From Unbiased Numerical Estimates to Unbiased Interval Estimates

Baokun Li[1], Gang Xiang[2],
Vladik Kreinovich[3], and Panagis Moschopoulos[4]

[1]Southwestern University of Finance and Economics
Chengdu, 610074, China, bali@swufe.edu.cn

[2]Applied Biomathematics, 100 North Country Rd.
Setauket, NY 11733, USA, gxiang@sigmaxi.net

[3]Department of Computer Science
[4]Department of Mathematical Sciences
University of Texas at El Paso
El Paso, Texas 79968, USA, vladik@utep.edu

Abstract

One of the main objectives of statistics is to estimate the parameters of a probability distribution based on a sample taken from this distribution. Of course, since the sample is finite, the estimate $\widehat{\theta}$ is, in general, different from the actual value θ of the corresponding parameter. What we can require is that the corresponding estimate is *unbiased*, i.e., that the mean value of the difference $\widehat{\theta} - \theta$ is equal to 0: $E\left[\widehat{\theta}\right] = \theta$. In some problems, unbiased estimates are not possible. We show that in some such problems, it is possible to have *interval* unbiased estimates, i.e., interval-valued estimates $\left[\underline{\widehat{\theta}}, \overline{\widehat{\theta}}\right]$ for which $\theta \in E\left[\underline{\widehat{\theta}}, \overline{\widehat{\theta}}\right] \stackrel{\text{def}}{=} \left[E\left[\underline{\widehat{\theta}}\right], E\left[\overline{\widehat{\theta}}\right]\right]$. In some such cases, it is possible to have *asymptotically sharp* estimates, for which the interval $\left[E\left[\underline{\widehat{\theta}}\right], E\left[\overline{\widehat{\theta}}\right]\right]$ is the narrowest possible.

Keywords: statistics, interval uncertainty, unbiased numerical estimates, unbiased interval estimates

1 Traditional Unbiased Estimates: A Brief Reminder

Estimating parameters of a probability distribution: a practical problem. Many real-life phenomena are random. This randomness often come from diversity: e.g., different plants in a field of wheat are, in general, of somewhat different heights. In practice, we observe a *sample* $x_1, \ldots, x_n$ of the corresponding values – e.g., we measure the heights of several plants, or we perform several temperature measurements. Based on this sample, we want to estimate the original probability distribution.

Let us formulate this problem in precise terms.

Estimating parameters of a probability distribution: towards a precise formulation of the problem. We want to estimate a probability distribution F that describes the actual values corresponding to possible samples $x = (x_1, \ldots, x_n)$. In other words, we need to estimate a probability distribution on the set $\mathbb{R}^n$ of all n-tuples of real numbers.

In statistics, it is usually assumed that we know the class $\mathcal{D}$ of possible distributions. For example, we may know that the distribution is normal, in which case $\mathcal{D}$ is the class of all normal distributions.

Usually, a distribution is characterized by several numerical characteristics – usually known as its *parameters*. For example, a normal distribution $N(\mu, \sigma^2)$ can be uniquely characterized by its mean μ and variance σ^2. In general, to describe a parameter θ means to describe, for each probability distribution F from the class $\mathcal{F}$, the numerical value $\theta(F)$ of this parameter for the distribution F. For example, when $\mathcal{D}$ is a family of all normal distributions $N(\mu, \sigma^2)$, then the parameter θ describing the mean assigns, to each distribution $F = N(\mu, \sigma^2)$ from the class $\mathcal{F}$, the value $\theta(F) = \mu$. Alternatively, we can have a parameter θ for which $\theta(N(\mu, \sigma^2)) = \sigma^2$, or a parameter for which $\theta(N(\mu, \sigma^2)) = \mu + 2\sigma$.

In general, a parameter can be defined as a mapping from the class $\mathcal{F}$ to real numbers. In these terms, to estimate a distribution means to estimate all relevant parameters.

In some cases, we are interested in learning the values of *all* possible parameters. In other situations, we are only interested in the values of *some* parameters. For example, when we analyze the possible effect of cold weather on the crops, we may be only interested in the lowest temperature. On the other hand, when we are interested in long-term effects, we may be only interested in the average temperature.

We need to estimate the value of this parameter based on the observations. Due to the random character of the sample $x_1, \ldots, x_n$, the resulting estimate $f(x_1, \ldots, x_n)$ is, in general different from the desired parameter $\theta(F)$. In principle, it is possible to have estimates that tend to overestimate $\theta(F)$ and estimates that tend to underestimate $\theta(F)$. It is reasonable to consider *unbiased* estimates, i.e., estimates for which the mean value $E_F\left[\hat{\theta}(x_1, \ldots, x_n)\right]$ coincides with $\theta(F)$.

