

Estimation of Location Parameter from Two Biased Samples

Leonid I. Piterbarg

Department of Mathematics, University of Southern California, Los Angeles, USA

Email: piter@usc.edu

Received June 5, 2013; revised July 5, 2013; accepted July 12, 2013

Copyright © 2013 Leonid I. Piterbarg. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

We consider a problem of estimating an unknown location parameter from two biased samples. The biases and scale parameters of the samples are not known as well. A class of non-linear estimators is suggested and studied based on the fuzzy set ideas. The new estimators are compared to the traditional statistical estimators by analyzing the asymptotical bias and carrying out Monte Carlo simulations.

Keywords: Biasness; Fuzzy Estimator; Location Parameter

1. Introduction

The problem is to estimate an unknown scalar parameter θ from two different independent samples of size n_1 and n_2

$$\begin{aligned} x_{1i} &= \theta + b_1 + \sigma_1 \xi_i, \quad i = 1, 2, \dots, n_1, \\ x_{2i} &= \theta + b_2 + \sigma_2 \eta_i, \quad i = 1, 2, \dots, n_2 \end{aligned} \quad (1)$$

where b_j and σ_j are the bias and scale parameter of the j -th sample respectively, $j = 1, 2$, and ξ_i, η_i are zero mean independent random noises.

A novelty in our set up is that the biases $b_j, j = 1, 2$ are assumed to be unknown which makes θ unidentifiable from the classical statistics viewpoint, e.g. [1]. Thus, traditional formulations of best estimation are not applicable in this situation which nevertheless often arises in applications. An important example is an assimilation problem in physical oceanography and meteorology where information on a certain parameter comes from both observations and a circulation model [2,3]. Typically such a model is biased for a variety reasons: uncertainties in a forcing and dissipation, boundary conditions, model parameters, etc. The bias in observations is mostly due to inaccurate measurements and time/space averaging intrinsically present in any measuring procedure. That type of bias sometimes can be excluded using a learning samples [4,5], however it is often difficult to justify the key assumption that learning and control samples are taken from the same ensemble.

One can simply ignore the biases and apply traditional

least square or maximum likelihood methods which would result in a biased estimate of θ . As an alternative, we suggest to use fuzzy set (possibility theory) ideas [6,7] to construct non-linear estimators for θ diminishing the bias comparing to the aforementioned approach. With biased observations the focus is naturally shifted from the variance of an estimator, which can be arbitrarily reduced by increasing a sample size, to its asymptotical bias. More exactly, in the traditional representation of the squared standard error (SE)

$$E\left\{\left(\hat{\theta} - \theta\right)^2\right\} = B_{\hat{\theta}}^2 + \text{Var}\left\{\hat{\theta}\right\}$$

our primary point of interest is the first term. Thus, we start with analyzing the asymptotical bias of the suggested estimators and then compare it to that of estimators traditionally used in statistics such as weighted mean or weighted median. Then, SE is addressed for small samples via Monte Carlo simulations when the second term is not negligible.

Worthy noting that in the simplest situation with unbiased observations $b_j = 0$ and known σ 's the unbiased estimator with the smallest SE (least square estimate) is given by the weighted mean, e.g. [8]

$$\hat{\theta}_{LS} = \frac{\sigma_2^2 \bar{x}_1 + \sigma_1^2 \bar{x}_2}{\sigma_1^2 + \sigma_2^2} \quad (2)$$

where \bar{x}_j is the sample mean of the j -th sample. Moreover, with normal noises it is the maximum likelihood estimator [1].

In the general formulation (1) with biased observations a choice of an appropriate measure of the estimation skill is a challenge because the bias $B_{\hat{\theta}}$ depends on unknown (nuisance) parameters b_1, b_2 which never can be identified from the available observations. We construct such a measure as follows.

For large samples one can efficiently estimate $\sigma_j^2, j=1,2$ and the bias difference $\Delta b = b_1 - b_2$ from the observations (1) by subtracting the second sample from the first one. Introduce

$$\alpha = \frac{b_1}{\Delta b} \tag{3}$$

Assume that we deal with a class of estimators $\hat{\theta}$ for which the asymptotical bias $(n_1, n_2 \rightarrow \infty)$ exists and thereby is a function of all the involved parameters $B(\alpha; \Delta b, \sigma_1^2, \sigma_2^2)$ (sub $\hat{\theta}$ is dropped as a matter of brevity).

Unlike $\Delta b, \sigma_1^2, \sigma_2^2$ one cannot estimate α at all under the given observations. In such a situation one of the ways to order estimators according to their biases is to accept that $\hat{\theta}^{(1)}$ is better than $\hat{\theta}^{(2)}$ (under the identifiable parameters being fixed) if

$$B^{(1)}(\pi) \leq B^{(2)}(\pi) \tag{4}$$

for a certain class of densities $\pi(\alpha)$ where

$$B^{(k)}(\pi) = \int_{-\infty}^{\infty} B^{(k)}(\alpha; \Delta b, \sigma_1^2, \sigma_2^2) \pi(\alpha) d\alpha$$

and $B^{(k)}$ is the asymptotical bias of $\hat{\theta}^{(k)}$.

In simple words under arbitrary distribution of α the absolute value of the first bias is not greater than that of the second one.

Our first finding and pretty surprising one is that the best estimator in sense (4) among a wide class including traditional statistical and fuzzy estimators is the simple arithmetic mean

$$\hat{\theta}_0 = \frac{\mu_1 + \mu_2}{2} \tag{5}$$

where μ_j is any consistent estimator of the center of x_i (i.e. $\mu_j \rightarrow \theta + b_j$ as $n_j \rightarrow \infty$) say the sample mean or median.

Does it mean that (5) is the best way to deal with biased observations? Of course it is not because after all a real matter of concern is SE which can be essentially less for the weighted mean (2) than for (5) under small enough biases b_j .

