Lower Bounds on Estimator Error and the Threshold Effect

Judah Jacobson
jsjacobs@fas.harvard.edu

Submitted to the Department of Mathematics
in partial fulfillment of the requirements
for the degree with honors
of Bachelor of Arts

Harvard College
April 5, 2004

Abstract

Rather than compute the exact error of a specific estimator, it is often more convenient to apply results which lower-bound the error of *any* estimator for a given problem. We present three such bounds and illustrate their applications to signal parameter estimation. The Cramèr-Rao bound is simple to compute and approximates the actual error under conditions of high signal-to-noise ratio. However, it fails to provide useful information for low signal-to-noise ratio conditions, and thus tighter bounds are required. The Barankin and Ziv-Zakai bounds approximate the Cramèr-Rao bound for high signal-to-noise ratio conditions but are significantly tighter than it for low signal-to-noise ratios. This situation demonstrates theoretically the well-known threshold effect for sonar time-delay estimation.

Acknowledgments

I would like to thank my advisor, Nathan Intrator, for his guidance, suggestions and support, and for motivating the study of the threshold effect for me through his own research on the subject. Additionally, I am grateful to Cliff Taubes for his helpful remarks on style. Finally, I issue a hearty thank you to my parents, family, fellow thesis-writers, and other friends for their support throughout this process; in particular, I am grateful to Paul Berman and Joel Norwood for their comments and suggestions.

Contents

Introduction

The performance of sonar systems under conditions of high noise is an important subject of current research, due to the relatively high noise levels at ultrasound frequencies. Early research on these types of systems was motivated primarily by radar, and thus tended to produce results which assume the low-noise conditions typical of those situations. These equations were also applied to sonar, when it first became a widely applicable field of study. However, they do not accurately reflect high-noise performance, and more sophisticated results are required.

In general, the error of a system such as radar or sonar is difficult to calculate directly. Several approaches exist for approximating the error; our method will be that of lower bounds on estimator error. An estimator is a system which attempts to predict a parameter value from one or more observations. For radar and sonar, the observation is a radio or sound wave which bounces off of a target, and the parameter is a property of the target such as its distance, angle of orientation, or velocity. The Cramèr-Rao bound is a well-known and easily computed lower limit on estimator error which is close to the actual error for low-noise situations. However, theoretical and empirical results demonstrate that it diverges significantly from the actual error under conditions of high noise, suggesting the need for more sophisticated bounds.

We present two such bounds, namely the Barankin and Ziv-Zakai bounds. The derivations and forms of these bounds are quite distinct; however, they have frequently been used to analyze similar types of estimation problems, and are therefore worth considering together. We present the derivation of each bound first in its form for a general estimation problem, and then as applied to the problem of signal parameter estimation. In particular, we compute their forms for the estimation of amplitude (signal energy) and time-delay (the point in time when the signal was received).

The Cramèr-Rao, Barankin and Ziv-Zakai bounds are equal for amplitude estimation at all noise levels. In the case of time-delay estimation, however, the Barankin and Ziv-Zakai bounds diverge from the Cramèr-Rao bound at high noise levels. The poor approximation by the Cramèr-Rao bound at high noise levels is known as the threshold effect. We will see that the Barankin and Ziv-Zakai bounds are significantly better methods for understanding the noise tolerance of sonar systems.

Chapter 1 introduces estimation theory, the framework of this paper.

We use this theory to set up the specific problem of signal parameter estimation. We present the ML and MAP estimators, general methods of estimation which are commonly used for radar and sonar. Finally, we define the threshold effect formally in terms of lower bounds on estimator error.

Chapter 2 describes the Cramèr-Rao bound. We will see that it is easily computed, and theoretically linked to the ML and MAP estimators. This bound is in fact achievable in the case of amplitude estimation. Furthermore, in nonlinear problems such as time-delay estimation, it approximates the error under conditions of high noise. However, it diverges from the actual error under low-noise conditions.

Chapter 3 presents the Barankin bound, which gives the minimum achievable error for an unbiased estimator. It is not possible to compute the exact Barankin bound for time-delay estimation; we turn instead to less optimal versions which are still significant improvements over the Cramèr-Rao bound. We apply this method to the signal parameter estimation problem and give a specific example for time-delay estimation. We show that the Barankin bound provides a direct explanation of the threshold effect in terms of secondary peaks of the autocorrelation function.

Chapter 4 presents the Ziv-Zakai bound, which connects estimator error to a simpler, related problem in detection theory. We compute the bound in its form for signal parameter estimation. We show that the Ziv-Zakai bound is equal to the Cramèr-Rao for amplitude estimation. For time-delay estimation, the bound exhibits a threshold effect as high noise levels spread the estimator error along the a priori interval.

Chapter 1

Estimation Theory

This chapter describes the mathematical framework which we use to describe sonar and radar systems. First, we present a simple explanation of radar and sonar systems. Then, we introduce the basic definitions of estimation theory. We present the specific example of signal parameter estimation, which is the motivating example throughout our discussion of estimation theory. We then present the ML and MAP estimators and give their forms for signal parameter estimation. Finally, we discuss the threshold effect and its connection to lower bounds on estimator error.

1.1 A Brief Explanation of Radar and Sonar

The typical setup for radar and sonar systems is as follows. A short pulse (of radio or sound waves, respectively) is emitted from the transmitter, bounces off of a target and is received back at the source. Typically, the transmitted pulse is a cosine packet of the form $s_0(t) = A_0(t)\cos(\omega_0 t - \tau_0)$, where $A_0(t)$ is a Gaussian envelope. The received pulse will be of the same form, but its parameters A_0, ω_0 and τ_0 may have changed, so that the signal becomes $s(t) = A(t)\cos(\omega t - \tau)$.

The parameters of the received signal $s(t)$ encode multiple forms of information about the target, such as distance, velocity, or bearing. For example, $\tau - \tau_0$ is the time taken by the transmitted pulse to travel to the target and back. Furthermore, if the speed of the transmitted wave is denoted by c, then the range (distance to the target) is $c(\tau - \tau_0)/2$. The velocity of the target in the direction of the receiver may be similarly computed as the Doppler shift between the angular frequencies ω_0 and ω. If an array of multiple receivers is set up, then the delay between the times when each receives the signal may be used to calculate the bearing of the target.

In practice, one cannot compute signal parameters exactly from the received waveform, as it also contains some noise. This noise can come from the environment, or possibly from the receiving equipment itself. We may relate the signal-to-noise ratio to the performance of a receiver; specifically, as the noise level increases relative to the signal energy, performance degrades due to greater uncertainty. If the signal-to-noise ratio is relatively high, then the error increases approximately linearly with the noise variance. However, below a certain threshold signal-to-noise ratio, one finds that the error increases significantly and can no longer be accurately characterized by the high signal-to-noise ratio equations.

Although one may assume high signal-to-noise ratio levels in the case of radar, for sonar the low signal-to-noise ratio situation cannot be ignored. More ambient noise is present at ultrasound frequencies; additionally, simply increasing the signal energy is not always a viable option, both for practical reasons and from a desire not to disturb the environment. As a result, the threshold effect is an important part of the theoretical understanding of sonar systems. We seek to develop equations which will more accurately describe the performance of a receiver at a wide range of conditions.

The general situation is modeled as follows: we receive a signal

$$x(t) = s(t, \theta) + n(t)$$

where θ is the signal parameter, s is a known function of t and θ, and $n(t)$, the noise, is a random process. We are given $x(t)$, and must estimate θ from this observation. The study of this type of problem is known as estimation theory.

1.2 An Introduction to Estimation Theory

The general setting for estimation theory problems consists of a *parameter space* Θ and an *observation space* $(\mathcal{X}, \mathcal{F}, \mu)$, a complete measure space. We assume throughout that Θ is an interval on the real line. For each $\theta \in \Theta$ we are given the *likelihood function* $p(x \,|\, \theta)$, a conditional probability density function on possible observations $x \in \mathcal{X}$ arising from this parameter value.

In a *nonrandom parameter* (or *local*) estimation problem, we fix a *true parameter* $\theta \in \Theta$ and compute expectations with respect to this value: for measurable $h : \mathcal{X} \to \mathbb{R}$,

$$E_\theta \left[h(x)\right] \doteq \int_{\mathcal{X}} h(x) p(x \,|\, \theta) \, d\mu(x). \tag{1.1}$$

In a *random parameter* (or *global*) estimation problem, we are also given $p(\theta)$, an *a priori* probability density function on Θ. Here, expectations are computed jointly over both $\mathcal{X}$ and Θ :

$$E \left[h(x)\right] \doteq \int_\Theta \int_{\mathcal{X}} h(x) p(x \,|\, \theta) \, d\mu(x) \, p(\theta) \, d\theta. \tag{1.2}$$

It will be made clear from the context which of the two situations is being discussed. Note in particular that if the distribution on Θ is uniform, then $E_\theta \left[h(x)\right] = E \left[h(x)\right]$ for all θ.

Definition 1. *Let Θ, $\mathcal{X}$ be as above, and let $g : \Theta \to \mathbb{R}$ be any measurable* *function. An* estimator of g *is any function* $f : \mathcal{X} \to \mathbb{R}$. *We say that* f *is an* estimator of Θ *if it is being considered as an estimator of the identity function* $g(\theta) = \theta$.

Note that in the above definition of an estimator, f does not explicitly depend on the function g which is to be estimated. However, a primary goal of estimation theory is to find an f which is a "good" estimator of g; that is, if $x \in \mathcal{X}$ is drawn from the distribution $p(x \mid \theta)$, we would like the estimated value $f(x)$ to approximate the true value $g(\theta)$ with high probability.

There are several ways to evaluate estimator performance:

Definition 2. *Let* $f : \mathcal{X} \to \mathbb{R}$ *be an estimator of* $g : \Theta \to \mathbb{R}$. *Let* $\theta \in \Theta$.

1. *The* local mean-squared-error *of* f *is* $\mathrm{Err}_\theta(f) \doteq E_\theta\left[\left(f(x) - g(\theta)\right)^2\right]$. *The* global-mean-squared-error *is* $\mathrm{Err}(f) \doteq E\left[\left(f(x) - g(\theta)\right)^2\right]$.
2. *The* variance *of* f *is* $\mathrm{Var}_\theta(f) \doteq E_\theta\left[\left(f(x) - E_\theta\left[f(x)\right]\right)^2\right]$.
3. *The* bias *of* f *at* θ *is* $b_f(\theta) \doteq E_\theta\left[f(x) - g(\theta)\right]$. *If* $b_f(\theta) = 0$, *we say that* f *is* unbiased at θ. *We say that* f *is* unbiased *if it is unbiased at all* $\theta \in \Theta$.

Note that $\mathrm{Err}_\theta(f) = \mathrm{Var}_\theta(f) + b_f(\theta)^2$. In particular, if f is unbiased at θ then the local mean-squared-error equals the variance. We will focus on the mean-squared-error as the primary measure of estimator performance.

One may also consider the estimation of a vector parameter, where $\Theta \subset \mathbb{R}^n$. Due to space constraints, we will not discuss those problems in detail. See [2, 13, 19] for extensions of the results in this paper to vector parameter estimation.

1.3 Example: Signal Parameter Estimation

Our motivating example throughout this discussion will be the problem of estimating signal parameters in the presence of additive noise, which forms the basis of much analysis of radar and sonar systems. The signal is modeled as $\{s(t, \theta)\}_{\theta \in \Theta}$, a collection of functions in $L^2[-T, T]$ parametrized over Θ. The noise is modeled as a stochastic process $n \in L^2[-T, T]$. The observation is a random process $x \in \mathcal{X} = L^2[-T, T]$ defined by

$$x(t) \doteq s(t, \theta) + n(t). \tag{1.3}$$

We will assume that both $\frac{\partial s(t,\theta)}{\partial \theta}$ and $\frac{\partial^2 s(t,\theta)}{\partial \theta^2}$ exist and are continuous for all θ.

To simplify computations, we also assume that the noise is a stationary, uncorrelated Gaussian process with zero mean and variance N_0, also known as white noise. In particular, we assume that the covariance function $E\left[n(t)n(s)\right] = N_0\delta(t - s)$. It should be noted that white noise is difficult to model formally as a stochastic process [18]. However, this simple situation allows us to present examples in concise forms which have a clear relation to the noise variance.

The simplest case of signal parameter estimation is *amplitude estimation.* Here we let s_0 be some C^2 function on $[-T, T]$, and define

$$s(t, A) = As_0(t). \tag{1.4}$$

We might consider the estimation of A itself, or of a function of A such as the signal energy $g(A) = A^2E$, where

$$E \doteq \int_{-T}^{T} s_0(t)^2 \, dt. \tag{1.5}$$

Another important example is *time delay estimation,* in which $s(t, \tau) = s_0(t - \tau)$. We let the parameter space be $[-T/2, T/2]$, and assume that s_0 is C^2 on $[-T/2, T/2]$ and $s_0(t) = 0$ for $|t| > T/2$. Then the support of $s_0(t - \tau)$ is contained in $[-T, T]$ for all τ, and thus the signal energy $E \doteq \int_{-T}^{T} s_0(t-\tau)^2 \, dt$ is the same for all τ. As mentioned before, finding the distance to a sonar target is equivalent to estimating the function $g(\tau) = c(\tau - \tau_0)/2$, where c is the speed of the transmitted wave and τ_0 is the initial time that the signal was transmitted. The quantity E/N_0 is called the *signal-to-noise ratio* in each of the above problems.

