Efficient Estimation of Scalar Parameters

Aleksandar Dogandžić

2017-01-31

Contents

Reading: §3 in the textbook and (Hero 2015, §4.4.4).

Review: Information Inequality and Cramér-Rao Bound

Information inequality

Consider statistics $T(X)$ that satisfy

$$\mathrm{var}_{X\,|\,\Theta}[T(X)\,|\,\theta] < +\infty$$

for all θ. Suppose that Assumptions 1 and 2 from handout `crb` hold and $0 < \mathcal{I}(\theta) < +\infty$, where $\mathcal{I}(\theta)$ is the Fisher information for θ. Define

$$\mathrm{E}_{X\,|\,\Theta}[T(X)\,|\,\theta] = \psi(\theta).$$

Then, for all θ,

$$\mathrm{var}_{X\,|\,\Theta}[T(X)\,|\,\theta] \geq \frac{|\psi'(\theta)|^2}{\mathcal{I}(\theta)} \tag{1a}$$

where

$$\psi'(\theta) = \frac{\mathrm{d}\psi(\theta)}{\mathrm{d}\theta}. \tag{1b}$$

Cramér-Rao bound

SUPPOSE that Assumptions 1 and 2 from handout crb hold and that $T(X)$ is an unbiased estimator of θ, i.e.,

$$\mathrm{E}_{X\,|\,\Theta}[T(X)\,|\,\theta] = \theta.$$

Then

$$\mathrm{var}_{X\,|\,\Theta}[T(X)\,|\,\theta] \geq \frac{1}{\mathcal{I}(\theta)} \tag{2}$$

where equality is attained if and only if, for some nonradom scalar c_θ,

$$\frac{\partial}{\partial\theta} \ln f_{X\,|\,\Theta}(x\,|\,\theta) = c_\theta\,[T(x) - \theta]. \tag{3}$$

Efficiency

AN unbiased estimator of θ that attains the Cramér-Rao bound (CRB) for θ for all θ in the parameter space sp_Θ is said to be efficient.

✻ NOTE: Efficient estimators are always minimum-variance unbiased (MVU), but not conversely.[1] Indeed, CRB is not always attainable by MVU estimators (particularly for finite samples, i.e., finite number of measurements N).

[1] (Hero 2015) gives a reference to a counterexample.

Under certain regularity conditions, maximum-likelihood (ML) estimators attain CRB asymptotically (for large N); hence they are *asymptotically efficient*.

Proof that efficiency implies MVU: For *any unbiased* estimator, its variance must be greater than or equal to the CRB. If there exists an unbiased estimator whose variance is equal to CRB for all $\boldsymbol{\theta}$ in the parameter space sp_Θ, then this estimator must be MVU. □

Example: i.i.d. Poisson measurements

CONSIDER i.i.d. measurements $\big(X[n]\big)_{n=0}^{N-1}$ from Poisson(λ) distribution:

$$p_{X\,|\,\Lambda}(x[n]\,|\,\lambda) = \frac{\lambda^{x[n]}}{x[n]!}\mathrm{e}^{-\lambda}.$$

We have derived the Fisher information for λ in handout crb:

$$\mathcal{I}(\lambda) = \frac{N}{\lambda}.$$

In this example,

$$\bar{X} = \frac{1}{N}\sum_{n=0}^{N-1} X[n]$$

sample mean

is the ML estimator of λ and it is unbiased.

For Poisson distribution,

$$\mathrm{var}_{X\,|\,\Lambda}(X[n]\,|\,\lambda) = \lambda$$

and

$$\begin{aligned}
\mathrm{var}_{\boldsymbol{X}\,|\,\Lambda}(\overline{X}\,|\,\lambda) &= \mathrm{var}_{\boldsymbol{X}\,|\,\Lambda}\Big(\frac{1}{N}\sum_{n=0}^{N-1} X[n]\,\Big|\,\lambda\Big) \\
&= \frac{1}{N^2}\,\mathrm{var}_{\boldsymbol{X}\,|\,\Lambda}\Big(\sum_{n=0}^{N-1} X[n]\,\Big|\,\lambda\Big) \qquad X[n] \text{ i.i.d.} \\
&= \frac{1}{N^2} N\lambda \\
&= \frac{\lambda}{N} \\
&= \frac{1}{\mathcal{I}(\lambda)} = \mathrm{CRB}(\lambda)
\end{aligned}$$

which implies that $\overline{X}$ attains the CRB for λ, see (2).