The next important result of this study is that the suggested fuzzy estimators are better than weighted estimators of type (2) in sense (4). Finally, to decide which estimator should be preferred in dealing with small samples we carry out Monte Carlo simulations and use SE averaged over the nuisance parameters as a measure of the

estimation skill

$$SE(\hat{\theta}) = \sqrt{\left\langle E\{\hat{\theta} - \theta\}^2 \right\rangle_{\alpha, \beta}} \tag{6}$$

where the angle brackets mean averaging over parameters α defined in (3) and the ratio

$$\beta = \frac{\sigma_1}{\sigma_1 + \sigma_2} \tag{7}$$

characterizing the difference in the noise level of two sources. The reason for including the averaging over identifiable parameter β is that we are interested in small samples $(n_1, n_2 \leq 10)$ for which it is not possible to efficiently estimate even identifiable parameters.

We then investigate dependence of (6) on the bias level Δb and noise scale $\sigma = \sigma_1 + \sigma_2$ for different estimators and suggest recommendations for sensible choice among them. In general, fuzzy estimators seem to be preferable for high values of Δb and σ in most of scenarios determined by different noise distribution (normal, Cauchy), non-probabilistic noise (logistic chaos), and by different estimates of the center (mean or median).

2. Estimators and Their Asymptotical Bias

First recall that a fuzzy set is a pair (A, P) where A is a set and $P: A \rightarrow [0, 1]$ is called the membership function, the value $P(x)$ characterizes the degree of membership x in A . The set $\{x \in A | P(x) > 0\}$ is called the support of (A, P) . For our purposes it is enough to consider real fuzzy sets, $A \subset \mathbf{R}$.

Regarding to the formulation (1) we consider A as a range of the observed random variable X . Let us introduce a class of fuzzy estimators similar to estimators based on the triangular membership function (possibility distribution) discussed in [9].

Let $F(x)$ be a cumulative distribution function, symmetric and increasing, i.e. $F(x) + F(-x) = 1, F'(x) > 0$.

Introduce the membership functions generated by each of two samples as follows

$$P_j(x) = \begin{cases} 2F\left(\frac{x - \mu_j}{s_j}\right) & \text{if } x \leq \mu_j \\ 2\left(1 - F\left(\frac{x - \mu_j}{s_j}\right)\right) & \text{if } x \geq \mu_j \end{cases}$$

and μ_j, s_j are consistent estimators of the center and spread of the j -th sample respectively, $j = 1, 2$. In other words we assume

$$\mu_j \rightarrow \theta + b_j, s_j \rightarrow \sigma_j$$

with probability 1 as $n_j \rightarrow \infty$.

We address the following estimator of θ henceforth called the fuzzy estimator

$$\hat{\theta}_f = \frac{\int_O x(P_1(x) \wedge P_2(x)) dx}{\int_O P_1(x) \wedge P_2(x) dx} \quad (8)$$

where O is the set of the Pareto optimal solutions and $a \wedge b$ is the minimum of a and b . Pareto set is defined by $x \in O$ if and only if

$$P_1'(x)P_2'(x) < 0$$

In other words, a Pareto optimal solution (Pareto optimal) is one in which any improvement of one objective function in the two objective optimization problem

$$P_1(x) \rightarrow \max, \quad P_2(x) \rightarrow \max$$

can be achieved only at the expense of another, e.g. [10].

A key point in the base of (8) is that we use a standard fuzzy logic aggregation procedure, but integration is carried out only over the set of x maximizing both memberships of x . Such an approach goes far beyond linear estimators traditionally used in statistics. However, the introduced class of estimators allows their analytical study at least regarding to the asymptotical bias.

Together with (8) we consider the class of weighted estimators

$$\hat{\theta}_p = w_p \mu_1 + (1 - w_p) \mu_2 \quad (9)$$

where μ_j is a sample mean or other consistent estimator of the center of j -th sample.

The weights in (9) are given by

$$w_p = s_2^p / (s_1^p + s_2^p)$$

and our primary focus is on the minimum least square estimator corresponding to $p = 2$, simple arithmetic

$$\hat{\theta}_f = \theta + \frac{\int_{(\mu_1 - \mu_2)/s_2}^{(\mu_1 - \mu_2)/(s_1 + s_2)} s_2 (\mu_2 + s_2 u) F(u) du + \int_{(\mu_2 - \mu_1)/(s_1 + s_2)}^{(\mu_2 - \mu_1)/s_1} s_1 (\mu_1 + s_1 u) (1 - F(u)) du}{s_2 \int_{(\mu_1 - \mu_2)/s_2}^{(\mu_1 - \mu_2)/(s_1 + s_2)} F(u) du + s_1 \int_{(\mu_2 - \mu_1)/(s_1 + s_2)}^{(\mu_2 - \mu_1)/s_1} (1 - F(u)) du}$$

Let us change u to $-u$ in the second integrals on the top and at the bottom. Then using the symmetry of $F(u)$ get

$$\hat{\theta}_f = \theta + \frac{\int_{(\mu_1 - \mu_2)/s_2}^{(\mu_1 - \mu_2)/(s_1 + s_2)} s_2 (\mu_2 + s_2 u) F(u) du + \int_{(\mu_1 - \mu_2)/(s_1 + s_2)}^{(\mu_1 - \mu_2)/s_1} s_1 (\mu_1 - s_1 u) F(u) du}{s_2 \int_{(\mu_1 - \mu_2)/s_2}^{(\mu_1 - \mu_2)/(s_1 + s_2)} F(u) du + s_1 \int_{(\mu_2 - \mu_1)/(s_1 + s_2)}^{(\mu_2 - \mu_1)/s_1} F(u) du}$$

Proceeding to the limit $n_1, n_2 \rightarrow \infty$, accounting for the consistency of estimates μ_j, s_j and symmetry of $F(x)$ obtain

$$B_f = \frac{\int_{\Delta b/\sigma_2}^{\Delta b/(\sigma_1 + \sigma_2)} \sigma_2 (b_2 + \sigma_2 u) F(u) du + \int_{\Delta b/\sigma_1}^{\Delta b/(\sigma_1 + \sigma_2)} \sigma_1 (b_1 - \sigma_1 u) F(u) du}{\sigma_2 \int_{\Delta b/\sigma_2}^{\Delta b/(\sigma_1 + \sigma_2)} F(u) du + \sigma_1 \int_{\Delta b/\sigma_1}^{\Delta b/(\sigma_1 + \sigma_2)} F(u) du}$$

mean ($p = 0$) and the weighted estimator with $p = 1$.

