University of Texas at El Paso
ScholarWorks@UTEP

Departmental Technical Reports (CS) | Computer Science

7-2011

Maximum Likelihood Approach to Pointwise Estimation in Statistical Data Processing under Interval Uncertainty

Nitaya Buntao
King Mongkut's Institute of Technology North Bangkok

Sa-aat Niwitpong
King Mongkut's Institute of Technology North Bangkok

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Follow this and additional works at: https://scholarworks.utep.edu/cs_techrep

 Part of the Computer Engineering Commons

Comments:

Technical Report: UTEP-CS-11-39

Recommended Citation

Buntao, Nitaya; Niwitpong, Sa-aat; and Kreinovich, Vladik, "Maximum Likelihood Approach to Pointwise Estimation in Statistical Data Processing under Interval Uncertainty" (2011). *Departmental Technical Reports (CS)*. 649.
https://scholarworks.utep.edu/cs_techrep/649

This Article is brought to you for free and open access by the Computer Science at ScholarWorks@UTEP. It has been accepted for inclusion in Departmental Technical Reports (CS) by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.

Maximum Likelihood Approach to Pointwise Estimation in Statistical Data Processing under Interval Uncertainty

Nitaya Buntao[1], Sa-aat Niwitpong[1], and Vladik Kreinovich[2]

[1]Department of Applied Statistics
King Mongkut's University of Technology North Bangkok
1518 Piboonsongkhram Road, Bangsue
Bangkok 10800 Thailand
taltanot@hotmail.com, snw@kmutnb.ac.th
[2]Computer Science Department, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu

Abstract

Traditional statistical estimates $C(x_1, \ldots, x_n)$ for different statistical characteristics (such as mean, variance, etc.) implicitly assume that we know the sample values $x_1, \ldots, x_n$ exactly. In practice, the sample values $\widetilde{x}_i$ come from measurements and are, therefore, in general, different from the actual (unknown) values x_i of the corresponding quantities. Sometimes, we know the probabilities of different values of the measurement error $\Delta x_i = \widetilde{x}_i - x_i$, but often, the only information that we have about the measurement error is the upper bound Δ_i on its absolute value – provided by the manufacturer of the corresponding measuring instrument. In this case, the only information that we have about the actual values x_i is that they belong to the intervals $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

In general, different values $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ lead to different values of the corresponding statistical characteristic $C(x_1, \ldots, x_n)$. In this case, it is desirable to find the set of all possible values of this characteristic. For continuous estimates $C(x_1, \ldots, x_n)$, this range is an interval.

The values of C are used, e.g., in decision making – e.g., in a control problem, to select an appropriate control value. In this case, we need to select a single value from the corresponding interval. It is reasonable to select a value which is, in some sense, the most probable. In this paper, we show how the Maximum Likelihood approach can provide such a value, i.e., how it can produce pointwise estimates in statistical data processing under interval uncertainty.

1 Formulation of the Problem

Prediction and decision making: deterministic vs. stochastic situations. In practice, we need to *predict* the future state of the system and to make decisions based on these predictions.

In some cases, such predictions are largely deterministic. For example, if we know the current position and velocity of a spaceship, we can use Newton's equations to predict its future trajectory. We can then use these predictions to design a trajectory that will lead the spaceship to the target – e.g., to the desired place on the Moon surface.

In other cases, only probabilistic predictions are possible. For example, even when we have a good record of meteorological parameters such as temperature, wind speed, humidity, rainfall, we can only predict future weather with a certain probability. Similarly, in reliability engineering, we cannot predict which exact component will fail, but we can predict the probabilities of different components failing during a given time interval. We then need to make a decision based on these predictions. For example, in the meteorological case, if there is a probability that the expected total rainfall will exceed the capacity of a dam, we need to reinforce this dam to prevent potential flooding. If there is a probability that both duplicate control systems will fail, then we need to add additional failsafe features to our system. In all these cases, we use some statistical characteristic of the corresponding distribution – such as mean (= expected value), probability, variance, etc. – to make the corresponding decisions.

Need for statistical data processing. Sometimes, we know the exact probability distribution for the desired quantities. However, in most probabilistic situations, we do not know the exact values of the corresponding probabilities. These values must be determined from the observations. In such situations, we use the observed data $\widetilde{x}_1, \ldots, \widetilde{x}_n$ to find the probability distribution that fits this data, and then we make a decision based on the corresponding characteristic of the resulting distribution.

How probability distributions are determined now. Usually, while we do not know the desired probabilities, but, based on the previous observations, we know a typical shape of the corresponding probability distributions. For example, we may not know the exact probability distribution of a rainfall at a given location, but we usually have a sample of rainfall distributions from various locations, and our experience has shown that the distribution at our location shall have the same shape. For each possible combination of the parameters $c_1, \ldots, c_k$, we know the corresponding probability density function $\rho(x, c_1, \ldots, c_k)$.

For example, often, we observe that the distribution of the sample values $x_1, \ldots, x_n$ follows the Gaussian (normal) distribution, with the probability density function

$$\rho(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

with mean μ and variance σ^2. In our notations, this means that we have a probability density function

$$\rho(x, c_1, c_2) = \frac{1}{\sqrt{2\pi} \cdot c_2} \cdot \exp\left(-\frac{(x - c_1)^2}{2c_2^2}\right)$$

with mean $c_1 = \mu$ and standard deviation $c_2 = \sigma$. The fact that many real-life random variables are normally distributed can be explained by the Central Limit Theorem, according to which, crudely speaking, the distribution of the sum of a large number of small random quantities is close to normal; see, e.g., [13].

