OPTIMAL MINMAX ESTIMATION AND THE DEVELOPMENT OF MINMAX ESTIMATION ERROR BOUNDS

James C. Preisig

Department of Applied Ocean Physics and Engineering
Woods Hole Oceanographic Institution
Woods Hole, MA 02543

ABSTRACT

It is often desired to create an optimal estimator of some parameter θ given the observation x. However, the relationship between θ and x may depend on another parameter ϕ which is unknown to the processor and not of direct interest. In this case, an estimator which performs well for one value of ϕ may perform poorly for another value of ϕ. One approach to dealing with this problem is to develop an estimator whose worst case performance evaluated over some range of ϕ is as good as possible. Such an optimal minmax estimator is derived. The derivation of this estimator also motivates an approach to developing lower bounds on the minmax estimation error achievable by any estimator.

1. INTRODUCTION

In many random parameter estimation problems, it is desired to develop an estimator to minimize

$$\varepsilon(g) = \mathrm{E}[|\, \theta - g(x)\,|^2],$$

where $\theta \in \mathbb{R}$ is the parameter to be estimated (for simplicity, only the real parameter case is considered), $x \in X$ is the observation, and $g : X \rightarrow \mathbb{R}$ is the estimator. The well-known solution to this problem is the conditional expectation

$$g_{mmse}(x) = \mathrm{E}[\theta \mid x].$$

However, in many scenarios, the relationship between the observation and the signal or parameter to be estimated depends on other parameters which are unknown to the processor and are not of direct interest. These are often referred to as nuisance parameters or environmental parameters and will be denoted by ϕ. For example, in the acoustic array processing problem in the ocean, θ may be the signal emitted by a source at the array focal point as received at a particular array sensor (i.e., the desired signal) and x may be the noisy vector time series received at the entire array. In this case, the relationship between the desired signal and the received signal depends on the propagation characteristics of the ocean environment between the source and the array. Thus, an estimator which may yield good results for one value of the environmental parameters may yield poor performance for another value of the environmental parameters.

One approach to dealing with the problem of unknown environmental parameters is to develop an estimator which minimizes the worst case conditional estimation error, where the estimation error is conditioned on the environmental parameter ϕ and the worst case is evaluated over the range of environmental conditions in which the processor is designed to operate. This range is denoted by the set Φ (again for simplicity, Φ is assumed to be a finite set $\Phi = \{\phi_1, \phi_2, \ldots, \phi_K\}$). This estimator, referred to as the optimal minmax estimator, is defined as

$$g_{opt} = \arg \min_{g: X \rightarrow \mathbb{R}} \max_{\phi \in \Phi} \varepsilon(g, \phi), \qquad (1)$$

where

$$\varepsilon(g, \phi) = \mathrm{E}[|\, \theta - g(x)\,|^2 \,|\, \phi].$$

Poor and others [1, 2, 3, 4] have developed approaches to solving this minmax estimation problem for the class of problems for which the solution is a saddlepoint of the error function in the $(g\ ,\ \phi)$ space. The work present here is applicable to a much broader class of problems and has two purposes. The first is to develop a general expression for the optimal minmax estimator, (i.e. the solution to (1)). The second purpose is to present an approach to developing lower bounds on the extremal (maximum) estimation error achievable by any processor. That is, an approach to developing lower bounds over g on

$$\triangle(g) = \max_{\phi \in \Phi} \varepsilon(g, \phi)$$

is sought, where g is any function mapping the observation space into the real numbers.

$\triangle(g)$ is referred to as the extremal value given the processor g, the values of the environmental parameter ϕ for which $\varepsilon(g, \phi) = \triangle(g)$ are referred to as extremal points, and the set of extremal points for a given processor is denoted by $M(g) = \{\phi \in \Phi \mid \varepsilon(g, \phi) = \triangle(g)\}$.