To analyze properties of $\hat{\theta}_f$, let us fix the bias difference $\Delta b = b_1 - b_2$, the observation error level $\sigma = \sigma_1 + \sigma_2$ and introduce dimensionless parameters $\alpha = b_1/\Delta b$, $\beta = \sigma_1/\sigma$, $a = |\Delta b|/\sigma$. Then the bias of (9) as $n = n_1 = n_2 \rightarrow \infty$ goes to

$$B_p = \Delta b (\alpha - v_p(\beta)), \quad v_p(\beta) = \frac{\beta^p}{\beta^p + (1 - \beta)^p} \quad (10)$$

The asymptotical bias for (8) is given in the next statement.

Proposition 1

Let $n = n_1 = n_2 \rightarrow \infty$, then for any fixed a the bias $\hat{B}_f = E\hat{\theta}_f - \theta$ goes to

$$B_f = \Delta b (\alpha - \lambda(\beta)) \quad (11)$$

where

$$\lambda(\beta) = \frac{ap(1 - \beta) + q(1 - \beta) - q(\beta)}{a(p(\beta) + p(1 - \beta))}$$

with

$$p(x) = x \int_{-a/x}^{-a} F(u) du,$$

$$q(x) = x^2 \int_{-a/x}^{-a} F(u) du$$

Proof

With no loss of generality assume $\mu_1 < \mu_2$, then the Pareto optimal set is simply an interval $O = (\mu_1, \mu_2)$ due to the condition $F'(x) > 0$. Since the symmetry the only intersection point of $P_1(x)$ and $P_2(x)$ in this interval is given by

$$x_0 = \frac{\mu_1 s_2 + \mu_2 s_1}{s_1 + s_2}$$

and after an appropriate variable change (8) becomes

Notice that in the considered case $\Delta b < 0$. Using dimensionless variables $\alpha = b_1/\Delta b$, $\beta = \sigma_1/\sigma$, $a = |\Delta b|/\sigma$ after some algebra arrive at (11).

Notice that

$$\lambda(x) + \lambda(1-x) = 1 \tag{12}$$

Proposition 2

If $0 < x < 1/2$, then

$$\lambda(x) < 1/2 \tag{13}$$

Proof

Since $p(x) > 0$ inequality (13) is equivalent to

$$ap(x) + 2q(x) > ap(1-x) + 2q(1-x)$$

The expression on the right hand side can be written as

$$\begin{aligned} S(x) &= x \int_{-a/x}^{-a} (a + 2xu) F(u) du \\ &= \frac{1}{2} \int_{-a}^{a(1-2x)} v F\left(\frac{v-a}{2x}\right) dv \end{aligned}$$

then the inequality takes form $S(x) > S(1-x)$ and is equivalent to

$$\int_{-a}^{a(1-2x)} v F\left(\frac{v-a}{2x}\right) dv > \int_{-a}^{a(2x-1)} v F\left(\frac{v-a}{2(1-x)}\right) dv \tag{14}$$

Let us break down the integral on LHS into integrals from $-a$ to 0 and from 0 to $a(1-2x)$ respectively, change v to $-v$ in the second integral, and move it to RHS. Next, make the same variable change on the RHS of (14) and move it to LHS. The goal is to have all the integrals over positive intervals. The result is

$$\begin{aligned} &\int_0^{a(1-2x)} v F\left(\frac{v-a}{2x}\right) dv + \int_{a(1-2x)}^a v F\left(-\frac{v+a}{2(1-x)}\right) dv \\ &> \int_0^a v F\left(-\frac{v+a}{2x}\right) dv \end{aligned}$$

Obviously $(v-a)/2x > -(v+a)/2x$ and $-(v+a)/2(1-x) > -(v+a)/2x$ whenever $0 < x < 1/2$ and hence the last inequality is true since $F(u)$ is increasing. The proof is over.

Further we restrict ourselves by distributions which decay fast enough as $u \rightarrow -\infty$

$$F(u) \leq \frac{C}{|u|} \tag{15}$$

From (15) it follows that $\lim_{x \rightarrow 0} q(x)/p(x) = 0$ and since $p(1) = q(1) = 0$ we get

$$\lambda(0) = 0, \quad \lambda(1) = 1 \tag{16}$$

Next we consider the class of all estimators for which the asymptotical bias is expressible in form

$$B^\mu = \Delta b(\alpha - \mu(\beta)) \tag{17}$$

where $\mu(x)$ satisfies

$$\mu(x) + \mu(1-x) = 1.$$

$$\mu(0) = 1, 0 \leq \mu(x) \leq 1/2 \text{ if } 0 \leq x \leq 1/2$$

Notice that both the weighted mean (9) and the fuzzy estimator (8) are in this class. For the former it is easy to check and for the latter the statement follows from (12, 13, 16).

Now we intend to order estimators from this class according to their asymptotical bias. Roughly speaking for fixed a and β one estimator is better than another if the asymptotical bias of the former averaged over α is less than that of the latter. Rigorously, for two estimator with biases $B^{\mu_k} \equiv B^{(k)}$ given by (17) ($k=1,2$) the first one is better than the second one if (3) holds true for any positive function $\pi(x)$ satisfying two conditions

$$\int_{-\infty}^{\infty} |x| \pi(x) dx < \infty, \quad \pi(0.5+x) = \pi(0.5-x) \tag{18}$$

Value $\alpha = 0.5$ is singled out since it corresponds to the case of balanced biases $b_1 + b_2 = 0$.