Rather than consider $p(x \mid \theta)$ directly, it will suffice to consider the scaled function

$$\ell(x \mid \theta) \doteq \frac{p(x \mid \theta)}{p_0(x)}, \tag{1.6}$$

where p_0 is the distribution of the noise process n on $L^2[-T, T]$. Several different techniques exist for computing this quantity; among them are the Karhunen-Loeve expansion [9] and properties of reproducing kernel Hilbert spaces [12]. These techniques may be applied to arbitrary noise covariance functions; however, a complete discussion is beyond the scope of this paper. The net result for our assumed situation of white noise is the following:

$$\ell(x \mid \theta) = \exp\left\{ \frac{1}{N_0} \int_{-T}^{T} x(t)s(t, \theta) \, dt - \frac{1}{2N_0} \int_{-T}^{T} s(t, \theta)^2 \, dt \right\}. \tag{1.7}$$

We will use this computation throughout this paper, applying general results from estimation theory to the specific case of signal parameter estimation.

1.4 The ML and MAP Estimators

The *maximum likelihood* (ML) and *maximum a posteriori* (MAP) estimators are two examples of general methods for constructing estimators. Their constructions depend only on the functions $p(x \mid \theta)$ or $p(\theta)$ and thus may be applied to a wide range of estimation problems. We will assume in this section that $p(x \mid \theta)$ and $p(\theta)$ are differentiable with respect to θ.

The ML estimator may be applied to any local estimation problem. Given an observation $x \in \mathcal{X}$, $\hat{\theta}_{\mathrm{ML}} : \mathcal{X} \to \Theta$ chooses θ so as to maximize the likelihood function; it thus satisfies

$$p(x \mid \hat{\theta}_{\mathrm{ML}}(x)) = \sup_{\theta \in \Theta} p(x \mid \theta). \tag{1.8}$$

We may equivalently maximize the log likelihood function $\log(p(x \mid \theta))$. Then $\hat{\theta}_{\mathrm{ML}}(x)$ solves the equation

$$\left. \frac{\partial \log p(x \mid \theta)}{\partial \theta} \right|_{\theta = \hat{\theta}_{\mathrm{ML}}(x)} = 0. \tag{1.9}$$

We say that any estimator $\hat{\theta}_{\mathrm{ML}}$ which satisfies equation (1.9) is an ML estimator. For any function $g : \theta \to \mathbb{R}$, we consider $g \circ \hat{\theta}_{\mathrm{ML}}$ to be an ML estimator of g.

Similarly, the MAP estimator applies to a global estimation problem. The estimator $\hat{\theta}_{\mathrm{MAP}}$ maximizes the *a posteriori* probability $p(\theta \mid x)$, which equals $p(x \mid \theta)p(\theta)/p(x)$ by Bayes' rule. As the denominator $p(x)$ does not depend on θ, we need only maximize $p(x \mid \theta)p(\theta)$. Then as before,

$$\left. \left\{ \frac{\partial \log p(x \mid \theta)}{\partial \theta} + \frac{\partial \log p(\theta)}{\partial \theta} \right\} \right|_{\theta = \hat{\theta}_{\mathrm{MAP}}(x)} = 0. \tag{1.10}$$

Any estimator satisfying equation (1.10) is called a MAP estimator. For the sake of brevity, we will focus primarily on local estimation problems and the ML estimator in this chapter and the next. Similar theorems exist throughout for the random parameter and MAP estimator.

We now consider the ML estimator in the context of the signal parameter estimation problem. It is easy to see from (1.8) and (1.9) that replacing $p(x \mid \theta)$ with $\ell(x \mid \theta)$ will not affect the computations of the ML and MAP estimators. In particular,

$$\begin{aligned} \frac{\partial \log p(x \mid \theta)}{\partial \theta} &= \frac{\partial \log \ell(x \mid \theta)}{\partial \theta} \\ &= \frac{1}{N_0} \int_{-T}^{T} [x(t) - s(t, \theta)] \frac{\partial s(t, \theta)}{\partial \theta} \, dt, \end{aligned} \tag{1.11}$$

and we may compute $\hat{\theta}_{\mathrm{ML}}$ by setting the above to zero and solving for θ. For example, recall that in amplitude estimation, $s(t, A) = As_0(t)$; then from (1.9) and (1.11) we obtain

$$\hat{\theta}_{\mathrm{ML}}(x) = \frac{\int_{-T}^{T} x(t)s_0(t) \, dt}{\int_{-T}^{T} s_0(t)^2 \, dt} = \frac{1}{E} \int_{-T}^{T} x(t)s_0(t) \, dt. \tag{1.12}$$

In many cases, however, it is easier to directly maximize $\ell(x \mid \theta)$ as in

(1.8). Recall that for time-delay estimation, $s(t, \tau) = s_0(t - \tau)$, where we assume that the signal energy $\int_{-T}^{T} s_0(t - \tau)^2\, dt$ does not depend on τ. In this case it follows from (1.7) that we must maximize the cross-correlation

$$\int_{-T}^{T} x(t) s_0(t - \tau)\, dt. \tag{1.13}$$

The estimator which chooses τ to maximize (1.13) is known as the *matched filter estimator* and is commonly used in practice. In general, the matched filter estimator is an ML estimator if the signal power does not vary with the parameter.

1.5 The Threshold Effect

In the absence of noise, the matched-filter estimator picks the maximum of the *autocorrelation function* $\psi(\tau) = \int_{-T}^{T} s_0(t) s_0(t - \tau)\, dt$. The signal $s_0(t)$ is typically a waveform multiplied by a Gaussian envelope; as a result, $\psi(\tau)$ has multiple peaks, the highest being at $\tau = 0$. At low noise levels, the matched-filter estimator tends to pick a point which is close to this global maximum. However, if the noise level is sufficiently high, then the true peak cannot be distinguished reliably from the other local maxima of ψ. In this case, the error becomes spread out along the Gaussian envelope, which is significantly wider than a single peak. At even lower signal-to-noise ratios, the noise obscures any meaningful information about the received signal itself, and the predictions of the estimator become spread throughout the *a priori* interval.

We will describe this phenomenon in terms of lower bounds on estimator error. First, we make precise a few terms about these bounds. We say that a lower bound is achievable for an estimation problem if there exists an estimator whose error equals that bound. Note that in that case, the estimator must have the minimum error of any estimator for that problem. We define the actual error to be the infimum of the errors of all estimators for that problem. We say that one lower bound is at least as tight as another bound if the former is at least as large as the latter, and that it is tighter if it is strictly larger.

Recall that the signal-to-noise ratio is defined as E/N_0, where E is the signal energy and N_0 is the noise. For high signal-to-noise ratios, the Cramèr-Rao, Barankin and Ziv-Zakai bounds are all approximately equal; specifically, we will demonstrate that the ratios between the Barankin and the Cramèr-Rao bounds and between the Ziv-Zakai and the Cramèr-Rao bounds both converge to one as $E/N_0 \to \infty$. For low signal-to-noise ratios (i.e., in the limit as $E/N_0 \to 0$), the Barankin and Ziv-Zakai bounds for time-delay estimation are significantly tighter than the Cramèr-Rao bound. Since each bound is a lower limit on the error of any estimator, this implies that no estimator can approximate the Cramèr-Rao bound for low signal-to-noise ratios.

The above phenomenon is known as the *threshold effect*, since in practice one observes a rapid jump in estimator error around a particular value between the low noise region, in which the Cramèr-Rao bound is approximately achievable, and the high noise region, in which this is no longer the case. One may use these bounds to compute the threshold signal-to-noise ratio at which the ratio between the tighter bound and the Cramèr-Rao bound exceeds some fixed value. Additionally, some work has been done recently on theoretically quantifying the threshold signal-to-noise ratio using other lower bounds on estimator error. In this paper, however, we will focus primarily on demonstrating the existence of the threshold effect by considering what occurs in the limits of arbitrarily large or small signal-to-noise ratios.

Chapter 2

The Cramèr-Rao Bound

The Cramèr-Rao bound has been frequently used to analyze estimator performance. Its popularity comes from its ease of computability and from its approximation of the ML estimator under high signal-to-noise ratio conditions. It does not provide a good characterization of sonar performance, however, as that situation generally involves low signal-to-noise ratios. For our purposes, this bound serves as a basis of comparison with the tighter but more involved bounds which will be presented later in this paper.

2.1 The Fisher Information

Assume that we are given a local estimation problem, where Θ is the parameter space, and $\mathcal{X}$ the observation space. We will assume that for all θ, $\frac{\partial p(x\,|\,\theta)}{\partial\theta}$ and $\frac{\partial^2 p(x\,|\,\theta)}{\partial\theta^2}$ exist μ-a.e. and are integrable. In order to simplify the discussion, we will also assume that for all θ, $p(x\,|\,\theta)$ is nonzero for a.e. x. For $\theta \in \Theta$, define the *Fisher information*

$$I_\theta \doteq E_\theta\left[\left(\frac{\partial \log p(x\,|\,\theta)}{\partial\theta}\right)^2\right]. \tag{2.1}$$

This quantity will be seen to be a main component of the Cramèr-Rao bound.

The following lemma simplifies the computation of the Fisher information:

Lemma 1.

$$E_\theta\left[\left(\frac{\partial \log p(x\,|\,\theta)}{\partial\theta}\right)^2\right] = -E_\theta\left[\frac{\partial^2 \log p(x\,|\,\theta)}{\partial\theta^2}\right]. \tag{2.2}$$

Proof. First, note that

$$\frac{\partial p(x\,|\,\theta)}{\partial\theta} = \frac{\partial \log p(x\,|\,\theta)}{\partial\theta} p(x\,|\,\theta). \tag{2.3}$$

Since $p(x \mid \theta)$ is a probability density function, $\int_{\mathcal{X}} p(x \mid \theta)\, d\mu(x) = 1$; then differentiating with respect to θ and applying (2.3) gives

$$\int_{\mathcal{X}} \frac{\partial p(x \mid \theta)}{\partial \theta}\, d\mu(x) = \int_{\mathcal{X}} \frac{\partial \log p(x \mid \theta)}{\partial \theta} p(x \mid \theta)\, d\mu(x) = 0. \tag{2.4}$$

Differentiating again with respect to θ gives

$$\int_{\mathcal{X}} \frac{\partial^2 \log p(x \mid \theta)}{\partial \theta^2}\, d\mu(x) + \int_{\mathcal{X}} \left(\frac{\partial \log p(x \mid \theta)}{\partial \theta}\right)^2 p(x \mid \theta)\, d\mu(x) = 0, \tag{2.5}$$

from which the result follows. □

Using this lemma, we may calculate I_θ for the signal parameter estimation problem. From Lemma 1 and (1.11), it follows that

$$I_\theta = -\frac{1}{N_0} E_\theta \left[\int_{-T}^{T} [x(t) - s(t,\theta)] \frac{\partial^2 s(t,\theta)}{\partial \theta^2}\, dt\right] + \frac{1}{N_0} E_\theta \left[\int_{-T}^{T} \left[\frac{\partial s(t,\theta)}{\partial \theta}\right]^2 dt\right]. \tag{2.6}$$

We see that the first term is zero by interchanging the expectation and the integral, since the noise $n(t) = x(t) - s(t,\theta)$ has zero mean. This is valid since $\int_{-T}^{T} \mathrm{E}\big[\,|n(t)|\,\big] \left|\frac{\partial^2 s(t,\theta)}{\partial \theta^2}\right| dt$ may be shown to be finite by the Cauchy-Schwarz inequality. Thus,

$$I_\theta = \frac{1}{N_0} \int_{-T}^{T} \left[\frac{\partial s(t,\theta)}{\partial \theta}\right]^2 dt. \tag{2.7}$$

For signal amplitude estimation, in which $s(t, A) = A s_0(t)$, it follows that $I_A = E/N_0$. Recall that this quantity is the signal-to-noise ratio.

We may express the Fisher information for time-delay estimation in two useful forms. Recall that $s(t,\tau) = s_0(t-\tau)$, and that by assumption the signal energy $E = \int_{-T}^{T} s_0(t-\tau)^2\, dt$ is the same for all τ. Define the *auto-correlation function*

$$\psi(\tau) \doteq \frac{1}{E} \int_{-T}^{T} s_0(t) s_0(t-\tau)\, dt. \tag{2.8}$$

Lemma 2. *For time-delay estimation,* $I_\tau = \frac{E}{N_0}\left[-\psi''(0)\right]$.