Thus, we arrive at the following definition.

Definition 1. *Let $n > 0$ be a positive integer, and let $\mathcal{F}$ be a class of probability distributions on $\mathbb{R}^n$.*

- *By a* parameter*, we mean a mapping $\theta : \mathcal{F} \to \mathbb{R}$.*
- *For each parameter θ, by its* unbiased estimate*, we mean a function $\widehat{\theta} : \mathbb{R}^n \to \mathbb{R}$ for which, for every $F \in \mathcal{F}$, we have*

$$E_F\left[\widehat{\theta}(x_1, \ldots, x_n)\right] = \theta(F).$$

Examples. One can easily check that when each distribution from the class $\mathcal{F}$ corresponds to n independent, identically distributed random variables, then the arithmetic average $\widehat{\mu}(x_1, \ldots, x_n) = \dfrac{x_1 + \ldots + x_n}{n}$ is an unbiased estimate for the mean μ of the individual distribution. When, in addition, the individual distributions are normal, the sample variance

$$\widehat{V}(x_1, \ldots, x_n) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \widehat{\mu})^2$$

is an unbiased estimate for the variance V of the corresponding distribution.

2 What If We Take Measurement Uncertainty into Account

Need to take measurement uncertainty into account. In the traditional approach, we assume that we know the exact sample values $x_1, \ldots, x_n$. In practice, measurements are never absolutely accurate: due to measurement imprecision, the observed values $\widetilde{x}_i$ are, in general, different from the actual values x_i of the corresponding quantities.

Since we do not know the exact values $x_1, \ldots, x_n$, we need to estimate the desired parameter $\theta(F)$ based on the observed values $\widetilde{x}_1, \ldots, \widetilde{x}_n$.

Towards a precise formulation of the problem. In addition to the probability distribution of possible values x_i, we also have, for each x_i, a probability distribution of possible values of the difference $\widetilde{x}_i - x_i$. In other words, we have a *joint* distribution J on the set of all possible tuples $(x_1, \ldots, x_n, \widetilde{x}_1, \ldots, \widetilde{x}_n)$.

The meaning of this joint distribution is straightforward:

- first, we use the distribution on the set of all tuples x to generate a random tuple $x \in \mathbb{R}^n$;
- second, for this tuple x, we use the corresponding probability distribution of measurement errors to generate the corresponding values $\widetilde{x}_i - x_i$, and thus, the values $\widetilde{x}_1, \ldots, \widetilde{x}_n$.

Similarly to the previous case, we usually have some partial information about the joint distribution – i.e., we know that the distribution J belongs to a known class $\mathcal{D}$ of distributions.

We are interested in the parameter $\theta(F)$ corresponding to the distribution F of all possible tuples $x = (x_1, \ldots, x_n)$. In statistical terms, F is a marginal distribution of J corresponding to x (i.e., obtained from J by averaging over $\widetilde{x} = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$): $F = J_x$. Thus, we arrive at the following definition.

Definition 2. *Let $n > 0$ be a positive integer, and let $\mathcal{D}$ be a class of probability distributions on the set $(\mathbb{R}^n)^2$ of all pairs $(x, \widetilde{x})$ of n-dimensional tuples. For each distribution $J \in \mathcal{D}$, we will denote the marginal distribution corresponding to x by J_x. The class of all such marginal distributions is denoted by $\mathcal{D}_x$.*

- *By a* parameter*, we mean a mapping $\theta : \mathcal{D}_x \to \mathbb{R}$.*
- *For each parameter θ, by its* unbiased estimate*, we mean a function $\widehat{\theta} : \mathbb{R}^n \to \mathbb{R}$ for which, for every $J \in \mathcal{D}$, we have*

$$E_J\left[\widehat{\theta}(\widetilde{x}_1, \ldots, \widetilde{x}_n)\right] = \theta(J_x).$$

Example. When the sample values are independent, identically distributed random variables, and the measurement errors have 0 mean, (i.e., $E[\widetilde{x}_i] = x_i$ for each i), then the arithmetic average $\widehat{\mu}$ is still an unbiased estimate for the mean.

What we show in this paper. In this paper, we show that in some real-life situations, it is not possible to have number-valued unbiased estimates, but we can have *interval-valued* estimates which are unbiased in some reasonable sense.