Proposition 3

If for some $0 < \beta < 1/2$ inequality $\mu_1(\beta) > \mu_2(\beta)$ is true, then the first estimator is better than the second one

$$B^{(1)}(\pi) \leq B^{(2)}(\pi) \tag{19}$$

Proof

By direct computations with using symmetry of π obtain for $z < 0.5$

$$\frac{d}{dz} \int_{-\infty}^{\infty} |x-z| \pi(x) dx = - \int_z^{1-z} \pi(x) dx < 0$$

Hence $\mu_1(\beta) > \mu_2(\beta)$ implies (19) because $\mu_j(\beta) < 0.5$.

A surprising result is readily derived from Proposition 3.

Corollary

The trivial estimator

$$\hat{\theta}_0 = \frac{\mu_1 + \mu_2}{2} \tag{20}$$

is the best estimator in the class (17)

This statement follows from the fact that for (20) $\mu(x) = \nu_0(x) \equiv 1/2$.

The problem with (20) is that for small biases it is essentially less efficient than $\hat{\theta}_2$ and $\hat{\theta}_1$. Let us compare the efficiency of (20) to that of $\hat{\theta}_2$ in the case of equal small biases $b_1 = b_2 = \varepsilon\sigma$. Introduce ratio of MSEs

$$R = \sqrt{\frac{E(\hat{\theta}_0 - \theta)^2}{E(\hat{\theta}_2 - \theta)^2}}$$

Proposition 4

For any large $C > 0$

$$R > C \tag{21}$$

whenever

$$\beta < \frac{1}{2} - \frac{1}{2} \sqrt{\frac{C-1}{C+1}} \text{ or } \beta > \frac{1}{2} + \frac{1}{2} \sqrt{\frac{C-1}{C+1}} \tag{22}$$

and

$$\varepsilon < \sqrt{\frac{(\beta^2 + (1-\beta)^2)^2 - 4C^2\beta^2(1-\beta)^2}{4(C^2-1)(\beta^2 + (1-\beta)^2)^2}} \tag{23}$$

Proof

Let $s = (\beta^2 + (1-\beta)^2)/4$ and

$$t = \beta^2(1-\beta)^2 / (\beta^2 + (1-\beta)^2)^2$$

then (21) becomes

$$\frac{\varepsilon^2 + s}{\varepsilon^2 + t} > C^2$$

or

$$\varepsilon^2 < \frac{s - tC^2}{C^2 - 1} \tag{24}$$

Inequality $s - tC^2 > 0$ is equivalent to (23) and then (24) turns in (22).

For example assume $C = 2$, then (20) holds true if $\beta > 0.7887$ or $\beta < 0.2113$. Then for $\beta = 0.1$ we have $R > 2$ if $\varepsilon < 0.2594$.

Thus it is worth to compare the suggested fuzzy estimators with $\hat{\theta}_2$ and $\hat{\theta}_1$. For that purpose we need another representation for λ appearing in (11)

$$\lambda(x) = x + \frac{T(1-x) - T(x)}{a(p(x) + p(1-x))} \tag{25}$$

where

$$T(x) = x^2 \int_{-a/x}^{-a} (a+u)F(u)du$$

Since the denominator in (25) is positive for all x the following statement gives conditions for for the fuzzy estimator to dominate $\hat{\theta}_1$.

Proposition 5

$$\lambda(x) > x, x \in (0,1/2) \tag{26}$$

if and only if

$$T(1-x) - T(x) > 0, x \in (0,1/2) \tag{27}$$

For example if the possibility distribution is given by

$$P(u) = \frac{1}{1+|u|} \tag{28}$$

which corresponds to $F(u) = 0.5/(1-u)$, $u < 0$, then (27) is fulfilled. Indeed, for (28)

$$t(x,b) \equiv T(1-x) - T(x)$$

$$= (1-x)^2 \ln\left(1 + \frac{bx}{1-x}\right) - x^2 \ln\left(1 + \frac{b(1-x)}{x}\right),$$

$$b = \frac{a}{a+1}$$

Then for any fixed $x \in (0,1/2)$ one gets $t(x,0) = 0$ and

$$\frac{\partial t}{\partial b} = \frac{b(1-2x)}{2(1-x+bx)(x+b(1-x))} > 0$$

which implies (27).

Condition (27) is not always fulfilled for the triangle membership function defined by the cumulative probability distribution function

$$F(u) = 0.5 \begin{cases} 1 + \frac{u}{q} & \text{if } -q \leq u \leq 0 \\ 0 & \text{if } u \leq -q \end{cases}$$

where q is a fixed quantile. Indeed, direct computations yield [11].

Proposition 6

$$\lambda(x) - x = \frac{1}{3} \begin{cases} \frac{(1-2x)((3\gamma+1)x(1-x)-2)}{x(1-x)(2\gamma+1)-1} & \text{if } 1/\gamma \leq x \leq 1-1/\gamma \\ \frac{x((\gamma^3-3\gamma^2)(1-x)+2)}{\gamma^2(1-x)-1} & \text{if } x \leq \min(1/\gamma, 1-1/\gamma) \\ \frac{(1-x)((\gamma^3-3\gamma^2)x+2)}{x\gamma^2-1} & \text{if } x \geq \max(1/\gamma, 1-1/\gamma) \\ (\gamma-1)(2x-1) & \text{if } 1-1/\gamma \leq x \leq 1/\gamma \end{cases}$$

where $\gamma = q/a$.

It can be derived from that statement that (26) holds true if and only if $\gamma > 1 + \sqrt{2}$ while for $1 < \gamma < 7/3$ the opposite relation holds true: $\lambda(x) < x$ for all $0 < x < 1/2$. In the intermediate case $7/3 < \gamma < 1 + \sqrt{2}$ the difference $\lambda(x) - x$ changes the sign once in $(0,1/2)$. If $\gamma < 1$ then two samples are incompatible.