Proof. Since $\int_{-T}^{T} s_0(t-\tau)^2\, dt = E$, differentiating both sides twice with respect to τ produces

$$\int_{-T}^{T} \left[s_0'(t-\tau)\right]^2 + \int_{-T}^{T} s_0''(t-\tau) s_0(t-\tau)\, dt = 0. \tag{2.9}$$

But

$$\psi''(\tau) = \frac{1}{E} \int_{-T}^{T} s_0''(t-\tau) s_0(t)\, dt, \tag{2.10}$$

and thus

$$\psi''(0) = \frac{1}{E}\int_{-T}^{T} s_0''(t)s_0(t)\,dt = \frac{1}{E}\int_{-T}^{T} s_0''(t-\tau)s_0(t-\tau)\,dt. \qquad (2.11)$$

Then the result follows from (2.7) and (2.9). □

Now, let $S(\omega)$ be the Fourier transform of $s_0(t)$. The *signal bandwidth* is defined as

$$\beta^2 \doteq \frac{\int_0^\infty |S(\omega)|^2\,\omega^2\,d\omega}{\int_0^\infty |S(\omega)|^2\,\omega^2\,d\omega}. \qquad (2.12)$$

Note that by Parseval's theorem, $E = 2\int_0^\infty |S(\omega)|^2\,\omega^2\,d\omega$; and by the correlation theorem,

$$\psi(t) = \frac{2}{E}\int_0^\infty |S(\omega)|^2\cos(\omega t)\,dt. \qquad (2.13)$$

It follows by differentiating both sides that $-\psi''(0) = \beta^2$. Thus, we may restate the above lemma as

$$I_\tau = \frac{E}{N_0}\beta^2. \qquad (2.14)$$

2.2 Deriving the Cramèr-Rao Bound

The importance of the Fisher information becomes apparent in the following theorem:

Theorem 1 (Cramèr-Rao Bound, local version). *Let Θ, $\mathcal{X}$ be as above, and let $g : \Theta \to \mathbb{R}$ be a measurable function. Let $\theta \in \Theta$ such that $g'(\theta)$ exists and $\frac{\partial p(x\,|\,\theta)}{\partial\theta}$ exists and is continuous μ-a.e. Assume that $I_\theta > 0$.*

If $f : \mathcal{X} \to \mathbb{R}$ is an estimator of g which is unbiased at θ, then

$$\mathrm{Err}_\theta(f) \geq \frac{g'(\theta)^2}{I_\theta}. \qquad (2.15)$$

Proof. Since f is unbiased, by definition for all $\theta \in \Theta$

$$\int_{\mathcal{X}} p(x\,|\,\theta)\,[f(x) - g(\theta)]\;d\mu(x) = 0. \qquad (2.16)$$

Taking the derivative of both sides with respect to θ,

$$\begin{aligned} 0 &= \frac{d}{d\theta}\int_{\mathcal{X}} p(x\,|\,\theta)\big(f(x) - g(\theta)\big)\,d\mu(x) && (2.17)\\ &= \int_{\mathcal{X}}\left[\frac{\partial p(x\,|\,\theta)}{\partial\theta}\big(f(x) - g(\theta)\big) - g'(\theta)p(x\,|\,\theta)\right]d\mu(x). && (2.18) \end{aligned}$$

Thus by (2.3),

$$\int_{\mathcal{X}} g'(\theta) p(x \mid \theta)\, d\mu(x) = \int_{\mathcal{X}} \frac{\partial \log p(x \mid \theta)}{\partial \theta} \big(f(x) - g(\theta)\big) p(x \mid \theta)\, d\mu(x). \quad (2.19)$$

The left-hand side is just $g'(\theta)$, since $\int_{\mathcal{X}} p(x \mid \theta)\, d\mu(x) = 1$. Then

$$\begin{aligned} g'(\theta) &= \int_{\mathcal{X}} \frac{\partial \log p(x \mid \theta)}{\partial \theta} \big(f(x) - g(\theta)\big) p(x \mid \theta)\, d\mu(x) && (2.20) \\ &= \int_{\mathcal{X}} \frac{\partial \log p(x \mid \theta)}{\partial \theta} \sqrt{p(x \mid \theta)} \big(f(x) - g(\theta)\big) \sqrt{p(x \mid \theta)}\, d\mu(x). && (2.21) \end{aligned}$$

Therefore by the Cauchy-Schwarz inequality,

$$\begin{aligned} g'(\theta)^2 &\leq \left[\int_{\mathcal{X}} \left(\frac{\partial \log p(x \mid \theta)}{\partial \theta}\right)^2 p(x \mid \theta)\, d\mu(x)\right] \\ &\quad \times \left[\int_{\mathcal{X}} \big(f(x) - g(\theta)\big)^2 p(x \mid \theta)\, d\mu(x)\right] && (2.22) \\ &= E_\theta \left[\left(\frac{\partial \log p(x \mid \theta)}{\partial \theta}\right)^2\right] \mathrm{Err}_\theta(f) \;=\; I_\theta \, \mathrm{Err}_\theta(f). && (2.23) \end{aligned}$$

Dividing both sides by I_θ completes the proof. □

It is shown in [12] that the bias $b_f(\theta)$ is differentiable with respect to θ. Using this fact, the following corollaries follow easily from Theorem 1:

Corollary 1. *Let f be an estimator of Θ with bias $b_f(\theta)$, and assume $I_\theta > 0$. Then*

$$\mathrm{Err}_\theta(f) \geq b_f(\theta)^2 + \frac{\big(1 + b_f'(\theta)\big)^2}{I_\theta}. \quad (2.24)$$

Proof. This follows immediately from (2.15), since f is an unbiased estimator of the function $g(\theta) = \theta + b_f(\theta)$, and as an estimator of g, $\mathrm{Err}(f) = \mathrm{Var}(f) + b_f(\theta)^2$. □

Corollary 2. *Let f be an estimator of Θ which is unbiased at θ. Then*

$$\mathrm{Err}_\theta(x) \geq I_\theta^{-1}. \quad (2.25)$$

Bounds may be derived in the global estimation case which are similar to those given above.

Note that the bound of Corollary 1 may be computed for an estimator with a known bias function, while Corollary 2 provides a lower bound on the error of *any* unbiased estimator f. These restrictions on the bias function are critical, as they allow us to disregard, for example, the trivial estimator which achieves zero local error by always taking the value of the true parameter regardless of the observation.

Our computations will focus on the unbiased version of the Cramèr-Rao bound. This begs the question, though, of whether an unbiased estimator of Θ even exists. For example, if Θ is a closed interval, then f is unbiased at the endpoints of Θ only if it has zero error at those points. Frequently, one may only assume the existence of an unbiased estimator at the middle point of the parameter space. However, it may be shown for the problems we are discussing that an estimator has small bias close to the center of the interval; and one sees from Corollary 1 that if $b_f(\theta)$ is small, then the biased version of the bound approaches the unbiased version. Additionally, we may sidestep the above concerns by using the Cramèr-Rao bound to discuss the variance of an estimator rather than its error, since any estimator f is an unbiased estimator of its own expectation function $E_\theta\left[f(x)\right]$. This approach is useful since, as mentioned before, the error and variance are similar if the bias is small.

2.3 Remarks and Computations

One naturally wonders what conditions are necessary for equality to hold in Theorem 1 or Corollaries 1 and 2. If an estimator f does achieve the Cramèr-Rao bound, it is called *efficient*. The Cauchy-Schwarz inequality (2.22) becomes an equality if and only if with probability one

$$\frac{\partial \log p(x\,|\,\theta)}{\partial\theta} = k(\theta)\left[f(x) - g(\theta)\right], \tag{2.26}$$

where $k(\theta)$ is a function not depending on x. Combining this with (1.9),

$$k\big(\hat{\theta}_{\mathrm{ML}}(x)\big)\left[f(x) - g\big(\hat{\theta}_{\mathrm{ML}}(x)\big)\right] = \left.\frac{\partial \log p(x\,|\,\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_{\mathrm{ML}}(x)} = 0. \tag{2.27}$$

We restrict ourselves to the case where the ML estimator depends on the observation. Then it follows that $f(x) = g(\hat{\theta}_{\mathrm{ML}}(x))$. Thus, if an efficient estimator exists, it will be an ML estimator.

The above remarks may be illustrated by computing the Cramèr-Rao bound for the signal parameter estimation problem. From (2.7) and Corollary 2, we see that if f is unbiased at θ then

$$\mathrm{Err}_\theta(f) \geq \frac{N_0}{\int_{-T}^{T}\left[\frac{\partial s(t,\theta)}{\partial\theta}\right]^2 dt}. \tag{2.28}$$

In the amplitude estimation problem, where $s(t,A) = A\,s_0(t)$, it follows that $\mathrm{Err}_\theta(f) \geq N_0/E$. Futhermore, recall from (1.11) and (1.12) that the ML estimator is $f(x) = \frac{1}{E}\int_{-T}^{T} x(t)s_0(t)\,dt$, and that

$$\frac{\partial \log p(x\,|\,\theta)}{\partial\theta} = \frac{1}{N_0}\left(\int_{-T}^{T} x(t)s_0(t)\,dt - E\right). \tag{2.29}$$

It follows that (2.26) is satisfied, which implies that the ML estimator is efficient; thus, in fact $\mathrm{Err}(f) = N_0/E$.

However, in most cases of interest, the ML estimator is not efficient. In general, if the signal-to-noise ratio $\frac{1}{N_0}\int_{-T}^{T} s(t,\theta)^2\,dt$ and I_θ are both large, then the Cramèr-Rao bound does provide a good approximation for the actual error [11]. For example, the bound for time-delay estimation is

$$\mathrm{Err}_\theta(f) \geq \frac{N_0}{E\left[-\psi''(0)\right]} = \frac{N_0}{E\beta^2}. \tag{2.30}$$

Intuitively, $-1/\psi''(0)$ may be thought of as measuring the variance of the autocorrelation function near its peak at 0. This bound accurately characterizes the error of the matched-filter receiver at high signal-to-noise ratios. However, as we shall see it diverges from the actual error when the estimator is likely to pick points that do not lie close to the true peak. For a better understanding of sonar performance, then, we must turn to more accurate bounds.

Chapter 3

The Barankin Bound

Although the Cramèr-Rao bound is easily computed, it provides little information about performance under conditions of low signal-to-noise ratio. This problem arises because although the Cramèr-Rao does lower-bound the error of any estimator, there is no guarantee that this bound approximates the actual error, and one must resort to propositions which assume high signal-to-noise ratio.

The Barankin bound, which uses ideas from the study of Banach spaces, provides a solution—with a few caveats. Given an estimation problem, it produces an achievable local lower bound on the error of any unbiased estimator. That is, we will prove the existence of an unbiased estimator whose error is in fact exactly the Barankin bound. This estimator has the minimum (local) error of all unbiased estimators.

In general, the minimum error estimator will depend on the true parameter and thus has less direct practical use than one might hope. Furthermore, the exact Barankin bound is difficult, if not impossible, to compute in general, and in applications one usually resorts to related bounds which are simpler to compute but no longer achievable. Still, the theory demonstrates a close link between the Barankin bound and the particulars of a given estimation problem. In particular, the threshold effect may be explained theoretically by comparing the Barankin and Cramèr-Rao bounds.

The organization of this chapter proceeds as follows. We first derive the Barankin bound and show that it is achievable under certain conditions. Since this bound is difficult to compute exactly, we turn our attention to less optimal forms. We present the Chapman-Robbins bound as part of the Barankin theory and show that it is at least as tight as the Cramèr-Rao bound. We then turn our attention to an approach which involves picking specific test parameters. We compute this bound for time-delay estimation, and demonstrate that it exhibits a threshold effect which may be explained by the presence of secondary peaks in the autocorrelation function.

3.1 Deriving the Barankin Bound

The following reformulation of local estimation problems was first made by Barankin [1]. As before, let Θ denote the parameter space, and let $(\mathcal{X}, \mathcal{F}, \mu)$ be the observation space. In order to simplify notation, we define the collection of functions $\{p_\theta\}_{\theta \in \Theta}$, where $p_\theta(x) = p(x \,|\, \theta)$ for all θ and x. We assume that each p_θ is $\mathcal{F}$-measurable. Throughout this chapter, we fix θ_0 as the true parameter.

Define the measure ν on $(\mathcal{X}, \mathcal{F})$ by

$$\nu(A) \doteq \int_A p_{\theta_0} \, d\mu, \; (A \subset \mathcal{X}). \tag{3.1}$$

Note that $(\mathcal{X}, \mathcal{F}, \nu)$ is a complete measure space. A function $h : \mathcal{X} \to \mathbb{R}$ is integrable with respect to ν if and only if $h \cdot p_{\theta_0}$ is integrable with respect to μ; furthermore

$$E_{\theta_0}[h] = \int_{\mathcal{X}} h \cdot p_{\theta_0} \, d\mu = \int_{\mathcal{X}} h \, d\nu. \tag{3.2}$$

Thus, we may relate expectation values over the true parameter to integrals over ν.

Let $L^2(\mathcal{X})$ denote the space of all $\mathcal{F}$-measurable functions with finite norm, where the norm $||\cdot||$ is defined with respect to ν:

$$||h|| = \left(\int_{\mathcal{X}} |h|^2 \, d\nu \right)^{1/2}. \tag{3.3}$$

It is well-known that $L^2(\mathcal{X})$ is complete.

Let $g : \Theta \to \mathbb{R}$ be the function to be estimated. If f is an unbiased estimator of g, then it follows from (3.2) and (3.3) that

$$\mathrm{Var}_{\theta_0}(f) = \mathrm{Err}_{\theta_0}(f) = ||f - g(\theta_0)||^2. \tag{3.4}$$

If $g \equiv g_0$ is constant, then the trivial estimator which always takes the value g_0 regardless of the observation will achieve zero error. Thus, we ignore this trivial case, and it will be assumed throughout that g is nonconstant.

For all $\theta \in \Theta$, define the *likelihood ratio*

$$L_\theta(x) \doteq \frac{p_\theta(x)}{p_{\theta_0}(x)}. \tag{3.5}$$

The fundamental assumption made in this chapter is that each L_θ has finite norm. We also assume that each L_θ is defined μ-a. e. The latter assumption does not reduce generality, as we are only concerned with expectations over p_{θ_0} and thus may remove from consideration all $x \in \mathcal{X}$ such that $p_{\theta_0}(x) = 0$.

The following lemma follows easily from the above definitions.