3 A Realistic Example In Which Unbiased Numerical Estimates Are Impossible

Description of an example. Let us assume that the actual values $x_1, \ldots, x_n$ are independent identically distributed (i.i.d.) normal variables $N(\mu, \sigma^2)$ for some unknown values μ and $\sigma^2 \geq 0$, and that the only information that we have about the measurement errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$ is that each of these differences is bounded by a known bound $\Delta_i > 0$: $|\Delta x_i| \leq \Delta_i$. The situation in which we only know the upper bound on the measurement errors (and we do not have any other information about the probabilities) is reasonably frequent in real life; see, e.g., [3].

In this case, $\mathcal{D}$ is the class of all probability distributions for which the marginal J_x corresponds to i.i.d. normal distributions, and $|\widetilde{x}_i - x_i| \leq \Delta_i$ for all i with probability 1. In other words, the variables $x_1, \ldots, x_n$ are i.i.d. normal, and $\widetilde{x}_i = x_i + \Delta x_i$, where Δx_i can have any distribution for which Δx_i is located on the interval $[-\Delta_i, \Delta_i]$ with probability 1 (the distribution of Δx_i may depend

on $x_1, \ldots, x_n$, as long as each difference Δx_i is located within the corresponding interval).

Let us denote the class of all such distributions by $\mathcal{I}$. By definition, the corresponding marginal distributions J_x correspond to i.i.d. normals. As a parameter, let us select the parameter μ of the corresponding normal distribution.

Proposition 1. *For the class $\mathcal{I}$, no unbiased estimate of μ is possible.*

Proof. Let us prove this result by contradiction. Let us assume that there is an unbiased estimate $\hat{\mu}(x_1, \ldots, x_n)$). By definition of the unbiased distribution, we must have $E_J\left[\widehat{\mu}(\widetilde{x}_1, \ldots, \widetilde{x}_n)\right] = \mu$ for all possible distributions $J \in \mathcal{I}$.

Let us take two distributions from this class. In both distributions, we take $\sigma^2 = 0$, meaning that all the values x_i coincide with μ with probability 1.

In the first distribution, we assume that each value Δx_i is equal to 0 with probability 1. In this case, all the values $\widetilde{x}_i = x_i + \Delta x_i$ coincide with μ with probability 1. Thus, the estimate $\widehat{\mu}(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ coincides with $\widehat{\mu}(\mu, \ldots, \mu)$ with probability 1. So, its expected value $E_J\left[\widehat{\mu}(\widetilde{x}_1, \ldots, \widetilde{x}_n)\right]$ is also equal to $\widehat{\mu}(\mu, \ldots, \mu)$ with probability 1, and thus, the equality that described that this estimate is unbiased takes the form

$$\widehat{\mu}(\mu, \ldots, \mu) = \mu.$$

In other words, for every real number x, we have

$$\widehat{\mu}(x, \ldots, x) = x.$$

In the second distribution, we select a number $\delta = \min\limits_i \Delta_i > 0$, and assume that each value Δx_i is equal to δ with probability 1. In this case, all the values $\widetilde{x}_i = x_i + \Delta x_i$ coincide with $\mu + \delta$ with probability 1. Thus, the estimate $\widehat{\mu}(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ coincides with $\widehat{\mu}(\mu + \delta, \ldots, \mu + \delta)$ with probability 1. So, its expected value $E_J\left[\widehat{\mu}(\widetilde{x}_1, \ldots, \widetilde{x}_n)\right]$ is also equal to $\widehat{\mu}(\mu + \delta, \ldots, \mu + \delta)$ with probability 1, and thus, the equality that described that this estimate is unbiased takes the form

$$\widehat{\mu}(\mu + \delta, \ldots, \mu + \delta) = \mu.$$

However, from $\widehat{\mu}(x, \ldots, x) = x$, we conclude that

$$\widehat{\mu}(\mu + \delta, \ldots, \mu + \delta) = \mu + \delta \neq \mu.$$

This contradiction proves that an unbiased estimate for μ is not possible.

