MEAN SQUARE ERROR PERFORMANCE OF SAMPLE MEAN AND SAMPLE MEDIAN ESTIMATORS

Adrià Gusi-Amigó, Pau Closas[*]

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Spain

Luc Vandendorpe, [†]

Institute of Information and Communication Technologies, Electronics and and Applied Mathematics (ICTEAM)
Belgium

ABSTRACT

Based on the Ziv-Zakai methodology to bound estimators, we derived an estimation bound able to predict the mean square error degradation due to model mismatches. In this article, we build upon this result to provide a performance comparison between mean and median estimators in the presence of outliers. The latter is well known to be statistically more robust than the mean in the presence of outliers. Here we show this superiority by comparing their theoretical error bounds. Analytical results are obtained, which are validated by computer simulations.

***Index Terms*—** MSE, sample median, robust statistics, model mismatch, Ziv-Zakai.

1. INTRODUCTION

Outliers can easily affect the performance of classical ML estimators when the presence of such atypical observations is unknown [1]. In this article we are interested in obtaining Mean Square Error (MSE) expressions for the canonical problem of estimating a deterministic, unknown parameter θ that is buried in noise. Under the assumption of perfect model knowledge, the optimal estimator (in the maximum likelihood (ML) sense) can be readily obtained as a weighted sum of the observations [2, 3]. The MSE of such estimator follows the theoretical estimation bounds, which can be derived following the Cramér-Rao Bound (CRB) or Ziv-Zakai Bound (ZZB) methodologies for instance [4]. However, MSE bounds for the case of wrongly specified models is a problem which is far to be solved.

In this contribution, we provide a MSE expression of the ML estimator under observation outliers. That is to say, when with some probability, the observations can be contaminated by larger errors than expected. As a consequence, an expression for the particular case of the sample mean is also derived and compared to the sample median MSE performance. In both cases, closed-form expressions are obtained and discussed. Section 2 formulates the problem. Sections 3 and 3.1 contain the main results and derivations. Section 4 discusses some computer simulations that validate the theoretical results.

2. SYSTEM MODEL

The unknown scalar random parameter of interest is denoted by $\theta \in \mathbb{R}$, and has a uniform probability density function (pdf) $p_\theta(\theta)$ in the interval $[0, T]$. The observations consist of a random $K \times 1$ vector $\mathbf{x}$ which is parametrized by the scalar θ. Let us consider that following observations

$$\mathbf{x}_* = \mathbf{h}_*\theta + \mathbf{n}_* \tag{1}$$

where $\mathbf{h}_*$ is a known vector and $\mathbf{n}_*$ is the correctly specified noise with the following Gaussian mixture distribution

$$p_*(\mathbf{n}_*) = \omega_1 p_{1*}(\mathbf{n}_*) + \omega_2 p_{2*}(\mathbf{n}_*) \tag{2}$$

where ω_1 and ω_2 are positive numbers and are such that $\omega_1 + \omega_2 = 1$, and $p_{1*}(\mathbf{n}_*)$ and $p_{2*}(\mathbf{n}_*)$ are multivariate Gaussian distributions with zero mean ($\boldsymbol{\mu}_{1*} = \mathbf{0}$, and $\boldsymbol{\mu}_{2*} = \mathbf{0}$) and covariance matrices $\boldsymbol{\Sigma}_{1*} = \sigma^2_{1*}\mathbf{I}$ and $\boldsymbol{\Sigma}_{2*} = \sigma^2_{2*}\mathbf{I}$, such that $\sigma^2_{2*} > \sigma^2_{1*}$, where $\mathbf{I}$ denotes the identity matrix. The distribution (2) is typically used to model noisy observations with outliers. In this case, the first distribution models the thermal noise and the other, with larger variance, the contribution of the outliers which occur with probability ω_2. We can define the correct statistical model as the set $\mathcal{M}_* = \{\mathbf{x}_* \sim p_*(\mathbf{x}|\theta) \; : \; \theta \in \mathbb{R}\}$, where $p_*(\mathbf{x}_*|\theta)$ is the pdf of the observations $\mathbf{x}_*$ as a function of the unknown parameter θ. On the other hand, consider the following assumed model for the observations

$$\mathbf{x} = \mathbf{h}_*\theta + \mathbf{n} \tag{3}$$

where $\mathbf{n}$ is a zero mean Gaussian noise process with covariance matrix $\boldsymbol{\Sigma}_{1*}$. Note that this model considers only the