We assume that all observations are independent, and that each of these observations comes from the same distribution – a distribution from the given family, with unknown values c_j that we need to determine. Due to the independence assumption, for each sample $x_1, \ldots, x_n$ and for each combination of parameters $c_1, \ldots, c_k$, we can thus describe the *likelihood* $L(c_1, \ldots, c_k)$ that the observe data come from the distribution corresponding to these parameters as the product of the corresponding probability densities, i.e., as

$$L(c_1, \ldots, c_k) = \prod_{i=1}^{n} \rho(x, c_1, \ldots, c_k).$$

Then, we select the values c_j for which this likelihood is the largest possible. Such a selection is known as the *Maximum Likelihood* approach; see, e.g., [13].

Example. For the Gaussian distribution, this expression takes the form

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Maximizing this expression is equivalent to minimizing an auxiliary expression

$$\psi(\mu, \sigma) \stackrel{\text{def}}{=} -\ln(L(\mu, \sigma)) = n \cdot \ln(\sqrt{2\pi} \cdot \sigma) + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}.$$

Differentiating this expression with respect to μ and equating the derivative to 0, we conclude that $\mu = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$. Differentiating with respect to σ, we conclude that $\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \mu)^2$. These are standard ways to estimate parameters of a normal distribution – except that often, a slightly different un-biased estimator is used to estimate σ^2, in which the factor $\frac{1}{n}$ is replaced by a slightly different factor $\frac{1}{n-1}$.

Similar formulas are used for many other distributions, e.g., for the lognormal distribution.

Need for interval uncertainty. Traditional statistical estimates $C(x_1, \ldots, x_n)$ for different statistical characteristics (such as mean, variance, etc.) implicitly assume that we know the sample values $x_1, \ldots, x_n$ exactly. In practice, the sample values $\widetilde{x}_i$ come from measurements and are, therefore, in general, different from the actual (unknown) values x_i of the corresponding quantities. Sometimes, we know the probabilities $\rho_i(\Delta x_i)$ of different values of the measurement error $\Delta x_i = \widetilde{x}_i - x_i$, but often, the only information that we have about the measurement error is the upper bound Δ_i on its absolute value – provided by the manufacturer of the corresponding measuring instrument. In this case, the only information that we have about the actual values x_i is that they belong to the intervals $\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$; see, e.g., [12].

In general, different values $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ lead to different values of the corresponding statistical characteristic $C(x_1, \ldots, x_n)$. In this case, it is desirable to find the range of all possible values of this characteristic:

$$\mathbf{C} = \{C(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

For continuous estimates $C(x_1, \ldots, x_n)$, this range is an interval $\mathbf{C}$.

For different statistical characteristics, there exist numerous efficient algorithms for computing the interval ranges of these characteristics under interval uncertainty; see, e.g., [2, 3, 5, 6, 7, 8, 9, 10, 11, 15].

Need for pointwise estimates. The values of C are used, e.g., in decision making – e.g., in a control problem, to select an appropriate control value. In this case, we need to select a single value from the corresponding interval of possible values of C.

For example, in many applications of meteorological predictions – e.g., in agriculture – decisions are based on the expected values of the average temperature, expected rainfall, average humidity, sometimes taking into account the corresponding variance. Thus, the users expect to see numbers representing these averages or variances – and they use these numbers to decide, e.g., how much fertilizer to buy or how much crop to expect.

Which point C from the interval $\mathbf{C}$ should we choose?

2 Main Idea

Analysis of the problem. In the traditional statistical analysis, the unknowns are the parameters $c_1, \ldots, c_k$ of the corresponding statistical distribution. According to the Maximum Likelihood approach, we select the "most probable" values of these parameters. Once these values are selected, the corresponding probability distribution is well defined. Based on these distributions, we can compute the estimates for the desired statistical characteristics.

In our case, in addition to not knowing the parameters $c_1, \ldots, c_k$ of the sample distribution, we also do not know the probability distributions of the

measurement errors $\Delta x_1, \ldots, \Delta x_n$ corresponding to all n measurements. To be more precise, we know that these distributions are located in the corresponding intervals $[-\Delta_i, \Delta_i]$, but the exact probability density functions $\rho_i(\Delta x_i)$ corresponding to these measurement errors are not known.

Idea. When we only had the parameters c_j as unknowns, we used the Maximum Likelihood method to find the values for which the likelihood is the largest. A natural idea is to use the Maximum Likelihood approach in our new case as well, i.e., to find *both* the values c_j *and* the distributions $\rho_i(\Delta x_i)$ from the condition that the corresponding likelihood is the largest possible.

How to implement this idea. For given values $c_1, \ldots, c_k$, the probability density for different actual values x is equal to $\rho(x, c_1, \ldots, c_k)$. We assume that different values from the sample are independent. So, if we have n actual values $x_1, \ldots, x_n$, then the corresponding probability density is equal to

$$\prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k).$$

Measurement errors are also assumed to be independent, so for any set of values $\widetilde{x}_i$ and x_i, the probability of the corresponding sequence of measurement errors is equal to

$$\prod_{i=1}^{n} \rho_i(\widetilde{x}_i - x_i).$$

The total probability density is given by the product:

$$\prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) \cdot \prod_{i=1}^{n} \rho_i(\widetilde{x}_i - x_i).$$

In practice, we do not know the actual values x_i, we only know the measurement results $\widetilde{x}_i$. The probability of getting the observed measurement results can be obtained by the formula of full probability, when we integrate over all possible values of x_i. As a result, we get the following formula for the probability of observing the values $\widetilde{x}_1, \ldots, \widetilde{x}_n$:

$$L = \int \ldots \int \prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) \cdot \prod_{i=1}^{n} \rho_i(\widetilde{x}_i - x_i)\, dx_1 \ldots\, dx_n. \qquad (1)$$

Our objective is, given the observations $\widetilde{x}_1, \ldots, \widetilde{x}_n$, to find the values c_j and the distributions $\rho_i(\Delta x_i)$ for which the value L is the largest possible.