These conclusions are illustrated in **Figure 1** by plotting the difference $d(x) = \lambda(x) - x$ for the triangular membership function (left and middle panel) and for membership (28) (right panel). If $1 < \gamma < 2$ then for sure $\gamma < 7/3$ and hence $d(x) < 0$ for $x < 1/2$ (left panel). If $2 < \gamma < 3$ then two options are possible either $d(x) > 0$ or $d(x)$ changes the sign once (central panel). Finally

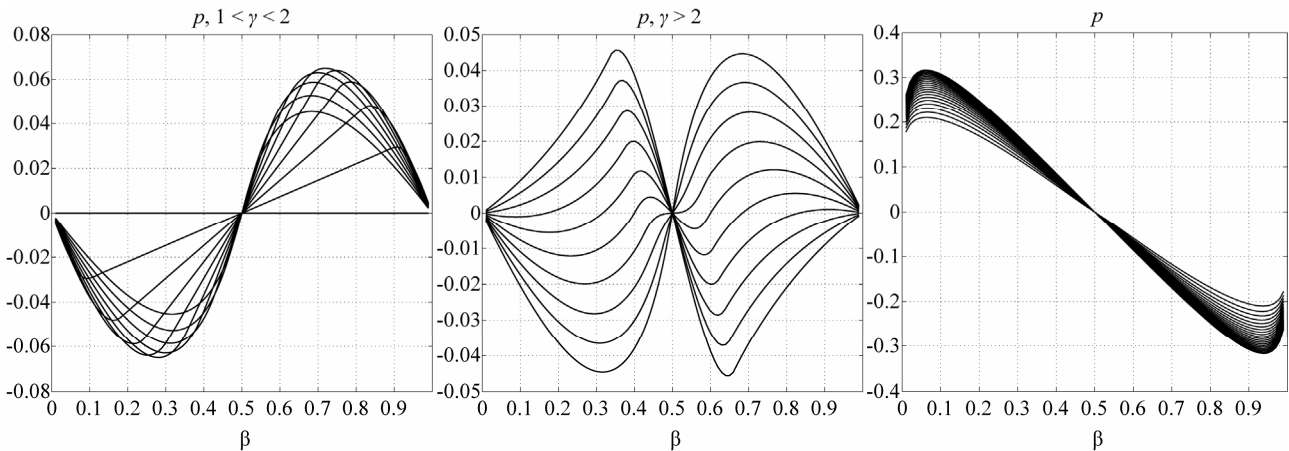


Figure 1. The difference $\lambda(x) - x$ vs x . 1) Triangle membership, $1 < \gamma < 2$; 2) Triangle membership, $2 < \gamma < 3$; 3) Membership given by (28), $0.5 \leq a \leq 5$.

for membership (2) $d(x) > 0$ for all a .

However, it can be shown [9], that $\lambda(x) > x^2$ when $x \in (0, 1/2)$ thereby the fuzzy triangle estimator is better than $\hat{\theta}_2$ in sense (19).

Next, define another quantitative characteristic for comparing asymptotical biases which is easy to evaluate, unlike $B(\pi)$. First define the space of the parameters of interest

$$\Pi = \{ \alpha, \beta \mid -1/2 < \alpha < 3/2, 0 < \beta < 1 \}$$

The range for α is founded in [9]. Then introduce the difference of biases for the mentioned estimators defined by functions $\mu_1(x)$ and $\mu_2(x)$ respectively

$$\Delta B(\alpha, \beta) = |B^{\mu_1}| - |B^{\mu_2}|$$

and define

$$G_{12} = \{ (\alpha, \beta) \in \Pi \mid \Delta B(\alpha, \beta) < 0 \}$$

Obviously the area of Π equals 2 and hence the first estimator can be viewed as better than the second one whenever the area of G_{12} exceeds 1. This excess is quantified in the next statement.

Proposition 7

If $\mu_1(x) > \mu_2(x)$, $0 < x < 1/2$ then

$$|G_{12}| = 1 + \int_0^{1/2} (1 - \mu_1(x) - \mu_2(x)) dx > 1$$

Indeed, the solution of $|\alpha - \mu_1(x)| < |\alpha - \mu_2(x)|$ is equivalent to $\alpha > (\mu_1(x) + \mu_2(x))/2$ from which the statement readily follows.

The following proposition gives a rough estimate of improving a fuzzy estimator comparing to any weighted mean (9). Set

$$G = \{ (\alpha, \beta) \in \Pi \mid |B_f(\alpha, \beta)| < |B_p(\alpha, \beta)| \}$$

Corollary

Let for some fixed a

$$\lambda(x) > v_p(x) \tag{29}$$

for all $0 < x < 1/2$. Then

$$|G| > 1 + \int_{1/2}^1 (v_p(x, a) - 1/2) dx$$

Thus under (29) the region G always occupies more than a half Π . In particular, for $v_1(x) = x$ the integral on the right hand side is equal to 0.125 and for $v_2(x) = x^2 / (x^2 + (1-x)^2)$ it is $\ln 2/4 \approx 0.1733$. The latter follows from

$$\int_0^{1/2} v_2(x) dx = (1 - \ln 2)/4$$

In particular, the area of a region in Π where the bias of the fuzzy estimator (28) is less than that of $\hat{\theta}_1$ is greater than 1.125. While the area of a region in Π where the fuzzy triangle estimator is better than $\hat{\theta}_2$ is greater than 1.1733.

3. Simulations

The goal of the performed simulations is to compare the efficiency of different estimators under different noise distributions. Now the asymptotical bias is not of primary concern, but rather the relative standard error

$$SE(\hat{\theta}) = \sqrt{\frac{1}{2\theta^2} \int_{-0.5}^{1.5} \int_0^1 E \{ \hat{\theta} - \theta \}^2 d\alpha d\beta}$$

is addressed under small samples.

We examine two fuzzy estimators (8), first, determined by (28) (Fuzzy 1) and, second, determined by $P(u) = 0.5e^{-|u|}$ (Fuzzy 2) and three weighted estimators $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$ given by (9).

Sample sizes $n_1 = n_2 = 10$ and number of Monte Carlo trials $M = 100$ for each mesh (α, β) are kept the same for all experiments.

