COMPSCI 650 Applied Information Theory Mar 8, 2016

Lecture 1

Instructor: Arya Mazumdar *Scribe: Arya Mazumdar*

Today we will spend sometime discussing about the upcoming midterm. Hence the length of the lecture is smaller than usual.

1 Secret Sharing Schemes

This is an example that we discuss in the class.

Suppose a secret S has to be divided between two users. The share of user 1 is X_1, and share of user 2 is X_2. We must have the following two properties:

1. $H(S|X_1) = H(S|X_2) = H(S)$.
2. $H(S|X_1, X_2) = 0$.

Let $S \in \mathcal{X} \equiv \{0, \dots, q-1\}$. Then choose X_1 randomly and uniformly from $\mathcal{X}$. Let $X_2 = (X_1 + S) \mod q$. This scheme clearly satisfies the above two properties.

Question: Can you think of a generalization?

2 Parameter Estimation

Consider an indexed family of distributions $\{f(x; \theta)\}$. X is the underlying random variable with sample space $\mathcal{X}$. That is we have,

- $f(x; \theta) \geq 0$.
- $\int f(x; \theta) dx = 1$.

Here $\theta \in \Theta$, the parameter set.

An estimator is a mapping

$$T : \mathcal{X}^n \to \Theta.$$

The estimation error is

$$E_\theta(T(X_1, \dots, X_n) - \theta),$$

where the θ in the subscript denote that the expect ion is taken with respect to $f(x; \theta)$. The estimator is called *unbiased* if the estimation error is 0 for all $\theta \in \Theta$. Note that we can ask for more out of an estimator; such as $P_\theta(|T(X_1, \dots, X_n) - \theta| > \epsilon) \to 0$ for any $\epsilon > 0$. An estimator T_1 dominates another estimator T_2 if for all θ,

$$E_\theta(T_1(X_1, \dots, X_n) - \theta)^2 \leq E_\theta(T_2(X_1, \dots, X_n) - \theta)^2.$$

This begs the question: what is the best estimator?

2.1 Score function

The score function is:

$$V(X) = \frac{\partial}{\partial \theta} \ln f(X; \theta) = \frac{\frac{\partial}{\partial \theta} f(X; \theta)}{f(X; \theta)}.$$

If $X_1, \dots, X_n$ are i.i.d. $f(x; \theta)$, then:

$$V(X_1, \dots, X_n) = \frac{\partial}{\partial \theta} \ln f(X_1, \dots, X_n; \theta) = \frac{\partial}{\partial \theta} \sum_i \ln f(X_i; \theta) = \sum_i V(X_i).$$

Also, we must have,

$$E_\theta V = \int f(x;\theta)\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}dx = \frac{\partial}{\partial\theta}\int f(x;\theta)dx = \frac{\partial}{\partial\theta}1 = 0.$$

Hence,

$$\mathrm{Var}_\theta(V) = E_\theta V^2.$$

2.2 Fisher Information

How to quantify the amount of information about θ that is present in the data?

Define,

$$J(\theta) = \mathrm{Var}_\theta(V) = E_\theta V^2.$$

If $X_1,\ldots,X_n$ are i.i.d. $f(x;\theta)$, then $\mathrm{Var}_\theta(\sum_i V(X_i)) = \sum_i \mathrm{Var}_\theta(V(X_i))$. Hence,

$$J_n(\theta) = nJ(\theta).$$

2.3 Cramer-Rao bound

Theorem 1 *The mean-square error of any unbiased estimator T is*

$$\mathrm{Var}_\theta(T) \geq \frac{1}{J(\theta)}.$$

Proof Just an application of the Cauchy-Schwartz inequality. Note,

$$E_\theta T = \theta$$

and

$$E_\theta V = 0.$$

Hence,

$$\begin{aligned}
E_\theta\Big((V - E_\theta V)(T - E_\theta T)\Big) &= E_\theta\Big(VT - V\theta\Big) \\
&= E_\theta\Big(VT\Big) \\
&= \int f(x;\theta)\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}T(x)dx \\
&= \frac{\partial}{\partial\theta}\int f(x;\theta)T(x)dx \\
&= \frac{\partial}{\partial\theta}\theta \\
&= 1.
\end{aligned}$$

But from Cauchy-Schwartz,

$$1 = \Big(E_\theta\Big((V - E_\theta V)(T - E_\theta T)\Big)\Big)^2 \leq \mathrm{Var}_\theta V \cdot \mathrm{Var}_\theta(T) = J(\theta)\cdot \mathrm{Var}_\theta(T).$$

■

Example: Consider $X_1,\ldots,X_n$ all i.i.d. Gaussian $\mathcal{N}(\theta,\sigma^2)$.

$$V(X) = \frac{\partial}{\partial\theta}\ln\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(X-\theta)^2}{2\sigma^2}} = \frac{X-\theta}{\sigma^2}.$$

Therefore,

$$J(\theta) = \mathrm{Var}_\theta V = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}.$$

SInce, $J_n(\theta) = nJ(\theta)$, from Cramer-Rao,

$$\mathrm{Var}_\theta(T) \geq \frac{\sigma^2}{n}.$$

Now consider the estimator

$$T(X_1, \ldots, X_n) = \frac{1}{n}\sum_i X_i.$$

We have,

$$E_\theta(T-\theta)^2 = \mathrm{Var}_\theta\Big(\frac{1}{n}\sum_i X_i\Big) = \frac{1}{n}\sum_i \mathrm{Var}_\theta(X_i) = \frac{\sigma^2}{n}.$$

So we can achieve the Cramer-Rao bound!