Estimation Accuracy of Non-Standard Maximum Likelihood Estimators

Nabil Kbayer, Jerome Galy, Eric Chaumette, François Vincent, Alexandre Renaux and Pascal Larzabal

Université de Toulouse/Tésa, Université de Montpellier 2/LIRMM, Université de Toulouse/ISAE-Supaéro, Université de Toulouse/ISAE-Supaéro, Université Paris-Sud/LSS, Université Paris-Sud/SATIE

Abstract

In many deterministic estimation problems, the probability density function (p.d.f.) parameterized by unknown deterministic parameters results from the marginalization of a joint p.d.f. depending on additional random variables. Unfortunately, this marginalization is often mathematically intractable, which prevents from using standard maximum likelihood estimators (MLEs) or any standard lower bound on their mean squared error (MSE). To circumvent this problem, the use of joint MLEs of deterministic and random parameters are proposed as being a substitute. It is shown that, regarding the deterministic parameters: 1) the joint MLEs provide generally suboptimal estimates in any asymptotic regions of operation yielding unbiased efficient estimates, 2) any representative of the two general classes of lower bounds, respectively the Small-Error bounds and the Large-Error bounds, has a "non-standard" version lower bounding the MSE of the deterministic parameters estimate.

Standard and Non-Standard deterministic estimation problems

- A model of the general deterministic estimation problem has the following four components: 1) a parameter space Θ_d, 2) an observation space $\mathcal{X}$, 3) a probabilistic mapping from parameter vector space Θ_d to observation space $\mathcal{X}$, that is the probability law that governs the effect of a parameter vector value $\boldsymbol{\theta}$ on the observation $\boldsymbol{x}$ and, 4) an estimation rule, that is the mapping of the observation space $\mathcal{X}$ into vector parameter estimates $\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$.
- Actually, in many estimation problems, the probabilistic mapping results from a two steps probabilistic mechanism involving an additional random vector $\boldsymbol{\theta}_r$, $\boldsymbol{\theta}_r \in \Theta_r \subset \mathbb{R}^{P_r}$, that is i) $\boldsymbol{\theta} \to \boldsymbol{\theta}_r \sim p(\boldsymbol{\theta}_r;\boldsymbol{\theta})$, ii) $(\boldsymbol{\theta},\boldsymbol{\theta}_r) \to \boldsymbol{x} \sim p(\boldsymbol{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta})$, and leading to a compound probability distribution:

$$p(\boldsymbol{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta})\, p(\boldsymbol{\theta}_r;\boldsymbol{\theta}), \tag{1a}$$

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \int_{\Theta_r} p(\boldsymbol{x},\boldsymbol{\theta}_r;\boldsymbol{\theta})\, d\boldsymbol{\theta}_r, \tag{1b}$$

  where $p(\boldsymbol{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta})$ is the conditional probability density function (p.d.f.) of $\boldsymbol{x}$ given $\boldsymbol{\theta}_r$, and $p(\boldsymbol{\theta}_r;\boldsymbol{\theta})$ is the prior p.d.f., parameterized by $\boldsymbol{\theta}$.
- Therefore, deterministic estimation problems can be divided into two subsets:
  1) the subset of "standard" deterministic estimation problems for which a closed-form expression of $p(\boldsymbol{x};\boldsymbol{\theta})$ is available,
  2) the subset of "non-standard" deterministic estimation problems for which only an integral form of $p(\boldsymbol{x};\boldsymbol{\theta})$ (1b) is available.

Non-standard maximum likelihood estimator for deterministic estimation (single parameter case)