First, normal noise is tested with $\mu_j = \bar{x}_j$, i.e. the

sample center is estimated by the sample mean (Figure 2).

At zero bias (left panel) $\hat{\theta}_2$ is best as expected, while Fuzzy 2 is only slightly worse. For modest biases (central panel) $\hat{\theta}_0$ is best for small level of noise while for modest and high level of noise again Fuzzy 2 is better. Finally for high biases (right panel) only two estimates are competing, $\hat{\theta}_0$ and Fuzzy 1. The former is better for small noise while the latter dominates for high noise. $\hat{\theta}_2$ and $\hat{\theta}_1$ are much worse in this case.

If the center is estimated by the sample median the conclusions remain basically the same, however the numbers are slightly different (experiments are not shown).

In the case of Cauchy distribution of the noises (Figure 3) only the median $\mu_j = \text{med}(x_j)$ was used for estimating the centers.

One can see that even for zero bias of noises Fuzzy 2 has the smallest SE along with $\hat{\theta}_1$. For modest biases all the estimators except $\hat{\theta}_0$ are approximately of the same accuracy for the whole range of σ . The primitive $\hat{\theta}_0$ appears to be essentially worse than any other estimate

for high values of σ . Finally for high biases again Fuzzy 1 is best for intermediate values of σ while $\hat{\theta}_0$ and Fuzzy 2 are slightly better for small and large σ 's respectively. Comparing to the normal noise the accuracy of all the estimators are somewhat lower.

In Figure 4 results are presented for logistic noise generated by

$$\xi_{k+1} = r\xi_k(1 - \xi_k)$$

with $r = 5.2$ for the first sample and $r = 3$ for the second one. The results are closer to the normal noise case rather than to the Cauchy case, however errors in general are smaller than in the normal case.

In summary for all three experiments, the classical weighted estimator $\hat{\theta}_2$ is worsening fast as the level of bias is increasing while the fuzzy estimators demonstrate a steady skill in all the scenarios and for whole range of σ and Δb . At this point it is hard to give a preference to either Fuzzy 1 or Fuzzy 2.

Similar conclusions can be drawn from Figures 5 and 6 where other experiments are presented in which the dependence of SE on Δb were studied for fixed σ . We

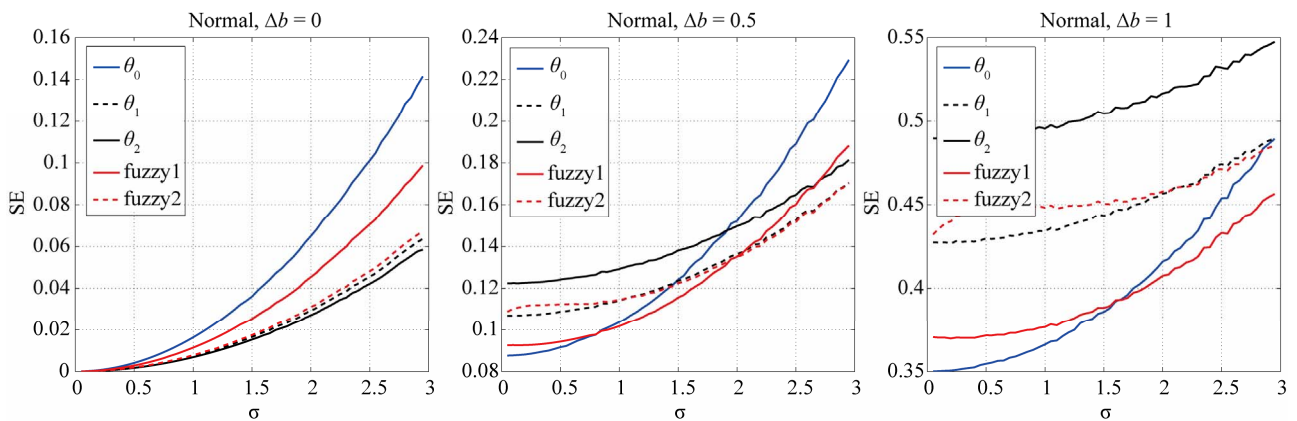


Figure 2. Dependence of standard error on the noise level σ for different values of bias scale Δb normal noise. The mean is taken as an estimate of center. 1) $\Delta b = 0$; 2) $\Delta b = 0.5$; 3) $\Delta b = 1$.

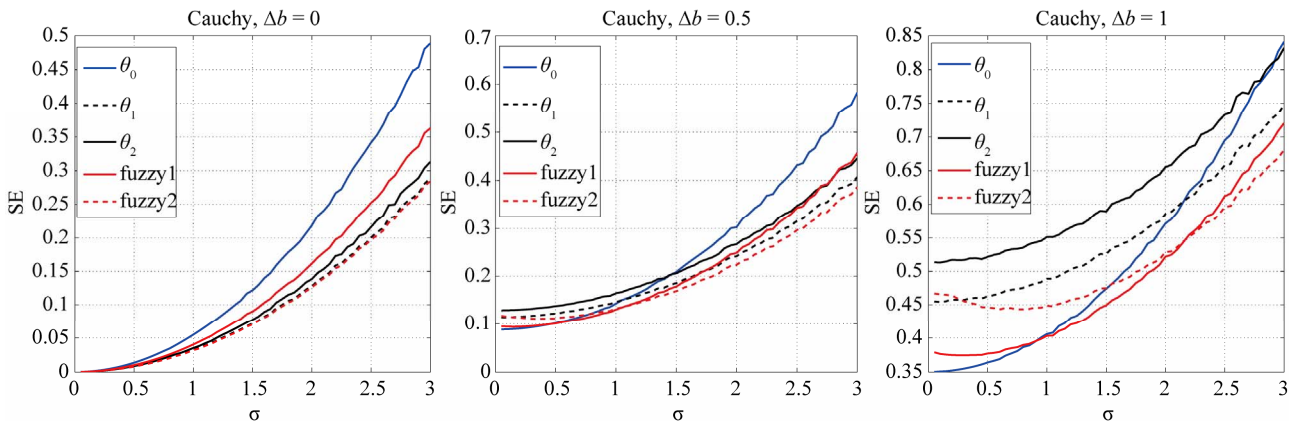


Figure 3. Dependence of standard error on the noise level σ for different values of bias scale Δb Cauchy noise. The median is taken as an estimate of center. 1) $\Delta b = 0$; 2) $\Delta b = 0.5$; 3) $\Delta b = 1$.

