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 $h(X) \leq \frac{1}{2} \log 2\pi e \sigma^2$. The equality holds iff $X \sim N(\mu, \sigma^2)$.

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

Sometimes we do not need the expected value.

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

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

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. $f(x) \geq 0$, 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 $1 \leq i \leq m$

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 Start from Heat Equation

Definition. In physics, the **heat equation** is a PDE:

$$\frac{\partial}{\partial t} f(x, t) = \frac{1}{2} \frac{\partial^2}{\partial x^2} f(x, t)$$

where $f(x, t)$ is the temperature at time t and position x.

There is some relation between the heat equation and normal random variables.

Fact. Let $f(x)$ be the p.d.f. of a random variable X, and $Z \sim N(0, 1)$ be independent of X. Then let $Y_t = X + \sqrt{t}Z$. The p.d.f. of Y_t

$$f(y; t) = \int f(x) \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} \mathrm{dx}$$

satisfies the heat equation.

2.2 Basic Definition

Definition. Let $f(x)$ be the p.d.f. of a random variable X. Then its **Fisher information** is defined as

$$I(X) = \int_{-\infty}^{\infty} f(x) \left[\frac{\frac{\partial}{\partial x} f(x)}{f(x)}\right]^2 \mathrm{dx}$$

Note. We write the partial differentiator to indicate that $f(x)$ may be parameterized.

Remark. Fisher information is more related to statistics.

2.3 Sign of Partial Derivatives of $h(Y_t)$

For Y_t, there is a special relation.

Theorem 4.

$$\frac{\partial}{\partial t}h(Y_t) = \frac{1}{2}I(Y_t)$$

There is some interesting properties about the sign of (high-order) derivatives of $h(Y_t)$.

Theorem 5.

$$\frac{\partial}{\partial t}h(Y_t) \geq 0, \frac{\partial^2}{\partial t^2}h(Y_t) \leq 0, \frac{\partial^3}{\partial t^3}h(Y_t) \geq 0, \frac{\partial^2 4}{\partial t^4}h(Y_t) \leq 0$$

For higher orders, it is still an open problem to determine the sign.

For a general X, it is very hard to find $h(Y_t)$. But for $X \sim N(0,1)$, we have

$$h(Y_t) = \frac{1}{2}\log\left[2\pi e(1+t)\right]$$

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 6. (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

For a linear inequality about entropies, we can define the net weight of each random variable.

Definition. Let $\alpha = \{\alpha_i\} \subset [n] = \{1, \cdots, n\}$. Let X_α (or $X(\alpha)$) be $\left(X_{\alpha_1}, \cdots, X_{\alpha_{|\alpha|}}\right)$. Then a linear inequality about entropies can be written as

$$\sum_\alpha w_\alpha H(X_\alpha) \geq 0$$

and the **net weight** of X_i is $\sum_\alpha w_\alpha [i \in \alpha]$.

Note. The conditions are not count into weight.

Definition. An inequality is called **balanced** if $\forall i \in [n]$, the net weight of X_i is 0.

With the balanced condition, we can bridge the inequalities on differential entropies and those on discrete entropies.

Theorem 7. $\sum_\alpha w_\alpha h(X_\alpha) \geq 0$ is valid iff $\sum_\alpha w_\alpha H(X_\alpha) \geq 0$ is valid and balanced.

An application of this theorem is the Han's inequality.

Theorem 8. (Han's inequality) Let

$$h_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{|S|=k} \frac{h(X_S)}{k}, g_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{|S|=k} \frac{h(X_S|X_{S^c})}{k}$$

where $S \subset [n]$. Then

$$h_1^{(n)} \geq h_2^{(n)} \geq \cdots \geq h_n^{(n)} = g_n^{(n)} \geq g_{n-1}^{(n)} \geq g_1^{(n)}$$

This inequality is balanced so it holds for both differential entropies and discrete entropies.

3.3 EPI and FII

Theorem 9. (Entropy-power Inequality, EPI) For two n-dimensional random vectors X, Y,

$$\exp\left[\frac{2}{n}h(X+Y)\right] \geq \exp\left[\frac{2}{n}h(X)\right] + \exp\left[\frac{2}{n}h(Y)\right]$$

This inequality will be useful in dealing with the Gaussian channel.

Theorem 10. (Fisher Information Inequality, FII)

$$\frac{1}{I(X+Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}$$

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.