Information Theory: Lecture Notes 9

zqy1018

2020. 4. 20

Contents

1 Maximum Entropy

How can we find a f under some constraints that maximize $h(f)$?

1.1 With Constraints on Variance

We will show that when the variance of a random variable is determined or limited, then the maximizing X obeys the normal distribution.

Theorem 1. $X \in \mathbb{R}, EX = \mu, DX = \sigma^2$. Then $h(X) \le \frac{1}{2} \log 2\pi e \sigma^2$. The equality holds iff $X \sim N(\mu, \sigma^2)$.

Proof. Take $X_G \sim N(\mu, \sigma^2)$. Then $D(X \| X_G) \ge 0 \implies \int f_X \log \frac{f}{f_{X_G}} \ge 0 \implies h(X) \le -\int f_X \log f_{X_G}$. Then

$$\begin{aligned} -\int f_X \log f_{X_G} &= \frac{1}{2} \log 2\pi\sigma^2 + \frac{1}{2\sigma^2} \int f_X (x - \mu^2) \mathrm{dx} \\ &= \frac{1}{2} \log 2\pi\sigma^2 + \frac{1}{2\sigma^2} DX \\ &= \frac{1}{2} \log 2\pi e \sigma^2 \end{aligned}$$

The equality holds iff $f_X = f_{X_G}$ (a.e.), that is, $X \sim N(\mu, \sigma^2)$. $\square$

Sometimes we do not need the expected value.

Theorem 2. $X \in \mathbb{R}, EX^2 \le \sigma^2$. Then $h(X) \le \frac{1}{2} \log 2\pi e \sigma^2$. The equality holds iff $X \sim N(0, \sigma^2)$.

Proof. $DX = EX^2 - (EX)^2 \le \sigma^2$. By theorem 1, we can see that the equality holds when $EX = 0$, and thus $X \sim N(0, \sigma^2)$. $\square$

Note. When using the theorems above, we should first check the constraint on the variance.

1.2 With Constraints on Integrals

We consider the following constraints:

1. $f(x) \ge 0$, with equality outside the support set S
2. $\int_S f(x) \mathrm{dx} = 1$
3. $\int_S f(x) r_i(x) \mathrm{dx} = \alpha_i \quad$ for $1 \le i \le m$

Many set of constraints can be formalized this way.

Theorem 3. There is a unique maximizing f^* satisfying the constraints above which is in the form

$$f^*(x) = \exp\left[\lambda_0 + \sigma_i \lambda_i r_i(x)\right]$$

where $\lambda_0, \cdots, \lambda_m$ are to be determined.

We give some useful examples here.

Example.

1. $S = [a,b], m = 0$. Then $f^*(x) = \frac{1}{b-a}$.
2. $S = [0,\infty), EX = \mu$. Then $f^*(x) = \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right)$.
3. $S = \mathbb{R}, EX = \alpha, DX = \beta$. Then $f^*(x)$ is the c.d.f. of $X \sim N(\alpha,\beta)$.

We can see that under different constraints, the maximizing f^* will be also different.

2 Fisher Information

2.1 Basic Definition

Note. Fisher information is more related to statistics.

3 Information Inequalities

3.1 Hadamard's Inequality

Since any semi-definite symmetric matrix can be the covariance matrix of a normal random vector, we can use the properties of differential entropies to prove some inequalities about semi-definite symmetric matrices.

Theorem 4. (Hadamard's inequality) K: semi-definite symmetric matrix. Then $\prod K_{ii} \geq |K|$. The equality holds iff K is a diagonal matrix.

Proof. Let $\mathbf{X} \sim N(0,K)$. Then $h(\mathbf{X}) = \frac{1}{2}\log(2\pi e)^n|K| \leq \sum h(X_i) = \frac{1}{2}\log(2\pi e)^n \prod K_{ii}$.

So $\prod K_{ii} \geq |K|$. The equality holds iff X_i's are independent, that is, K is a diagonal matrix. □

3.2 Balanced Inequalities

3.3 EPI and FII

4 Review

Acknowledgment

The contents are mainly based on the course materials of CS258, 2020 Spring, Shanghai Jiao Tong University and *Elements of Information Theory* by Thomas M. Cover and Joy A. Thomas.