.
- You also have some **testable information** about it.

↳ things like expectation, variance, etc.

- The principle of maximum entropy states that we should assign to X the probability distribution that maximizes the entropy subject to the constraints imposed by the testable information.

PREDICTIVE SCIENCE LABORATORY

Mathematical definition of testable information

$$\mathbb{E}[f_k(X)] = F_k$$

known function of X

known value

$k=1,\dots,K$

Is this definition broad enough?

I = "the expected value of X is μ"

$$\mathbb{E}[X] = \mu$$

$$K = 1, \quad f_1(x) = x, \quad F_1 = \mu.$$

PREDICTIVE
SCIENCE LABORATORY

Is this definition broad enough?

I = “the expected value of X is μ and the variance of X is σ^2”

$$\boxed{\mathbb{E}[X]=\mu}\ ;\quad \mathbb{V}[X]=\sigma^2$$

$$\sigma^2=\mathbb{V}[X]=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=\mathbb{E}[X^2]-\mu^2$$

$$\Rightarrow \boxed{\mathbb{E}[X^2]=\sigma^2+\mu^2}$$

Mathematical statement of the principle of maximum entropy

You should assign to X the pmf $p(x)$ that

$$\max\left\{\mathbb{H}[p(X)]\right\} = \max\left\{-\sum_{i=1}^{N} p(x_i)\log p(x_i)\right\}$$

subject to

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

$$\mathbb{E}[f_k(X)] = \sum_{i=1}^{N} f_k(x_i)\,p(x_i)$$

and

$$\sum_{i=1}^{N} p(x_i) = 1$$

PREDICTIVE
SCIENCE LABORATORY

The general solution to the maximum entropy problem

Lagrange Multipliers

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

need to be det.

normalization constant

$$Z = \sum_{i=1}^{N} \exp\left\{ \sum_{k=1}^{K} \lambda_k f_k(x_i) \right\}$$

$$\boxed{F_k = \frac{\partial Z}{\partial \lambda_k}}$$

solve for

Example 1

- X takes N different values (no other constraints)

$$p(X=x_i) = \frac{1}{N}$$

Example 2

- X takes two values 0 and 1.

- $\mathbb{E}[X] = \theta$.

$X \sim \text{Bernoulli}(\theta)$

PREDICTIVE
SCIENCE LABORATORY

Example 3

- X takes values $0,1,2,\ldots,N$.
- $\mathbb{E}[X] = \mu$.
- X is the number of successful trials in N sequential experiments (potentially correlated)/

$$X \sim B\left(N, \frac{\mu}{N}\right)$$

PREDICTIVE
SCIENCE LABORATORY

Example 4

- X takes values $0,1,2,\dots$
- $\mathbb{E}[X] = \mu$.
- X is the number of successful trials in an infinite number of sequential experiments (potentially correlated).

$$X \sim \text{Poisson}(\mu)$$

PREDICTIVE
SCIENCE LABORATORY