

# Bounds on Estimation

Arye Nehorai<sup>\*a</sup> and Gongguo Tang<sup>b</sup>

<sup>a</sup>Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA

<sup>b</sup>Department of Electrical Engineering & Computer Science, Colorado School of Mines, Golden, CO, USA

## Abstract

We review several universal lower bounds on statistical estimation, including deterministic bounds on unbiased estimators such as Cramér-Rao bound and Barankin-type bound, as well as Bayesian bounds such as Ziv-Zakai bound. We present explicit forms of these bounds, illustrate their usage for parameter estimation in Gaussian additive noise, and compare their tightness.

## Keywords

Statistical estimation, Mean-squared error, Cramér-Rao bound, Barankin-type bound, Ziv-Zakai bound

## Introduction

Statistical estimation involves inferring the values of parameters specifying a statistical model from data. The performance of a particular statistical algorithm is measured by the error between the true parameter values and those estimated by the algorithm. However, explicit forms of estimation error are usually difficult to obtain except for the simplest statistical models. Therefore, performance bounds are derived as a way of quantifying estimation accuracy while maintaining tractability.

In many cases, it is beneficial to quantify performance using universal bounds that are independent of the estimation algorithms and rely only upon the model. In this regard, universal lower bounds are particularly useful as it provides means to assess the difficulty of performing estimation for a particular model and can act as benchmarks to evaluate the quality of any algorithm: the closer the estimation error of the algorithm to the lower bound, the better the algorithm. In the following, we review three widely used universal lower bounds on estimation: Cramér-Rao bound (CRB), Barankin-type bound (BTB), and Ziv-Zakai bound (ZZB). These bounds find numerous applications in determining the performance of sensor arrays, radar, and nonlinear filtering; in benchmarking various algorithms; and in optimal design of systems.

---

<sup>\*</sup>E-mail: nehorai@ese.wustl.edu

This work was supported in part by NSF Grants CCF-1014908 and CCF-0963742, ONR Grant N000141310050, AFOSR Grant FA9550-11-1-0210.

## Statistical Model and Related Concepts

To formalize matters, we define a statistical model for estimation as a family of parameterized probability density functions in  $\mathbb{R}^N$ :  $\{p(x; \theta) : \theta \in \Theta \subset \mathbb{R}^d\}$ . We observe a realization of  $x \in \mathbb{R}^N$  generated from a distribution  $p(x; \theta)$ , where  $\theta \in \Theta$  is the true parameter to be estimated from data  $x$ . Though we assume a single observation  $x$ , the model is general enough to encompass multiple independent, identically distributed samples (i.i.d.) by considering the joint probability distribution.

An estimator of  $\theta$  is a measurable function of the observation  $\hat{\theta}(x) : \mathbb{R}^N \rightarrow \Theta$ . An unbiased estimator is one such that

$$\mathbb{E}_{\theta} \{\hat{\theta}(x)\} = \theta, \forall \theta \in \Theta. \quad (1)$$

Here we used the subscript  $\theta$  to emphasize that the expectation is taken with respect to  $p(x; \theta)$ . We focus on the performance of unbiased estimators in this entry. There are various ways to measure the error of the estimator  $\hat{\theta}(x)$ . Two typical ones are the error covariance matrix:

$$\mathbb{E}_{\theta} \{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} = \text{Cov}(\hat{\theta}), \quad (2)$$

where the equation holds only for unbiased estimators, and the mean-squared error (MSE):

$$\mathbb{E}_{\theta} \{\|\hat{\theta}(x) - \theta\|_2^2\} = \text{trace} \left( \mathbb{E}_{\theta} \{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} \right). \quad (3)$$

*Example 1 (Signal in additive Gaussian noise (SAGN)).* To illustrate the usage of different estimation bounds, we use the following statistical model as a running example:

$$x_n = s_n(\theta) + w_n, n = 0, \dots, N - 1. \quad (4)$$

Here  $\theta \in \Theta \subset \mathbb{R}$  is a scalar parameter to be estimated and the noise  $w_n$  follows i.i.d. Gaussian distribution with mean 0 and *known* variance  $\sigma^2$ . Therefore, the density function for  $x$  is

$$\begin{aligned} p(x; \theta) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_n - s_n(\theta))^2}{2\sigma^2} \right\} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp \left\{ -\sum_{n=0}^{N-1} \frac{(x_n - s_n(\theta))^2}{2\sigma^2} \right\}. \end{aligned}$$

In particular, we consider the frequency estimation problem where  $s_n(\theta) = \cos(2\pi n\theta)$  with  $\Theta = [0, \frac{1}{4})$ .