[*]This work has been partially supported by the Spanish Ministry of Economy and Competitiveness through project TEC2015-69868-C2-2-R (ADVENTURE) and by the Government of Catalonia under Grant 2014–SGR–1567.

[†]The author would like to thank BELSPO for funding the IAP BESTCOM network

presence of thermal noise by means of $\mathbf{\Sigma}_{1*}$, also present in (1), and does not incorporate the appearance of outliers. The pdf for the assumed distributions is denoted as $p(\mathbf{x}|\theta)$ and the assumed statistical model is therefore defined as $\mathcal{M} = \{\mathbf{x} \sim p(\mathbf{x}|\theta) \; : \; \theta \in \mathbb{R}\}$.

With the different observation models already defined, the estimators of θ that will be analyzed can be introduced. Let us consider first a ML estimator for the true observations in (1) with instantaneous knowledge of the noise distribution (either $p_{1*}(\mathbf{n}_*)$ or $p_{2*}(\mathbf{n}_*)$) found in a given sample. Note that for this particular case of Gaussian mixture, where $p_{1*}(\mathbf{n}_*)$ and $p_{2*}(\mathbf{n}_*)$ are zero mean, uncorrelated and identically distributed, one can construct a diagonal matrix $\mathbf{\Sigma}_c$ containing the corresponding values of σ_{1*}^2 or σ_{2*}^2 along its diagonal. The estimator of θ with the instantaneous information $\mathbf{\Sigma}_c$ can be expressed as

$$\hat{\theta}_{\text{ML}}(\mathcal{M}_*) = \frac{\mathbf{h}_*^\top \mathbf{\Sigma}_c^{-1} \mathbf{x}_*}{\mathbf{h}_*^\top \mathbf{\Sigma}_c^{-1} \mathbf{h}_*}. \tag{4}$$

On the other hand, we can consider a ML estimator for the assumed model in (3). This estimator leaves out the contribution of the outliers and can be expressed as

$$\hat{\theta}_{\text{ML}}(\mathcal{M}) = \frac{\mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1} \mathbf{x}_*}{\mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1} \mathbf{h}_*} = \frac{\mathbf{h}_*^\top \mathbf{x}_*}{\mathbf{h}_*^\top \mathbf{h}_*}. \tag{5}$$

Last, we consider the sample median estimator, that is, the numerical value that separates the higher half of the observations from the lower half. The sample median is a known robust estimator that is not as much influenced by outliers in certain conditions [5]. The median is applicable with unidimensional random variables and thus its use makes sense when $\mathbf{h}_* = \mathbf{1}$ holds, where $\mathbf{1}$ is the $K \times 1$ all-ones vector. We define the sample median estimator as

$$\hat{\theta}_{\text{med}} = \text{median}\left(\mathbf{x}_*\right). \tag{6}$$

The ZZB for the ML estimator with instantaneous information in (4) is given by

$$\text{ZZB}(\mathcal{M}_*) = \frac{1}{\mathbf{h}_*^\top \mathbf{\Sigma}_c^{-1} \mathbf{h}_*}. \tag{7}$$

Note that in the case of amplitude estimation with a parameter being uniformly distributed, the ZZB is equivalent to the CRB [6], which has a well known expression [3, 4].