To describe the most general probability distributions $\rho_i(\Delta x_i)$, including the ones that are located at a single point, we must consider not only continuous functions $\rho_i(\Delta x_i)$, but also generalized functions – e.g., a delta-function $\delta(\Delta x_i)$ that represents a distribution located at a single point 0 – i.e., a random variable which is equal to 0 with probability 1.

Thus, we arrive at the following formulation:

Precise formulation of the problem.

- *Given:* n values $\widetilde{x}_1, \ldots \widetilde{x}_n$, and a function $\rho(x, c_1, \ldots, c_k)$.
- *Find:* the values c_j and the distributions $\rho_i(\Delta x_i)$ for which the expression (1) attains the largest possible values.

Comment. Once the values $c_1, \ldots, c_k$ are determined, we use the corresponding probability distribution $\rho(x, c_1, \ldots, c_k)$ to determine the pointwise estimates of all the desired statistical characteristics.

3 Main Result

Discussion. The original optimization problem (1) is difficult to solve since in this problem, we need to optimize not only over all possible *values* $c_1, \ldots, c_k$, but also over all possible *functions* $\rho_i(\Delta x_i)$. To make this problem easier to solve, we will reduce it to the standard optimization problem, in which there are only unknown values, but no unknown functions.

Main result: formulation. We will prove that the largest value of the expression (1) is attained for the values $c_1, \ldots, c_k$ for which the auxiliary expression

$$\widetilde{L} = \prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) \tag{2}$$

attains its largest possible value as a function of the variables $x_1 \in \mathbf{x}_1$, $\ldots$, $x_n \in \mathbf{x}_n$, c_1, $\ldots$, c_k.

Proof. Let $x_1^{\rm opt} \in \mathbf{x}_1$, $\ldots$, $x_n^{\rm opt} \in \mathbf{x}_n$, $c_1^{\rm opt}$, $\ldots$, $c_k^{\rm opt}$ be the values for which the expression $\widetilde{L}$ attains its largest value $L^{\rm opt}$:

$$L^{\rm opt} = \prod_{i=1}^{n} \rho(x_i^{\rm opt}, c_1^{\rm opt}, \ldots, c_k^{\rm opt}) =$$

$$\max\left\{\prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n, c_1, \ldots, c_k\right\}.$$

Then, for the values $c_j = c_j^{\rm opt}$ and for the distributions

$$\rho_i(\Delta x_i) = \delta((\Delta x_i + \widetilde{x}_i) - x_i^{\rm opt}) = \delta(x_i - x_i^{\rm opt})$$

which are located, with probability 1, on the values $x_i = x_i^{\rm opt}$, the expression (1) takes the form

$$L = \int \ldots \int \prod_{i=1}^{n} \rho(x_i, c_1^{\rm opt}, \ldots, c_k^{\rm opt}) \cdot \prod_{i=1}^{n} \delta(x_i - x_i^{\rm opt})\, dx_1 \ldots\, dx_n.$$

Here, as usual, $\delta(x)$ denotes Dirac's delta-function (a generalized function that describes the probability density of a random variable which is located at point 0 with probability 1).

By definition of the delta-function, this means that this value L is equal to

$$\prod_{i=1}^{n} \rho(x_i^{\rm opt}, c_1^{\rm opt}, \ldots, c_k^{\rm opt}),$$

i.e., to $L^{\rm opt}$. Since we have selected $x_i^{\rm opt}$ and $c_j^{\rm opt}$ as the values for which the expression (2) attains its maximum, we can thus conclude that

$$L^{\rm opt} = \max\left\{\prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n, c_1, \ldots, c_k\right\}.$$

To prove our result, it is sufficient to show that for every other values c_j and for every other set of distributions $\rho_i(\Delta x_i)$, the value L (defined by the expression (1)) is smaller than or equal to $L^{\rm opt}$.

Indeed, for each function, its expected value is smaller than or equal than its maximum. The expression (1) is the expected value of the expression (2) under the distributions for $x_1, \ldots, x_n$ with probability density $\prod\limits_{i=1}^{n} \rho_i(\Delta x_i)$. Thus, for every set of values $c_1, \ldots, c_k$, and for every set of probability distributions $\rho_i(\Delta x_i)$, we have

$$L \le \max\left\{\prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\right\}.$$

The right-hand side of this expression, in turn, is bounded by the largest value of this expression over all possible parameters $c_1, \ldots, c_k$:

$$\max\left\{\prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\right\} \le$$

$$\max\left\{\prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n, c_1, \ldots, c_k\right\}.$$

By definition of $L^{\rm opt}$, this means that

$$\max\left\{\prod_{i=1}^{n} \rho(x_i, c_1, \ldots, c_k) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\right\} \le L^{\rm opt}.$$

This means that $L \le L^{\rm opt}$,, i.e., that the value L of the expression (1) is indeed smaller than or equal to $L^{\rm opt}$.

Thus, the value $L^{\rm opt}$ is indeed the largest possible values of the maximum likelihood expression (1). The statement is proven.

Observation. Once the values $x_1^{\rm opt}, \ldots, x_n^{\rm opt}$ are determined, the values $c_1, \ldots, c_k$ can be found from the condition that the likelihood is the largest – i.e., from the standard Maximum Likelihood method.

4 First Example: Normal Distributions

Description of the example. Let us first apply our main idea to the case when the shape function $\rho(x, c_1, c_2)$ describes normal distribution.

Analysis of the problem. In this case, once we fix the optimal values $x_1, \ldots, x_n$, the corresponding values of the mean $\mu = c_1$ and of the standard deviation $\sigma = c_2$ are determined by the usual Maximum Likelihood approach.