2. THE OPTIMAL MINMAX ESTIMATOR

The optimal minmax estimator is developed by artifically assigning a probability mass function, $\underline{p}$, to the environmental parameter set Φ (i.e., $p_i = \mathrm{Prob}[\overline{\phi} = \phi_i]$). It is then shown that, for a particular pmf referred to the least favorable pmf ($\underline{p}_{lf}$), the minimum mean-squared error estimator

This work was done while the author was with the RLE Digital Signal Processing Group at the Massachusetts Institute of Technology. It was supported in part by the U.S. Navy-Office of Naval Research under Grant No. N00014-91-J-1628 and in part by a General Electric Foundation Graduate Fellowship in Electrical Engineering.

equals the optimal minmax estimator. The application of this approach to the development of optimal minmax estimators is discussed more fully in [5].

For any pmf, $\underline{p}$, and estimator, g, the mean squared estimation error is given by

$$\begin{aligned} \varepsilon(g,\underline{p}) &= \mathrm{E}[\mid \theta - g(x) \mid^2] \\ &= \sum_{i=1}^{K} p_i \mathrm{E}[\mid \theta - g(x) \mid^2 \mid \phi = \phi_i] \\ &= \sum_{i=1}^{K} p_i \varepsilon(g, \phi_i). \end{aligned} \tag{2}$$

$\varepsilon(g,\underline{p})$ is the average of $\varepsilon(g,\phi)$ taken over all ϕ. Therefore, for any pmf, $\underline{p}$,

$$\varepsilon(g,\underline{p}) \leq \max_{\phi\in\Phi} \varepsilon(g,\phi) = \Delta(g). \tag{3}$$

Taking the minimum of both sides of (3) over all possible estimators yields

$$\min_{g:X\to\mathbb{R}} \varepsilon(g,\underline{p}) \leq \min_{g:X\to\mathbb{R}} \max_{\phi\in\Phi} \varepsilon(g,\phi) = \min_{g:X\to\mathbb{R}} \Delta(g). \tag{4}$$

Let $g_{mmse}(x,\underline{p})$ be the minimum mean-squared error estimator of θ given the pmf $\underline{p}$ on ϕ. That is,

$$\begin{aligned} g_{mmse}(x,\underline{p}) &= \arg \min_{g:X\to\mathbb{R}} \varepsilon(g,\underline{p}) \\ &= \sum_{i=1}^{K} p_i \mathrm{E}[\theta \mid x, \phi = \phi_i] \end{aligned} \tag{5}$$

Then (4) implies that

$$\varepsilon(g_{mmse}(x,\underline{p}),\underline{p}) \leq \min_{g:X\to\mathbb{R}} \Delta(g). \tag{6}$$

Suppose that a pmf, $\underline{\tilde{p}}$, could be found such that

$$\varepsilon(g_{mmse}(x,\underline{\tilde{p}}),\underline{\tilde{p}}) = \Delta(g_{mmse}(x,\underline{\tilde{p}})). \tag{7}$$

Then, evaluating (6) at $\underline{p} = \underline{\tilde{p}}$ and noting that

$$\min_{g:X\to\mathbb{R}} \Delta(g) \leq \Delta(g_{mmse}(x,\underline{\tilde{p}})),$$

yields

$$\varepsilon(g_{mmse}(x,\underline{\tilde{p}}),\underline{\tilde{p}}) \leq \min_{g:X\to\mathbb{R}} \Delta(g) \leq \Delta(g_{mmse}(x,\underline{\tilde{p}})). \tag{8}$$

Combining (7) and (8) yields

$$\varepsilon(g_{mmse}(x,\underline{\tilde{p}}),\underline{\tilde{p}}) = \min_{g:X\to\mathbb{R}} \Delta(g).$$

Therefore, if such a distribution could be found, the solution to (1) would be

$$g_{opt}(x) = g_{mmse}(x,\underline{\tilde{p}}). \tag{9}$$

For some insight into how such a distribution can be found, consider the following example:

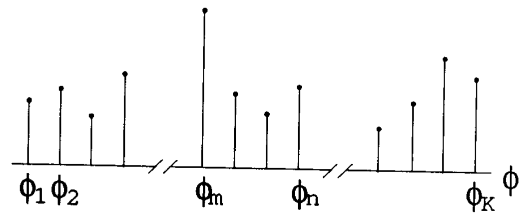

Figure 1: $\varepsilon(g_{mmse}(x,\underline{p}_o),\phi)$ vs ϕ

Assume that ϕ is a real, scalar variable and for some pmf $\underline{p}_o$, $\varepsilon(g_{mmse}(x,\underline{p}_o),\phi)$ is as shown in Figure 1. In this case,

$$\Delta(g_{mmse}(x,\underline{p}_o)) = \varepsilon(g_{mmse}(x,\underline{p}_o),\phi_m).$$