4 Unbiased Interval Estimates: From Idea to Definition

Analysis of the problem. In the above example, the reason why we did not have an unbiased estimate is that the estimate $\widehat{\theta}$ depends only on the distribution of the values $\widetilde{x}_1, \ldots, \widetilde{x}_n$, i.e., only on the marginal distribution $J_{\tilde{x}}$. On the

other hand, what we try to reconstruct is the characteristic of the marginal distribution J_x. In the above example, even if we know $J_{\tilde{x}}$, we cannot uniquely determine J_x, because there exists another distribution J' for which $J'_{\tilde{x}} = J_{\tilde{x}}$ but for which $J'_x \neq J_x$ and, moreover, $\theta(J'_x) \neq \theta(J_x)$. In this case, we cannot uniquely reconstruct $\theta(J_x)$ from the sample $\widetilde{x}_1, \ldots, \widetilde{x}_n$ distributed according to the distribution $J_{\tilde{x}}$.

From numerical to interval-valued estimates. While we cannot uniquely reconstruct the value $\theta(J_x)$ – because we may have distributions J' with the same marginal $J'_{\tilde{x}} = J_{\tilde{x}}$ for which the value $\theta(J'_x)$ is different – we can try to reconstruct the set of all possible values $\theta(J'_x)$ corresponding to such distributions J'.

Often, for every distribution J, the class $\mathcal{C}$ of all distributions J' for which $J'_{\tilde{x}} = J_{\tilde{x}}$ is connected, and the function that maps a distribution J' into a parameter $\theta(J'_x)$ is continuous. In this case, the resulting set $\{\theta(J'_x) : J' \in \mathcal{C}\}$ is also connected, and is, thus, an interval (finite or infinite). In such cases, it is reasonable to consider interval-valued estimates, i.e., estimates $\widehat{\theta}$ that map each sample $\widetilde{x}$ into an interval $\widehat{\theta}(\widetilde{x}) = \left[\underline{\widehat{\theta}}(\widetilde{x}), \overline{\widehat{\theta}}(\widetilde{x})\right]$.

How to define expected value of an interval estimate. On the set of all intervals, addition is naturally defined as

$$\mathbf{a} + \mathbf{b} \stackrel{\text{def}}{=} \{a + b : a \in \mathbf{a}, b \in \mathbf{b}\},$$

which leads to component-wise addition $[\underline{a}, \overline{a}] + [\underline{b}, \overline{b}] = [\underline{a} + \underline{b}, \overline{a} + \overline{b}]$. Similarly, we can define an arithmetic mean of several intervals $[\underline{a}_1, \overline{a}_1], \ldots, [\underline{a}_n, \overline{a}_n]$, and it will be equal to the interval $[\underline{a}_{\text{av}}, \overline{a}_{\text{av}}]$, where $\underline{a}_{\text{av}} \stackrel{\text{def}}{=} \dfrac{\underline{a}_1 + \ldots + \underline{a}_n}{n}$ and $\overline{a}_{\text{av}} \stackrel{\text{def}}{=} \dfrac{\overline{a}_1 + \ldots + \overline{a}_n}{n}$. Thus, it is natural to define the expected value $E[\mathbf{a}]$ of an interval-valued random variable $\mathbf{a} = [\underline{a}, \overline{a}]$ component-wise, i.e., as an interval formed by the corresponding expected values $E[\mathbf{a}] \stackrel{\text{def}}{=} [E[\underline{a}], E[\overline{a}]]$.

When is an interval-valued estimate unbiased? Main idea. It is natural to say that an interval-valued estimate $\widehat{\theta}(\widetilde{x})$ is unbiased if the actual value of the parameter $\theta(J_x)$ is contained in the interval $E\left[\widehat{\theta}(\widetilde{x})\right]$.

Let us take into account that the expected value is not always defined. The above idea seems a reasonable definition, but it may be a good idea to make this definition even more general, by also considering situations when, e.g., the expected value $E[\underline{a}]$ is not defined – i.e., when the function $\underline{a}$ is not integrable. In this case, instead of the exactly defined integral $E[\underline{a}]$, we have a *lower integral* $\underline{E}[\underline{a}]$ and an *upper integral* $\overline{E}[\underline{a}]$. Let us remind what these notions mean.

Lower and upper integrals: a brief reminder. These notions are known in calculus, where we often first define an integral of simple functions $s(x)$ (e.g., piece-wise constant ones).

To define the integral of a general function, we can then use the fact that if $s(x) \leq f(x)$ for all x, then $\int s(x)\,dx \leq \int f(x)\,dx$. Thus, the desired integral

$\int f(x)\,dx$ is larger than or equal to the integrals of all simple functions $s(x)$ for which $s(x) \le f(x)$. Hence, the desired integral is larger than or equal to the supremum of all such integrals $\int s(x)\,dx$.