As we have shown, this means that we take $\mu = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$ and $\sigma^2 = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} (x_i - \mu)^2$. The corresponding value of the likelihood is equal to

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \frac{1}{(\sqrt{2\pi})^n \cdot \sigma^n} \cdot \exp\left(-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Because of our selection of σ^2, we have

$$\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} = \frac{n}{2\sigma^2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 = \frac{n}{2\sigma^2} \cdot \sigma^2 = \frac{n}{2}.$$

Thus, the above expression for L takes the form

$$L = \frac{1}{(\sqrt{2\pi})^n \cdot \sigma^n} \cdot \exp\left(-\frac{n}{2}\right).$$

This expression is the largest if and only the variance $V = \sigma^2$ is the smallest. Thus, to find the corresponding values c_j, we first need to find the values $x_i \in \mathbf{x}_i$ for which the sample variance takes the smallest possible value.

Resulting algorithm. We have shown that for the normal distribution, the maximum likelihood approach to pointwise estimation implies that we select the values $x_1 \in \mathbf{x}_1 = [\underline{x}_1, \overline{x}_1]$, $\ldots$, $x_n \in \mathbf{x}_n = [\underline{x}_n, \overline{x}_n]$ for which the sample variance $\sigma^2 = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} (x_i - \mu)^2$ attains its smallest possible value.

Algorithms for computing the smallest possible values of the sample variance under such interval uncertainty – and of finding the values $x_i \in \mathbf{x}_i$ for which this minimal variance is attained – are given in [5, 6, 7, 9, 10, 15]. These algorithms are based on the fact (proved in these papers) that when the minimum of the

variance is attained, for each interval $\mathbf{x}_i$, the corresponding minimizing value x_i is the closest to the mean $\mu = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$.

In particular, in [15], based on this fact, we present a sophisticated linear-time algorithm for computing these values. The easiest to explain is the following slightly slower $O(n \cdot \log(n))$ algorithm:

- First, we sort all $2n$ endpoints $\underline{x}_i$ and $\overline{x}_i$ in increasing order, into a sequence $r_1 \le r_2 \le \ldots \le r_{2n}$. To this sequence, we add values $r_0 = -\infty$ and $r_{2n+1} = +\infty$; as a result, we divide the real line into $2n+1$ *zones* $[r_k, r_{k+1}]$, $k = 0, 1, \ldots, 2n$.
- For each zone $[r_k, r_{k+1}]$, and for each i, we find the value $x_{i,k} \in \mathbf{x}_i$ which is minimizing under the assumption that the corresponding mean μ_k belongs to this zone:
  - when the interval $\mathbf{x}_i$ is fully to the left of the zone, i.e., when $\overline{x}_i \le r_k$, we take $x_{i,k} = \overline{x}_i$;
  - when the interval $\mathbf{x}_i$ is fully to the right of the zone, i.e., when $r_{k+1} \le \underline{x}_i$, we take $x_{i,k} = \underline{x}_i$;
  - finally, in the remaining cases, i.e., when the interval contains the zone, we mark $x_{i,k}$ as equal to the (still to be determined) value μ_k.

  Then, we find the value μ_k from the condition that it is equal to the mean of all selected values, i.e., that

$$\sum_{i:\overline{x}_i \le r_k} \overline{x}_i + \sum_{i:r_{k+1} \le \underline{x}_i} \underline{x}_i + \sum_{\text{other } i} \mu_k = n \cdot \mu_k.$$

  The resulting value is equal to $\mu_k = \frac{N_k}{D_k}$, where

$$N_k = \sum_{i:\overline{x}_i \le r_k} \overline{x}_i + \sum_{i:r_{k+1} \le \underline{x}_i} \underline{x}_i;$$

$$D_k = \#\{i : \overline{x}_i \le r_k\} + \#\{i : r_{k+1} \le \underline{x}_i\}.$$

  If this value μ_k does not belong to the zone $[r_k, r_{k+1}]$, this means that our initial assumption that μ_k is within this zone is inconsistent, so we move to the next zone. If the resulting value μ_k is within the zone, we use the selected values x_i to compute the sample variance $V_k = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_{i,k} - \mu_k)^2$. To simplify the computations, we can use the known fact that, in general, the variance $V = \sigma^2$ can be represented as $V = M - \mu^2$, where $M = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2$. In our case, we have

$$V_k = \frac{1}{n} \cdot (M_k + (n - D_k) \cdot \mu_k^2) - \mu_k^2,$$

where we denoted

$$M_k = \sum_{i:\overline{x}_i \le r_k} (\overline{x}_i)^2 + \sum_{i:r_{k+1} \le \underline{x}_i} (\underline{x}_i)^2.$$

Thus, we conclude that $V_k = \dfrac{1}{n} \cdot M_k - \dfrac{D_k}{n} \cdot \mu_k^2$.

- The smallest of the resulting values V_k is then returned as the smallest possible value of the variance, and the corresponding value μ_k is then returned as the corresponding value of the mean.

Computational comment. To speed up computations, when we move from one zone to the next, we do not re-compute the values N_k, D_k, and M_k from scratch: we use the previous values and only add and delete terms that changed. As we go from $k = 0$ to $k = 2n$, each value $x_{i,k}$ changes at most twice: from $\underline{x}_i$ to μ_k and then from μ_k to $\overline{x}_i$. Thus, each term has to be recomputed only twice, thus, after sorting (which takes time $O(n \cdot \ln(n))$) we only need linear time to find all the values N_k, D_k, and M_k.

Comment. To avoid confusion, we should mention that our description is slightly different from the description given in [5, 6, 7, 9, 10, 15], since in these papers, we were only interested in the smallest value of the variance, but now, we are also interested in the values x_i where this smallest value is attained – and in the corresponding value of the mean.