✻ An alternative argument: Note that

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\ln p_{\boldsymbol{X}\,|\,\Lambda}(\boldsymbol{x}\,|\,\lambda) &= \frac{\sum_{n=0}^{N-1} x[n]}{\lambda} - N \\
&= \frac{N}{\lambda}(\overline{x} - \lambda)
\end{aligned} \tag{4}$$

satisfies the CRB tightness condition (3); hence $\overline{X}$ attains the CRB for its expectation λ:

$$\begin{aligned}
\mathrm{var}_{\boldsymbol{X}\,|\,\Lambda}(\overline{X}\,|\,\lambda) &= \frac{1}{\mathcal{I}(\lambda)} \\
&= \mathrm{CRB}(\lambda) && \text{(5a)} \\
\mathrm{E}_{\boldsymbol{X}\,|\,\Lambda}(\overline{X}\,|\,\lambda) &= \lambda. && \text{(5b)}
\end{aligned}$$

Example: i.i.d. Gaussian measurements with unknown mean

Consider i.i.d. measurements $(X[n])_{n=0}^{N-1}$ from $\mathcal{N}(\mu, \sigma^2)$, conditional on μ. Here, μ is the unknown parameter and σ^2 is a known constant.

Then, the score function

$$\frac{\partial \ln f_{\boldsymbol{X}\,|\,\mu}(\boldsymbol{x}\,|\,\mu)}{\partial\mu} = \frac{N}{\sigma^2}(\overline{x} - \mu)$$

satisfies the CRB tightness condition (3); hence $\overline{X}$ attains the CRB for its expectation μ:

$$\begin{aligned}
\mathrm{var}_{\boldsymbol{X}\,|\,\mu}(\overline{X}\,|\,\mu) &= \frac{1}{\mathcal{I}(\mu)} \\
&= \mathrm{CRB}(\mu) && \text{(6a)} \\
\mathrm{E}_{\boldsymbol{X}\,|\,\mu}(\overline{X}\,|\,\mu) &= \mu. && \text{(6b)}
\end{aligned}$$

Observe that

$$
\begin{aligned}
\mathrm{E}_{\boldsymbol{X}|\mu}(\,\overline{X} \mid \mu) &= \mu \\
\mathrm{var}_{\boldsymbol{X}|\mu}(\,\overline{X} \mid \mu) &= \frac{1}{N^2}\,\mathrm{var}_{\boldsymbol{X}|\mu}\Big(\sum_{n=0}^{N-1} X[n] \,\Big|\, \mu\Big) \\
&= \frac{1}{N^2} N\, \mathrm{var}_{X|\mu}(X[n] \mid \mu) \\
&= \frac{\sigma^2}{N} \\
&= \frac{1}{\mathcal{I}(\mu)} \\
&= \mathrm{CRB}(\mu).
\end{aligned}
$$

$X[n]$ i.i.d.

One-parameter Canonical Exponential Family

In handout `expon_family`, we introduced the one-parameter canonical exponential family:

$$f_{\boldsymbol{X}|\eta}(\boldsymbol{x} \mid \eta) = h(\boldsymbol{x}) \exp[\eta T(\boldsymbol{x}) - A(\eta)]. \tag{7}$$

The score function is

$$\frac{\partial \ln f_{\boldsymbol{X}|\eta}(\boldsymbol{x} \mid \eta)}{\partial \eta} = T(\boldsymbol{x}) - \frac{\mathrm{d}A(\eta)}{\mathrm{d}\eta}. \tag{8}$$

We have shown in handout `crb` that

$$\mathrm{E}_{\boldsymbol{X}|\eta}[T(\boldsymbol{X}) \mid \eta] = \frac{\mathrm{d}A(\eta)}{\mathrm{d}\eta}, \qquad \mathrm{var}_{\boldsymbol{X}|\eta}(T(\boldsymbol{X}) \mid \eta) = \frac{\mathrm{d}^2 A(\eta)}{\mathrm{d}\eta^2} \tag{9}$$

see also (5) in handout `expon_family`. Since (8) has the form in (3), we also know that $T(\boldsymbol{X})$ is an efficient estimate of its expectation $\mathrm{E}_{\boldsymbol{X}|\eta}[T(\boldsymbol{X}) \mid \eta] = \mathrm{d}A(\eta)/\mathrm{d}\eta$.