Similarly, if $f(x) \le s(x)$ for all x, then $\int f(x)\,dx \le \int s(x)\,dx$. So, the integral $\int f(x)\,dx$ is smaller than or equal to the integrals of all simple functions $s(x)$ for which $s(x) \le f(x)$. Thus, the desired integral is smaller than or equal to the infimum of all such integrals $\int s(x)\,dx$.

For well-behaving functions, both the supremum of the values $\int s(x)\,dx$ for all $s(x) \le f(x)$ and the infimum of the values $\int s(x)\,dx$ for all $s(x) \ge f(x)$ coincide – and are equal to the integral. For some functions, however, these supremum and infimum are different. The supremum – which is known to be smaller than or equal to the desired integral $\int f(x)\,dx$ – is called the *lower integral*, and the infimum – which is known to be larger than or equal to the desired integral $\int f(x)\,dx$ – is called the *upper integral*.

For the expected value $E[a] \stackrel{\text{def}}{=} \int x \cdot \rho(x)\,dx$, the corresponding lower and upper integrals are called lower and upper expected values, and denoted by $\underline{E}[a]$ and $\overline{E}[a]$.

Towards the final definition. In the case of an integrable estimate, we would like to require that $E\left[\underline{\widehat{\theta}}\right] \le \theta(J_x)$ and that $\theta(J_x) \le E\left[\overline{\widehat{\theta}}\right]$. When the estimate $\underline{\widehat{\theta}}$ is not integrable, this means, crudely speaking, that we do not know the expected value $E\left[\underline{\widehat{\theta}}\right]$, we only know the lower and upper bounds $\underline{E}\left[\underline{\widehat{\theta}}\right]$ and $\overline{E}\left[\underline{\widehat{\theta}}\right]$ for this mean value. When we know that $E\left[\underline{\widehat{\theta}}\right] \le \theta(J_x)$, we cannot conclude anything about the upper bound, but we can conclude that $\underline{E}\left[\underline{\widehat{\theta}}\right] \le \theta(J_x)$.

Similarly, crudely speaking, we do not know the expected value $E\left[\overline{\widehat{\theta}}\right]$, we only know the lower and upper bounds $\underline{E}\left[\overline{\widehat{\theta}}\right]$ and $\overline{E}\left[\overline{\widehat{\theta}}\right]$ for this mean value. When we know that $\theta(J_x) \le E\left[\overline{\widehat{\theta}}\right]$, we cannot conclude anything about the lower bound, but we can conclude that $\theta(J_x) \le \overline{E}\left[\overline{\widehat{\theta}}\right]$.

Thus, we conclude that $\underline{E}\left[\underline{\widehat{\theta}}\right] \le \theta(J_x) \le \overline{E}\left[\overline{\widehat{\theta}}\right]$, i.e., that

$$\theta(J_x) \in \left[\underline{E}\left[\underline{\widehat{\theta}}\right], \overline{E}\left[\overline{\widehat{\theta}}\right]\right].$$

So, we arrive at the following definition:

Definition 3. *Let $n > 0$ be a positive integer, and let $\mathcal{D}$ be a class of probability distributions on the set $(\mathbb{R}^n)^2$ of all pairs $(x, \widetilde{x})$ of n-dimensional tuples. For each distribution $J \in \mathcal{D}$, we will denote:*

- *the marginal distribution corresponding to x by J_x, and*
- *the marginal distribution corresponding to $\widetilde{x}$ by $J_{\widetilde{x}}$.*

The classes of all such marginal distributions are denoted by $\mathcal{D}_x$ and $\mathcal{D}_{\widetilde{x}}$.

- *By a* parameter*, we mean a mapping $\theta : \mathcal{D}_x \to \mathbb{R}$.*
- *For each parameter θ, by its* unbiased interval estimate*, we mean a function $\widehat{\theta} : \mathbb{R}^n \to \mathbb{I}$ that maps $\mathbb{R}^n$ into the set $\mathbb{I}$ of all intervals for which, for every $J \in \mathcal{D}$, we have*

$$\theta(J_x) \in \left[\underline{E}_J\left[\underline{\widehat{\theta}}(\widetilde{x}_1, \ldots, \widetilde{x}_n)\right], \overline{E}_J\left[\overline{\widehat{\theta}}(\widetilde{x}_1, \ldots, \widetilde{x}_n)\right]\right].$$

Comment. When the interval-values estimate $\widehat{\theta}(\widetilde{x}) = \left[\underline{\widehat{\theta}}(\widetilde{x}), \underline{\widehat{\theta}}(\widetilde{x})\right]$ is integrable, and its expected value is well-defined, the above requirement takes a simpler form

$$\theta(J_x) \in E_J\left[\widehat{\theta}(\widetilde{x}_1, \ldots, \widetilde{x}_n)\right].$$

5 Unbiased Interval Estimates are Often Possible when Unbiased Numerical Estimates are Not Possible

Let us show that for examples similar to the one presented above – for which unbiased *numerical* estimates are not possible – it is possible to have unbiased *interval* estimates.

