Lower Bounds on Exponential Moments of the Quadratic Error in Parameter Estimation

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 3200004, Israel

ISIT 2018, Vail, Colorado, U.S.A., June 2018.

Objectives

We consider the problem of estimating a parameter θ based on an observation, $\boldsymbol{Y}$.

Instead of the MSE, we seek non–trivial lower bounds on:

Bayesian regime: $\mathbf{E}\exp\{\alpha[\hat{\theta}(\boldsymbol{Y})-\Theta]^2\}$

Non-Bayesian regime: $\mathbf{E}_\theta\exp\{\alpha[\hat{\theta}(\boldsymbol{Y})-\theta]^2\}, \quad \hat{\theta}(\cdot)$ is unbiased

where $\alpha > 0$ is a given constant.

By "non–trivial" lower bounds, we mean something more sophisticated than applying Jensen's inequality:

$$\mathbf{E}\exp\{\alpha[\hat{\theta}(\boldsymbol{Y})-\Theta)^2\} \geq \exp\{\alpha\mathbf{E}[\hat{\theta}(\boldsymbol{Y})-\theta]^2\} \geq \exp\{\alpha[\text{MSE lower bound}]\}.$$

Motivations

- Risk–sensitivity: penalizing large errors (optimization & stoch. control).
- Robustness to model uncertainty.
- The MGF/CGF of $(\hat{\theta} - \theta)^2$ as a function of α is more informative.
- Related to large deviations performance, $\Pr\{|\hat{\theta} - \theta| \geq \delta\}$.
- Tail behavior of $\epsilon = \hat{\theta} - \theta$:

$$\alpha \geq \alpha_{\text{c}} \;\rightarrow\; \mathbf{E}\exp\{\alpha[\hat{\theta}(\boldsymbol{Y}) - \Theta]^2\} \text{ diverges } \;\rightarrow\; \text{tail of } f(\epsilon) > \exp\{-\alpha_{\text{c}}\epsilon^2\}.$$

Difficulty

In the Bayesian regime, the optimal estimator

$$\hat{\theta}(\boldsymbol{y}) = \arg\min_{\eta} \mathbf{E}\left[\exp\{\alpha(\Theta - \eta)^2\} | \boldsymbol{Y} = \boldsymbol{y}\right],$$

which is given by the solution, η, to the equation,

$$\eta = \frac{\mathbf{E}\left[\Theta \exp\{\alpha(\Theta - \eta)^2\} | \boldsymbol{Y} = \boldsymbol{y}\right]}{\mathbf{E}\left[\exp\{\alpha(\Theta - \eta)^2\} | \boldsymbol{Y} = \boldsymbol{y}\right]},$$

is even more difficult to calculate than the MMSE estimator,

$$\hat{\theta}(\boldsymbol{y}) = \mathbf{E}\{\Theta | \boldsymbol{Y} = \boldsymbol{y}\},$$

hence the quest for lower bounds is even more crucial here.

Generic Lower Bounds

Based on the Laplace principle,

$$\ln \mathbf{E}_P e^Z \geq \mathbf{E}_Q Z - D(Q\|\|P),$$

we have the following generic bounds,

Bayesian:

$$\forall Q(\theta, \boldsymbol{y}) : \quad \ln \mathbf{E} \exp\{\alpha[\hat{\theta}(\boldsymbol{Y}) - \Theta]^2\} \geq \alpha L_{\mathsf{B}}(Q) - D(Q\|P),$$

where $L_{\mathsf{B}}(Q)$ is an arbitrary Bayesian MSE lower bound under Q.

Non–Bayesian:

$$\forall \theta' : \quad \ln \mathbf{E}_\theta \exp\{\alpha[\hat{\theta}(\boldsymbol{Y}) - \theta]^2\} \geq \alpha L_{\mathsf{NB}}(\theta') + \alpha(\theta' - \theta)^2 - D(P_{\theta'}\|P_\theta),$$

where $L_{\mathsf{NB}}(\theta')$ is an arbitrary non–Bayesian MSE lower bound under θ'.