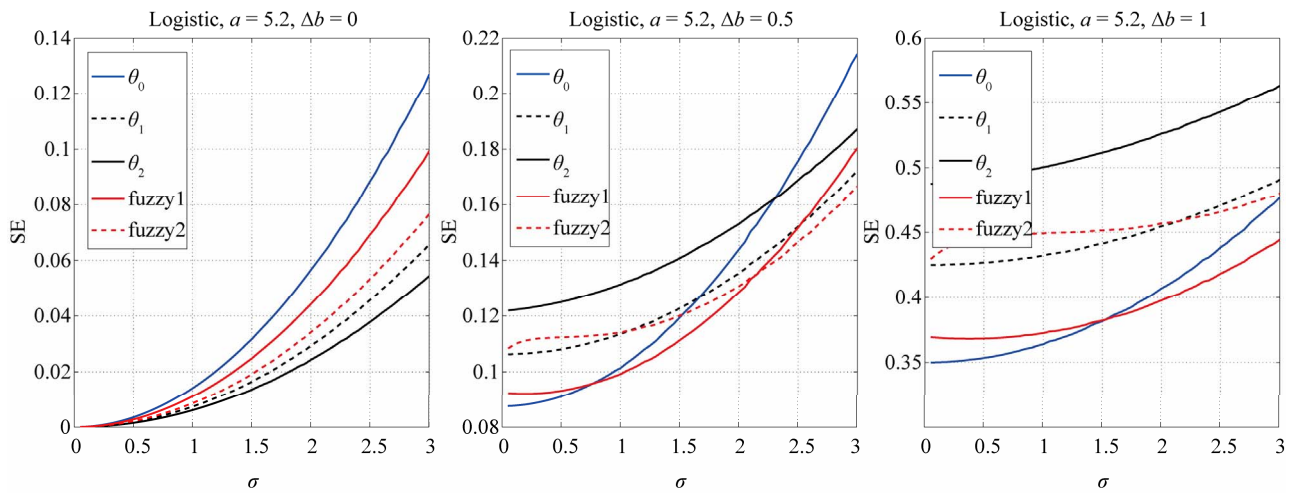


Figure 4. Dependence of standard error on the noise level σ for different values of bias scale Δb logistic noise. The median is taken as an estimate of center. 1) $\Delta b = 0$; 2) $\Delta b = 0.5$; 3) $\Delta b = 1$.

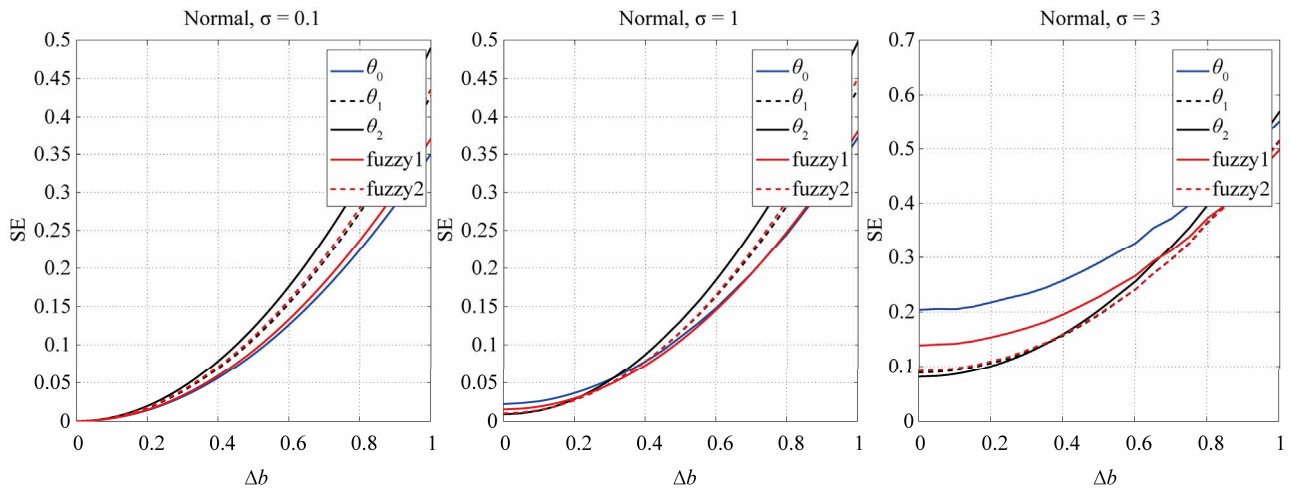


Figure 5. Dependence of standard error on the bias scale Δb for different values of noise level σ normal noise. The median is taken as an estimate of center. 1) $\sigma = 0.1$; 2) $\sigma = 1$; 3) $\sigma = 3$.

