Bounds on the Bias of Signal Parameter Estimators

By JACOB ZIV

(Manuscript received December 20, 1968)

Any estimator which is constrained to take values in a finite range is, in general, biased. Many times the bias is unknown; furthermore, in some cases the bias may become the main contributor to the mean square error of an estimator. This paper derives upper and lower bounds on the bias of a finite-range, signal parameter estimator.

I. INTRODUCTION AND MAIN RESULTS

1.1 *Introduction*

Let the parameter be denoted by a and let a take values in $[-\alpha, \alpha]$. We refer to 2α as the *a priori* range (or space) of a. We assume that there exists probabilistic mapping from the parameter space to an observation space, that is, a probability law that governs the effect of a on the observation.[1] This probability law will be referred to as the "channel." After observing the "outcome" which is a point in the observation space, we estimate the value of a. Let this estimate be denoted by $\hat{a}$. Clearly, $\hat{a}$ is a random variable.

We assume, throughout this paper, that $\hat{a}$ takes values in $[-A, A]$. Let the bias be defined

$$b(a) \triangleq E_a[\hat{a} - a] = \int (\hat{a} - a)\, dp(\hat{a} \mid a) \tag{1}$$

where $p(\hat{a} \mid a)$ is the probability distribution function of $\hat{a}$ given a. Assume that we are now told that the true value of the parameter a is either a_1 or $-a_1$ with equal probabilities. Let H_{a_1} be the hypothesis that $a = a_1$ and let H_{-a_1} be the hypothesis that $a = -a_1$. The minimum probability of error is (dropping the subscript 1 from a_1):

$$P_e\{a, -a\} \triangleq \text{Min}\, \{\tfrac{1}{2}[Pr\, \{a \mid -a\} + \Pr\, \{-a \mid a\}]\}$$

where $\Pr\{a \mid -a\}$ is the probability that the decision will be a, given that $-a$ is transmitted, and where the minimization is carried over all possible decision rules. (A decision rule is a mapping from the observation space to the set $\{-a; a\}$.) Then we show in Section II that

$$\tfrac{1}{2}[b(a) - b(-a)] \leqq -AP_e\{-a; a\} + (A - a); \qquad a \geqq 0; \tag{2a}$$

$$\tfrac{1}{2}[b(a) - b(-a)] \geqq AP_e\{-a; a\} - (A + a); \qquad a \geqq 0. \tag{2b}$$

By equation (2a) we have

$$\tfrac{1}{2}[| b(-a) | + | b(a) |] \geqq AP_e\{-a; a\} - (A - a); \qquad a \geqq 0.$$

Hence, an estimator must have a nonzero bias if

$$\frac{| a |}{A} > 1 - P_e\{-a; a\}. \tag{3}$$

The bounds of equation (2) are the main result of this paper.

If we assume that for any a, $b(a) = -b(-a)$ we have by equation (2) that

$$b(a) \leqq -AP_e\{-a; a\} + (A - a) \tag{4a}$$

and

$$b(a) \geqq AP_e(-a; a) - (A + a). \tag{4b}$$

These bounds are sketched in Fig. 1.

If, in addition, we assume a symmetry around a in the sense that

$$\int_{-A+|a| \leqq \hat{a}-a \leqq A-|a|} (\hat{a} - a)\, dp(\hat{a} \mid a) = 0. \tag{5a}$$

We show (see Section II) that in this case

$$b(a) \geqq -aP_e\{-a; a\}; \qquad -A \leqq a \leqq -\frac{A}{2} \tag{5b}$$

$$(A - a)P_e(-a; a) \geqq b(a) \geqq 0; \qquad -\frac{A}{2} \leqq a \leqq 0 \tag{5c}$$

and

$$b(a) \leqq -aP_e\{-a; a\}; \qquad A \geqq a \geqq \frac{A}{2} \tag{5d}$$

$$-(A + a)P_e(-a; a) \leqq b(a) \leqq 0; \qquad \frac{A}{2} \geqq a \geqq 0. \tag{5e}$$