Lemma 3. *Let f be an estimator of g. Then f is unbiased if and only if* *for all* $\theta \in \Theta$,

$$\int_{\mathcal{X}} (f - g(\theta_0))\, L_\theta \, d\nu = g(\theta) - g(\theta_0). \tag{3.6}$$

Proof. If $h : \mathcal{X} \to \mathbb{R}$ is any measurable function, then $E_\theta[h] = \int_{\mathcal{X}} h \cdot p_\theta \, d\mu = \int_{\mathcal{X}} h \cdot L_\theta \, d\nu$. Thus it follows that $\int_{\mathcal{X}} L_\theta \, d\nu = 1$, and that f is unbiased if and only if for all $\theta \in \Theta$, $\int_{\mathcal{X}} f \cdot L_\theta \, d\nu = g(\theta)$. Combining these two equations produces (3.6). □

Let $\mathcal{M}$ be the set of unbiased estimators of g with finite variance. We have the following theorem, due to Barankin:

Theorem 2 (Barankin Bound). *1. Let* $f \in \mathcal{M}$. *Then*

$$\mathrm{Var}_{\theta_0}(f) \geq \sup_{n, a_i, \theta_i} \frac{\left[\sum_{i=1}^n a_i \left(g(\theta_i) - g(\theta_0)\right)\right]^2}{\left|\left|\sum_{i=1}^n a_i L_{\theta_i}\right|\right|^2}, \tag{3.7}$$

where the supremum is taken over all $n \in \mathbb{N}$, $a_1 \ldots a_n \in \mathbb{R}$, $\theta_1 \ldots \theta_n \in \Theta$, *such that* $||\sum_{i=1}^n a_i L_{\theta_i}|| \neq 0$.

2. If $\mathcal{M}$ *is nonempty, then there exists a unique (up to a set of* ν*-measure zero) unbiased estimator* f *which achieves the above bound; i.e.,*

$$\mathrm{Var}_{\theta_0}(f) = \sup_{n, a_i, \theta_i} \frac{\left[\sum_{i=1}^n a_i \left(g(\theta_i) - g(\theta_0)\right)\right]^2}{\left|\left|\sum_{i=1}^n a_i L_{\theta_i}\right|\right|^2}. \tag{3.8}$$

Proof of part 1. For any $n \in \mathbb{N}$, parameter values $\theta_1, \ldots, \theta_n \in \Theta$, and real numbers $a_1, \ldots, a_n$, it follows from Lemma 3 that

$$\sum_{i=1}^n a_i \left(g(\theta) - g(\theta_0)\right) = \int_{\mathcal{X}} (f - g(\theta_0)) \sum_{i=1}^n a_i L_{\theta_i} \, d\nu. \tag{3.9}$$

Then by the Cauchy-Schwarz inequality,

$$\left|\sum_{i=1}^n a_i \left(g(\theta) - g(\theta_0)\right)\right| \leq ||f - g(\theta_0)|| \cdot \left|\left|\sum_{i=1}^n a_i L_{\theta_i}\right|\right|. \tag{3.10}$$

Thus

$$||f - g(\theta_0)|| \geq \frac{\left|\sum_{i=1}^n a_i \left(g(\theta) - g(\theta_0)\right)\right|}{\left|\left|\sum_{i=1}^n a_i L_{\theta_i}\right|\right|} \tag{3.11}$$

(assuming that the denominator is nonzero), and equation (3.7) follows by taking the supremum of the right-hand side. □

Proof of part 2. Define

$$C \doteq \sup_{n, a_i, \theta_i} \frac{\left[\sum_{i=1}^n a_i \left(g(\theta_i) - g(\theta_0)\right)\right]^2}{\left|\left|\sum_{i=1}^n a_i L_{\theta_i}\right|\right|^2}. \tag{3.12}$$

Assume that $\mathcal{M}$ is nonempty, i.e. that there exists an unbiased estimator f with finite variance: then it follows from (3.11) that $C \leq ||f - g(\theta_0)||$ and thus is finite. It remains to show the existence of a unique unbiased estimator whose variance is C. First, we consider uniqueness; let $f_1, f_2 \in \mathcal{M}$ be two unbiased estimators such that $||f_1 - g(\theta_0)|| = ||f_2 - g(\theta_0)|| = C$. Then $1/2\,(f_1 + f_2) \in \mathcal{M}$, so $1/2\,||f_1 + f_2 - 2g(\theta_0)|| \geq C$ from the above result. But by the triangle inequality,

$$\frac{1}{2}\,||f_1 + f_2 - 2g(\theta_0)|| \leq \frac{1}{2}\,(||f_1 - g(\theta_0)|| + ||f_2 - g(\theta_0)||) = C, \tag{3.13}$$

so in fact

$$||f_1 + f_2 - 2g(\theta_0)|| = ||f_1 - g(\theta_0)|| + ||f_2 - g(\theta_0)||\,. \tag{3.14}$$

This implies that $f_1 - g(\theta_0) = d\,(f_2 - g(\theta_0))$ ν-a.e., where d is some positive constant. But since $f_1 - g(\theta_0)$ and $f_2 - g(\theta_0)$ have equal norms, $d = 1$ and $f_1 = f_2$ ν-a.e. Thus, if any unbiased estimator achieves (3.8), it is unique.

It remains to prove the existence of an unbiased estimator f_0 such that $||f_0 - g(\theta_0)|| = C$. Let $\mathcal{B} \doteq \{L_\theta : \theta \in \Theta\}$. Consider the functional F on $\mathcal{B}$, defined by

$$F(L_\theta) \doteq g(\theta) - g(\theta_0). \tag{3.15}$$

We may extend this to a linear functional on $[\mathcal{B}]$, the linear subspace of $L^2(\mathcal{X})$ spanned by $\mathcal{B}$.

We now use two results from functional analysis [15]. The *operator norm* of F is defined by

$$||F|| \doteq \sup_{h \in [\mathcal{B}]} \frac{|F(h)|}{||h||}. \tag{3.16}$$

It follows from (3.12) that $||F||^2 = C$. The Hahn-Banach theorem states that F may be extended to a linear functional G on all of $L^2(\mathcal{X})$, such that $F \equiv G$ on $[\mathcal{B}]$, and $||G|| = ||F|| = C^{1/2}$.

Then, the Riesz representation theorem asserts that there exists a function $\phi \in L^2(\mathcal{X})$ such that $||\phi|| = ||G|| = C^{1/2}$, and $G(h) = \int_{\mathcal{X}} \phi\, h\, d\nu$ for all $h \in L^2(\mathcal{X})$. Let $f = \phi + g(\theta_0)$. For all $\theta \in \Theta$, it follows that

$$\int_{\mathcal{X}} (f - g(\theta_0))\, L_\theta\, d\nu = G(L_\theta) = g(\theta) - g(\theta_0), \tag{3.17}$$

so by Lemma 3, f is an unbiased estimator of g. Furthermore, $||f - g(\theta_0)||^2 = ||\phi||^2 = C$, as desired. □

Part 2 of the above theorem states that for local bounds on the error of unbiased estimators, we can do no better than the above bound. Note that the substitution of *any* fixed n, a_i, and θ_i into equation (3.11) gives a (possibly not achievable) lower bound on the error of unbiased estimators. As a result, the above theorem yields a whole class of bounds which, although not optimal, are usually easier to compute than the exact bound (3.8). We will elaborate on this point in the next section.

3.2 Computing the Barankin Bound

Although the above theorem gives an explicit formula for the tightest possible local bound on unbiased estimator error, it is difficult to compute the quantity (3.7) exactly for many problems of interest. One must take a supremum over all possible n, θ_i and a_i, a daunting task. Note, though, that the substitution of *any* choice of n, θ_i and a_i into (3.11) provides a lower bound. Thus, if we restrict the supremum to a subset of all possible n, θ_i and a_i, we still produce a lower bound, albeit one which loses the guarantee of achievability by an unbiased estimator. This may be summarized as follows:

Corollary 3. *Fix $n \in \mathbb{N}$, and let $\Theta_1, \ldots, \Theta_n \subset \Theta$, $A_1, \ldots, A_n \subset \mathbb{R}$. Then for all $f \in \mathcal{M}$,*

$$\mathrm{Var}_{\theta_0}(f) \geq \sup_{\substack{\theta_i \in \Theta_i, \\ a_i \in A_i}} \frac{\left[\sum_{i=1}^{n} a_i \left(\theta_i - \theta_0\right)\right]^2}{\left|\left|\sum_{i=1}^{n} a_i L_{\theta_i}\right|\right|^2}. \tag{3.18}$$

As we shall see, a proper choice of Θ_i and A_i produces a more easily computed bound which is still reasonably tight. However, one must keep in mind that any bound obtained in such a manner is not guaranteed to be achievable. Thus, in applications it is important to specify which specific Barankin bound one uses. We will see that it is possible to produce bounds from the above theorem which are more easily computed than the exact Barankin bound, but still demonstrate useful information about the problems to which they are applied.

3.3 The Chapman-Robbins Bound

The simplest useful example of the above class of Barankin bounds is the Chapman-Robbins bound [5]. Although it was originally derived directly, its status as a Barankin bound was noted in that paper, and it is currently considered to be part of the larger Barankin theory. This bound has primarily been applied to bearing estimation [7, 17], although recently it has been shown to demonstrate a threshold effect for frequency estimation [10]. We will restrict ourselves to discussing its derivation and relationship to the Cramèr-Rao bound.

To derive the Chapman-Robbins bound, we set $n = 2$, $A_1 = \{1\}$, $A_2 =$

$\{-1\}$, $\Theta_2 = \{\theta_0\}$, and $\Theta_1 = \Theta$. Then from Corollary 3,

$$\begin{aligned}
\mathrm{Var}(f) &\geq \sup_{\theta_1 \in \Theta} \frac{(\theta_1 - \theta_0)^2}{\int_{\mathcal{X}} (L_{\theta_1}(x) - L_{\theta_0}(x))^2 \, d\nu} && (3.19)\\
&= \sup_{\theta_1 \in \Theta} \frac{(\theta_1 - \theta_0)^2}{\int_{\mathcal{X}} (p(x \,|\, \theta_1) - p(x \,|\, \theta_0))^2 / p(x \,|\, \theta_0) d\mu} && (3.20)\\
&= \sup_{\theta_1 \in \Theta} (\theta_1 - \theta_0)^2 \left\{ \int_{\mathcal{X}} p(x \,|\, \theta_1)^2 / p(x \,|\, \theta_0) \, d\mu \right.\\
&\qquad \left. - \int_{\mathcal{X}} \big(2\, p(x \,|\, \theta_1) - p(x \,|\, \theta_0)\big) \, d\mu \right\}^{-1} && (3.21)\\
&= \sup_{\theta_1 \in \Theta} \frac{(\theta_1 - \theta_0)^2}{\int_{\mathcal{X}} p(x \,|\, \theta_1)^2 / p(x \,|\, \theta_0) \, d\mu - 1}. && (3.22)
\end{aligned}$$

Letting $\lambda = \theta_1 - \theta_0$, we obtain the usual statement of this bound:

Theorem 3 (Chapman-Robbins bound). *Let f be an unbiased estimator. Then*

$$\mathrm{Var}(f) \geq \sup_{\lambda} \left\{ \frac{1}{\lambda^2} \left[\int_{\mathcal{X}} \frac{p(x \,|\, \theta_0 + \lambda)^2}{p(x \,|\, \theta_0)} \, d\mu - 1 \right] \right\}^{-1}. \qquad (3.23)$$

If we assume that $\frac{\partial p(x \,|\, \theta)}{\partial \theta}$ exists and is continuous for all $\theta \in \Theta$, it may be verified that the Cramèr-Rao bound for unbiased estimators is equivalent to taking the limit as $\lambda \to 0$: using the equality between (3.22) and (3.20),

$$\begin{aligned}
&\lim_{\lambda \to 0} \frac{\lambda^2}{\int_{\mathcal{X}} p(x \,|\, \theta_0 + \lambda)^2 / p(x \,|\, \theta_0) d\mu - 1} && (3.24)\\
&= \left(\lim_{\lambda \to 0} \int_{\mathcal{X}} \left[\frac{1}{p(x \,|\, \theta_0)} \frac{p(x \,|\, \theta_0 + \lambda) - p(x \,|\, \theta_0)}{\lambda} \right]^2 p(x \,|\, \theta_0) \, d\mu \right)^{-1} && (3.25)\\
&= \left(\int_{\mathcal{X}} \left[\left. \frac{\partial \log p(x \,|\, \theta)}{\partial \theta} \right|_{\theta = \theta_0} \right]^2 p(x \,|\, \theta_0) \, d\mu \right)^{-1} = I_\theta^{-1}. && (3.26)
\end{aligned}$$

Thus, the Chapman-Robbins bound is at least as tight as the Cramèr-Rao. Note that since the Cramèr-Rao bound is achievable for signal parameter estimation, the Chapman-Robbins bound equals the Cramèr-Rao bound in that problem. More generally, these results demonstrate that the Cramèr-Rao bound may be derived as part of the larger Barankin theory.

3.4 The Probable Outliers Approach

Another common method for evaluating the Barankin bound [14, 16], developed specifically for the signal parameter estimation problem, fixes particular $\theta_1, \ldots, \theta_n \in \Theta$ and takes the supremum over all possible values for the a_i. That is, using the previous notation we let $\Theta_i = \{\theta_i\}$ and $A_i = \mathbb{R}$ for $i = 1, \ldots, n$. The θ_i are usually chosen to be specific test points, i.e., parameter values, of interest. For example, recall that in the signal parameter estimation problem under conditions of high noise the receiver is likely to mistake the secondary peaks of the cross-correlation function with the true peak. These probable outliers are a good intuitive choice for the θ_i.

For $\theta_1, \theta_2 \in \Theta$, define

$$G_{\theta_0}(\theta_1, \theta_2) \doteq \int_{\mathcal{X}} L_{\theta_1} L_{\theta_2}\, d\nu = \int_{\mathcal{X}} \frac{p_{\theta_1} p_{\theta_2}}{p_{\theta_0}}\, d\mu. \tag{3.27}$$

By the Cauchy-Schwarz inequality,

$$|G_{\theta_0}(\theta_1, \theta_2)| = \left| \int_{\mathcal{X}} L_{\theta_1} L_{\theta_2}\, d\nu \right| \leq ||L_{\theta_1}|| \cdot ||L_{\theta_2}|| \,. \tag{3.28}$$

Furthermore, $G_{\theta_0}(\theta, \theta) = ||L_\theta||^2$. Thus, $G_{\theta_0}(\theta_i, \theta_j)$ is finite for all $\theta_i, \theta_j \in \Theta$ if and only if L_θ has finite norm for all $\theta \in \Theta$.