## Cramér-Rao Bound

The Cramér-Rao bound (CRB) (Kay 2001a; Van Trees 2001; Stoica and Nehorai 1989) is arguably the most well-known lower bounds on estimation. Define the Fisher information matrix  $I(\theta)$  via

$$\begin{aligned} I_{i,j}(\theta) &= \mathbb{E}_{\theta} \left\{ \frac{\partial}{\partial \theta_i} \log p(x; \theta) \frac{\partial}{\partial \theta_j} \log p(x; \theta) \right\} \\ &= -\mathbb{E}_{\theta} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(x; \theta) \right\}. \end{aligned}$$

Then for *any* unbiased estimator  $\hat{\theta}$ , the error covariance matrix is bounded by

$$\mathbb{E}_{\theta} \left\{ (\hat{\theta} - \theta)(\hat{\theta} - \theta)^T \right\} \geq [I(\theta)]^{-1}, \quad (5)$$

where  $A \succeq B$  for two symmetric matrices means  $A - B$  is positive semidefinite. The inverse of the Fisher information matrix  $\text{CRB}(\theta) = [I(\theta)]^{-1}$  is called the Cramér-Rao bound.

When  $\theta$  is scalar,  $I(\theta)$  measures the expected sensitivity of the density function with respect to changes in the parameter. A density family that is more sensitive to parameter changes (larger  $I(\theta)$ ) will generate observations that look more different when the true parameter varies, making it easier to estimate (smaller error).

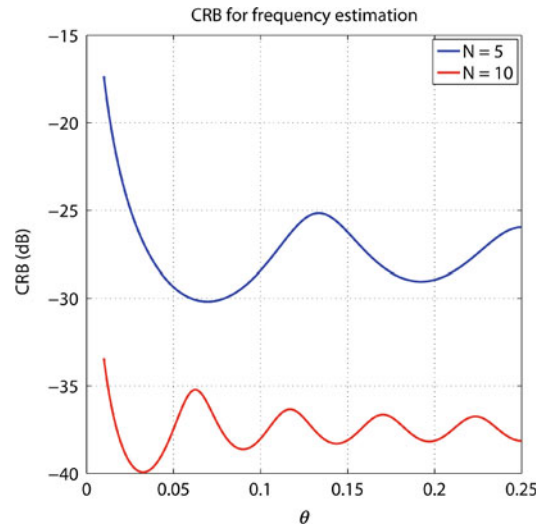
*Example 2.* For the SAGN model (4), the CRB is

$$\text{CRB}(\theta) = I(\theta)^{-1} = \frac{\sigma^2}{\sum_{n=0}^{N-1} \left[ \frac{\partial s_n(\theta)}{\partial \theta} \right]^2}. \quad (6)$$

The inverse dependence on the  $\ell_2$  norm of signal derivative suggests that signals more sensitive to parameter change are easier to estimate.

For the frequency estimation problem with  $s_n(\theta) = \cos(2\pi n\theta)$ , the CRB as a function of  $\theta$  is plotted in Fig. 1.

There are many modifications of the basic CRB such as the posterior CRB (Van Trees 2001; Tichavsky et al. 1998), the hybrid CRB (Rockah and Schultheiss 1987), the modified CRB (D'Andrea et al. 1994), the concentrated CRB (Hochwald and Nehorai 1994), and constrained CRB (Stoica and Ng 1998; Gorman and Hero 1990; Marzetta 1993). The posterior CRB takes into account the prior information of the parameters when they are modeled as random variables, while the hybrid CRB considers the case that the parameters contain both random and deterministic parts. The modified CRB and the concentrated CRB focus on handling nuisance parameters in a tractable manner. The application of these CRBs requires a regular parameter space (e.g., an open set in  $\mathbb{R}^d$ ). However, in many cases, the parameter space  $\Theta$  is a low-dimensional manifold in  $\mathbb{R}^d$  specified by equalities and inequalities. In this case, the constrained CRB provides tighter lower bounds by incorporating knowledge of the constraints.



**Fig. 1** Cramér-Rao bound on frequency estimation:  $N = 5$  vs.  $N = 10$

## Barankin Bound

CRB is a local bound in the sense that it involves only local properties (the first or second order derivatives) of the log-likelihood function. So if two families of log-likelihood functions coincide at a region near  $\theta^0$ , the CRB at  $\theta^0$  would be the same, even if they are drastically different in other regions of the parameter space.