5 Second Example: Lognormal Distribution

Description of the example. Let us now apply our main idea to the case when the shape function $\rho(x, c_1, c_2)$ describes a lognormal distribution, with the probability density function

$$\rho(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot x} \cdot \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right).$$

Analysis of the problem. For lognormal distribution, likelihood has the form

$$L(\mu, \sigma, x_1, \ldots, x_n) = \prod_{i=1}^{n} \rho(x_i, \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot x_i} \cdot \exp\left(-\frac{(\ln(x_i) - \mu)^2}{2\sigma^2}\right).$$

According to our approach, we need to find both the optimal values of the parameters μ and σ and the optimal values $x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n$ from the requirement that the resulting value of the likelihood L is the largest possible. In particular, this means that once we fix the optimal values $x_1, \ldots, x_n$, then the corresponding values of the parameters μ and σ are determined by the usual Maximum Likelihood approach.

Maximizing the likelihood $L(\mu, \sigma, x_1, \ldots, x_n)$ is equivalent to minimizing the auxiliary function

$$\psi(\mu, \sigma, x_1, \ldots, x_n) = -\ln(L(\mu, \sigma, x_1, \ldots, x_n)) =$$

$$n \cdot \ln(\sqrt{2\pi} \cdot \sigma) + \sum_{i=1}^{n} \ln(x_i) + \sum_{i=1}^{n} \frac{(\ln(x_i) - \mu)^2}{2\sigma^2}.$$

Differentiating this expression with respect to μ and equating the derivative to 0, we conclude that $\mu = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln(x_i)$. Differentiating with respect to σ and equating derivative to 0, we conclude that $\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (\ln(x_i) - \mu)^2$. Substituting these values into the expression for ψ and taking into account that $\sum_{i=1}^{n} \ln(x_i) = n \cdot \mu$, we conclude that

$$\psi = n \cdot \ln(\sqrt{2\pi} \cdot \sigma) + n \cdot \mu + \frac{n}{2} = n \cdot \ln(\sqrt{2\pi}) + n \cdot \ln(\sigma) + n \cdot \mu + \frac{n}{2}.$$

We are interested in the values $x_1, \ldots, x_n$ for which this expression is the smallest possible. The location of the minimum does not change if we subtract the constant $n \cdot \ln(\sqrt{2\pi}) + \frac{n}{2}$ from the objective function and divide the resulting expression by a constant n. Thus, the values $x_i \in [\underline{x}_i, \overline{x}_i]$ can be found from the condition that the following expression attains its smallest possible value: $\varphi = \mu + \ln(\sigma)$, i.e.,

$$\varphi = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln(x_i) + \frac{1}{2} \cdot \ln\left(\frac{1}{n} \cdot \sum_{i=1}^{n} (\ln(x_i))^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \ln(x_i)\right)^2\right).$$

This expression can simplified if instead of the original unknowns $x_i \in [\underline{x}_i, \overline{x}_i]$ we consider the new unknowns $y_i = \ln(x_i)$ for which $y_i \in [\underline{y}_i, \overline{y}_i]$, where $\underline{y}_i = \ln(\underline{x}_i)$ and $\overline{y}_i = \ln(\overline{y}_i)$. In terms of these new unknowns y_i, the minimized expression φ takes the form

$$\varphi = \mu + \ln(\sigma),$$

where

$$\mu = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i$$

and

$$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} y_i\right)^2.$$

With respect to reach unknown x_i, the minimum of the function φ on the interval $[\underline{y}_i, \overline{y}_i]$ is attained either inside the interval, or at one of its endpoints.

- If the minimum is attained inside the interval, then, according to calculus, the corresponding partial derivative is equal to 0: $\dfrac{\partial \varphi}{\partial y_i} = 0$.
- If the minimum is attained for $y_i = \underline{y}_i$, then we should have $\dfrac{\partial \varphi}{\partial y_i} \geq 0$ – because if $\dfrac{\partial \varphi}{\partial y_i} < 0$, then, for small $h > 0$, the value $\varphi(y_1, \ldots, y_{i-1}, y_i + h, y_{i+1}, \ldots, y_n)$ would be smaller than the minimal value $\varphi(y_1, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_n)$.
- Similarly, if the minimum is attained for $y_i = \overline{y}_i$, then we should have $\dfrac{\partial \varphi}{\partial y_i} \leq 0$.

To use this property, let us find the explicit expression for this partial derivative. We have

$$\frac{\partial \mu}{\partial y_i} = \frac{\partial}{\partial y_i}\left(\frac{1}{n} \cdot \sum_{i=1}^{n} y_i\right) = \frac{1}{n}.$$

For $\sigma^2 = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} y_i^2 - \left(\dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} y_i\right)^2$, we have

$$\frac{\partial \sigma^2}{\partial y_i} = \frac{1}{n} \cdot (2 \cdot y_i) - 2 \cdot \mu \cdot \frac{1}{n} = \frac{2}{n} \cdot (y_i - \mu).$$

Thus, due to the chain rule, for $\ln(\sigma) = \dfrac{1}{2} \cdot \ln(\sigma^2)$, the derivative is equal to

$$\frac{\partial \ln(\sigma)}{\partial y_i} = \frac{1}{2} \cdot \frac{\partial \ln(\sigma^2)}{\partial y_i} = \frac{1}{2} \cdot \frac{1}{\sigma^2} \cdot \frac{\partial \sigma^2}{\partial y_i} = \frac{1}{2} \cdot \frac{1}{\sigma^2} \cdot \frac{2}{n} \cdot (y_i - \mu) = \frac{y_i - \mu}{n \cdot \sigma^2}.$$

So, for $\varphi = \mu + \ln(\sigma)$, the partial derivative is equal to

$$\frac{\partial \varphi}{\partial y_i} = \frac{\partial \mu}{\partial y_i} + \frac{\partial \ln(\sigma)}{\partial y_i} = \frac{1}{n} + \frac{y_i - \mu}{n \cdot \sigma^2} = \frac{y_i - (\mu - \sigma^2)}{n \cdot \sigma^2}.$$

Since $n \cdot \sigma^2 > 0$, the sign of the derivative coincides with the sign of the numerator $z \stackrel{\text{def}}{=} \mu - \sigma^2$:

- $\dfrac{\partial \varphi}{\partial y_i} = 0$ if and only if $y_i - z = 0$, i.e., if and only if $y_i = z$;
- $\dfrac{\partial \varphi}{\partial y_i} \geq 0$ if and only if $y_i - z \geq 0$, i.e., if and only if $y_i \geq z$; and
- $\dfrac{\partial \varphi}{\partial y_i} \leq 0$ if and only if $y_i - z \leq 0$, i.e., if and only if $y_i \leq z$.