Regarding the sample median estimator, for a random variable x with pdf $f_x(x)$ and median ξ, it is known that the distribution of its estimates converges asymptotically to a normal distribution with mean $\mu_{\text{med}} = \xi$ and variance $\sigma_{\text{med}}^2 = \frac{1}{4 f_x(\mu_{\text{med}})^2 K}$, if $f_x(\mu_{\text{med}}) \neq 0$ and $f_x'(x)$ is continuous in the neighborhood of $x = \xi$ [7]. Given that the noise process in (2) is i.i.d. and assuming that $\mathbf{h}_* = \mathbf{1}$, we have that $\text{median}\left(\mathbf{x}_*\right) = \theta$ asymptotically, and that

$$\text{MSE}(\hat{\theta}_{\text{med}}) = \frac{1}{4\left(\frac{\omega_1}{\sigma_{1*}\sqrt{2\pi}} + \frac{1-\omega_1}{\sigma_{2*}\sqrt{2\pi}}\right)^2 K}. \tag{8}$$

3. ZIV ZAKAI BOUND FOR A MISMATCHED ML ESTIMATOR

The scope here is to provide a lower bound for the mismatched ML estimator $\hat{\theta}_{\text{ML}}(\mathcal{M})$ in (5). The ZZB provides a bayesian bound on the MSE over the *a priori* distribution of the parameter. The bound was first derived in [8] for scalar parameters and subsequently adapted to vector parameters in [9]. We are interested in a lower bound on the mean square estimation error of the time delay

$$\mathbb{E}_*\{\epsilon_*^2\} = \mathbb{E}_*\{\left(\hat{\theta} - \theta\right)^2\} \tag{9}$$

where $\mathbb{E}_*\{\cdot\}$ denotes the expected value with respect to the true observations $\mathbf{x}_*$ in (1) and θ. $\hat{\theta}$ is an estimator of θ. The ZZB can be obtained from the identity

$$\mathbb{E}_*\{\epsilon_*^2\} = \frac{1}{2}\int_0^\infty \mathbb{P}\left(|\epsilon_*| \geq \frac{h}{2}\right) h \, dh \tag{10}$$

and lower bounding $\mathbb{P}\left(|\epsilon_*| \geq \frac{h}{2}\right)$, given that h and $\mathbb{P}\left(|\epsilon_*| \geq \frac{h}{2}\right)$ are nonnegative. The expression $\mathbb{P}\left(|\epsilon_*| \geq \frac{h}{2}\right)$, with $h \geq 0$, is related to a binary detection scheme with equally probable hypotheses

$$\begin{aligned} &\mathcal{H}_1 : \theta = \theta_o; &&\mathbf{x}_* = \mathbf{h}_*\theta + \mathbf{n}_*|\theta = \theta_o \\ &\mathcal{H}_2 : \theta = \theta_o + h; &&\mathbf{x}_* = \mathbf{h}_*\theta + \mathbf{n}_*|\theta = \theta_o + h. \end{aligned} \tag{11}$$

when considering a suboptimal decision scheme where the parameter is first estimated and a nearest-neighbor decision is made afterward

$$\hat{\mathcal{H}} = \begin{cases} \mathcal{H}_1, & \text{if } \hat{\tau} \leq \theta_o + \frac{h}{2} \\ \mathcal{H}_2, & \text{if } \hat{\tau} > \theta_o + \frac{h}{2}. \end{cases} \tag{12}$$

The probability of error for this suboptimum detector can be lower bounded by the minimum error probability $\mathbb{P}_{e_*}(\theta_o, \theta_o + h)$ given by the likelihood ratio (LR) test

$$\Lambda_* = \frac{p_*(\mathbf{x}_*|\theta_o)}{p_*(\mathbf{x}_*|\theta_o + h)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{13}$$

The term $\mathbb{P}\left(|\epsilon| \geq \frac{h}{2}\right)$ can be shown [9] to be greater or equal to

$$\int_{-\infty}^{\infty} \left(p_\theta\left(\theta_o\right) + p_\theta\left(\theta_o + h\right)\right) \mathbb{P}_{e_*}(\theta_o, \theta_o + h) d\theta_o. \tag{14}$$