This subclass of Barankin bounds may be stated as follows:

Lemma 4. *Fix $n \in \mathbb{N}$, and pick $\theta_1, \ldots, \theta_n \in \Theta$. If f is an unbiased estimator of Θ, then*

$$\mathrm{Var}_{\theta_0}(f) \geq \sup_{a_1, \ldots, a_n} \frac{\left[\sum_{i=1}^n a_i(\theta_i - \theta_0)\right]^2}{\sum_{i=1}^n \sum_{j=1}^n a_i a_j G_{\theta_0}(\theta_i, \theta_j)}. \tag{3.29}$$

Proof. The numerator of the right hand side is the same as in Corollary 3. The denominator was

$$\begin{aligned}
\left|\left|\sum_{i=1}^n a_i L_{\theta_i}\right|\right|^2 &= \int_{\mathcal{X}} \left(\sum_{i=1}^n a_i L_{\theta_i}\right)^2 d\nu &(3.30)\\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \int_{\mathcal{X}} L_{\theta_i} L_{\theta_j}\, d\nu &(3.31)\\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j G_{\theta_0}(\theta_i, \theta_j). &(3.32)
\end{aligned}$$

□

If we add more test points, the bound of Lemma 4 becomes at least as tight (and possibly tighter) by maximizing over a larger set. By modifying the proof of Theorem 2, part 2, it may be shown that this bound corresponds to the minimum achievable error for estimators which are unbiased at the points $\theta_1, \ldots, \theta_n$. Furthermore, if we take the limit of this bound as the set of test points becomes dense in Θ, then this bound converges to the bound of Theorem 2, i.e., the minimum achievable error for estimators which are unbiased on all of Θ [8]. Thus, although suboptimal, this bound has significant theoretical justification.

Happily, we may evaluate the supremum in Lemma 4 explicitly. We first rewrite the above bound in terms of matrix multiplication. Define the n-dimensional column vector N and $n \times n$ matrix B by $N_i = \theta_i - \theta_0$, $B_{ij} = G_{\theta_0}(\theta_i, \theta_j)$ for $i, j = 1, \ldots, n$. Then the bound becomes

$$\mathrm{Var}_{\theta_0}(f) \geq \sup_{A \in \mathbb{R}^n} \frac{\left(N^t A\right)^2}{A^t B A}. \tag{3.33}$$

We now use the following lemma:

Lemma 5. *Let N be an n-dimensional real vector, and B a positive definite symmetric real $n \times n$ matrix. Then*

$$\sup_{A \in \mathbb{R}^m} \frac{\left(N^t A\right)^2}{A^t B A} = N^t B^{-1} N. \tag{3.34}$$

Proof. Since B is positive-definite and symmetric, it has a positive-definite symmetric square root $B^{1/2}$. Then $N^t A = \left(B^{-1/2} N\right)^t B^{1/2} A$ for all $A \in \mathbb{R}^n$, so

$$\frac{\left(N^t A\right)^2}{A^t B A} = \frac{\left[\left(B^{-1/2} N\right)^t B^{1/2} A\right]^2}{(B^{1/2} A)^t B^{1/2} A}. \tag{3.35}$$

Thus the Cauchy-Schwarz inequality implies that

$$\begin{aligned} \frac{\left(N^t A\right)^2}{A^t B A} &\leq \frac{\left(B^{-1/2} N\right)^t B^{-1/2} N \left(B^{1/2} A\right)^t B^{1/2} A}{\left(B^{1/2} A\right)^t B^{1/2} A} && (3.36) \\ &= N^t B^{-1} N, && (3.37) \end{aligned}$$

with equality if and only if $B^{-1/2} A = \lambda B^{1/2} N$ for some scalar λ; i.e., if $A = \lambda B^{-1} N$. □

Recall that $B_{ij} \doteq \int_{\mathcal{X}} L_{\theta_i} L_{\theta_j} \, d\nu$. As a result, B may be thought of as a matrix of inner products and is therefore positive semidefinite [14]. If it is not positive definite, we may remove some of the θ_i from consideration without affecting the bound. Thus, the above lemma may be applied without loss of generality. We therefore have the following lower bound on the variance, which may be computed explicitly by inverting the matrix B:

$$\mathrm{Var}_{\theta_0}(f) \geq N^t B^{-1} N. \tag{3.38}$$

A problem with this bound is that it is not immediately clear whether this bound is as tight as the Cramèr-Rao bound. In the remainder of this section, we will improve upon the above bound by modifying its derivation.

Let $L(x)$ be the function on Θ which assigns the value $L_\rho(x)$ to each $\rho \in \Theta$. Then for $\theta \in \Theta$, define

$$L'_\theta(x) \doteq \left. \frac{\partial L(x)}{\partial \rho} \right|_{\rho = \theta}. \tag{3.39}$$

Note that

$$L'_\theta(x) = \frac{1}{p(x\,|\,\theta)} \frac{\partial p(x\,|\,\theta)}{\partial \theta} = \frac{\partial \log p(x\,|\,\theta)}{\partial \theta}. \tag{3.40}$$

We have the following:

Lemma 6. *Assume that L'_{θ_0} is defined almost everywhere (with respect to ν), has finite norm, and that*

$$\lim_{\theta \to \theta_0} \left\| \frac{L_\theta - L_{\theta_0}}{\theta - \theta_0} - L'_{\theta_0} \right\| = 0. \tag{3.41}$$

Fix $\theta_1, \ldots, \theta_n \in \Theta$. Then for all unbiased estimators f of Θ,

$$\mathrm{Var}_{\theta_0}(f) \geq \sup_{a_0, \ldots, a_n \in \mathbb{R}} \frac{\left[\sum_{i=1}^n a_i(\theta_i - \theta_0) + a_0\right]^2}{\left\| \sum_{i=1}^n a_i L_\theta + a_0 L'_{\theta_0} \right\|^2}. \tag{3.42}$$

Proof. Consider the bound from Corollary 3, with $n+2$ terms in the summations. Fix $\theta \neq \theta_0$, and let $\theta_{n+1} = \theta_0$, $\theta_{n+2} = \theta$, $a_{n+1} = a_0/(\theta - \theta_0)$, $a_{n+2} = a_0/(\theta - \theta_0)$. Then

$$\mathrm{Var}_{\theta_0}(f) \geq \lim_{\theta \to \theta_0} \frac{\left[\sum_{i=1}^n a_i(\theta_i - \theta_0) + a_0\right]^2}{\left\| \sum_{i=1}^n a_i L_\theta + a_0 \frac{L_\theta - L_{\theta_0}}{\theta - \theta_0} \right\|^2}. \tag{3.43}$$

From the assumption (3.41), it follows that the element $\sum_{i=1}^n a_i L_\theta + a_0 \frac{L_\theta - L_{\theta_0}}{\theta - \theta_0}$ converges in the norm to $\sum_{i=1}^n a_i L_\theta + a_0 L'_{\theta_0}$, as $\theta \to \theta_0$. As a result, the denominator of (3.43) converges to that of (3.42); taking the supremum completes the proof. □

We may compute the maximization of the above bound in the same way as was done for Lemma 4. Define the $n+1$-dimensional vector N' and $(n+1) \times (n+1)$-dimensional matrix B' by

$$\begin{aligned}
N'_i &= \theta_i - \theta_0, && 1 \leq i \leq n, && (3.44)\\
N'_0 &= 1, && && (3.45)\\
B'_{ij} &= G_{\theta_0}(\theta_i, \theta_j), && 1 \leq i, j \leq n && (3.46)\\
B'_{i0} = B'_{0i} &= \int_{\mathcal{X}} L'_{\theta_0} L_{\theta_i}\, d\nu && && (3.47)\\
B'_{00} &= \int_{\mathcal{X}} (L'_{\theta_0})^2\, d\nu. && && (3.48)
\end{aligned}$$

It is easy to see that

$$B'_{i0} = \frac{\partial}{\partial \theta} G_{\theta_0}(\theta, \theta_i)\Big|_{\theta = \theta_0}, \qquad 1 \leq i \leq n \tag{3.49}$$

and, by (3.40),

$$B'_{00} = I_{\theta_0}. \tag{3.50}$$

Then we may restate (3.42) as

$$\begin{aligned}
\mathrm{Var}_{\theta_0}(f) &\geq \sup_{a_0,\ldots,a_n} \frac{\left[a_0 + \sum_{i=1}^n a_i(\theta_i - \theta_0)\right]^2}{\sum_{i=0}^n \sum_{j=0}^n B'_{ij} a_i a_j} && (3.51)\\
&= \sup_{A \in \mathbb{R}^{n+1}} \frac{\left((N')^t A\right)^2}{A^t B' A} && (3.52)\\
&= (N')^t (B')^{-1} N', && (3.53)
\end{aligned}$$

where the last step follows from Lemma 5.

In particular, if we set $a_i = 0$ for $1 \leq i \leq n$, this bound becomes

$$\mathrm{Var}_{\theta_0}(f) \geq (B'_{00})^{-1} = I_{\theta_0}^{-1}, \tag{3.54}$$

which is just the Cramèr-Rao bound for unbiased estimators. Thus, as with the Chapman-Robbins bound, this class of bounds contains the Cramèr-Rao bound as a special case. In the next section, we will see that we may compute bounds which are tighter than the Cramèr-Rao bound, exhibiting a threshold effect for time-delay estimation.

3.5 Computing the Bound for Signal Parameter Estimation

We will now apply the bound (3.53) to the signal parameter estimation problem. Assume for the moment that we have chosen the test points $\theta_1, \ldots, \theta_n$; we will return to the question of how to choose them.

From the previous chapter,

$$\begin{aligned}
L_\theta(x) = \frac{\ell(x \,|\, \theta)}{\ell(x \,|\, \theta_0)} = \exp\Bigg\{ &\frac{1}{N_0} \int_{-T}^{T} x(t) \left(s(t,\theta) - s(t,\theta_0)\right) \, dt \\
&- \frac{1}{2N_0} \int_{-T}^{T} \left(s(t,\theta)^2 - s(t,\theta_0)^2\right) \, dt \Bigg\}.
\end{aligned} \tag{3.55}$$

We use this result to compute $G_{\theta_0}(\theta_i, \theta_j)$ as follows. First consider the following lemma:

Lemma 7. *Let Y be a Gaussian random variable with zero mean and variance σ^2. Then $E\left[e^Y\right] = e^{\sigma^2/2}$.*

Proof. First, note that

$$E\left[e^Y\right] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{y - \frac{y^2}{2\sigma^2}} \, dy. \tag{3.56}$$

By completing the square we see that

$$E\left[e^Y\right] = e^{\sigma^2/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\sigma^2)^2/2\sigma^2}\, dy, \tag{3.57}$$

and the integral evaluates to 1. □

Lemma 8. *For the signal parameter estimation problem,*

$$G_{\theta_0}(\theta_i,\theta_j) = \exp\left\{\frac{1}{N_0}\int_{-T}^{T}\left[s(t,\theta_i)-s(t,\theta_0)\right]\left[s(t,\theta_j)-s(t,\theta_0)\right]\, dt\right\}. \tag{3.58}$$

Proof. Recall that $G_{\theta_0}(\theta_i,\theta_j) = E_{\theta_0}\left[L_{\theta_i}L_{\theta_j}\right]$. Letting $x(t) = s(t,\theta_0) + n(t)$ in (3.55), since we are computing the expectation with respect ot the parameter θ_0,

$$\begin{aligned}
G_{\theta_0}(\theta_i,\theta_j) = {} & \exp\left\{-\frac{1}{2N_0}\int_{-T}^{T}\left[s(t,\theta_i)^2 + s(t,\theta_j)^2 - 2s(t,\theta_0)^2\right]\, dt\right\} \\
& \times \exp\left\{\frac{1}{N_0}\int_{-T}^{T} s(t,\theta_0)\left[s(t,\theta_i)+s(t,\theta_j)-2s(t,\theta_0)\right]\, dt\right\} \\
& \times E_{\theta_0}\left[\exp\left\{\frac{1}{N_0}\int_{-T}^{T} n(t)\left[s(t,\theta_i)+s(t,\theta_j)-2s(t,\theta_0)\right]\, dt\right\}\right].
\end{aligned} \tag{3.59}$$

Let $Y = \frac{1}{N_0}\int_{-T}^{T} n(t)\left[s(t,\theta_i)+s(t,\theta_j)-2s(t,\theta_0)\right]\, dt$. Since n is a Gaussian process with zero mean and variance N_0, it follows that Y is a Gaussian random variable with zero mean and variance

$$\frac{1}{N_0}\int_{-T}^{T}\left(s(t,\theta_i)+s(t,\theta_j)-2s(t,\theta_0)\right)^2\, dt. \tag{3.60}$$

Therefore by the previous lemma,

$$E_{\theta_0}\left[e^Y\right] = \exp\left\{\frac{1}{N_0}\int_{-T}^{T}\left(s(t,\theta_i)+s(t,\theta_j)-2s(t,\theta_0)\right)^2\right\}\, dt. \tag{3.61}$$

Substituting this into (3.59) and then expanding the terms within the integrals in that equation produces the desired result. □

It follows from the above lemma that

$$B_{j0} = \frac{1}{N_0}\int_{-T}^{T} s'(t,\theta_0)\left[s(t,\theta_j)-s(t,\theta_0)\right] dt, \tag{3.62}$$

$$B_{00} = \frac{1}{N_0}\int_{-T}^{T} s'(t,\theta_0)^2 dt, \tag{3.63}$$

where

$$s'(t,\theta_0) = \left.\frac{\partial s(t,\theta)}{\partial\theta}\right|_{\theta=\theta_0}. \tag{3.64}$$

These results may be used along with Lemma 8 to compute the bound (3.53) for any choice of test points.