So, the above calculus-related property take the following form:

- if the minimum is attained strictly inside the interval $[\underline{y}_i, \overline{y}_i]$, then $\dfrac{\partial \varphi}{\partial y_i} = 0$ and thus, the corresponding values of y_i is equal to z;
- if the minimum is attained at $y_i = \underline{y}_i$, then $\dfrac{\partial \varphi}{\partial y_i} \geq 0$ and thus, we have $y_i \geq z$;
- finally, if the minimum is attained at $y_i = \overline{y}_i$, then $\dfrac{\partial \varphi}{\partial y_i} \leq 0$ and thus, we have $y_i \leq z$.

Thus, if the interval $[\underline{y}_i, \overline{y}_i]$ is located completely to the left of the point z, i.e., if $\overline{y}_i < z$ and thus, $y_i < z$ for all $y_i \in [\underline{y}_i, \overline{y}_i]$, then:

- the minimum cannot be attained strictly inside the interval $[\underline{y}_i, \overline{y}_i]$, because then we would have $y_i = z$ for some y_i from this interval;
- the minimum also cannot be attained for $y_i = \underline{y}_i$, because then we would have $\underline{y}_i = y_i \geq z$.

So, in this case, the minimum is attained when $y_i = \overline{y}_i$.

Similarly, if the interval $[\underline{y}_i, \overline{y}_i]$ is located completely to the right of the point z, i.e., if $z < \underline{y}_i$ and thus, $z < y_i$ for all $y_i \in [\underline{y}_i, \overline{y}_i]$, then:

- the minimum cannot be attained strictly inside the interval $[\underline{y}_i, \overline{y}_i]$, because then we would have $y_i = z$ for some y_i from this interval;
- the minimum also cannot be attained for $y_i = \overline{y}_i$, because then we would have $\overline{y}_i = y_i \leq z$.

So, in this case, the minimum is attained when $y_i = \underline{y}_i$.

Finally, if the interval $[\underline{y}_i, \overline{y}_i]$ contains the point z inside, then:

- if the minimum is attained at a point y_i which is strictly inside this interval, then $y_i = z$;
- if the minimum is attained at a point $y_i = \underline{y}_i$, then we should have $\underline{y}_i \geq z$; since $z \in [\underline{y}_i, \overline{y}_i]$, this inequality is only possible when $y_i = z$;
- similarly, if the minimum is attained at a point $y_i = \overline{y}_i$, then we should have $\overline{y}_i \leq z$; since $z \in [\underline{y}_i, \overline{y}_i]$, this inequality is only possible when $y_i = z$.

In all three cases, we have $y_i = z$.

So, we can make the following conclusion about the point y_i at which the minimum is attained:

- if $\overline{y}_i < z$, then the minimum is attained at the point $y_i = \overline{y}_i$;
- if $z < \underline{y}_i$, then the minimum is attained at the point $y_i = \underline{y}_i$;
- if $\underline{y}_i \leq z \leq \overline{y}_i$, then the minimum is attained at the point $y_i = z$.

In all these cases, the minimum is attained at a point which is the closest to z among all the points of the interval $[\underline{y}_i, \overline{y}_i]$.

Once we know where z is with respect to all the endpoints $\underline{y}_i$ and $\overline{y}_i$, we can uniquely determine all the minimizing values y_i – on the condition that we know the corresponding value $z = \mu - \sigma^2$. This value z can be found from the condition that $z = \mu - \sigma^2 = \mu - M + \mu^2$, where $M = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i^2$. In terms of the notations that we used for the case of the normal distribution, we get $\mu = \dfrac{1}{n} \cdot (N_k + (n - D_k) \cdot z)$ and $M = \dfrac{1}{n} \cdot (M_k + (n - D_k) \cdot z^2)$. Thus, the above equation takes the form

$$z = \frac{1}{n} \cdot (N_k + (n - D_k) \cdot z) - \frac{1}{n} \cdot (M_k + (n - D_k) \cdot z^2) + \frac{1}{n^2} \cdot (N_k + (n - D_k) \cdot z)^2.$$

This equation is quadratic in z, so it is easy to solve and find two possible values of z. Specifically, if we open the parentheses, multiply both sides by n^2, and move all the terms to the left-hand side, we get

$$z^2(n \cdot (n - D_k) - (n - d_k)^2) - z \cdot (n \cdot (n - D_k) + 2N_k \cdot (n - D_k)) -$$

$$(n \cdot N_k + n \cdot M_k + N_k^2) = 0,$$

i.e., simplifying,

$$a \cdot z^2 - b \cdot z - c = 0,$$

where $a = D_k \cdot (n - D_k)$, $b = (n + 2N_k) \cdot (n - D_k)$, and $c = n \cdot (N_k + M_k) + N_k^2$. Thus,

$$z_{\pm} = \frac{b \pm \sqrt{b^2 + 4a \cdot c}}{2a}.$$

As a result, we arrive at the following algorithm.

Resulting algorithm. First, we sort all $2n$ endpoints $\underline{y}_i$ and $\overline{y}_i$ in increasing order, into a sequence $r_1 \le r_2 \le \ldots \le r_{2n}$. To this sequence, we add values $r_0 = -\infty$ and $r_{2n+1} = +\infty$; as a result, we divide the real line into $2n + 1$ *zones* $[r_k, r_{k+1}]$, $k = 0, 1, \ldots, 2n$.