Proposition 2. *Let $\mathcal{D}_0$ be a class of probability distributions on $\mathbb{R}^n$, let θ be a parameter, let $\widehat{\theta}(x_1, \ldots, x_n)$ be a continuous function which is an unbiased numerical estimate for θ, and let $\Delta_1, \ldots, \Delta_n$ be positive real numbers. Let $\mathcal{D}$ denote the class of all distributions J on $(x, \widetilde{x})$ for which the marginal J_x belongs to $\mathcal{D}_0$ and for which, for all i, we have $|x_i - \widetilde{x}_i| \le \Delta_i$ with probability 1. Then, the following interval-values function is an unbiased interval estimate for θ:*

$$\widehat{\theta}_r(\widetilde{x}_1, \ldots, \widetilde{x}_n) \stackrel{\text{def}}{=}$$

$$\left\{\widehat{\theta}(x_1, \ldots, x_n) : x_1 \in [\widetilde{x}_1 - \Delta_1, \widetilde{x}_1 + \Delta_1], \ldots, x_n \in [\widetilde{x}_n - \Delta_n, \widetilde{x}_n + \Delta_n]\right\}.$$

Comment. Since the function $\widehat{\theta}(x_1, \ldots, x_n)$ is continuous, its range $\widehat{\theta}_r(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ on the box $[\widetilde{x}_1 - \Delta_1, \widetilde{x}_1 + \Delta_1] \times \ldots \times [\widetilde{x}_n - \Delta_n, \widetilde{x}_n + \Delta_n]$ is an interval. Method of estimating this intervals are known as methods of *interval computations*; see, e.g., [1, 2].

Proof. For every tuple $x = (x_1, \ldots, x_n)$, since $|x_i - \widetilde{x}_i| \le \Delta_i$, we have $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. Thus,

$$\theta(x_1, \ldots, x_n) \in$$

$$\left\{\widehat{\theta}(x_1, \ldots, x_n) : x_1 \in [\widetilde{x}_1 - \Delta_1, \widetilde{x}_1 + \Delta_1], \ldots, x_n \in [\widetilde{x}_n - \Delta_n, \widetilde{x}_n + \Delta_n]\right\} =$$

$$\widehat{\theta}_r(\widetilde{x}) = \left[\underline{\widehat{\theta}_r}(\widetilde{x}), \overline{\widehat{\theta}_r}(\widetilde{x})\right]$$

and thus,

$$\underline{\widehat{\theta}_r}(\widetilde{x}) \le \theta(x) \le \overline{\widehat{\theta}_r}(\widetilde{x}).$$

It is known that if $f(x) \le g(x)$, then $\underline{E}(f) \le \underline{E}[g]$ and $\overline{E}[f] \le \overline{E}[g]$. Thus, we get

$$\underline{E}\left[\underline{\widehat{\theta}_r}(\widetilde{x})\right] \le \underline{E}[\theta(x)] \text{ and } \overline{E}[\theta(x)] \le \overline{E}\left[\overline{\widehat{\theta}_r}(\widetilde{x})\right].$$

We have assumed that θ is an unbiased estimate; this means that the man $E[\theta(x)]$ is well defined and equal to $\theta(J_x)$. Since the mean is well-defined, this means that $\underline{E}[\theta(x)] = \overline{E}[\theta(x)] = E[\theta(x)] = \theta(J_x)$. Thus, the above two inequalities take the form

$$\underline{E}\left[\underline{\widehat{\theta}_r}(\widetilde{x})\right] \le \theta(J_x) \le \overline{E}\left[\overline{\widehat{\theta}_r}(\widetilde{x})\right].$$

This is exactly the inclusion that we want to prove. The proposition is thus proven.

6 Case When We Can Have Sharp Unbiased Interval Estimates

Need for sharp unbiased interval estimates. All we required in our definition of an unbiased interval estimate (Definition 3) is that the the actual value θ of the desired parameter is contained in the interval obtained as an expected value of the interval-valued estimates $\widehat{\theta}_r(x)$.