However, the entire parametric space should play a role in determining the difficulty of parameter estimation. To see this, imagine that there are two statistical models. In the first model there is another point  $\theta^1 \in \Theta$  such that the likelihood family  $p(x; \theta)$  behaves similarly around  $\theta^0$  and  $\theta^1$ , but these two points are not in neighborhoods of each other. Then it would be difficult to distinguish these two points for any estimation algorithm, and the estimation performance for the first statistical model would be bad (an extreme case is  $p(x; \theta^0) \equiv p(x; \theta^1)$  in which case the model is non-identifiable; more discussions on identifiability and Fisher information matrix can be found in Hochwald and Nehorai (1997)). In the second model, we remove the point  $\theta^1$  and its near neighborhood from  $\Theta$ , then the performance should get better. However, CRB for both models would remain the same whether we exclude  $\theta^1$  from  $\Theta$  or not. As a matter of fact,  $\text{CRB}(\theta^0)$  uses only the fact that the estimator is unbiased in a neighborhood of the true parameter  $\theta^0$ .

Barankin bound addresses CRB's shortcoming of not respecting the global structure of the statistical model by introducing finitely many test points  $\{\theta^i, i = 1, \dots, M\}$  and ensures that the estimator is unbiased at the neighborhood of  $\theta^0$  as well as these test points (Forster and Larzabal 2002). The original Barankin bound (Barankin 1949) is derived for scalar parameter  $\theta \in \Theta \subset \mathbb{R}$  and any unbiased estimator  $\widehat{g(\theta)}$  for a function  $g(\theta)$ :

$$\mathbb{E}_{\theta}(\widehat{g(\theta)} - g(\theta))^2 \geq \sup_{M, \theta^i, a^i} \frac{\left[ \sum_{m=1}^M a^i (g(\theta^i) - g(\theta)) \right]^2}{\mathbb{E}_{\theta} \left[ \sum_{m=1}^M a^i \frac{p(x; \theta^i)}{p(x; \theta)} \right]^2} \quad (7)$$

Using (7), we can derive a Barankin-type bound on the error covariance matrix of any unbiased estimator  $\hat{\theta}(x)$  for a vector parameter  $\theta \in \Theta \subset \mathbb{R}^d$  (Forster and Larzabal 2002):

$$\mathbb{E}_{\theta} \left\{ (\hat{\theta} - \theta)(\hat{\theta} - \theta)^T \right\} \geq \Phi(B - 11^T)^{-1} \Phi^T, \quad (8)$$

where the matrices are defined via

$$B_{i,j} = \mathbb{E}_{\theta} \left\{ \frac{p(x; \theta^i)}{p(x; \theta)} \frac{p(x; \theta^j)}{p(x; \theta)} \right\}, 1 \leq i, j \leq M,$$

$$\Phi = [\theta^1 - \theta \dots \theta^M - \theta]$$

and  $1$  is the vector in  $\mathbb{R}^M$  with all ones. Note that we have used  $\theta^i$  with a superscript to denote different points in  $\Theta$ , while  $\theta_i$  with a subscript to denote the  $i$ th component of a point  $\theta$ .

Since the bound (8) is valid for any  $M$  and any choice of test points  $\{\theta^i\}$ , we obtain the tightest bound by taking the supremum over all finite families of test points. Note that when we have  $d$  test points that approach  $\theta$  in  $d$  linearly independent directions, the Barankin-type bound (8) converges to the CRB. If we have more than  $d$  test points, however, the Barankin-type bound is always not worse than the CRB. Particularly, the Barankin-type bound is much tighter in the regime of low signal-to-noise ratio (SNR) and small number of measurements, which allows one to investigate the “threshold” phenomena as shown in the next example.

*Example 3.* For the SAGN model, if we have  $M$  test points, the elements of matrix  $B$  are of the following form:

$$B_{i,j} = \exp \left\{ \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [s_n(\theta^i) - s_n(\theta)][s_n(\theta^j) - s_n(\theta)] \right\}$$

In most cases, it is extremely difficult to derive an analytical form of the Barankin bound by optimizing with respect to  $M$  and the test points  $\{\theta^j\}$ . In Fig. 2, we plot the Barankin-type bounds for  $s_n(\theta) = \cos(2\pi n\theta)$  for  $M = 10$  randomly selected test points. We observe that Barankin-type bound is tighter than the CRB when SNR is small. There is a SNR region around 0 dB that the Barankin-type bound drops drastically. This is usually called the “threshold” phenomenon. Practical systems operate much better in the region above the threshold.