These generic bounds offer considerable freedom...

Bayesian Regime – Conditions for Tightness

If you can guess an estimator $g(\boldsymbol{y})$ and a reference measure Q, such that both:

- $g(\cdot)$ minimizes $\mathbf{E}_Q[\hat{\theta}(\boldsymbol{Y}) - \Theta]^2$.
- $Q(\theta|\boldsymbol{y}) \propto P(\theta|\boldsymbol{y}) \exp\{\alpha[g(\boldsymbol{y}) - \theta]^2\}$.

then g minimizes $\mathbf{E} \exp\{\alpha[\hat{\theta}(\boldsymbol{Y}) - \Theta]^2\}$.

Condition for tightness of the Bayesian Cramér–Rao lower bound (BCRLB), $L_{\mathsf{B}}(Q)$:

$$Q(\theta|\boldsymbol{y}) = \mathcal{N}(g(\boldsymbol{y}), L_{\mathsf{BCR}}(Q)).$$

Combining with the above, we then should have

$$P(\theta|\boldsymbol{y}) = \mathcal{N}\left(g(\boldsymbol{y}), \frac{L_{\mathsf{BCR}}(Q)}{1 - 2\alpha L_{\mathsf{BCR}}(Q)}\right).$$

Example: Gaussian Linear Models

Under P, let $\Theta \sim \mathcal{N}(0, \sigma^2)$ and for a given $\Theta = \theta$, let

$$y(t) = \theta s(t) + n(t), \qquad 0 \leq t \leq T$$

where $n(t)$ is AWGN, with psd $N_0/2$, and $s(t)$ is with energy E_s.

Denoting $z = \int_0^T y(t)s(t)\mathrm{d}t$, we get

$$P(\theta|\boldsymbol{y}) = P(\theta|z) \propto \exp\left\{-\frac{N_0 + 2\sigma^2 E_s}{2\sigma^2 N_0}\left(\theta - \frac{\sigma^2}{\sigma^2 E_s + N_0/2} \cdot z\right)^2\right\},$$

and the optimality conditions are satisfied by the estimator,

$$\hat{\theta} = \frac{\sigma^2}{\sigma^2 E_s + N_0/2} \cdot z.$$

Example: Gaussian Linear Models (Cont'd)

The estimator,

$$\hat{\theta} = \frac{\sigma^2}{\sigma^2 E_s + N_0/2} \cdot z$$

achieves

$$\inf_g \ln \mathbf{E} \exp\{\alpha[g(\boldsymbol{Y}) - \Theta]^2\} = \frac{1}{2} \ln \frac{1}{1 - \alpha/\alpha_{\mathsf{c}}}, \qquad 0 < \alpha < \alpha_{\mathsf{c}} \triangleq \frac{1}{2\sigma^2} + \frac{E_s}{N_0}.$$

No estimator g can have an estimation error $\epsilon = g(\boldsymbol{Y}) - \Theta$ with a pdf with tail that decays faster than

$$\exp\left\{-\left(\frac{1}{2\sigma^2} + \frac{E_s}{N_0}\right)\epsilon^2\right\}.$$

Non–Linear Model & Reference Gaussian–Linear Model

Under P, $\Theta \sim \mathcal{N}(0, \sigma^2)$ and given $\theta = \theta$, $y(t) = x(t, \theta) + n(t)$.

Under Q, $\Theta \sim \mathcal{N}(0, \tilde{\sigma}^2)$, and $y(t) = \theta s(t) + n(t)$.

$$\ln \mathbf{E} \exp\{\alpha[g(\boldsymbol{Y}) - \Theta]^2\} \geq \frac{\alpha \tilde{\sigma}^2 N_0}{N_0 + 2\tilde{\sigma}^2 E_s} - D(\tilde{\sigma}^2 \| \sigma^2) - \frac{1}{N_0} \int_0^T \mathbf{E}[x(t, \Theta) - \Theta s(t)]^2 \mathrm{d}t.$$