So, if, instead of the original interval-valued estimate $\widehat{\theta}_r(x) = \left[\underline{\widehat{\theta}_r}(x), \overline{\widehat{\theta}_r}(x)\right]$, we take a wider enclosing interval, e.g., an interval $\left[\underline{\widehat{\theta}_r}(x) - 1, \overline{\widehat{\theta}_r}(x) + 1\right]$, this wider interval estimate will also satisfy our definition.

It is therefore desirable to come up with the *narrowest possible* ("sharpest") unbiased interval estimates.

A realistic example where sharp unbiased interval estimates are possible. Let us give a realistic example in which a sharp unbiased interval estimate is possible. This example will be a (slight) generalization of the example on which we showed that an unbiased numerical estimate is not always possible.

Specifically, let us assume that the actual values $x_1, \ldots, x_n$ have a joint normal distribution $N(\mu, \Sigma)$ for some unknown means $\mu_1, \ldots, \mu_n$ and an unknown covariance matrix Σ, and that the only information that we have about the measurement errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$ is that each of these differences is bounded by a known bound $\Delta_i > 0$: $|\Delta x_i| \le \Delta_i$. As we have mentioned earlier, the situation in which we only know the upper bound on the measurement errors (and we do not have any other information about the probabilities) is reasonably frequent in real life.

In this case, $\mathcal{D}$ is the class of all probability distributions for which the marginal distribution J_x is normal, and $|\widetilde{x}_i - x_i| \le \Delta_i$ for all i with probability 1. In other words, the tuple $(x_1, \ldots, x_n)$ is normally distributed, and $\widetilde{x}_i = x_i + \Delta x_i$, where Δx_i can have any distribution for which Δx_i is located on the interval $[-\Delta_i, \Delta_i]$ with probability 1 (the distribution of Δx_i may depend on $x_1, \ldots, x_n$, as long as each Δx_i is located within the corresponding interval).

Let us denote the class of all such distributions by $\mathcal{I}'$. By definition, the corresponding marginal distributions J_x correspond to n-dimensional normal distribution. As a parameter, let us select the average

$$\beta \stackrel{\text{def}}{=} \frac{\mu_1 + \ldots + \mu_n}{n}.$$

For the class of all marginal distributions J_x, there is an unbiased numerical estimate: namely, we can take $\widehat{\beta}(x_1, \ldots, x_n) = \dfrac{x_1 + \ldots + x_n}{n}$. Indeed, one can easily check that since the expected value of each variable x_i is equal to μ_i, the expected value of the estimate $\widehat{\beta}(x)$ is indeed equal to β. Due to Proposition 2, we can conclude that the range

$$\widehat{\beta}_r(\widetilde{x}_1, \ldots, \widetilde{x}_n) \stackrel{\text{def}}{=}$$

$$\left\{\widehat{\beta}(x_1, \ldots, x_n) : x_1 \in [\widetilde{x}_1 - \Delta_1, \widetilde{x}_1 + \Delta_1], \ldots, x_n \in [\widetilde{x}_n - \Delta_n, \widetilde{x}_n + \Delta_n]\right\} =$$

$$\left\{\frac{x_1 + \ldots + x_n}{n} : x_1 \in [\widetilde{x}_1 - \Delta_1, \widetilde{x}_1 + \Delta_1], \ldots, x_n \in [\widetilde{x}_n - \Delta_n, \widetilde{x}_n + \Delta_n]\right\}$$

is an unbiased interval estimate for the parameter β.

This range can be easily computed if we take into account that the function $\widehat{\beta}(x_1, \ldots, x_n) = \dfrac{x_1 + \ldots + x_n}{n}$ is an increasing function of all its variables. Thus:

- the smallest value of this function is attained when each of the variables x_i attains its smallest possible value $x_i = \widetilde{x}_i - \Delta_i$, and
- the largest value of this function is attained when each of the variables x_i attains its largest possible value $x_i = \widetilde{x}_i + \Delta_i$.

So, the range has the form

$$\widehat{\beta}_r(\widetilde{x}_1, \ldots, \widetilde{x}_n) =$$

$$\left[\frac{(\widetilde{x}_1 - \Delta_1) + \ldots + (\widetilde{x}_n - \Delta_n)}{n}, \frac{(\widetilde{x}_1 + \Delta_1) + \ldots + (\widetilde{x}_n + \Delta_n)}{n}\right].$$

Let us show that this unbiased interval estimate is indeed sharp.

