EE6367: Topics in Data Storage and Communications **2023**

Lecture 3: 11 September 2023

Instructor: Shashank Vatedka *Scribe: Gautam Singh*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

3.1 Cramer-Rao Inequality

This inequality shows a lower bound on the variance of an estimator. In particular, if the estimator is unbiased, it gives a lower bound on the mean squared error of the estimator.

Theorem 3.1 (Cramer-Rao Lower Bound)**.** *Suppose $f_{X^n|\theta}$ is a known probability density on $\mathbb{R}^n$, and $g\left(X^n\right)$ is any estimator satisfying*

$$\frac{\partial}{\partial\theta}\int_{x^n\in\mathbb{R}^n} f_{X^n|\theta}\left(x^n\right)g\left(x^n\right)dx^n = \int_{x^n\in\mathbb{R}^n} g\left(x^n\right)\frac{\partial}{\partial\theta}\left(f_{X^n|\theta}\left(x^n\right)\right)dx^n. \tag{3.1}$$

Then,

$$\mathrm{Var}\left(g\left(X^n\right)\right) \geqslant \frac{\left(\frac{\partial}{\partial\theta}\mathbb{E}\left[g\left(X^n\right)\right]\right)^2}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_{X^n|\theta}\left(X^n\right)\right)^2\right]}. \tag{3.2}$$

Proof. We prove (3.2) using the *Cauchy-Schwarz Inequality*

$$\left(\mathrm{Cov}\left(U,V\right)\right)^2 \leqslant \mathrm{Var}\left(U\right)\mathrm{Var}\left(V\right). \tag{3.3}$$

Setting $U = g\left(X^n\right)$ and $V = \frac{\partial}{\partial\theta}\log f_{X^n|\theta}\left(X^n\right)$, and noting that

$$\begin{aligned}
\mathbb{E}\left[V\right] &= \int_{x^n\in\mathbb{R}^n} f_{X^n|\theta}\left(x^n\right)\frac{\partial}{\partial\theta}\log f_{X^n|\theta}\left(x^n\right)dx^n && (3.4)\\
&= \int_{x^n\in\mathbb{R}^n}\frac{\partial}{\partial\theta}f_{X^n|\theta}\left(x^n\right)dx^n && (3.5)\\
&= \frac{\partial}{\partial\theta}\int_{x^n\in\mathbb{R}^n} f_{X^n|\theta}\left(x^n\right)dx^n && (3.6)\\
&= \frac{\partial}{\partial\theta}\left(1\right) = 0, && (3.7)
\end{aligned}$$

where (3.5) is obtained by setting $g\left(x^n\right) = 1$ in (3.1) and (3.7) is obtained by using the definition of a probability density function. Using (3.7),

$$\mathrm{Cov}\left(U,V\right) = \mathbb{E}\left[\left(U - \mathbb{E}\left[U\right]\right)\left(V - \mathbb{E}\left[V\right]\right)\right] \tag{3.8}$$

$$= \mathbb{E}\left[UV\right] - \mathbb{E}\left[U\right]\mathbb{E}\left[V\right] = \mathbb{E}\left[UV\right] \tag{3.9}$$

$$= \int_{x^n \in \mathbb{R}^n} g\left(x^n\right)\left(\frac{\partial}{\partial\theta} f_{X^n|\theta}\left(x^n\right)\right) dx^n. \tag{3.10}$$

Thus, using (3.10), (3.7) and rearranging (3.3), we obtain (3.2), as desired. □

An important corollary of Theorem 3.1 is obtained when the estimator $g\left(X^n\right)$ is unbiased. In this case, we have $\mathbb{E}\left[g\left(X^n\right)\right] = \theta$ and $\mathrm{Var}\left(g\left(X^n\right)\right) = \mathrm{MSE}\left(g\right)$, which gives us Corollary 3.2.

Corollary 3.2 (Cramer-Rao Lower Bound for Unbiased Estimators)**.** *In addition to the conditions imposed on $f_{X^n|\theta}$ and $g\left(X^n\right)$ in Theorem 3.1, if g is an unbiased estimator, then*

$$\mathrm{MSE}\left(g\right) \geqslant \frac{1}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_{X^n|\theta}\left(X^n\right)\right)^2\right]}. \tag{3.11}$$

Additionally, the denominator of the right-hand side of (3.11) *is called the* ***Fisher information*** *that X^n carries about θ.*

Example 3.1. *Calculate the Fisher information carried by N independent and identically distributed samples $X_i \sim \mathcal{N}\left(\theta, \sigma^2\right),\ 1 \leqslant i \leqslant n$.*

Solution: Note that

$$\frac{\partial}{\partial\theta}\log f_{X^n|\theta}\left(X^n\right) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\left[-\frac{\left(X_i - \theta\right)^2}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)\right] = \sum_{i=1}^{n}\frac{X_i - \theta}{\sigma^2}. \tag{3.12}$$

Since the samples are iid, the Fisher information is given by

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_{X^n|\theta}\left(X^n\right)\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n}\frac{X_i - \theta}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^4}\sum_{i=1}^{n}\mathrm{Var}\left(X_i\right) = \frac{n}{\sigma^2}. \tag{3.13}$$

■

3.2 Estimation Protocols

Consider a situation where independent and identically distributed samples are present at various locations. These sampls need to be sent to a central server that computes the mean of these samples. Communication constraints are now involved, such as the addition of noise or constraint on the number of bits that can be reliably sent (quantization error). Therefore, we need to replace estimators with *estimation protocols*. Some use cases of estimation protocols include sensor networks and federated learning.