The bias $b(a)$ is unknown, in general. However, the probability

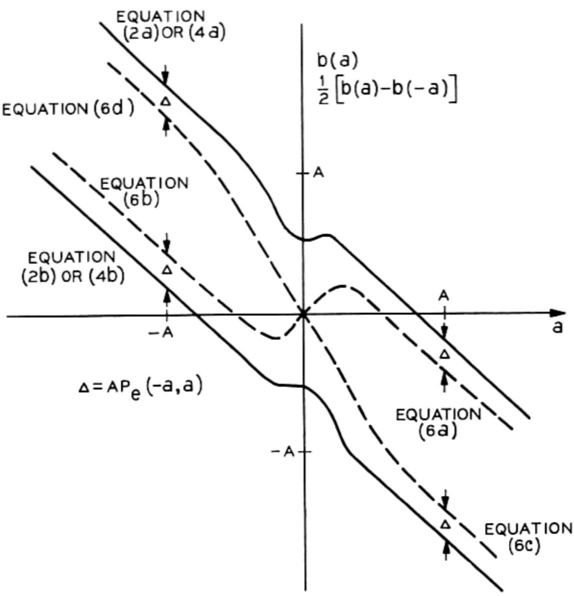

Fig. 1 — Bounds on the bias of an estimator.

$P_e\{a;\ -a\}$ is known for many important cases. The bounds derived here depend on the channel probability law through $P_e(-a;\ a)$ only, and therefore are easy to compute in many cases.

1.2 *Sharpness*

Section III shows that, for one special case, $b(a)$ is given by

$$b(a) = -2AP_e\{-a;a\} + (A - a); \qquad a \geqq 0, \tag{6a}$$

$$b(a) = 2AP_e\{-a;a\} - (A + a); \qquad a \leqq 0. \tag{6b}$$

Section III also shows that, for another special case, $b(a)$ is given by

$$b(a) = 2AP_e\{-a;a\} - (A + a); \qquad a \geqq 0, \tag{6c}$$

$$b(a) = -2AP_e\{-a;a\} + (A - a); \qquad a < 0. \tag{6d}$$

The comparison of equation (6) with the bounds of equations (4) and (2) indicates the degree of sharpness of these bounds (see Fig. 1).

1.3 *Examples*

Let the received message be a sample function of the random process

$$r(t) = s(t - a) + n(t), \tag{7a}$$

where a is an unknown parameter constrained to take values in $(-\alpha;\ \alpha)$. The term $n(t)$ is assumed to be white gaussian noise with (double sided) spectral density N_o.

Let

$$\int_{-\infty}^{\infty} s^2(t)\, dt = E, \tag{7b}$$

$$\rho(2a) = \frac{1}{E}\int_{-\infty}^{\infty} s(t-a)s(t+a)\, dt, \tag{7c}$$

$$q = ([1 - \rho(2a)]E/2N_o)^{\frac{1}{2}}. \tag{7d}$$

Hence, in this case,[2]

$$\begin{aligned} P_e(-a;\ a) &= (2\pi)^{-\frac{1}{2}} \int_q^{\infty} \exp(-x^2/2)\, dx \\ &\geqq (2\pi)^{-\frac{1}{2}} \int_{(E/N_o)^{\frac{1}{2}}}^{\infty} \exp(-x^2/2)\, dx. \end{aligned} \tag{8a}$$

Hence, by equation (3)

$$|\, b(a)\, | > 0 \quad \text{for any} \quad \frac{|\, a\, |}{A} > 1 - (2\pi)^{-\frac{1}{2}} \int_{(E/N_o)^{\frac{1}{2}}}^{\infty} \exp(-x^2/2)\, dx. \tag{8b}$$