Given that $p_\theta\left(\theta\right)$ follows a uniform distribution in the interval $[0, T]$, the lower bound on the estimation error can then be expressed as

$$\mathbb{E}\{\epsilon_*^2\} \geq \text{ZZB}_* = \frac{1}{T}\int_0^T h \int_0^{T-h} \mathbb{P}_{e_*}(\theta_o, \theta_o + h) d\theta_o dh. \tag{15}$$

Moreover, when $\mathbb{P}_{e_*}(\theta_o, \theta_o + h)$ is independent of θ_o we can write $\mathbb{P}_{e_*}(h)$ instead. Assuming this, the ZZB is given by

$$\text{ZZB}_* = \frac{1}{T}\int_0^T h(T-h)\mathbb{P}_{e_*}(h)dh. \quad (16)$$

With $\hat{\theta}_{\text{ML}}(\mathcal{M})$ shown in (5), we are then interested in a lower bound for

$$\mathbb{E}_*\{\epsilon^2\} = \mathbb{E}_*\{\left(\hat{\theta}_{\text{ML}}(\mathcal{M}) - \theta\right)^2\}. \quad (17)$$

Note that the expectation is still with respect to the true model of the observations. The corresponding binary detection problem can now be expressed as

$$\begin{aligned} &\mathcal{H}_1 : \theta = \theta_o; \qquad \mathbf{x}_* = \mathbf{h}_*\theta + \mathbf{n}|\theta = \theta_o \\ &\mathcal{H}_2 : \theta = \theta_o + h; \quad \mathbf{x}_* = \mathbf{h}_*\theta + \mathbf{n}|\theta = \theta_o + h \end{aligned} \quad (18)$$

and the error probability for this suboptimum detection scheme can be lower bounded by the mismatched likelihood ratio (MLR) test [10]

$$\Lambda = \frac{p(\mathbf{x}_*|\theta_o)}{p(\mathbf{x}_*|\theta_o + h)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} 1. \quad (19)$$

From the MLR test we can compute the minimum probability of error $\mathbb{P}_e(\theta_o, \theta_o + h)$ for the mismatched ML estimator. This probability can replace $\mathbb{P}_{e_*}(\theta_o, \theta_o + h)$ in (15) or (16) to obtain the corresponding bound.

3.1. Derivation of the error probability

The minimum error probability for the MLR test in (19) can be computed as

$$\begin{aligned} \mathbb{P}_e(\theta_o, \theta_o + h) &= \mathbb{P}(\ln\Lambda < 0|\mathcal{H}_1)\,\mathbb{P}(\mathcal{H}_1) \\ &\quad + \mathbb{P}(\ln\Lambda > 0|\mathcal{H}_2)\,\mathbb{P}(\mathcal{H}_2) \\ &= \frac{1}{2}\mathbb{P}(\ln\Lambda < 0|\mathcal{H}_1) \\ &\quad + \frac{1}{2}\mathbb{P}(\ln\Lambda > 0|\mathcal{H}_2), \end{aligned} \quad (20)$$

where equally likely hypotheses are assumed for the second equality. The remaining probabilities can be obtained as

$$\mathbb{P}(\ln\Lambda < 0|\mathcal{H}_1) = \mathbb{P}(\mathcal{L}(\theta_o) - \mathcal{L}(\theta_o + h) < 0|\theta = \theta_o) \quad (21)$$

$$\mathbb{P}(\ln\Lambda > 0|\mathcal{H}_2) = \mathbb{P}(\mathcal{L}(\theta_o) - \mathcal{L}(\theta_o + h) > 0|\theta = \theta_o + h). \quad (22)$$

where $\mathcal{L}(\theta) = \ln p(\mathbf{x}_*|\theta)$ is the log-likelihood function of θ neglecting the irrelevant constant terms. Evaluating the log-likelihood at θ_o, we have that