- The widespread use of MLEs originates from the fact that, under reasonably general conditions on the observation model, the MLEs are asymptotically uniformly unbiased, Gaussian distributed and efficient when the number of independent observations tends to infinity.
- If a closed-form of $p(\boldsymbol{x};\theta)$ does not exist and the standard MLE of θ:

$$\widehat{\theta}_{ML}(\boldsymbol{x}) = \arg\max_{\theta\in\Theta_d}\{p(\boldsymbol{x};\theta)\}, \tag{2}$$

  cannot be derived, a sensible solution in the search of a realizable estimator based on the ML principle is to look for:

$$\left(\underline{\widehat{\boldsymbol{\theta}}_r}(\boldsymbol{x}), \underline{\widehat{\theta}}(\boldsymbol{x})\right) = \arg\max_{\theta\in\Theta_d,\boldsymbol{\theta}_r\in\Theta_r}\{p(\boldsymbol{x}|\boldsymbol{\theta}_r,\theta)\}. \tag{3}$$

  referred to as **"non-standard" MLEs (NSMLEs)**. The underlying idea is that in many estimation problems the closed-form of $p(\boldsymbol{x}|\boldsymbol{\theta}_r;\theta)$ is known and the NSMLEs (3) take advantage not only of the MLEs properties, but also of the extensive open literature on MLE closed-form expressions or approximations.
- The NSMLE is more attractive than:
  1) the joint maximum a posteriori-maximum likelihood estimate (JMAPMLE) of the hybrid parameter vector $\left(\boldsymbol{\theta}_r^T,\theta\right)$:

$$\left(\widehat{\boldsymbol{\theta}}_{rJ}(\boldsymbol{x}), \widehat{\theta}_J(\boldsymbol{x})\right) = \arg\max_{\theta\in\Theta_d,\boldsymbol{\theta}_r\in\Theta_r}\{p(\boldsymbol{x},\boldsymbol{\theta}_r;\theta)\}, \tag{4}$$

  which is most times biased and inconsistent whatever the number of independent observations.
  2) the expectation-maximization (EM) algorithm which, in non-standard estimation, consists in the following iterative procedure:

$$\theta_{n+1} = \arg\max_{\theta\in\Theta_d}\left\{E_{\boldsymbol{\theta}_r|\boldsymbol{x};\theta_n}\left[\ln p(\boldsymbol{x},\boldsymbol{\theta}_r;\theta)\right]\right\}, \tag{5}$$

  which is unlikely to be of practical use in many estimation problems of interest where $p(\boldsymbol{\theta}_r;\theta)$ is not a conjugate prior for the likelihood function $p(\boldsymbol{x}|\boldsymbol{\theta}_r;\theta)$ and $p(\boldsymbol{\theta}_r|\boldsymbol{x};\theta)$ is not computable.

Non-standard lower bounds

- Let us denote $\boldsymbol{\phi} = \begin{pmatrix}\theta\\ \boldsymbol{\theta}_r\end{pmatrix}$, $p(\boldsymbol{x}|\boldsymbol{\phi}) \triangleq p(\boldsymbol{x}|\boldsymbol{\theta}_r;\theta)$, $E_{\boldsymbol{x}|\boldsymbol{\phi}}[\,] \triangleq E_{\boldsymbol{x}|\boldsymbol{\theta}_r;\theta}[\,]$. Then any unbiased estimator $\widehat{\boldsymbol{\phi}}(\boldsymbol{x}) = \begin{pmatrix}\widehat{\theta}(\boldsymbol{x})\\ \widehat{\boldsymbol{\theta}}_r(\boldsymbol{x})\end{pmatrix}$ of $\boldsymbol{\phi}$ verify:

$$E_{\boldsymbol{\theta}_r;\theta}\left[\Xi(\boldsymbol{\Phi}^N)\, \boldsymbol{R}_{\boldsymbol{\upsilon}_\phi}^{-1}(\boldsymbol{\Phi}^N)\, \Xi(\boldsymbol{\Phi}^N)^T\right] \leq E_{\boldsymbol{x},\boldsymbol{\theta}_r;\theta}\left[\left(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\right)\left(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\right)^T\right], \tag{6}$$