Proposition 3. *For the class $\mathcal{I}'$, if $\widehat{\beta'}_r(\widetilde{x})$ is an unbiased interval estimate for β, then for every tuple $\widetilde{x}$, we have $\widehat{\beta}_r(\widetilde{x}) \subseteq \widehat{\beta'}_r(\widetilde{x})$.*

Comment. So, the above interval estimate $\widehat{\beta}_r(\widetilde{x})$ is indeed the narrowest possible.

Proof. Let us pick an arbitrary tuple $\widetilde{y} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)$, and let us show that for this tuple, the interval $\widehat{\beta}_r\,(\widetilde{y}\,)$ is contained in the interval $\widehat{\beta'}_r\,(\widetilde{y}\,)$. To prove this, it is sufficient to prove that both endpoints of the interval $\widehat{\beta}_r\,(\widetilde{y}\,)$ are contained in the interval $\widehat{\beta'}_r\,(\widetilde{y}\,)$. Without losing generality, let us consider the left endpoint $\dfrac{(\widetilde{y}_1 - \Delta_1) + \ldots + (\widetilde{y}_n - \Delta_n)}{n}$ of the interval $\widehat{\beta}_r\,(\widetilde{y}\,)$; for the right endpoint $\dfrac{(\widetilde{y}_1 + \Delta_1) + \ldots + (\widetilde{y}_n + \Delta_n)}{n}$ of this interval, the proof is similar.

To prove that $\dfrac{(\widetilde{y}_1 - \Delta_1) + \ldots + (\widetilde{y}_n - \Delta_n)}{n} \in \widehat{\beta'}_r\,(\widetilde{y}\,)$, we will use the fact that the function $\widehat{\beta'}_r(\widetilde{x})$ is an unbiased interval estimate. Let us consider a distribution $J \in \mathcal{I}'$ for which each value x_i is equal to $\mu_i = \widetilde{y}_i - \Delta_i$ with probability 1, and each value $\widetilde{x}_i$ is equal to $\widetilde{y}_i$ with probability 1. One can easily see that here, $|\widetilde{x}_i - x_i| \le \Delta_i$ and therefore, this distribution indeed belongs to the desired class $\mathcal{I}'$.

For this distribution, since $\mu_i = \widetilde{y}_i - \Delta_i$, the actual value

$$\beta(J_x) = \frac{\mu_1 + \ldots + \mu_n}{n}$$

is equal to

$$\beta(J_x) = \frac{(\widetilde{y}_1 - \Delta_1) + \ldots + (\widetilde{y}_n - \Delta_n)}{n}.$$

On the other hand, since $\widetilde{x}_i = \widetilde{y}_i$ with probability 1, we have $\widehat{\beta'}_r(\widetilde{x}) = \widehat{\beta'}_r\,(\widetilde{y}\,)$ with probability 1, and thus, the expected value of $\widehat{\beta'}_r(\widetilde{x})$ also coincides with the interval $\widehat{\beta'}_r\,(\widetilde{y}\,)$: $E_J\left[\widehat{\beta'}_r(\widetilde{x})\right] = \widehat{\beta'}_r\,(\widetilde{y}\,)$.

So, from the condition that $\beta(J_x) \in E_J\left[\widehat{\beta'}_r(\widetilde{x})\right]$, we conclude that

$$\frac{(\widetilde{y}_1 - \Delta_1) + \ldots + (\widetilde{y}_n - \Delta_n)}{n} \in \widehat{\beta'}_r\,(\widetilde{y}\,),$$

i.e., that the left endpoint of the interval $\widetilde{\beta}_r\,(\widetilde{y}\,)$ indeed belongs to the interval $\widehat{\beta'}_r\,(\widetilde{y}\,)$. We can similarly prove that the right endpoint of the interval $\widetilde{\beta}_r\,(\widetilde{y}\,)$ belongs to the interval $\widehat{\beta'}_r\,(\widetilde{y}\,)$. Thus, the whole interval $\widetilde{\beta}_r\,(\widetilde{y}\,)$ is contained in the interval $\widehat{\beta'}_r\,(\widetilde{y}\,)$. The proposition is proven.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, and by a grant on F-transforms from the Office of Naval Research.

The authors are greatly thankful to John N. Mordeson for his encouragement.

References

[1] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[2] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM Press, Philadelphia, Pennsylvania, 2009.

[3] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York, 2005.