Now, let us consider our specific examples of amplitude and time-delay estimation. In the previous chapter, we saw that the Cramèr-Rao bound is the actual error for amplitude estimation. As a result, no threshold effects exist in this case, and the bound (3.53) is equal to the Cramèr-Rao bound.

The situation is more complicated for time-delay estimation, however. Recall that in this case, $s(t,\tau) = s_0(t-\tau)$. As before, we assume that the signal energy $E \doteq \int_{-T}^{T} s(t-\tau)^2\, dt$ does not depend on τ. The signal-to-noise ratio is defined to be E/N_0.

Denote the the true parameter by τ_0. Recall that the autocorrelation function is

$$\psi(\tau) \doteq \frac{1}{E}\int_{-T}^{T} s_0(t)s_0(t-\tau)\, dt. \tag{3.65}$$

If there is no noise, this will be the output of the matched filter receiver. The autocorrelation function has a maximum peak at $\tau = 0$ as well as various secondary peaks. The threshold effect occurs because under low signal-to-noise ratio conditions, the secondary peaks are mistaken for the true maximum. Thus, these secondary peaks are a good choice for the test points. By using them to compute the Barankin bound, we will see that reducing the magnitude of the secondary peaks lessens the threshold effect.

First, note that since the test points τ_i are in $[-T/2, T/2]$ and $s_0(t) = 0$ for $|t| > T/2$, the autocorrelation function satisfies

$$\psi(\tau_i - \tau_0) = \frac{1}{E}\int_{-T}^{T} s_0(t-\tau_i)s_0(t-\tau_0)\, dt. \tag{3.66}$$

Using the previous results, we see that

$$\begin{aligned} B_{ij} &= \exp\left\{\frac{1}{N_0}\int_{-T}^{T} [s_0(t-\tau_i) - s_0(t-\tau_0)]\right. \\ &\qquad\qquad \left.\times [s_0(t-\tau_j) - s_0(t-\tau_0)]\ dt\right\} && (3.67) \\ &= \exp\left\{\frac{E}{N_0}[\psi(\tau_i-\tau_j) - \psi(\tau_i-\tau_0) - \psi(\tau_j-\tau_0) + 1]\right\}. && (3.68) \end{aligned}$$

Differentiating $\psi(\tau_i - \tau_0)$ with respect to τ_0 gives

$$\psi'(\tau_i - \tau_0) = \frac{1}{E}\int_{-T}^{T} s_0(t-\tau_i)s_0'(t-\tau_0)\, dt. \tag{3.69}$$

Also, since $\int_{-T}^{T} s_0(t-\tau)^2\, dt = E$ for all $|\tau| \le T/2$, it follows that

$$\int_{-T}^{T} s_0(t-\tau)s_0'(t-\tau)\, dt = 0, \tag{3.70}$$

so in particular $\psi'(0) = 0$. Therefore,

$$\begin{aligned} B'_{i0} = B'_{0i} &= -\frac{1}{N_0}\int_{-T}^{T}\left[s_0(t-\tau_i) - s_0(t-\tau_0)\right] s'_0(t-\tau_0)\,dt && (3.71) \\ &= -\frac{E}{N_0}\left[\psi'(\tau_i-\tau_0) + \psi'(0)\right] && (3.72) \\ &= -\frac{E}{N_0}\psi'(\tau_i-\tau_0). && (3.73) \end{aligned}$$

Finally, using (3.50), and the computation from the previous chapter, we see that

$$B'_{00} = \frac{E}{N_0}\left[-\psi''(0)\right]. \tag{3.74}$$

As we have noted, a good intuitive choice for the test parameters τ_i are the critical points of the autocorrelation function. Thus, we pick $\tau_1, \ldots, \tau_n$ such that

$$\psi'(\tau_i - \tau_0) = 0, \tag{3.75}$$

where the points $\tau_i - \tau_0$ yield the n largest secondary peaks of ψ. It then follows that $B'_{i0} = B'_{0i} = 0$ for $i = 1, \ldots, n$. Thus, for this choice of test points we get

$$\mathrm{Var}(f) \geq \frac{1}{B'_{00}} + N^t B^{-1} N. \tag{3.76}$$

To see how this is affected by the signal-to-noise ratio, consider the following simple case from [14]. Assume that the autocorrelation function is symmetric about the origin, and pick δ such that $\pm\delta$ are its two largest secondary peaks. We choose $\tau_1 = \tau_0 + \delta$ and $\tau_2 = \tau_0 - \delta$ as the test points. Then

$$\begin{aligned} B_{11} = B_{22} &= \exp\left\{\frac{2E}{N_0}\left[1 - \psi(\delta)\right]\right\} && (3.77) \\ B_{12} = B_{21} &= \exp\left\{\frac{E}{N_0}\left[1 - 2\psi(\delta) + \psi(2\delta)\right]\right\}, && (3.78) \end{aligned}$$

and $N^t = (\delta, -\delta)$.

Here we may easily invert B, obtaining

$$\begin{aligned} \mathrm{Var}(f) &\geq \frac{1}{B'_{00}} + \frac{2\delta^2}{B_{11} - B_{12}} && (3.79) \\ &= \frac{N_0}{E\left[-\psi''(0)\right]} + 2\delta^2 \frac{\exp\left\{-\frac{2E}{N_0}\left[1-\psi(\delta)\right]\right\}}{1 - \exp\left\{-\frac{E}{N_0}\left[1-\psi(2\delta)\right]\right\}}. && (3.80) \end{aligned}$$

As $E/N_0 \to \infty$, the second term approaches zero. Thus, in high signal-to-noise conditions this bound approximates the Cramèr-Rao bound. In low signal-to-noise conditions, the second term will dominate, producing the threshold effect. This term is larger when $\psi(\delta)$ and $\psi(2\delta)$ are closer to 1. Therefore, the threshold effect will occur sooner when the secondary peaks of the cross-correlation are close to the height of the primary peak. A common method of reducing the threshold effect is to pick signals with small secondary peaks. Our analysis of the threshold effect using the above bound bound justifies that method theoretically. Note that adding more test points will result in a tighter bound, which must thus also exhibit the threshold effect.

We have presented the exact Barankin bound along with suboptimal but computationally useful forms, and applied them to signal parameter estimation problems. In the next chapter, we will see that the Ziv-Zakai bound takes a completely different approach, but also describes a threshold effect for time-delay estimation.

Chapter 4

The Ziv-Zakai Bound

The Ziv-Zakai bound was originally published in [21] and subsequently improved upon by [3] and [6]. Unlike the Cramèr-Rao and Barankin bounds, the its presentation was originally motivated explicitly by signal parameter estimation, specifically time-delay estimation—in particular, none of the above authors gives the statement of the bound for a general estimation problem. The earliest presentation found by the author for general estimation problems was in [20], which applies the bound to bearing estimation. Additionally, it has only very recently [2] been applied to the problem of estimating vector parameters.

In contrast to the Barankin bound, the Ziv-Zakai bound is global; that is, it bounds the error over all possible parameter values. This allows the Ziv-Zakai bound to include any *a priori* information on the parameter space (for example, that the parameter lies within a finite interval). Furthermore, in contrast to local bounds which we have seen, it does not require the estimator to be unbiased. Unlike the Barankin bound, however, the Ziv-Zakai bound is not guaranteed to be approximately achievable.

The organization of this chapter is as follows. We first give a brief introduction to detection theory, which is central to the derivation of the Ziv-Zakai bound. We then present the Ziv-Zakai bound, and give its explicit forms for amplitude and time-delay estimation. We see that for amplitude estimation it is equal to the Cramèr-Rao bound, and that for time-delay estimation it exhibits a threshold effect as the *a priori* information become the dominant term in the bound at low signal-to-noise ratios.

4.1 Results from Detection Theory

Whereas estimation theory is concerned with the prediction of a continuous parameter, *detection theory* is concerned with the prediction of a parameter taking values in a finite set. As a result of this simplification, many results about detection problems are easier to produce than their counterparts in estimation theory. In particular, the minimal error for a detection problem may be calculated explicitly. As the motivation for this discussion of detection theory is its applicability to the Ziv-Zakai bound, we will focus on a restricted version of the theory. A more general treatment may be found in [19].

The setup of a detection problem is similar to that of estimation problems. The *observation space* is a complete measure space $(\mathcal{X}, \mathcal{F}, \mu)$. The *hypothesis space* is a finite set $\mathcal{H} = \{H_0, \ldots, H_k\}$; for each $H_i \in \mathcal{H}$ we are given both the conditional likelihood $p(x \mid H_i)$, a probability distribution on $\mathcal{X}$, and the *a priori* probabilities $p(H_i)$.

A *detector* for the above problem is a measurable function $f : \mathcal{X} \to \mathcal{H}$; or equivalently, a partition of $\mathcal{X}$ into measurable sets $Z_0, \ldots, Z_k$. The error is the probability that the detector misclassifies a given observation. It may be defined as

$$\mathrm{P_E}(f) \doteq \sum_{i=0}^{k} p(H_i) \int_{\mathcal{X}-Z_i} p(x \mid H_i)\, d\mu, \tag{4.1}$$

where $Z_i = f^{-1}(H_i)$ for $i = 0, \ldots, k$.

The discussion will be restricted to binary detection problems, in which $\mathcal{H} = \{H_0, H_1\}$. Here, the error for a given detection scheme f is

$$\mathrm{P_E}(f) = p(H_0) \int_{Z_1} p(x \mid H_0)\, d\mu + p(H_1) \int_{Z_0} p(x \mid H_1)\, d\mu. \tag{4.2}$$

We now derive an optimal detector. Assume that we are given a binary detection problem. We wish to find a detector f which minimizes $\mathrm{P_E}(f)$. Let Z_0, Z_1 be as above. Since $\int_{\mathcal{X}} p(x \mid H_0)\, d\mu = 1$, we may rewrite (4.2) as

$$\mathrm{P_E}(f) = p(H_0) + \int_{Z_0} \left[p(H_1)p(x \mid H_1) - p(H_0)p(x \mid H_0)\right] d\mu. \tag{4.3}$$

As the first term $p(H_0)$ is constant, it suffices to find a set $Z_0 \subset \mathcal{X}$ which minimizes the second term. Consider the quantity inside the integral, $p(H_1)p(x \mid H_1) - p(H_0)p(x \mid H_0)$. It is not hard to see that we may minimize the integral by assigning to Z_0 all $x \in \mathcal{X}$ such that this quantity is negative and assigning to $Z_1 = \mathcal{X} - Z_0$ all x such that the quantity is positive. (If this quantity is zero, then it does not matter which of Z_0 or Z_1 we choose; for the sake of argument, assume that we assign all of those elements to Z_1.)

Therefore, the optimal detector may be written as

$$\left[p(H_0)p(x \mid H_0) - p(H_1)p(x \mid H_1)\right] \underset{H_0}{\overset{H_1}{\gtrless}} 0. \tag{4.4}$$

The above notation refers to the fact that given $x \in \mathcal{X}$, this detector chooses H_1 if the left-hand side is greater than the right-hand side and chooses H_0 if the right-hand side is greater than the left. Equivalently, assuming that $p(x \mid H_0)$ and $p(H_1)$ are nonzero,

$$\frac{p(x \mid H_1)}{p(x \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p(H_0)}{p(H_1)}. \tag{4.5}$$

The quantity on the left-hand side of (4.5) is the likelihood ratio which we encountered in the previous chapter; this detector is known as the *likelihood ratio test*. Note that taking the logarithm of both sides will not affect the performance. This detector is simple to use in practice, as it chooses between hypotheses by comparing the likelihood ratio to a fixed threshold $\eta = p(H_0)/p(H_1)$. A detector might compare the likelihood ratio to a different choice of η if the cost of falsely predicting H_0 is different than that of H_1; for example, one might prefer false positives to false negatives [19].

We denote the error of the likelihood ratio test by $\mathrm{P}_{\mathrm{min}}$. Since we have shown that (4.5) is the minimum error detector, $\mathrm{P}_{\mathrm{min}}$ is the minimum achievable error for the detection problem.

4.2 Binary Signal Detection

The binary signal detection problem has many applications, for example in digital system communications. As before, the setup will be restricted so that we may more easily produce concrete results. We receive one of two signals in the presence of additive noise, and must distinguish between them. The model is as follows:

$$x(t) = s_0(t) + n(t) \qquad (H_0), \tag{4.6}$$

$$x(t) = s_1(t) + n(t) \qquad (H_1), \tag{4.7}$$

where H_0, H_1 are the two hypotheses between which we must distinguish. Here s_0 and s_1 are known, nonrandom functions in $L^2[-T,T]$, and n is an independent Gaussian random process with zero mean and variance N_0. In order to simplify the argument we will assume that the two hypotheses H_0, H_1 are equally likely; the generalization may be found in [19].

Let the signal energies be

$$E_i \doteq \int_{-T}^{T} s_i(t)^2 \, dt \qquad i = 0, 1. \tag{4.8}$$

We are concerned with applying the optimal detection scheme of the previous section to this problem and computing its rate of error.