Exponential Family of Distributions, Mean-Value Parameterization

Exponential family of distributions plays a special role regarding efficiency. In particular, if X is a measurement from a density in the exponential family with scalar parameter θ

$$f_{X|\Theta}(x \mid \theta) = h(x) \exp[\eta(\theta) T(x) - B(\theta)]$$

having the mean-value parameterization[2], then

$$\mathrm{E}_{X|\Theta}[T(X) \mid \theta] = \theta \tag{10}$$

and the Fisher information $\mathcal{I}(\theta)$ for θ given X is

$$\mathcal{I}(\theta) = \frac{1}{\mathrm{var}_{X|\boldsymbol{\theta}}[T(X) \mid \theta]}.$$

[2] Recall discussion in handout `expon_family`. (10) holds by the definition of mean-value parameterization

If we have i.i.d. measurements $(X[n])_{n=0}^{N-1}$ from this distribution, then

$$\widehat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} T(X[n])$$

is an unbiased and efficient estimator of θ. HW: Prove this.

Existence of efficient estimators

EFFICIENT estimators exist only when measurements come from the exponential family of distributions with mean-value parametrization.

Without loss of generality we specialize to the case of a single measurement ($N = 1$) and parameter space over the entire real line $\mathrm{sp}_\Theta = (-\infty, +\infty)$. Recall the CRB tightness condition (3):

$$\frac{\partial}{\partial \theta} \ln f_{X \mid \Theta}(x \mid \theta) = c_\theta [T(x) - \theta]$$

where $T(X)$ is an unbiased estimator of the parameter θ. For fixed θ_0, integrate the left-hand side (LHS) over $\theta \in [\theta_0, \theta']$:

$$\int_{\theta_0}^{\theta'} \frac{\partial}{\partial \theta} \ln f_{X \mid \Theta}(x \mid \theta) \, \mathrm{d}\theta = \ln f_{X \mid \Theta}(x \mid \theta') - \ln f_{X \mid \Theta}(x \mid \theta_0)$$

and integrate the right-hand side (RHS) over $\theta \in [\theta_0, \theta']$:

$$\int_{\theta_0}^{\theta'} c_\theta \, [T(x) - \theta] \, \mathrm{d}\theta = T(x) \underbrace{\int_{\theta_0}^{\theta'} c_\theta \, \mathrm{d}\theta}_{\eta(\theta')} - \underbrace{\int_{\theta_0}^{\theta'} \theta c_\theta \, \mathrm{d}\theta}_{B(\theta')} .$$

Combine the integrals of RHS and LHS:

$$f_{X \mid \Theta}(x \mid \theta) = \underbrace{f_{X \mid \Theta}(x \mid \theta_0)}_{h(x)} \exp[\eta(\theta) T(x) - B(\theta)].$$

✻ POISSON example. $(X[n])_{n=0}^{N-1}$ are i.i.d. Poisson(λ):

$$\begin{aligned} p_{\boldsymbol{X} \mid \Lambda}(\boldsymbol{x} \mid \lambda) &= \frac{\lambda^{\sum_{n=0}^{N-1} x[n]}}{\prod_{n=0}^{N-1} x[n]!} \exp(-N\lambda) \\ &= \frac{1}{\prod_{n=0}^{N-1} x[n]!} \exp\Bigg(\underbrace{\frac{1}{N} \sum_{n=0}^{N-1} x[n]}_{T(\boldsymbol{x})} \; N \ln \lambda - N\lambda \Bigg) \end{aligned}$$

where $\boldsymbol{x} = (x[n])_{n=0}^{N-1}$. Note that this $p_{\boldsymbol{X} \mid \Lambda}(\boldsymbol{x} \mid \lambda)$ belongs to the exponential family of distributions and the last expression above is the mean-value parametrization for λ. Hence, the natural sufficient statistic

$$T(\boldsymbol{x}) = \frac{1}{N}\sum_{n=0}^{N-1} X[n]$$

is an efficient estimator of

$$\begin{aligned} \mathrm{E}_{\boldsymbol{X}|\Lambda}[T(\boldsymbol{x}) \mid \lambda] &= \mathrm{E}_{\boldsymbol{X}|\Lambda}\left(\frac{1}{N}\sum_{n=0}^{N-1} X[n] \,\middle|\, \lambda\right) \\ &= \lambda. \end{aligned}$$

Acronyms

CRB Cramér-Rao bound. 2, 3, 5

i.i.d. independent, identically distributed. 1–3, 5

LHS left-hand side. 5

ML maximum-likelihood. 2, 3

MVU minimum-variance unbiased. 2

RHS right-hand side. 5

References

Hero, Alfred O. (2015). *Statistical Methods for Signal Processing*. Lecture notes. Univ. Michigan, Ann Arbor, MI.