Furthermore, if the channel is that of equation (7) and if $\hat{a}$ is a maximum likelihood estimator, then it follows from equation (7) that the conditions of equation (5a) are satisfied, since the maximum likelihood procedure for estimating a is to evaluate

$$\begin{aligned} \lambda(a^*) &= \int_{-\infty}^{\infty} r(t)s(t-a^*)\, dt \\ &= \int_{-\infty}^{\infty} s(t-a)s(t-a^*)\, dt + \int_{-\infty}^{\infty} n(t)s(t-a^*)\, dt \end{aligned}$$

and to set $\hat{a}$ to the value of $a^*(-A \leqq a^* \leqq A)$ for which $\lambda(a^*)$ is maximum. Hence, the statistics of $\lambda(a_1^*)$ are the same as those of $\lambda(a_2^*)$ if $\frac{1}{2}(a_1^* + a_2^*) = a$; also, $b(a) = -b(-a)$. Therefore by equations (5) and (8a),

$$b(a) \geqq -a(2\pi)^{-\frac{1}{2}} \int_{(E/N_o)^{\frac{1}{2}}}^{\infty} \exp(-x^2/2)\, dx; \qquad -A \leqq a \leqq -\frac{A}{2} \tag{9a}$$

$$b(a) \geqq 0; \qquad -\frac{A}{2} \leqq a \leqq 0 \tag{9b}$$

$$b(a) \leqq -a(2\pi)^{-\frac{1}{2}} \int_{(E/N_o)^{\frac{1}{2}}}^{\infty} \exp(-x^2/2)\, dx; \qquad \frac{A}{2} \leqq a \leqq A \tag{9c}$$

$$b(a) \leqq 0; \qquad 0 < a \leqq \frac{A}{2}. \tag{9d}$$

II. DERIVATION OF THE BOUNDS[3]

$$\begin{aligned} b(a) &\triangleq \int (\hat{a} - a)\, dp(\hat{a} \mid a) \\ &= \int_{\hat{a}>0} (\hat{a} - a)\, dp(\hat{a} \mid a) + \int_{\hat{a}\leqq 0} (\hat{a} - a)\, dp(\hat{a} \mid a). \end{aligned} \tag{10}$$

Now,

$$\int_{\hat{a}>0} (\hat{a} - a)\, dp(\hat{a} \mid a) \geqq -a \Pr\{\hat{a} > 0 \mid a\} \tag{11a}$$

$$\int_{\hat{a}>0} (\hat{a} - a)\, dp(\hat{a} \mid a) \leqq (A - a) \Pr\{\hat{a} > 0 \mid a\} \tag{11b}$$

$$\int_{\hat{a}\leqq 0} (\hat{a} - a)\, dp(\hat{a} \mid a) \geqq -(A + a) \Pr\{\hat{a} \leqq 0 \mid a\} \tag{11c}$$

$$\int_{\hat{a}\leqq 0} (\hat{a} - a)\, dp(\hat{a} \mid a) \leqq -a \Pr\{\hat{a} \leqq 0 \mid a\}. \tag{11d}$$

Also, we have that

$$\Pr\{\hat{a} > 0 \mid a\} = 1 - \Pr\{\hat{a} \leqq 0 \mid a\}. \tag{12}$$

Inserting equations (11) and (12) into equation (10), we have

$$b(a) \geqq A \Pr\{\hat{a} > 0 \mid a\} - (A + a) \tag{13}$$

$$b(a) \leqq -A \Pr\{\hat{a} \leqq 0 \mid a\} + (A - a). \tag{14}$$

Consider the following detection problem. Assume that a and $-a$ ($a > 0$) are used as two signals for equiprobable binary signalling; decide on a if $\hat{a} > 0$ and decide on $-a$ if $\hat{a} \leqq 0$. The probability of error associated with this detection procedure is given by

$$P_a \triangleq \tfrac{1}{2} \Pr\{\hat{a} > 0 \mid -a\} + \tfrac{1}{2} \Pr\{\hat{a} \leqq 0 \mid a\}; \qquad a > 0. \tag{15}$$

The error probability P_a is lower bounded by $P_e(-a;\, a)$ which is the probability of error that is associated with the optimal binary detection scheme for this detection problem. In the same way P_a is upper bounded by $1 - P_e(-a;\, a)$.