  where $\boldsymbol{\Phi}^N = \left[\boldsymbol{\phi}^1 \ldots \boldsymbol{\phi}^N\right]$, $\Xi(\boldsymbol{\Phi}^N) = \left[\boldsymbol{\phi}^1 - \boldsymbol{\phi} \ldots \boldsymbol{\phi}^N - \boldsymbol{\phi}\right]$,
  $\boldsymbol{R}_{\boldsymbol{\upsilon}_\phi}(\boldsymbol{\Phi}^N) = E_{\boldsymbol{x}|\boldsymbol{\phi}}\left[\boldsymbol{\upsilon}_{\boldsymbol{\phi}}(\boldsymbol{\Phi}^N)\, \boldsymbol{\upsilon}_{\boldsymbol{\phi}}^T(\boldsymbol{\Phi}^N)\right]$,
  $\boldsymbol{\upsilon}_{\boldsymbol{\phi}}(\boldsymbol{\Phi}^N) \triangleq \boldsymbol{\upsilon}_{\boldsymbol{\phi}}(\boldsymbol{x};\boldsymbol{\Phi}^N) = \left(\upsilon_{\boldsymbol{\phi}}(\boldsymbol{x};\boldsymbol{\phi}^1), \ldots, \upsilon_{\boldsymbol{\phi}}(\boldsymbol{x};\boldsymbol{\phi}^N)\right)^T$ and $\upsilon_{\boldsymbol{\phi}}(\boldsymbol{x};\boldsymbol{\phi}') = \frac{p(\boldsymbol{x}|\boldsymbol{\phi}')}{p(\boldsymbol{x}|\boldsymbol{\phi})}$.
- In any asymptotic region of operation of NSMLEs, since NSMLEs is an unbiased estimate of $\boldsymbol{\phi}$, (6) is a LB on the covariance matrix of NSMLEs.
- In the same vein, any Barankin bound approximation (BBA) on the MSE of unbiased estimator $\widehat{\boldsymbol{\phi}}(\boldsymbol{x})$ results from a linear transformation of the NSMSB (6) and defines a non-standard BBA (NSBBA). As well as the NSMSB (6), any NSBBA lower bounds the MSE of NSMLEs in any asymptotic region of operation.
- Note that in general, the NSBBAs cannot be arranged in closed form due to the presence of the statistical expectation. They however can be evaluated by numerical integration or Monte Carlo simulation.

Non-standard lower bound Examples

A typical example is the CRB obtained for $N=2$, where $\boldsymbol{\phi}^1 = \left(\theta,\boldsymbol{\theta}_r^T\right)^T$ and $\boldsymbol{\phi}^2 = \left(\theta + d\theta, \boldsymbol{\theta}_r^T\right)^T$. Then by letting $d\theta$ be infinitesimally small, (6) becomes:

$$(6) \underset{d\theta\to 0}{\to} NSCRB \triangleq E_{\boldsymbol{\theta}_r|\theta}\left[E_{\boldsymbol{x}|\boldsymbol{\phi}}\left[\left(\frac{\partial \ln p(\boldsymbol{x}|\boldsymbol{\phi})}{\partial\theta}\right)^2\right]^{-1}\right], \tag{7}$$

that is the Miller and Chang bound.
Following the rationale introduced by Fraser and Guttman, a straightforward extension of (7) is obtained for $\boldsymbol{\Phi}^N = \left[\boldsymbol{\phi}^1 \ \ldots \ \boldsymbol{\phi}^N\right]$, $\boldsymbol{\phi}^n = \left(\theta + (n-1)\,d\theta, \boldsymbol{\theta}_r^T\right)^T$, $1 \leq n \leq N$. Then by letting $d\theta$ be infinitesimally small, (6) becomes the NS Battacharayya bounds of order $N-1$:

$$(6) \underset{d\theta\to 0}{\to} NSBaB \triangleq E_{\boldsymbol{\theta}_r|\theta}\left[\boldsymbol{e}_1^T E_{\boldsymbol{x}|\boldsymbol{\phi}}\left[\boldsymbol{b}(\boldsymbol{x};\boldsymbol{\phi})\, \boldsymbol{b}^T(\boldsymbol{x};\boldsymbol{\phi})\right]^{-1}\boldsymbol{e}_1\right], \tag{8}$$

where $\boldsymbol{b}(\boldsymbol{x};\boldsymbol{\phi}) = \frac{1}{p(\boldsymbol{x}|\boldsymbol{\phi})}\left(\frac{\partial p(\boldsymbol{x}|\boldsymbol{\phi})}{\partial\theta}, \ldots, \frac{\partial^{N-1}p(\boldsymbol{x}|\boldsymbol{\phi})}{\partial^{N-1}\theta}\right)^T$, $\boldsymbol{e}_1 = (1,0,\ldots,0)^T$.