Then, if an extremal point $\phi_m \in M(g_{mmse}(x,\underline{p}_o))$ and any non-extremal point $\phi_n \notin M(g_{mmse}(x,\underline{p}_o))$ for which $p_{o_n} > 0$ are selected, an incremental increase in p_{o_m} and an incremental decrease in p_{o_n} can be made. The new pmf, $\underline{p}(\delta)$, is given by

$$\begin{aligned} p_m(\delta) &= p_{o_m} + \delta, \\ p_n(\delta) &= p_{o_n} - \delta, \\ p_i(\delta) &= p_{o_i} \quad i \notin \{m,n\}, \end{aligned}$$

for an arbitrarily small, positive $\delta \in \mathbb{R}$.

The effects of this incremental change which are of principal interest are the differences between $\varepsilon(g_{mmse}(x,\underline{p}_o),\phi_m)$ and $\varepsilon(g_{mmse}(x,\underline{p}(\delta)),\phi_m)$, between $\varepsilon(g_{mmse}(x,\underline{p}_o),\phi_n)$ and $\varepsilon(g_{mmse}(x,\underline{p}(\delta)),\phi_n)$, between $\Delta(g_{mmse}(x,\underline{p}_o))$ and $\Delta(g_{mmse}(x,\underline{p}(\delta)))$, and finally between $\varepsilon(g_{mmse}(x,\underline{p}_o),\underline{p}_o)$ and $\varepsilon(g_{mmse}(x,\underline{p}(\delta)),\underline{p}(\delta))$.

From (5), $g_{mmse}(x,\underline{p})$ is the weighted sum of conditional expectations and the weighting is determined by the value of $\underline{p}$ for each value of ϕ. Therefore, $g_{mmse}(x,\underline{p}(\delta))$ has a larger weight on $\mathrm{E}[\theta \mid x, \phi = \phi_m]$ than does $g_{mmse}(x,\underline{p}_o)$ and has a smaller weight on $\mathrm{E}[\theta \mid x, \phi = \phi_n]$ than does $g_{mmse}(x,\underline{p}_o)$. Since $\mathrm{E}[\theta \mid x, \phi = \phi_i]$ is the estimator which minimizes $\varepsilon(g,\phi_i)$, it is therefore reasonable to expect that

$$\varepsilon(g_{mmse}(x,\underline{p}(\delta)),\phi_m) < \varepsilon(g_{mmse}(x,\underline{p}_o),\phi_m)$$

and

$$\varepsilon(g_{mmse}(x,\underline{p}(\delta)),\phi_n) > \varepsilon(g_{mmse}(x,\underline{p}_o),\phi_n).$$

Therefore, since the error at the extremal point is reduced, the extremal value will be reduced. That is,

$$\Delta(g_{mmse}(x,\underline{p}(\delta))) < \Delta(g_{mmse}(x,\underline{p}_o)).$$

Finally, from (2) $\varepsilon(g,\underline{p})$ is the weighted sum of the conditional mean-squared estimation errors $\varepsilon(g,\phi_i)$ and, as before, the weighting is determined by the value of $\underline{p}$ for each value of ϕ. Therefore, since

$$\varepsilon(g_{mmse}(x,\underline{p}_o),\phi_m) > \varepsilon(g_{mmse}(x,\underline{p}_o),\phi_n)$$

and the weighting on ϕ_m is increased and the weighting on ϕ_n is decreased, it is reasonable to expect that

$$\varepsilon(g_{mmse}(x, \underline{p}(\delta)), \underline{p}(\delta)) \; > \; \varepsilon(g_{mmse}(x, \underline{p}_o), \underline{p}_o).$$

If fact, it can be shown [5] that

$$\frac{\partial \varepsilon(g_{mmse}(x, \underline{p}(\delta)), \underline{p}(\delta))}{\partial \delta} \Big|_{\delta=0} \; = \; \varepsilon(g_{mmse}(x, \underline{p}_o), \phi_m) - \varepsilon(g_{mmse}(x, \underline{p}_o), \phi_n).$$

Since ϕ_m is an extremal point and ϕ_n is not,

$$\frac{\partial \varepsilon(g_{mmse}(x, \underline{p}(\delta)), \underline{p}(\delta))}{\partial \delta} \Big|_{\delta=0} > 0.$$

Therefore, for each incremental change the mean-squared estimation error averaged over the all ϕ will increase.

