

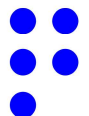
Lecture 11: Selecting prior information

Professor Ilias Bilonis

The principle of maximum entropy for discrete random variables

Prequel to the principle of maximum entropy

- You have a discrete random variable X .
- You know what values it takes, say x_1, \dots, 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.



Mathematical definition of testable information

$$E[f_k(x)] = f_k$$

known function of x

known value

$k=1, \dots, K$

Is this definition broad enough?

I = “the expected value of X is μ ”

$$E[X] = \mu$$

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

Is this definition broad enough?

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

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

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

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

Mathematical statement of the principle of maximum entropy

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

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

subject to

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

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

and

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

The general solution to the maximum entropy problem

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

Lagrange Multipliers

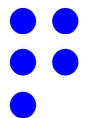
normalization constant

need to be det.

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

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

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

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

Example 3

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

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

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).

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