Adapting our computation of the likelihood ratio from the previous chapter, we find that:

$$\log \frac{p(x \mid H_1)}{p(x \mid H_0)} = \frac{1}{N_0} \left[\int_{-T}^{T} x(t)s_1(t)\, dt - \int_{-T}^{T} x(t)s_0(t)\, dt - \frac{E_1 - E_0}{2} \right]. \tag{4.9}$$

It follows that the log likelihood ratio test (4.5) is

$$\int_{-T}^{T} x(t)s_1(t)\, dt - \int_{-T}^{T} x(t)s_0(t)\, dt \underset{H_0}{\overset{H_1}{\gtrless}} \frac{E_1 - E_0}{2}. \tag{4.10}$$

In particular, if the signal energies are equal, then this test picks whichever signal of s_0 and s_1 has greater correlation with x. Recall that the ML and MAP estimators similarly pick the parameter value with the highest correlation.

As mentioned above, the error of this detector may be computed explicitly. Define

$$\Phi(x) \doteq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-y^2/2}\, dy. \tag{4.11}$$

Lemma 9. *The error of the likelihood ratio detector for the above signal parameter estimation problem is*

$$\mathrm{P}_{\min} = \Phi\left(\sqrt{\frac{E_0 + E_1 - 2\rho}{2N_0}}\right), \tag{4.12}$$

where $\rho \doteq \int_{-T}^{T} s_0(t)s_1(t)\, dt$.

Proof. Let P_i be the probability of error under hypothesis H_i. Then

$$\mathrm{P}_{\min}(f) = p(H_0)P_0 + p(H_1)P_1. \tag{4.13}$$

Under H_0, $x(t) = s_0(t)+n(t)$; thus using (4.10), $P_0 = \Pr\left(X < (E_0 - E_1)/2\right)$, where

$$X = \int_{-T}^{T} [s_0(t) + n(t)]\, s_0(t)\, dt - \int_{-T}^{T} [s_0(t) + n(t)]\, s_1(t)\, dt. \tag{4.14}$$

Since n has zero mean and is uncorrelated with variance N_0,

$$\begin{aligned}
E[X] &= \int_{-T}^{T} s_0(t)^2\, dt - \int_{-T}^{T} s_0(t)s_1(t)\, dt = E_0 - \rho, &(4.15)\\
E\left[X^2\right] &= E\Bigg[\int_{-T}^{T}\int_{-T}^{T} [s_0(t) + n(t)]\,[s_1(t) - s_0(t)] \\
&\qquad\qquad \times [s_0(u) + n(u)]\,[s_1(u) - s_0(u)]\; dt\, du\Bigg] &(4.16)\\
&= \left(\int_{-T}^{T} s_0(t)\,[s_0(t) - s_1(t)]\; dt\right)^2 \\
&\quad + \int_{-T}^{T}\int_{-T}^{T} [s_1(t) - s_0(t)]\,[s_1(u) - s_0(u)] \\
&\qquad\qquad \times N_0\delta(t-u)\, dt\, du &(4.17)\\
&= (E_0 - \rho)^2 + N_0(E_0 + E_1 - 2\rho). &(4.18)
\end{aligned}$$

Thus, X is a Gaussian random variable with mean $m \doteq E_0 - \rho$ and variance $\sigma^2 \doteq N_0(E_0 + E_1 - 2\rho)$. Then

$$\begin{aligned} P_0 = \Pr\left(X < \frac{E_0 - E_1}{2} \mid H_0 \right) & \\ = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\frac{E_0-E_1}{2}} e^{-(y-m)^2/2\sigma^2}\, dX & \quad (4.19) \\ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{(E_0+E_1-2\rho)/2N_0}} e^{-y^2/2}\, dy & \quad (4.20) \\ = \Phi\left(\sqrt{\frac{E_0 + E_1 - 2\rho}{2N_0}} \right). & \quad (4.21) \end{aligned}$$

.

By a similar argument, P_1 also evaluates to (4.21), and therefore $\mathrm{P_{min}}$ does as well. □

Let us consider how this applies to the analogues of amplitude and time-delay estimation. For amplitude detection, let $s_0(t) = A_0 s(t)$ and $s_1(t) = A_1 s(t)$, and let $E = \int_{-T}^{T} s(t)^2\, dt$. Then $E_0 = A_0^2 E$, $E_1 = A_1^2 E$, and $\rho = A_0 A_1 E$, so

$$\mathrm{P_{min}}(A_0, A_1) = \Phi\left((A_1 - A_0) \sqrt{\frac{E}{2N_0}} \right). \tag{4.22}$$

For time-delay detection, $s_0(t) = s(t - \tau_0)$, $s_1(t) = s(t - \tau_1)$, and $E_1 = E_2 = \int_{-T}^{T} s(t)^2\, dt$. Then, $\rho = E\psi(\tau_1 - \tau_0)$, where ψ is the autocorrelation function of s. Thus

$$\mathrm{P_{min}}(\tau_0, \tau_1) = \Phi\left(\sqrt{\frac{E\,(1 - \psi(\tau_1 - \tau_0))}{N_0}} \right). \tag{4.23}$$

Note that in both cases, $\mathrm{P_{min}}$ only depends on the differences $A_1 - A_0$ and $\tau_1 - \tau_0$. This fact greatly simplifies the computation of the Ziv-Zakai bound.

4.3 The Ziv-Zakai Bound

Let $\mathcal{X}$ be the observation space and Θ be the parameter space in a global estimation problem. Fix $\theta_0, \theta_1 \in \Theta$, and assume that $\theta_0 < \theta_1$. Consider the binary detection problem in which we must decide between θ_0 and θ_1, defined as follows: the observation space is $\mathcal{X}$, and the hypothesis space is $\mathcal{H} = \{H_0, H_1\}$, with probability distributions

$$p(x \mid H_i) = p(x \mid \theta_i), \qquad i = 0, 1, \tag{4.24}$$

$$p(H_i) = \frac{p(\theta_i)}{p(\theta_0) + p(\theta_1)}, \qquad i = 0, 1. \tag{4.25}$$

(In the above equations, the quantities $p(x \mid \theta_i)$ and $p(\theta_i)$ are those from the estimation problem.)

Let $f : \mathcal{X} \to \Theta$ be an estimator of Θ. The *estimation theorist's approach* for solving the above detection problem picks whichever hypothesis is closer to the value estimated by f. In other words,

$$f(x) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\theta_0 + \theta_1}{2}. \tag{4.26}$$

If $f(x) = (\theta_0 + \theta_1)/2$, we arbitrarily assign x to H_1.

Let $\mathrm{P}_f(\theta_0, \theta_1)$ denote the error of the estimation theorist's approach, and let $\mathrm{P}_{\min}(\theta_0, \theta_1)$ denote the error of the likelihood ratio test. Since the likelihood ratio test is optimal,

$$\mathrm{P}_{\min}(\theta_0, \theta_1) \leq \mathrm{P}_f(\theta_0, \theta_1). \tag{4.27}$$

The following lemma derives easily from the definitions (4.2) and (4.26):

Lemma 10. *Let $\theta_0, \theta_1 \in \Theta$, where $\theta_0 < \theta_1$, and let f be an estimator of Θ. Then*

$$\begin{aligned} \mathrm{P}_f(\theta_0, \theta_1) = \frac{1}{p(\theta_0) + p(\theta_1)} \Bigg[& p(\theta_0) \mathrm{Pr} \left(f(x) > \frac{\theta_0 + \theta_1}{2} \mid \theta_0 \right) \\ & + p(\theta_1) \mathrm{Pr} \left(f(x) \leq \frac{\theta_0 + \theta_1}{2} \mid \theta_1 \right) \Bigg]. \end{aligned} \tag{4.28}$$

The Ziv-Zakai bound is derived by comparing the likelihood ratio test to the suboptimal estimation theorist's approach.

Theorem 4 (Ziv-Zakai Bound). *Let $\mathcal{X}$ be the observation space and $\Theta = [\alpha, \beta]$ be the parameter space in a global estimation problem. Let $f : \mathcal{X} \to \Theta$ be an estimator of Θ. Then*

$$\mathrm{Err}(f) \geq \int_0^{\beta - \alpha} \int_\alpha^{\beta - h} \frac{p(\theta) + p(\theta + h)}{2} \mathrm{P}_{\min}(\theta, \theta + h)\, d\theta\, h\, dh. \tag{4.29}$$

Proof. First, combining (4.27) and Lemma 10 shows that for all $0 \leq h \leq \beta - \alpha$, and for all $\theta \in [\alpha, \beta - h]$,

$$\begin{aligned} & [p(\theta) + p(\theta + h)]\, \mathrm{P}_{\min}(\theta, \theta + h) \\ & \quad \leq \; p(\theta)\, \mathrm{Pr} \left(f(x) - \theta > \frac{h}{2} \mid \theta \right) \\ & \quad + \; p(\theta + h)\, \mathrm{Pr} \left(f(x) - \theta \leq \frac{h}{2} \mid \theta + h \right). \end{aligned} \tag{4.30}$$

Integrating both sides with respect to θ,

$$\begin{aligned}
\int_{\alpha}^{\beta-h} & [p(\theta)+p(\theta+h)]\,\mathrm{P}_{\min}(\theta,\theta+h)\,d\theta \\
&\le \int_{\alpha}^{\beta-h} p(\theta)\Pr\left(f(x)-\theta>\frac{h}{2}\mid\theta\right)d\theta \\
&\quad+\int_{\alpha+h}^{\beta} p(\theta_1)\Pr\left(f(x)-\theta_1\le-\frac{h}{2}\mid\theta_1\right)d\theta_1, \qquad (4.31)
\end{aligned}$$

where the last term follows via a change of variables, namely $\theta_1=\theta+h$.

Combining the two terms, we see that

$$\begin{aligned}
\int_{\alpha}^{\beta-h} & [p(\theta)+p(\theta+h)]\,\mathrm{P}_{\min}(\theta,\theta+h)\,d\theta \\
&\le \int_{\Theta} p(\theta)\Pr\left(|f(x)-\theta|\ge\frac{h}{2}\mid\theta\right)d\theta \\
&= \Pr\left(|f(x)-\theta|\ge\frac{h}{2}\right), \qquad (4.32)
\end{aligned}$$

where the last probability is taken jointly over θ and x.

We can relate this inequality to the error of f using the following identity [4]:

Lemma 11. *Let X be a random variable. Then*

$$E[X]=\int_0^{\infty}\Pr(X>t)\;dt. \qquad (4.33)$$

By change of variables, this becomes

$$E\left[X^2\right]=\frac{1}{2}\int_0^{\infty}\Pr\left(X^2>\frac{h^2}{4}\right)h\,dh. \qquad (4.34)$$

Therefore,

$$\begin{aligned}
\mathrm{Err}(f) &= E\left[|f(x)-\theta|^2\right] \\
&\ge \int_0^{\infty}\int_{\alpha}^{\beta-h}\frac{p(\theta)+p(\theta+h)}{2}\mathrm{P}_{\min}(\theta,\theta+h)\,d\theta\,h\,dh. \qquad (4.35)
\end{aligned}$$

However, since θ and $\theta+h$ must both be in $[\alpha,\beta]$, the integral over h may be restricted to the interval $[0,\beta-\alpha]$. □

We may improve the above bound as follows [3]. Note in the above proof that $\Pr(|f(x)-\theta|>h/2)$ is nondecreasing as a function of h. As a result, we may apply the *valley-filling function* $\mathcal{V}\{\cdot\}$ to the left-hand side of (4.32), which produces a nondecreasing function by filling in any valleys:

$$\mathcal{V}\{F(h)\} \doteq \sup_{h' \geq h} F(h'). \tag{4.36}$$

This yields

$$\mathcal{V}\left\{\int_{\alpha}^{\beta-h} [p(\theta)+p(\theta+h)]\, \mathrm{P}_{\min}(\theta,\theta+h)\, d\theta\right\} \leq \Pr\left(|f(x)-\theta| \geq \frac{h}{2}\right). \tag{4.37}$$

Using this equation to complete the proof as before, we obtain:

Corollary 4.

$$\mathrm{Err}(f) \geq \int_0^{\beta-\alpha} \mathcal{V}\left\{\int_{\alpha}^{\beta-h} \frac{p(\theta)+p(\theta+h)}{2} \mathrm{P}_{\min}(\theta,\theta+h)\, d\theta\right\} h\, dh. \tag{4.38}$$

This bound will often be strictly tighter [3, 20] than Theorem 4. However, it turns out that for the cases we will consider, it does not affect the computation of the bound [2]. Thus, we will ignore this extension in our computations.

4.4 Computing the Ziv-Zakai Bound

Although the Ziv-Zakai bound as given above applies to any global estimation problem, in practice one usually assumes that $\mathrm{P}_{\min}(\theta,\theta+h) = \mathrm{P}_{\min}(h)$ does not depend on θ. This restriction greatly simplifies the computation of the inner integral of the bound and is met by most problems of interest. In particular, we have seen that this restriction holds for both amplitude and time-delay estimation. The following corollary follows easily from Theorem 4:

Corollary 5. *Let $\Theta = [\alpha,\beta]$ be the parameter space in a global estimation problem, and let f be an estimator of Θ. Assume that the distribution on Θ is uniform, and that $\mathrm{P}_{\min}(\theta,\theta+h)$ does not depend on θ. Then*

$$\mathrm{Err}(f) \geq \int_0^{\beta-\alpha} \frac{\beta-\alpha-h}{\beta-\alpha} \mathrm{P}_{\min}(h) h\, dh. \tag{4.39}$$

4.4.1 Amplitude estimation

One could substitute the computation (4.22) of $\mathrm{P}_{\min}(h)$ for amplitude estimation into the above equation and evaluate the integrals directly. However, we take a more indirect approach. Recall that the ML estimator for amplitude estimation was shown to be efficient; that is, its error is equal to the Cramèr-Rao lower bound for that problem. Furthermore, in the previous chapter we derived Barankin bounds which are at least as tight as the Cramèr-Rao bound, and thus for efficient estimators are equal to the

Cramèr-Rao bound. A natural question to ask, then, is whether this is also the case for the Ziv-Zakai bound. We may answer in the affirmative.