For each zone $[r_k, r_{k+1}]$, and for each i, we find the value $y_{i,k} \in \mathbf{y}_i$ which is minimizing under the assumption that the corresponding value z belongs to this zone:

- when the interval $\mathbf{y}_i$ is fully to the left of the zone, i.e., when $\overline{y}_i < r_k$, we take $y_{i,k} = \overline{y}_i$;
- when the interval $\mathbf{y}_i$ is fully to the right of the zone, i.e., when $r_{k+1} < \underline{y}_i$, we take $y_{i,k} = \underline{y}_i$;
- finally, in the remaining cases, i.e., when the interval contains the zone, we mark $y_{i,k}$ as equal to the (still to be determined) value z.

To find the value z, we compute

$$N_k = \sum_{i:\overline{y}_i < r_k} \overline{y}_i + \sum_{i:r_{k+1} < \underline{y}_i} \underline{y}_i;$$

$$D_k = \#\{i : \overline{y}_i < r_k\} + \#\{i : r_{k+1} < \underline{y}_i\};$$

$$M_k = \sum_{i:\overline{y}_i < r_k} (\overline{y}_i)^2 + \sum_{i:r_{k+1} < \underline{y}_i} (\underline{y}_i)^2.$$

Then, we compute the values $a_k = D_k \cdot (n - D_k)$, $b_k = (n + 2N_k) \cdot (n - D_k)$, and $c_k = n \cdot (N_k + M_k) + N_k^2$, and

$$z_{k,\pm} = \frac{b_k \pm \sqrt{b_k^2 + 4a_k \cdot c_k}}{2a_k}.$$

If both values are outside the zone $[r_k, r_{k+1}]$, this means that our initial assumption that z is within this zone is inconsistent, so we move to the next zone. If one or both of the values $z_{k,\pm}$ is within the zone, we compute the values

$$\mu_{k,\pm} = \frac{1}{n} \cdot (N_k + (n - D_k) \cdot z_{k,\pm}); \quad M_{k,\pm} = \frac{1}{n} \cdot (M_k + (n - D_k) \cdot z_{k,\pm}^2),$$

$$\sigma_{k,\pm} = \sqrt{M_{k,\pm} - mu_{k,\pm}^2},$$

and

$$\psi_{k,\pm} = \mu_{k,\pm} + \ln(\sigma_{k,\pm}).$$

We then take the smallest of the resulting values $\psi_{k,\pm}$; the corresponding values $\mu_{k,\pm}$ and $\sigma_{k,\pm}$ are then returned as the desired pointwise estimates of the corresponding parameters of the lognormal distribution.

Computational comment. Similarly to the case of the normal distribution, when we move from one zone to the next, we do not re-compute the values N_k, D_k, and M_k from scratch: we use the previous values and only add and delete terms that changed. As we go from $k = 0$ to $k = 2n$, each value $y_{i,k}$ changes at most twice: from $\underline{y}_i$ to z and then from z to $\overline{y}_i$. Thus, each term has to be recomputed only twice, thus, after sorting (which takes time $O(n \cdot \ln(n))$ we only need linear time to find all the values N_k, D_k, and M_k. So, overall, our algorithm takes time $O(n \cdot \ln(n))$.

6 Third Example: Delta-Lognormal Distribution

Need for delta-lognormal distributions. In many practical applications, e.g., in medical applications and in meteorology, a quantity can take any non-negative values but have a positive probability of 0 values. In many such cases,

the probabilities are described by the *delta-lognomal* distribution, in which with a given probability $d > 0$, we get a value 0, and with the remaining probability $1 - d$, we get a lognormal distribution; see, e.g., [1, 4, 14].

In medical applications, in distribution of test costs, zeros correspond to the cases when a patient refused a test. In environmental applications, zeros correspond to the case when the actual concentration of the analyzed chemical is below the detection limit. In biological applications, e.g., in distribution of certain species in different geographic areas, zeros correspond to areas with are unsuitable for these species, etc.

The corresponding probability density has the form

$$\rho(x, \mu, \sigma, d) = d \cdot \delta(x) + (1 - d) \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot x} \cdot \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right),$$

where, as before, $\delta(x)$ denotes Dirac's delta-function (a generalized function that describes the probability density of a random variable which is located at point 0 with probability 1).

Formulation of the problem. As in the previous two examples, we have n intervals $[\underline{x}_i, \overline{x}_i]$ that contain the (unknown) values x_i which are delta-lognormally distributed, and we need to find the "most probable" values of the parameters μ, σ, and δ which are consistent with the given interval observations.

Challenge. We would like to use the above maximum likelihood approach to find these parameters, i.e., to look for the values that maximize the likelihood

$$L(\mu, \sigma, d, x_1, \ldots, x_n) = \prod_{i=1}^{n} \rho(x_i, \mu, \sigma, d).$$

The problem with this idea is that the delta-function is a generalized function, its value for $x = 0$ is infinite. As a result, when one of the values x_i is equal to 0, we get an infinite value of the probability density and thus, the infinite value of the likelihood. In other words, if we have two different combinations of values in both of which one of the values of x_i is 0, we cannot select which one is better, like for both combinations, the likelihood is infinite.

Solution. A natural solution to this problem comes from the recalling that one of the reasons for the ubiquity of delta-lognormal distribution is that when the actual (positive) value is below the sensor's detection limit ℓ, we record the measurement result as 0. It is therefore reasonable, instead of the actual delta-function which is located at exactly 0 with probability 1, to consider the approximate distribution $\delta_\ell(x)$ – e.g., uniform on the interval $[0, \ell]$ – and then tend ℓ to 0. This is, by the way, how delta-functions are often interpreted: as limits of distributions which are located on smaller and smaller intervals $[0, \ell]$ when the width ℓ of these intervals tends to 0. Let us describe what will result from this idea.