Degrees of freedom: $\tilde{\sigma}^2$ and the reference signal, $s(t)$. Optimal signal:

$$s^*(t) \propto \mathbf{E}\left\{\Theta \cdot x(t, \Theta)\right\}.$$

For $x(t, \theta) \propto \cos(\omega t + \theta)$,

$$\ln \mathbf{E} \exp\{\alpha[g(\boldsymbol{Y}) - \Theta]^2\} \geq \frac{1}{2} \ln \frac{1}{1 - \alpha/\alpha_{\mathsf{c}}} - \frac{E_x}{N_0}, \quad \alpha_{\mathsf{c}} = \frac{1}{2\sigma^2}.$$

Non–Linear Model and Reference

Under P, Θ has a general prior, P_Θ and $y(t) = x(t-\theta) + n(t)$.

Under Q, Θ has a general prior, Q_Θ and $y(t) = s(t-\theta) + n(t)$.

$$\text{Bound} = \frac{\alpha}{I(Q_\Theta) + \frac{2}{N_0}\int_0^T [\dot{s}(t)]^2 \mathrm{d}t} - D(Q_\Theta \| P_\Theta) - \frac{1}{N_0}\int_0^T [x(t) - s(t)]^2 \mathrm{d}t.$$

Degrees of freedom: Q_Θ and $s(t)$. The optimal $s(t)$ is the solution to

$$s(t) - \frac{\ddot{s}(t)}{\lambda} = x(t) \qquad \dot{s}(0) = \dot{s}(T) = 0$$

where λ controls the trade off.

For $x(t,\theta) = a(t)\sin(\theta t) + b(t)\cos(\theta t)$, the optimal $s(t)$ is

$$s(t,\theta) = \frac{\lambda x(t,\theta)}{t^2 + \lambda}.$$

The Non–Bayesian Regime

For the linear mmodel, $y(t) = \theta \cdot s(t) + n(t)$, we obtain,

$$\begin{aligned}\text{Bound} &= \sup_{\theta'} \left[\frac{\alpha N_0}{2E_s} + \left(\alpha - \frac{E_s}{N_0} \right) (\theta' - \theta)^2 \right] \\ &= \begin{cases} \frac{\alpha N_0}{2E_s} & \alpha \leq \frac{E_s}{N_0} \\ \infty & \alpha > \frac{E_s}{N_0} \end{cases}\end{aligned}$$

which means that $\alpha_c \leq E_s/N_0$.

The ML estimator achieves E_s/N_0, and so, $\alpha_c = E_s/N_0$

$$\ln \mathbf{E}_\theta \exp\{\alpha[\hat{\theta}_{\mathsf{ML}} - \theta]^2\} = -\frac{1}{2} \ln \left(1 - \frac{\alpha N_0}{E_s} \right).$$

The bound is achieves for small α and/or large SNR.

Extension to the Vector Case (Non–Bayesian)

For

$$x(t,\theta) = \sum_{i=1}^{k} \theta_i s_i(t),$$

define Γ be the $k \times k$ matrix of correlations with entries given by

$$\gamma_{ij} = \frac{1}{E_s} \int_0^T s_i(t) s_j(t) \mathrm{d}t.$$

$$\begin{aligned} \ln \mathbf{E}_\theta \exp\left\{ \left[\alpha^T(\hat{\theta} - \theta)\right]^2 \right\} \quad &\geq \quad \frac{N_0 \alpha^T \Gamma^{-1} \alpha}{2E_s} + \sup_{\theta'} (\theta' - \theta)^T \left(\alpha\alpha^T - \frac{E_s}{N_0}\Gamma \right) (\theta' - \theta) \\ &= \quad \begin{cases} \frac{N_0 \alpha^T \Gamma^{-1} \alpha}{2E_s} & \alpha^T \Gamma^{-1} \alpha \leq \frac{E_s}{N_0} \\ \infty & \text{elsewhere} \end{cases} \end{aligned}$$

More in the Paper

- Explicit bounds for various signal models.
- MMSE lower obunds other than the BCRLB, e.g., the W-W bound.
- Extension the Laplace principle to Rényi divergences.
- Phase transitions under the CGF criterion,