Lemma 12. *The Ziv-Zakai and Cramèr-Rao bounds are equal in the case of amplitude estimation, where the distribution over the parameter space is uniform.*

Proof. Looking back at the proof of the Ziv-Zakai bound, we see that the inequality arises by comparing the error $\mathrm{P}_{\min}(A_0, A_1)$ of the likelihood detector, which is optimal, to the error $\mathrm{P}_f(A_0, A_1)$ of the suboptimal estimation theorist's approach. Thus, it suffices to show that these two detectors have the same probability of error. In particular, we show that for all $A_0, A_1 \in \Theta$, the likelihood detector and estimation theorist's approach produce the same output (that is, choose the same hypothesis) for each $x \in \mathcal{X}$.

Recall that for amplitude estimation, the transmitted signal is $As(t)$, where A, the amplitude, is the parameter value. Fix A_0 and A_1. The MAP estimator (1.12) is

$$f(x) = \frac{1}{E} \int_{-T}^{T} x(t)s(t)\,dt, \tag{4.40}$$

where $E = \int_{-T}^{T} s(t)^2\,dt$. Let E_0 and E_1 be the energy of the corresponding signals; then $E_i = A_i^2 E$ for $i = 1, 2$. The estimation theorist's approach (4.26) is

$$\int_{-T}^{T} x(t)s(t)\,dt \underset{H_0}{\overset{H_1}{\gtrless}} \frac{E(A_0 + A_1)}{2}. \tag{4.41}$$

The likelihood ratio test (4.10) is

$$(A_1 - A_0) \int_{-T}^{T} x(t)s(t)\,dt \underset{H_0}{\overset{H_1}{\gtrless}} \frac{E_1 - E_0}{2}. \tag{4.42}$$

However, $(E_1 - E_0) = E(A_1^2 - A_0^2)$, so $(E_1 - E_0)/(A_1 - A_0) = E(A_1 + A_0)$. Thus, it follows that the detectors (4.41) and (4.42) choose the same hypothesis for all $x \in \mathcal{X}$. We need not worry about which hypothesis is assigned to observations causing the left and right sides of (4.41) or (4.42) to be equal, as it was shown in the derivation of the likelihood ratio test that they do not affect the overall error. □

The author is not aware of any previous computations of the Ziv-Zakai bound for amplitude estimation. Most exposition on the Ziv-Zakai bound has focused on its applications to time-delay and related problems, and connections to the Cramèr-Rao bound have only been demonstrated for those individual problems. An interesting avenue of future research would be to further explore the connections between these two bounds as generalized to arbitrary estimation problems.

4.4.2 Time-Delay Estimation

We now apply the Ziv-Zakai bound to time-delay estimation, and demonstrate the existence of a threshold effect. We will assume that the signal bandwidth β^2 (2.12) is finite. From our previous computations,

$$\mathrm{P}_{\min}(h) = \Phi\left(\sqrt{\frac{E(1-\psi(h))}{N_0}}\right). \tag{4.43}$$

Recall that

$$\psi(h) = \frac{\int_0^\infty 2\cos(\omega h)\left|S(\omega)\right|^2\, d\omega}{\int_0^\infty 2\left|S(\omega)\right|^2\, d\omega}, \tag{4.44}$$

where $S(\omega)$ is the Fourier transform of $s_0(t)$. Since $1-\cos(v) = 2\sin^2(v/2) \leq v^2/2$, it follows that

$$1-\psi(h) \leq \frac{1}{E}\int_0^\infty \left|S(\omega)\right|^2 \omega^2 h^2\, d\omega = \frac{\beta^2 h^2}{2}. \tag{4.45}$$

Since Φ is a decreasing function, we may lower-bound $\mathrm{P}_{\min}(h)$:

$$\mathrm{P}_{\min}(h) \geq \Phi\left(h\beta\sqrt{\frac{E}{2N_0}}\right). \tag{4.46}$$

Applying this to (4.39) produces a lower bound on estimator which is potentially weaker than the original Ziv-Zakai bound, but permits a direct computation:

$$\mathrm{Err}(f) \geq \int_0^T \left(h - \frac{h^2}{T}\right)\Phi\left(h\beta\sqrt{\frac{E}{2N_0}}\right)\, dt. \tag{4.47}$$

We will see that this bound exhibits the threshold effect. Computer simulations have shown that it provides a good characterization of the actual error of the matched-filter estimator at a wide range of signal-to-noise ratio conditions [2].

Integrating by parts, we find that

$$\begin{aligned}
\mathrm{Err}(f) \geq\ & \left[\left(\frac{h^2}{2} - \frac{h^3}{3T}\right)\right]_0^T \\
& + \frac{1}{2\sqrt{\pi}}\beta\sqrt{\frac{E}{N_0}}\int_0^T \left(\frac{h^2}{2} - \frac{h^3}{3T}\right) e^{-h^2\beta^2 E/4N_0}\, dt \qquad (4.48) \\
=\ & \frac{T^2}{6}\Phi\left(T\beta\sqrt{\frac{E}{2N_0}}\right) \\
& + \sqrt{\frac{N_0}{\pi\beta^2 E}}\int_0^{\frac{\beta^2 T^2 E}{4N_0}} \left(\sqrt{\frac{N_0}{\beta^2 E}}v^{1/2} - \frac{4N_0}{3TE\beta^2}v\right) e^{-v}\, dv. \qquad (4.49)
\end{aligned}$$

In order to express this in more digestible terms, define the regularized gamma function by

$$\Gamma_a(z) \doteq \frac{1}{\Gamma(a)} \int_0^z e^{-v} v^{a-1}\, dv, \tag{4.50}$$

where $\Gamma(a)$ is the usual gamma function. In particular, $\Gamma(3/2) = \sqrt{\pi}/2$, and $\Gamma(2) = 1$. Note that $\lim_{z\to\infty} \Gamma_a(z) = 1$, and $\lim_{z\to 0} \Gamma_a(z) = 0$.

Expressing the bound in terms of these functions produces

$$\begin{aligned} \mathrm{Err}(f) \quad \geq \quad & \frac{T^2}{6}\Phi\left(T\beta\sqrt{\frac{E}{2N_0}}\right) + \frac{N_0}{E\beta^2}\Gamma_{3/2}\left(\frac{\beta^2 T^2 E}{4N_0}\right) \\ & - \frac{4}{3T\sqrt{\pi}}\left(\frac{N_0}{\beta^2 E}\right)^{3/2}\Gamma_2\left(\frac{\beta^2 T^2 E}{4N_0}\right). \end{aligned} \tag{4.51}$$

This is the Ziv-Zakai bound for time-delay estimation.

Note that $\lim_{h\to\infty} \Gamma_{3/2}(h) = \lim_{h\to\infty} \Gamma_2(h) = 1$, and $\lim_{h\to\infty} \Phi(h) = 0$. It follows that the ratio of this bound to the Cramèr-Rao bound, which is $N_0/E\beta^2$, converges to 1 as as $E/N_0 \to \infty$; that is, this bound is approximately equal to the Cramèr-Rao bound for high signal-to-noise ratios.

For large noise levels, however, this bound is tighter than the Cramèr-Rao bound. It may be shown using l'Hospital's rule that $\lim_{h\to 0} \Gamma_{3/2}(h)/h = \lim_{h\to 0} \Gamma_2(h)/h^{3/2} = 0$. Furthermore, $\Phi(0) = 1/2$. Therefore, as $E/N_0 \to 0$, this bound converges to $T^2/3$, which is the *a priori* variance of the parameter in the interval $[T/2, T/2]$. This situation occurs because for very large levels of noise, the signal is obscured in the observation, and an estimator can do little better than randomly guess the true parameter.

Conclusion

In this paper, we have presented the Cramèr-Rao, Barankin and Ziv-Zakai lower bounds on estimator error. We have applied these bounds both to the general case of signal parameter estimation, and to the specific cases of amplitude and time-delay estimation. Amplitude estimation is completely characterized by the Cramèr-Rao bound. We have shown that time-delay estimation exhibits a threshold effect since the Barankin and Ziv-Zakai bounds approximate the Cramèr-Rao bound for high signal-to-noise ratios, but are significantly tighter than it at low signal-to-noise ratios.

The Barankin bound is the minimum achievable local error for unbiased estimators. Since it is difficult to compute this bound exactly, we resort to suboptimal bounds which nevertheless have significant theoretical backing. In particular, they approximate the Cramèr-Rao bound for high signal-to-noise ratios. Furthermore, the well-known explanation of the threshold effect in terms of secondary peaks of the autocorrelation function is justified theoretically. However, the bound does not approximate the true error at very low signal-to-noise ratios, as the existence of an approximately unbiased estimator is no longer true.

The Ziv-Zakai bound applies results from detection theory to lower bound the global error of any estimator. Unlike the Barankin bound, it does not extend the Cramèr-Rao bound in a simple, generic fashion. We have showed, however, that the two bounds are equal for amplitude estimation. We have also demonstrated that the Ziv-Zakai bound exhibits a threshold effect for time-delay estimation; at high signal-to-noise ratios it approximately equals the Cramèr-Rao bound, and at low signal-to-noise ratios it reflects a priori knowledge about the parameter.

There are several interesting topics for future research. One might examine whether the results of this paper for the specific case of time-delay estimation may be generalized to arbitrary estimation problems. In particular, we hypothesize that in arbitrary signal parameter estimation problems, the Ziv-Zakai bound is approximately equal to the Cramèr-Rao bound at high signal-to-noise ratios. Developing the theory behind the Ziv-Zakai bound to the level which exists for the Barankin bound would supplement its current widespread use with a deeper understanding of its overall properties. Additionally, in this paper we have only treated the threshold effect superficially; a more thorough treatment might try to quantify the threshold signal-to-noise ratio in terms of properties of the bounds.

Bibliography

[1] E. W. Barankin. Locally best unbiased estimates. *Ann. Math. Statist.*, 20:477–501, 1949.

[2] Kristine L. Bell, Yossef Steinberg, Yariv Ephraim, and Harry L. Van Trees. Extended Ziv-Zakai lower bound for vector parameter estimation. *IEEE Trans. Inform. Theory*, 43(2):624–637, Mar. 1997.

[3] S. Bellini and G. Tartara. Bounds on error in signal parameter estimation. *IEEE Trans. Communications*, 22(3):340–342, Mar. 1974.

[4] Erhan Çinlar. *Introduction to Stochastic Processes.* Prentice-Hall, 1975.

[5] Douglas G. Chapman and Herbert Robbins. Minimum variance estimation without regularity assumptions. *Ann. Math. Statist*, 22(4):581–586, Dec. 1951.

[6] Dan Chazan, Moshe Zakai, and Jacob Ziv. Improved lower bounds on signal parameter estimation. *IEEE Trans. Inform. Theory*, 21(1):90–93, Jan. 1975.

[7] Siu-Kay Chow and Peter M. Schultheiss. Delay estimation using narrow-band processes. *IEEE Trans. Acoust. Speech Signal Process.*, ASSP-29(3):478–484, June 1981.

[8] Frederick E. Glave. A new look at the Barankin bound. *IEEE Trans. Inform. Theory*, IT-18(3):349–356, May 1972.

[9] U. Grenander. Stochastic processes and statistical inference. *Ark. Mat.*, 1:195–277, 1950.

[10] Luc Knockaert. The Barankin bound and threshold behavior in frequency estimation. *IEEE Trans. Signal Process.*, 45(9):2378–2401, Sep. 1997.

[11] V. A. Kotel'nikov. *The Theory of Optimum Noise Immunity.* McGraw-Hill, New York, 1959.

[12] Yu. A. Kutoyants. *Parameter Estimation for Stochastic Processes.* Heldermann, Berlin, 1984.

[13] Robert J. McAulay and Edward M. Hofstetter. Barankin bounds on parameter estimation. *IEEE Trans. Inform. Theory*, IT-17(6):669–676, Nov. 1971.

[14] Robert J. McAulay and Lawrence P. Seidman. A useful form of the Barankin lower bound and its application to PPM threshold analysis. *IEEE Trans. Inform. Theory*, IT-15(2):273–279, Mar. 1969.

[15] H. L. Royden. *Real Analysis*. Prentice Hall, New Jersey, 1988.

[16] P. Swerling. Parameter estimation for waveforms in additive Gaussian noise. *J. Soc. Indust. Appl. Math.*, 7(2):152–166, June 1959.

[17] Joseph Tabrikian and Jeffrey L. Krolik. Barankin bounds for source localization in an uncertain ocean environment. *IEEE Trans. Signal Process.*, 47(11):2917–2927, Nov. 1999.

[18] Petar Todorovic. *An Introduction to Stochastic Processes and Their Applications*. Springer-Verlag, New York, 1992.

[19] Harry L. Van Trees. *Detection, Estimation and Modulation Theory, Part 1*. John Wiley & Sons, Inc., New York, 1968.

[20] Anthony J. Weiss and Ehud Weinstein. Fundamental limitations in passive time delay estimation–part 1: Narrow-band systems. *IEEE Trans. Acoust. Speech Signal Process.*, ASSP-31(2):472–485, Apr. 1983.

[21] Jacob Ziv and Moshe Zakai. Some lower bounds on signal parameter estimation. *IEEE Trans. Inform. Theory*, IT-15(3):386–391, May 1969.