The basic CRB and BTB belong to the family of deterministic “covariance inequality” bounds in the sense that the unknown parameter is assumed to be a *deterministic* quantity (as opposed to a *random* quantity). Additionally, both bounds work only for unbiased estimators, making them inappropriate performance indicators for biased estimators such as many regularization-based estimators.

## Ziv-Zakai Bound

In this section, we introduce the Ziv-Zakai bound (ZZB) (Bell et al. 1997) that is applicable to any estimator (not necessarily unbiased). Unlike the CRB and BTB, the ZZB is a Bayesian bound and the errors are averaged by the prior distribution  $p_\theta(\phi)$  of the parameter. For any  $a \in \mathbb{R}^d$ , the ZZB states that

$$a^T \mathbb{E} \left\{ (\hat{\theta}(x) - \theta)(\hat{\theta}(x) - \theta)^T \right\} a \geq \frac{1}{2} \int_0^\infty \mathcal{V} \left\{ \max_{\delta: a^T \delta = h} \left[ \int_{\mathbb{R}^d} (p_\theta(\phi) + p_\theta(\phi + \delta)) \times P_{\min}(\phi, \phi + \delta) d\phi \right] \right\} h dh,$$

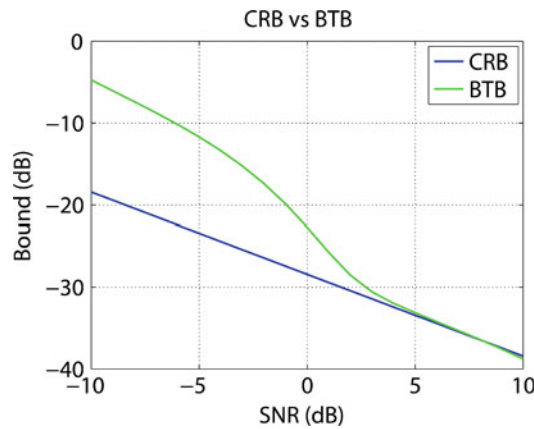
where the expectation is taken with respect to the joint disunity  $p(x; \theta)p_\theta(\phi)$ ,  $\mathcal{V}\{q(h)\} = \max_{r \geq 0} q(h + r)$  is the valley-filling function, and  $P_{\min}(\phi, \phi + \delta)$  is the minimal probability of error for the following binary hypothesis testing problem:

$$\begin{aligned} H_0 : \theta &= \phi; & x &\sim p(x; \phi) \\ H_1 : \theta &= \phi + \delta; & x &\sim p(x; \phi + \delta) \end{aligned}$$

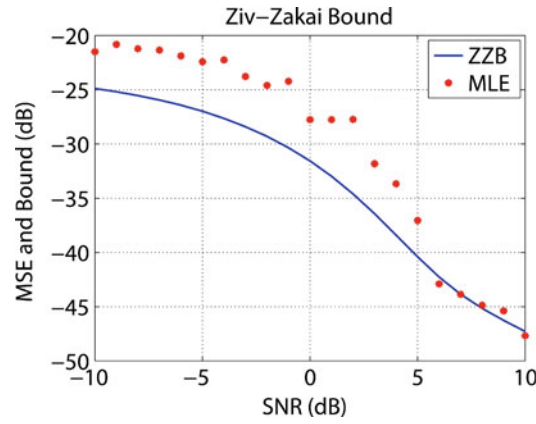
with

$$\begin{aligned} \Pr(H_0) &= \frac{p_\theta(\phi)}{p_\theta(\phi) + p_\theta(\phi + \delta)} \\ \Pr(H_1) &= \frac{p_\theta(\phi + \delta)}{p_\theta(\phi) + p_\theta(\phi + \delta)}. \end{aligned}$$

*Example 4.* For the ASGN model, we assume a uniform prior probability, i.e.,  $p_\theta(\phi) = 4, \phi \in [0, 1/4)$ . The ZZB simplifies to



**Fig. 2** Cramér-Rao bound vs. Barankin-type bound on frequency estimation when  $\theta^0 = 0.1$ . The BTB is obtained using  $M = 10$  uniform random points