This process of increasing p_m for extremal points ϕ_m and decreasing p_n for non-extremal points ϕ_n can continue until p_n equals zero for all non-extremal points. At each step, the extremal value $\Delta(g_{mmse}(x, \underline{p}))$ will decrease and the minimum mean-squared estimation error $\varepsilon(g_{mmse}(x, \underline{p}), \underline{p})$ will increase.

When the probabilities assigned to the non-extremal points all equal zero, any further incremental changes in the assigned probabilities require that the probability assigned to an extremal point be decreases while that assigned to some other point (extremal or non-extremal) be increased. This will result in no change or a reduction in $\varepsilon(g_{mmse}(x, \underline{p}), \underline{p})$. Therefore, the pmf for which the probability assigned to each non-extremal points equals zero ($\underline{p}_{lf}$) is that which maximizes the minimum mean-squared estimation error. Thus, $\underline{p}_{lf}$ is referred to as the least-favorable pmf. Furthermore, since p_{lf_n} equals zero for all non-extremal points, the mean-squared estimation error averaged over the environmental parameters will equal the extremal value of the conditional mean-squared estimation errors. That is,

$$\varepsilon(g_{mmse}(x, \underline{p}_{lf}), \underline{p}_{lf}) \; = \; \Delta(g_{mmse}(x, \underline{p}_{lf})). \qquad (10)$$

Therefore, $\underline{p}_{lf}$ satisfies (7). Then, by (9),

$$g_{opt}(x) \; = \; g_{mmse}(x, \underline{p}_{lf}). \qquad (11)$$

The insight brought out by this example is formalized in the following theorem, a proof of which is contained in [5] where it is referred to as Theorem 6.

Theorem 1 *Let* $\Phi = \{\phi_1, \ldots, \phi_K\}$ *and*

$$P = \{\underline{p} \in \mathbb{R}^K \mid \underline{p} \geq 0 \text{ and } \underline{e}^t \, \underline{p} = 1\}$$

where $\underline{e}$ *is the column vector of all ones. Let* $\varepsilon(g, \phi) = \mathrm{E}[(\theta - g(x))^2 \mid \phi]$ *and let* $g_{mmse} : X \times P \to \mathbb{R}$ *be given by*

$$g_{mmse}(x, \underline{p}) \; = \; \arg \min_{g: X \to \mathbb{R}} \sum_{i=1}^{K} p_i \, \varepsilon(g, \phi_i) \; = \; \frac{\sum_{i=1}^{K} p_i \, p_{x|\phi}(x \mid \phi_i) \mathrm{E}[\theta \mid x, \phi_i]}{\sum_{i=1}^{K} p_i \, p_{x|\phi}(x \mid \phi_i)},$$

where $p_{x|\phi}(x \mid \phi_i)$ *is the conditional pdf or pmf of the observation* x *given that the environmental parameter* $\phi = \phi_i$.

Let the least favorable pmf $\underline{p}_{lf} \in P$ *be defined as*

$$\underline{p}_{lf} \; = \arg \max_{\underline{p} \in P} \sum_{i=1}^{K} p_i \; \varepsilon(g_{mmse}(x, \underline{p}), \phi_i).$$

Then

$$g_{mmse}(x, \underline{p}_{lf}) \; = \; g_{opt}(x) \; = \; \arg \min_{g: X \to \mathbb{R}} \max_{\phi \in \Phi} \varepsilon(g, \phi).$$

It is interesting to note that the pmf which yields the optimal minmax estimator (i.e., the least favorable pmf) is non-zero for only those environmental parameter values ϕ which are extremal points. This contrasts sharply with the commonly used method of dealing with environmental (nuisance) parameters by assigning a uniform pmf to the parameter values and then using an optimal estimator for that pmf.