$$\begin{aligned} \mathcal{L}(\theta_o) &= -\frac{1}{2}(\mathbf{x}_* - \mathbf{h}_*\theta_o)^\top \mathbf{\Sigma}_{1*}^{-1}(\mathbf{x}_* - \mathbf{h}_*\theta_o) \\ &= -\frac{1}{2}\left(\mathbf{x}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{x}_* + (\mathbf{h}_*\theta_o)^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*\theta_o\right) \\ &\quad + \mathbf{x}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*\theta_o, \end{aligned} \quad (23)$$

where

$$\begin{aligned} \mathbf{x}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*\theta_o &= (\mathbf{h}_*\theta + \mathbf{n}_*)^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*\theta_o \\ &= (\mathbf{h}_*\theta)^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*\theta_o + \mathbf{n}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*\theta_o. \end{aligned} \quad (24)$$

The same can be done with $\mathcal{L}(\theta_o + h)$ and one can then compute $\ln\Lambda = \mathcal{L}(\theta_o) - \mathcal{L}(\theta_o + h)$ as

$$\begin{aligned} \ln\Lambda &= \frac{1}{2}\mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_* h^2 \\ &\quad + (\mathbf{h}_*(\theta_o - \theta))^\top \mathbf{\Sigma}_{1*}^{-1}(\mathbf{h}_* h) - \mathbf{n}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_* h \end{aligned} \quad (25)$$

The probability in (21) yields

$$\mathbb{P}(\ln\Lambda(\mathbf{x}) < 0|\mathcal{H}_1) = \mathbb{P}(S(h) + n < 0) \quad (26)$$

where

$$S(h) = \frac{1}{2}\mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_* h^2 \quad (27)$$

and

$$n = -\mathbf{n}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_* h. \quad (28)$$

In the same way, the probability in (22) leads to

$$\mathbb{P}(\ln\Lambda(\mathbf{x}) > 0|\mathcal{H}_2) = \mathbb{P}(S(h) - n < 0). \quad (29)$$

The goal is now to determine which is the impact of a Gaussian mixture on the noise variable n. Note that the variable n also follows a Gaussian mixture. The distribution of n is then as $p_n(n) = \omega_1 p_{n_1}(n) + \omega_2 p_{n_2}(n)$, where $p_{n_i}(n)$ is associated with $p_{i*}(\mathbf{n}_*)$. The probability in (26) can be computed as

$$\mathbb{P}(S(h) + n < 0) = \omega_1 Q\left(\frac{S(h)}{\sigma_{n_1}}\right) + \omega_2 Q\left(\frac{S(h)}{\sigma_{n_2}}\right) \quad (30)$$

where $Q(x) = (1/\sqrt{2\pi})\int_x^\infty \exp(-t^2/2))dt$ is the Q-function, expressed in terms of the complementary error function as $Q(x) = (1/2)\text{erfc}(x/\sqrt{2})$,

$$\sigma_{n_1}^2 = \mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_* h^2. \quad (31)$$

and

$$\sigma_{n_2}^2 = \mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{\Sigma}_{2*}\mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_* h^2. \quad (32)$$

The probability in (29) is equal to the one in (30).The probability of error is then

$$\mathbb{P}_e(h) = \omega_1 Q\left(\frac{S(h)}{\sigma_{n_1}}\right) + \omega_2 Q\left(\frac{S(h)}{\sigma_{n_2}}\right). \quad (33)$$

where

$$\frac{S(h)}{\sigma_{n_1}} = \frac{1}{2}\sqrt{\mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*}\,h. \quad (34)$$

and

$$\frac{S(h)}{\sigma_{n_2}} = \frac{\mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_* h}{2\sqrt{\mathbf{h}_*^\top \mathbf{\Sigma}_{1*}^{-1}\mathbf{\Sigma}_{2*}\mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*}}. \quad (35)$$