A new look at Gaussian observation models

A simple and well known instantiation of the Gaussian observation model is:

$$\boldsymbol{x}_t = \boldsymbol{s}(\tau)\, a_t + \boldsymbol{n}_t, \quad 1 \leq t \leq T, \tag{9}$$

where $a_1,\ldots,a_t$ are the complex amplitudes of the signal, $\boldsymbol{s}(\,)$ is a vector of M parametric functions depending on a single deterministic parameter τ, $\boldsymbol{n}_t \sim \mathcal{CN}\left(\boldsymbol{0},\sigma_{\boldsymbol{n}}^2\boldsymbol{I}_M\right)$, $1 \leq t \leq T$, are i.i.d. Gaussian complex circular noises independent of the signal of interest. Additionally if $\boldsymbol{a} \sim \mathcal{CN}\left(\boldsymbol{0},\sigma_a^2\boldsymbol{I}_T\right)$, then (9) is an unconditional observation model parameterized by $\boldsymbol{\theta} = \left(\tau,\sigma_a^2,\sigma_{\boldsymbol{n}}^2\right)^T$, and the MLE of τ is the UMLE $\widehat{\tau}$. The associated CRB is the UCRB:

$$UCRB_\tau = \sigma_{\boldsymbol{n}}^2\left(2h(\tau)\,T\sigma_a^2\right)^{-1}\left(1 + SNR^{-1}\right), \tag{10}$$

$$SNR = \frac{\sigma_a^2\left\|\boldsymbol{s}(\tau)\right\|^2}{\sigma_{\boldsymbol{n}}^2},\ h(\tau) = \frac{\partial\boldsymbol{s}(\tau)}{\partial\tau}^H \Pi_{\boldsymbol{s}(\tau)}^{\perp}\frac{\partial\boldsymbol{s}(\tau)}{\partial\tau},$$

where $\Pi_{\boldsymbol{a}}^{\perp} = \boldsymbol{I}_M - \boldsymbol{a}\boldsymbol{a}^H\left\|\boldsymbol{a}\right\|^{-2}$. The NSMLE (3) of τ is actually the CMLE $\underline{\widehat{\tau}}$ and the associated NSCRB is:

$$NSCRB_\tau = E_{\boldsymbol{a}|\sigma_a^2}\left[CCRB_\tau(\boldsymbol{a})\right], CCRB_\tau(\boldsymbol{a}) = \sigma_{\boldsymbol{n}}^2\left(2h(\tau)\left\|\boldsymbol{a}\right\|^2\right)^{-1},$$

where $CCRB_\tau$ is associated to the CMLE. First, it has been shown, in the case of a vector of unknown parameters $\boldsymbol{\tau}$, that asymptotically where $T\to\infty$:

$$\boldsymbol{C}_{\boldsymbol{\theta}}\left(\underline{\widehat{\boldsymbol{\tau}}}\right) \geq \boldsymbol{C}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\tau}}\right) = \boldsymbol{UCRB}_{\boldsymbol{\tau}} \geq \boldsymbol{CCRB}_{\boldsymbol{\tau}}, \tag{11}$$

which illustrates that the act of resorting to the NSMLE (here the CMLE) is in general an asymptotic suboptimal choice in the MSE sense. However, in the case of single unknown parameter τ, (11) becomes:

$$C_{\boldsymbol{\theta}}\left(\underline{\widehat{\tau}}\right) = C_{\boldsymbol{\theta}}\left(\widehat{\tau}\right) = UCRB_\tau,$$

which highlights that in some particular cases the NSMLE may be asymptotically equivalent to the MLE in the MSE sense. Second, if $T \geq 2$:

$$\frac{NSCRB_\tau}{UCRB_\tau} = \frac{NSCRB_\tau}{C_{\boldsymbol{\theta}}\left(\widehat{\tau}\right)} = \frac{T}{T-1}\frac{SNR}{SNR+1} \tag{12}$$

which illustrates the facts that NSLB are not in general neither upper bounds on the MSE of MLEs nor on any of its LBs