3. MINMAX ESTIMATION ERROR BOUNDS

While the implementation of the optimal minmax estimator (i.e., the minimum mean-squared error estimator for the least favorable pmf) is often impractical for computational reasons, the form of this estimator does motivate an approach to developing lower bounds on

$$\Delta(g) \; = \; \max_{\phi \in \Phi} \; \varepsilon(g, \phi).$$

From (6), it is clear that for any pmf assigned to the environmental parameter ϕ, any lower bound on the mean-squared estimation error for the parameter θ is also a lower bound on $\Delta(g)$. However, this fact is not very useful unless for some pmf, the mean-squared estimation error bound is a reasonably tight lower bound on $\Delta(g)$. The pmf which should yield a reasonably tight bound is the least favorable pmf $\underline{p}_{lf}$.

From (10) and (11),

$$\varepsilon(g_{mmse}(x, \underline{p}_{lf}), \underline{p}_{lf}) \; = \; \min_{g: X \to \mathbb{R}} \Delta(g).$$

Therefore, the minimum mean-squared estimator error for θ given the pmf $\underline{p}_{lf}$ is an achievable lower bound on $\Delta(g)$. However, this lower bound may be difficult to evaluate. Therefore, it may be preferable to use another mean-squared estimation error bound (e.g., the Cramer-Rao bound, the Weiss-Weinstein bound, etc.) to bound $\Delta(g)$.

Assume that the Cramer-Rao bound is used and denote the bound for any pmf $\underline{p}$ by $\mathrm{CR}(\underline{p})$. Then, $\mathrm{CR}(\underline{p}_{lf})$ will be as tight a bound of $\Delta(g)$ as it is for the mean-squared estimation error $\varepsilon(g, \underline{p}_{lf})$. $\underline{p}_{lf}$ is the pmf which maximizes the minimum achievable mean-squared estimation error. However, it may not maximize the Cramer-Rao bound. Therefore, if the least favorable bounding pmf is defined by

$$\underline{p}_{lfb} \; = \; \arg \max_{\underline{p} \in P} \mathrm{CR}(\underline{p}),$$

$CR(\underline{p}_{lfb})$ will be at least as tight a bound on $\Delta(g)$ as the Cramer-Rao bound is on the mean-squared estimation error. In fact, $CR(\underline{p}_{lfb})$ may be a tighter bound on $\Delta(g)$ than the Cramer-Rao bound is on the mean-squared estimation error.

This procedure can be applied using any mean-squared estimation error bound. Choose the bound and maximize the bound by adjusting the pmf assigned to ϕ. This will yield a minmax estimation error bound which is at least as tight as the mean-squared estimation error bound.

4. CONCLUSIONS

The optimal minmax estimator has been derived using the framework of the least favorable pmf. While this estimator is often impractical to implement, it does motivate the development of an approach to computing reasonable lower bounds on the achievable minmax estimation error. This framework can also be used to develop optimal minmax estimators of restricted form (e.g., linear estimators) for certain classes of problems [5].

REFERENCES

[1] K.S. Vastola, H.V. Poor, "Robust Wiener-Kolmogorov Theory", *IEEE Trans. Inform. Theory*, IT-30, no. 2, March 1984, pp. 316-327.

[2] S. Verdu, H.V. Poor, "On Minmax Robustness: A General Approach and Applications", *IEEE Trans. Inform. Theory*, IT-30, no. 2, March 1984, pp. 328-340.

[3] H.V. Poor, "Uncertainty Tolerance in Underwater Acoustic Signal Processing", *IEEE J. Oceanic Eng.*, OE-12, no. 1, January 1987, pp. 48-65.

[4] S.A. Kassam, T.L. Lim, "Robust Wiener Filters", *Journal of the Franklin Institute*, Vol. 304, No. 4/5, October/November 1977, pp. 171-185.

[5] J.C. Preisig, "*Adptive Matched Field Processing in an Uncertain Propagation Environment*", Ph.D. Thesis, Massachusetts Institute of Technology, 1992.