3.2. Main result

Note that the result is independent of θ_0 and the bound given in (16) applies. Moreover, when the argument of the Q-function depends linearly on h, as in

$$\text{ZZB} = \frac{1}{T}\int_0^T h(T-h)Q(\gamma h)\,dh. \tag{36}$$

the integral can be solved integrating by parts

$$\begin{aligned}\text{ZZB} &= \frac{T}{6}Q(T\gamma) + \frac{1}{4\gamma^2}\Gamma_{3/2}\left(\frac{T^2\gamma^2}{2}\right)\\ &\quad - \frac{2}{3T\sqrt{2\pi}\gamma^3}\Gamma_2\left(\frac{T^2\gamma^2}{2}\right)\end{aligned} \tag{37}$$

where $\Gamma_a(x)$ is the incomplete gamma function. For an interval of the *a priori* distribution of θ, $p_\theta(\theta)$, satisfying $T >> \frac{1}{2\gamma}$, the bound reduces to ZZB $= \frac{1}{4\gamma^2}$. As $\mathbb{P}_e(h)$ depends linearly on h, one can write $\frac{S(h)}{\sigma_{n_1}} = \gamma_1 h$, and $\frac{S(h)}{\sigma_{n_1 2}} = \gamma_2 h$, so that the ZZB is given by $\omega_1\frac{1}{4\gamma_1^2} + \omega_2\frac{1}{4\gamma_2^2}$

$$\begin{aligned}\text{ZZB}(\mathcal{M}) &= \omega_1\frac{1}{\mathbf{h}_*^\top\mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*} + (1-\omega_1)\frac{\mathbf{h}_*^\top\mathbf{\Sigma}_{1*}^{-1}\mathbf{\Sigma}_{2*}\mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*}{(\mathbf{h}_*^\top\mathbf{\Sigma}_{1*}^{-1}\mathbf{h}_*)^2}\\ &= \omega_1\frac{\sigma_{1*}^2}{\mathbf{h}_*^\top\mathbf{h}_*} + (1-\omega_1)\frac{\sigma_{2*}^2}{\mathbf{h}_*^\top\mathbf{h}_*}\end{aligned} \tag{38}$$

for $T >> \frac{1}{2\gamma}$. This is the bound corresponding to the mismatched ML estimator in (5).

4. NUMERICAL RESULTS

A test is carried out to assess the performance of the bound under a Gaussian mixture model mismatch. The length of the observation vector $\mathbf{x}$ is $K = 50000$. The linear function of θ is set to $\mathbf{h}_* = \mathbf{1}$. The covariance of the noise processes are given by $\mathbf{\Sigma}_{1*} = \sigma_{1*}^2\mathbf{I}$ and $\mathbf{\Sigma}_{2*} = \sigma_{2*}^2\mathbf{I}$, where $\sigma_{1*} = 1$ and $\sigma_{2*} = 5$. Figure 1 illustrates the MSE as a function of $(1-\omega_1)$. The latter presents a sweeping from 0 to 1 in order to accommodate all possible cases. The ZZB and the MLE for the mismatch model scenario described above are shown together with their corresponding version for matching models. The performance under the misspecified model is always worse than its matching counterpart, except for the extremes cases of $\omega_1 = 1$ and $\omega_1 = 0$, where they are equivalent. The performance of the sample median estimator is also shown. It appears from the figure that the sample median outperforms the MLE under the misspecified model, except when we are close to the limit cases.

One can find the values of $(1-\omega_1)$ for which the sample median estimator performance is above or below the mismatched ML upon solving ZZB$(\mathcal{M})$ = MSE$(\hat{\theta}_{\text{med}})$. If we consider the following ratio $\phi = \sigma_{2*}/\sigma_{1*}$ we can write

$$\omega_1 + (1-\omega_1)\phi^2 = \frac{1}{4\left(\frac{\omega_1}{\sqrt{2\pi}} + \frac{1-\omega_1}{\phi\sqrt{2\pi}}\right)^2}. \tag{39}$$