Hence,

$$1 - P_e(-a; a) \geqq \tfrac{1}{2}[\Pr\{\hat{a} \geqq 0 \mid -a\} + \Pr\{\hat{a} \leqq 0 \mid a\}] \geqq P_e(-a; a); \qquad a \geqq 0. \tag{16}$$

By equations (13), (14), and (16) we get equations (2a) and (2b).

Now, if

$$\int_{-A+|a| \leqq \hat{a}-a \leqq A-|a|} (\hat{a} - a)\, dp(\hat{a} \mid a) = 0$$

then

$$b(a) = \int_{2a+A}^{A} (\hat{a} - a)\, dp(\hat{a} \mid a) \geqq 0; \qquad -A \leqq a \leqq 0. \tag{17a}$$

Hence

$$b(a) > -a \Pr[\hat{a} > 0 \mid a]; \qquad -A \leqq a \leqq -\frac{A}{2} \tag{17b}$$

$$b(a) < (A - a) \Pr[\hat{a} > 0 \mid a]; \qquad -\frac{A}{2} \leqq a \leqq 0. \tag{17c}$$

Also

$$b(a) = \int_{-A}^{2a-A} (\hat{a} - a)\, dp(\hat{a} \mid a) \leqq 0; \qquad A \geqq a \geqq 0. \tag{17d}$$

Hence

$$b(a) < -a \Pr[\hat{a} \leqq 0 \mid a]; \qquad \frac{A}{2} \leqq a \leqq A \tag{17e}$$

$$b(a) > -(A + a) \Pr[\hat{a} \leqq 0 \mid a]; \qquad 0 \leqq a \leqq \frac{A}{2}. \tag{17f}$$

Equation (5) follows from equations (17) and (11).

III. THE SHARPNESS OF THE BOUNDS

In order to check the sharpness of the bounds on b(a), let us discuss the following example.

Let $\hat{a}$ be some estimation of the parameter a.

Let $\hat{\hat{a}}$ be defined as

$$\hat{\hat{a}} = A \qquad \text{if} \qquad \hat{a} > 0$$

$$\hat{\hat{a}} = -A \qquad \text{if} \qquad \hat{a} \leqq 0.$$

Now, regard $\hat{\hat{a}}$ as an estimation of a. The bias of $\hat{\hat{a}}$ is given by

$$b(a) = (A - a) \Pr [\hat{a} > 0 \mid a] + (-A - a)[1 - \Pr (\hat{a} > 0 \mid a]$$
$$= 2A \Pr \{\hat{a} > 0 \mid a\} - (A + a); \qquad a \leqq 0 \tag{18}$$

and also

$$b(a) = -2A \Pr \{\hat{a} \leqq 0 \mid a\} + (A - a); \qquad a > 0. \tag{19}$$

Compare equations (18) and (19) with equations (13) and (14).

In the special case where $\hat{a}$ is a maximum likelihood estimator and the channel is the one given by equation (7), we have that

$$\Pr \{\hat{a} \leqq 0 \mid a\} = \Pr \{\hat{a} > 0 \mid -a\}$$
$$= P_e\{-a; a\}; \qquad a \geqq 0. \tag{20}$$

Inserting equation (20) into equations (18) and (19) yields equation (6a) and (6b). By making $\hat{\hat{a}} = -A$ if $\hat{\hat{a}} > 0$ and $\hat{\hat{a}} = A$ if $\hat{a} \leqq 0$ we get equations (6c) and (6d) in a similar way.