**Fig. 3** Ziv-Zakai bound vs. maximum likelihood estimator for frequency estimation

$$\mathbb{E} \left\{ \|\hat{\theta}(x) - \theta\|_2^2 \right\} \geq \frac{1}{2} \int_0^{\frac{1}{4}} \mathcal{V} \left\{ \left[ \int_0^{\frac{1}{4}-h} 8P_{\min}(\phi, \phi + h) d\phi \right] \right\} h dh.$$

The binary hypothesis testing problem is to decide which one of two signals is buried in additive Gaussian noise. The optimal detector with minimal probability of error is the minimum distance receiver (Kay 2001b), and the associated probability of error is

$$P_{\min}(\phi, \phi + h) = Q \left( \frac{1}{2} \sqrt{\frac{\sum_{n=0}^{N-1} (s_n(\phi) - s_n(\phi + h))^2}{\sigma^2}} \right),$$

where  $Q(h) = \int_h^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ . For the frequency estimation problem, we numerically estimate the integral and plot the resulting ZZB in Fig. 3 together with the mean-squared error for the maximum likelihood estimator (MLE).

## Summary and Future Directions

We have reviewed several important performance bounds on statistical estimation problems, particularly, the Cramér-Rao bound, the Barankin-type bound, and the Ziv-Zakai bound. These bounds provide a universal way to quantify the performance of statistically modeled physical systems that is independent of any specific algorithm.

Future directions of performance bounds on estimation include deriving tighter bounds, developing computational schemes to approximate existing bounds in a tractable way, and applying them to practical problems.

## Cross-References

- [Estimation in Systems and Control Particle Filtering](#)

## Recommended Reading

Kay SM (2001a), Chapter 2 and 3; Stoica P, Nehorai A (1989); Van Trees HL (2001), Chapter 2.7; Forster and Larzabal (2002); Bell et al. (1997).

## Bibliography

- Barankin EW (1949) Locally best unbiased estimates. *Ann Math Stat* 20(4):477–501
- Bell KL, Steinberg Y, Ephraim Y, Van Trees HL (1997) Extended Ziv-Zakai lower bound for vector parameter estimation. *IEEE Trans Inf Theory* 43(2):624–637
- D’Andrea AN, Mengali U, Reggiannini R (1994) The modified Cramér-Rao bound and its application to synchronization problems. *IEEE Trans Commun* 42(234):1391–1399
- Forster P, Larzabal P (2002) On lower bounds for deterministic parameter estimation. In: *IEEE international conference on acoustics, speech, and signal processing (ICASSP)*, 2002, Orlando, vol 2. IEEE, pp II–1141
- Gorman JD, Hero AO (1990) Lower bounds for parametric estimation with constraints. *IEEE Trans Inf Theory* 36(6):1285–1301
- Hochwald B, Nehorai A (1994) Concentrated Cramér-Rao bound expressions. *IEEE Trans Inf Theory* 40(2):363–371
- Hochwald B, Nehorai A (1997) On identifiability and information-regularity in parametrized normal distributions. *Circuits Syst Signal Process* 16(1):83–89
- Kay SM (2001a) *Fundamentals of statistical signal processing, volume 1: estimation theory*. Prentice Hall, Upper Saddle River, NJ
- Kay SM (2001b) *Fundamentals of statistical signal processing, volume 2: detection theory*. Prentice Hall, Upper Saddle River, NJ
- Marzetta TL (1993) A simple derivation of the constrained multiple parameter Cramér-Rao bound. *IEEE Trans Signal Process* 41(6):2247–2249
- Rockah Y, Schultheiss PM (1987) Array shape calibration using sources in unknown locations – part I: far-field sources. *IEEE Trans Acoust Speech Signal Process* 35(3):286–299
- Stoica P, Nehorai A (1989) MUSIC, maximum likelihood, and Cramér-Rao bound. *IEEE Trans Acoust Speech Signal Process* 37(5):720–741
- Stoica P, Ng BC (1998) On the Cramér-Rao bound under parametric constraints. *IEEE Signal Process Lett* 5(7):177–179
- Tichavsky P, Muravchik CH, Nehorai A (1998) Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. *IEEE Trans Signal Process* 46(5):1386–1396
- Van Trees HL (2001) *Detection, estimation, and modulation theory: part 1, detection, estimation, and linear modulation theory*. Jhon Wiley & Sons, Hoboken, NJ