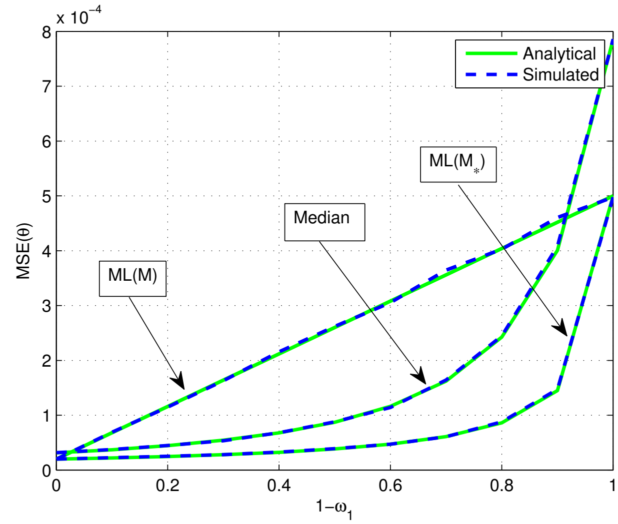

Fig. 1. MSE with respect to $(1-\omega_1)$, the outlier probability.

Ratio (ϕ)	$(1-\omega_1)_1$	$(1-\omega_1)_2$
2.2	–	–
2.25	0.35	0.53
2.4	0.23	0.66
3	0.10	0.81
5	0.03	0.92
7	0.01	0.95

Table 1. Solutions of $(1-\omega_1)$ with respect to $\phi = \sigma_{2*}/\sigma_{1*}$

This equation yields 3 solutions for $(1-\omega_1)$. In Table 1 we show the two solutions $(1-\omega_1)_1$ and $(1-\omega_1)_2$, if existing, for different values of ϕ. Note that for the range between $(1-\omega_1)_1$ and $(1-\omega_1)_2$ the sample median estimator outperforms the mismatched ML. For low values of ϕ, e.g. $\phi \leq 2.2$, the mismatched ML estimator always achieves lower MSE.

5. CONCLUSIONS

In this article, the MSE performance of the sample mean and the sample median estimators was compared in the presence of outliers. Particularly, we derived closed-form expressions for the MSE of these estimators and compared them both analytically and through computer simulations. The superiority of the sample median against outliers is well known in the robust statistics literature. Here we provide an analysis of the sample mean under outliers, providing the tools to compare both estimators under these circumstances.

6. REFERENCES

[1] A.M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts," *Signal Processing Magazine, IEEE*, vol. 29, no. 4, pp. 61–80, July 2012.

[2] L. Scharf, *Statistical Signal Processing. Detection, Estimation and Time series analysis*, Addison–Wesley, 1991.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing. Estimation Theory*, Prentice Hall, 1993.

[4] Harry L. Van Trees and Kristine L. Bell, Eds., *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, Wiley Interscience, 2007.

[5] R. Maronna, D. Martin, and V. Yohai, *Robust Statistics - Theory and Methods*, John Wiley & Sons, 1 edition, 2006.

[6] Judah Jacobson, *Lower Bounds on Estimator Error and the Threshold Effect*, Ph.D. thesis, Harvard College, 2004.

[7] John T Chu, "On the distribution of the sample median," *The Annals of Mathematical Statistics*, pp. 112–116, 1955.

[8] D. Chazan, M. Zakai, and J. Ziv, "Improved Lower Bounds on Signal Parameter Estimation," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 90–93, 1975.

[9] K.L. Bell, Y. Steinberg, Y. Ephraim, and H.L. Van Trees, "Extended Ziv-Zakai lower bound for vector parameter estimation," *Information Theory, IEEE Transactions on*, vol. 43, no. 2, pp. 624 –637, mar 1997.

[10] Wen Xu, A.B. Baggeroer, and K.L. Bell, "A bound on mean-square estimation error with background parameter mismatch," *Information Theory, IEEE Transactions on*, vol. 50, no. 4, pp. 621–632, April 2004.