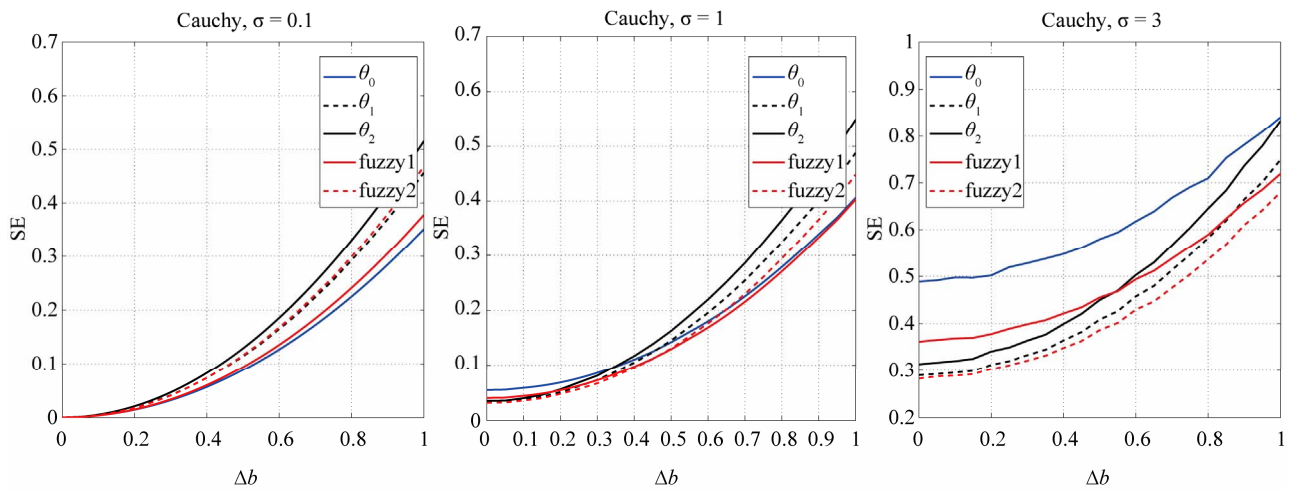


Figure 6. Dependence of standard error on the bias scale Δb for different values of noise level σ Cauchy noise. The median is taken as an estimate of center. 1) $\sigma = 0.1$; 2) $\sigma = 1$; 3) $\sigma = 3$.

want to stress two of them, first, for high σ Fuzzy 2 is uniformly better than any other estimate and, second, $\hat{\theta}_0$ is not appropriate for such values of σ regardless the noise distribution.

4. Conclusions and Discussion

A majority of studies in estimating a location parameter address, first, linear functionals of either the original sample or its ranking, e.g. [11], and, second, unbiased observations. Here a class of essentially nonlinear estimators is suggested based on the fuzzy set theory ideas to handle biased observations coming from two different sources. Because any analytical investigation of the standard error for highly non-linear functions of sample is hard, we concentrated on, first, analytical studying the asymptotical bias of the suggested estimators and, second, Monte Carlo simulations for small samples with the traditional standard error as an efficiency measure.

Conclusions from both studies are similar. Regarding the symptotical bias, the considered fuzzy estimators are uniformly better than the classical least square estimator. As for small samples, the new estimators are of higher accuracy than the traditional least square estimator and its modifications for essential bias and high level of noise. Moreover, even for small bias, the fuzzy estimators are only slightly worse than the optimal one.

Unexpectedly, the primitive arithmetic mean successfully competes with fuzzy estimators for essential biases and modest noise level, but it is of little help for high level of noise or negligible biases.

Notice that the developed fuzzy estimators are highly non-linear functionals of observations and apparently are more appropriate for estimating parameters in non-linear complex systems.

We tested only two estimators from a wide class introduced here. It would be interesting to find an optimal one among them. Another important problem is to address a similar problem for more than two information sources.

5. Acknowledgements

The support of the Office of Naval Research under grant

N00014-11-1-0369 and NSF under grant CMG-1025453 is greatly appreciated.

REFERENCES

- [1] G. Casella and R. Berger, "Statistical Inference," 2nd Edition, Brooks/Cole, Pacific Grove, 1990.
- [2] M. Ghil, "Meteorological Data Assimilation for Oceanographers. Part 1; Description and Theoretical Framework," *Dynamics of Atmospheres and Oceans*, Vol. 13, No. 2, 1989, pp. 171-218. [doi:10.1016/0377-0265\(89\)90040-7](https://doi.org/10.1016/0377-0265(89)90040-7)
- [3] P. Malanotte-Rizzoli, "Modern Approaches to Data Assimilation in Ocean Modeling," Elsevier, Amsterdam, 1996.
- [4] T. Krishnamurti, C. Kishtawal, Z. Zhang, T. Larow, D. Bachiochi and E. Williford, "Multimodel Ensemble Forecasts for Weather and Seasonal Climate," *Journal of Climate*, Vol. 13, No. 8, 2000, pp. 4196-4216. [doi:10.1175/1520-0442\(2000\)013<4196:MEFFWA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2)
- [5] F. J. Lenartz, M. Beckers, J. Chiggiato, B. Mourre, C. Troupin, L. Vandenbulcke and M. Rixen, "Super Ensemble Techniques Applied to Wave Forecast: Performance and Limitations," *Ocean Science*, Vol. 6, No. 3, 2010, pp. 595-604. [doi:10.5194/os-6-595-2010](https://doi.org/10.5194/os-6-595-2010)
- [6] D. Dubois and H. Prade, "Possibility Theory," Plenum Press, New York, London, 1988. [doi:10.1007/978-1-4684-5287-7](https://doi.org/10.1007/978-1-4684-5287-7)
- [7] G. Shafer, "A Mathematical Theory of Evidence," Princeton University Press, Hoboken, 1976.
- [8] T. P. Hill and J. Miller, "How to Combine Independent Data Sets for the Same Quantity," *Chaos*, Vol. 21, No. 3, 2011, p. 3. [doi:10.1063/1.3593373](https://doi.org/10.1063/1.3593373)
- [9] L. I. Piterbarg, "Parameter Estimation from Small Biased Samples: Statistics vs Fuzzy Logic," *Fuzzy Sets and Systems*, Vol. 170, No. 1, 2011, pp. 1-21. [doi:10.1016/j.fss.2011.01.010](https://doi.org/10.1016/j.fss.2011.01.010)
- [10] K. Deb, "Multi-Objective Optimization Using Evolutionary Algorithms," Willey, Princeton, 2001.
- [11] R. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, Vol. 35, No. 1, 1964, pp. 73-101. [doi:10.1214/aoms/1177703732](https://doi.org/10.1214/aoms/1177703732)