IV. APPLICATIONS

4.1 *Postdetection Integration*

Assume that one makes n independent, equally distributed, estimations of a: $\hat{a}_1, \hat{a}_2, \hat{a}_3, \cdots, \hat{a}_i, \cdots, \hat{a}_n$, and let

$$\tilde{a} = \frac{1}{n} \sum_{i=1}^{n} \hat{a}_i \ ;$$

$\tilde{a}$ is sometimes called the "postdetection estimation of a". Such an estimator appears in many applications: radar range estimation, postdetection diversity combiners in communication systems, and so on. Now

$$\epsilon_a^2 \triangleq E\{(\tilde{a} - a)^2 \mid a\} = E[(\tilde{a} - b(a) - a)^2 \mid a] + b^2(a)$$
$$= \frac{1}{n} \sigma_a^2 + b^2(a)$$

where σ_a^2 is the variance of $\hat{a}_i$ (for any i), given a. Clearly, if the estimator $\hat{a}$ is unbiased, the mean square error that is associated with $\tilde{a}$ can be made arbitrarily small by making n large enough. However, if $\hat{a}$ is biased, then, for any n, the mean square error is lower bounded by

$$\epsilon_a^2 \geqq b^2(a) \geqq b_L^2(a) \tag{21}$$

where $b_L(a)$ is the lower bound on $\mid b(a) \mid$ given by equations (2), (4), or (5).

Example: Let the channel be given by equation (7) and let $\hat{a}$ be a maximum likelihood estimator; then by equations (9a), (9c), and (21)

$$\epsilon_a^2 \geqq a^2 \frac{1}{2\pi} \left[\int_{(E/N_0)^{\frac{1}{2}}}^{\infty} \exp(-x^2/2)\, dx \right]^2; \qquad A \geqq |\, a \,| \geqq \frac{A}{2}.$$

Assume that the *a priori* range of a, is smaller than $(-A, A)$. Then

$$\epsilon_a^2 \geqq \frac{a^2}{2\pi} \left[\int_{(E/N_0)^{\frac{1}{2}}}^{\infty} \exp(-x^2/2)\, dx \right]^2; \qquad \frac{A}{2} \leqq |\, a \,| \leqq \alpha.$$

Now, let

$$\hat{\epsilon}^2 \triangleq \max_a \epsilon_a^2 .$$

Then, unless $A \geqq 2\alpha$ (that is, unless the range of $\hat{a}$ is at least twice as large as the *a priori* range of a), we have that

$$\hat{\epsilon}^2 \geqq \frac{\alpha^2}{2\pi} \left[\int_{(E/N_0)^{\frac{1}{2}}}^{\infty} \exp(-x^2/2)\, dx \right]^2$$

even if $n \to \infty$.

4.2 *Predetection Integration*

Let the channel be the one given by equation (7). Assume that the estimation is based on n repeated measurements; namely, the received signal is given by

$$r(t) = n(t) + \sum_{i=0}^{n-1} s(t - a - i2A).$$

In this case, an estimation is being made only after observing the complete received signal ("predetection integration"). It then follows that for a maximum likelihood estimation of a

$$b^2(a) \geqq a^2 \frac{1}{2\pi} \left[\int_{(nE/N_0)^{\frac{1}{2}}}^{\infty} \exp(-x^2/2)\, dx \right]^2; \qquad \frac{A}{2} \leqq |\, a \,| \leqq A$$

which is the same as for single measurement except for E being replaced by nE. In this case, unlike the previous case, the lower bound vanishes as n tends to infinity.

REFERENCES

1. Van Trees, H. L., "Detection, Estimation and Modulation Theory," Part 1, New York: John Wiley and Sons, 1968, p. 53.
2. Viterbi, A., "Principles of Coherent communication," New York: McGraw-Hill, 1966, Chapter 7.
3. Ziv, J., and Zakai, M., "Some Lower Bounds on Signal Parameter Estimation," IEEE Trans. Information Theory, *IT-15,* No. 3 (May 1969), pp. 386–391.