Analysis of the problem. For the uniform distribution on the interval $[0, \ell]$, the probability density $\delta_\ell(x)$ is equal to $1/\ell$ for values from this interval and to 0 outside. Thus, the corresponding approximation to the delta-lognormal distribution has the following form:

$$\rho_\ell(x_i, \mu, \sigma, d) = d \cdot \delta_\ell(x_i) + (1-d) \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot x_i} \cdot \exp\left(-\frac{(\ln(x_i) - \mu)^2}{2\sigma^2}\right),$$

When $x_i > 0$, then, for sufficiently small ℓ, the first term becomes 0 when $\ell < x_i$. Thus, for sufficiently small ℓ, the probability density is equal to

$$\rho_\ell(x_i, \mu, \sigma, d) = (1-d) \cdot \rho_{LN}(x_i, \mu, \sigma, d),$$

where

$$\rho_{LN}(x_i, \mu, \sigma) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot x_i} \cdot \exp\left(-\frac{(\ln(x_i) - \mu)^2}{2\sigma^2}\right)$$

is the probability density of the lognormal distribution.

When $x_i = 0$, then for the uniform distribution on the interval $[0, \ell]$, we have $\delta_\ell(x_i) = \dfrac{1}{\ell}$, while the lognormal term is equal to 0. Thus, for $x_i = 0$, we have

$$\rho_\ell(x_i, \mu, \sigma, d) = \frac{d}{\ell}.$$

Therefore,

$$L(\mu, \sigma, d, x_1, \ldots, x_n) = \prod_{i=1}^{n} \rho(x_i, \mu, \sigma, d) =$$

$$\prod_{i: x_i = 0} \rho(x_i, \mu, \sigma, d) \cdot \prod_{i: x_i \neq 0} \rho(x_i, \mu, \sigma, d) =$$

$$\left(\frac{d}{\ell}\right)^z \cdot \prod_{i: x_i \neq 0} ((1-d) \cdot \rho_{LN}(x_i, \mu, \sigma, d)),$$

where z is the number of zeros, i.e., the number of indices i for which $x_i = 0$.

We want to select the values μ, σ, d, x_1, $\ldots$, x_n for which, for sufficiently small ℓ, this value is the largest. When $z = 0$, the above value of $L(\mu, \sigma, d, x_1, \ldots, x_n)$ does not depend on ℓ at all. When $z > 0$, this value tends to ∞ as $\text{const} \cdot \ell^{-z}$.

When $z > z'$, then, for sufficiently small ℓ, we have

$$\text{const} \cdot \ell^{-z} > \text{const}' \cdot \ell^{-z'}.$$

This inequality is easy to prove: multiplying both sides by ℓ^z, we get an equivalent form $\text{const} > \text{const}' \cdot \ell^{z-z'}$, which is true since $\text{const} > 0$ and, for $z > z'$, we have $\ell^{z-z'} \to 0$ and thus, $\text{const}' \cdot \ell^{z-z'} < \text{const}$ for sufficiently small ℓ.

So, among all combinations with different values z, according to our criterion, we select the one with the largest possible value z, i.e., with the largest possible

number of indices for which $x_i = 0$. For each index i, its possible values are in the interval $[\underline{x}_i, \overline{x}_i]$. So, to get the largest possible value of zeros, for each interval that contains 0, we take $x_i = 0$.

Delta-lognormal distribution describes non-negative variables, hence we have $\underline{x}_i \geq 0$. The resulting interval $[\underline{x}_i, \overline{x}_i]$ contains 0 if and only $\underline{x}_i = 0$.

The next question is how to determine the values of the parameters μ, σ, and d from the maximum likelihood idea. Let us start with d. According to the above formula, we have $L = c \cdot d^z \cdot (1-d)^{n-z}$, where c does not depend on d. Maximizing L is equivalent to minimizing $\psi = -\ln(c) - z \cdot \ln(d) - (n-z) \cdot \ln(1-d)$. Differentiating this expression by d and equating the derivative to 0, we get

$$0 = -\frac{z}{d} + \frac{n-z}{1-d},$$

hence

$$\frac{z}{d} = \frac{n-z}{1-d}.$$

Multiplying both sides by $d \cdot (1-d)$, we get $z \cdot (1-d) = (n-z) \cdot d$. Moving all terms containing d to the right-hand side, we get $z = z \cdot d + (n-z) \cdot d = n \cdot d$, hence $d = \dfrac{z}{n}$.

With respect to μ and σ, the optimized expression has the form

$$L(\mu, \sigma, d, x_1, \ldots, x_n) = c \cdot \prod_{i: x_i \neq 0} \rho_{LN}(x_i, \mu, \sigma, d),$$

where c does not depend on μ or σ. Thus, the problem of finding the values μ and σ that maximize this expression is equivalent to the problem of finding the values μ and σ that maximize the product $\prod\limits_{i: x_i \neq 0} \rho_{LN}(x_i, \mu, \sigma, d)$. This is the problem that we considered in the previous section, and for which we already have the algorithm. Thus, we arrive at the following algorithm.

Resulting algorithm. We start with n intervals $[\underline{x}_i, \overline{x}_i]$ with $\underline{x}_i \geq 0$. For each index i for which $\underline{x}_i = 0$, we take $x_i = 0$. Then, we take $d = \dfrac{z}{n}$, where z is the number of indices for which we selected $x_i = 0$.

If $d = 1$, then we simply have a distribution which is located at 0 with probability 1. If $d < 1$, then we process the set of all remaining intervals (for which $\underline{x}_i < 0$) by applying the lognormal-related algorithm described in the previous section, and come up with the corresponding values of μ and σ.

Comment. Our arguments were given for the case when we approximate the original delta-function by a uniform distribution on an interval $[0, \ell]$ whose width ℓ tends to 0. One can check that when $\ell \to 0$, the optimization results will be the same if we use non-uniform distributions on these intervals.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and DUE-0926721 and by Grant 1 T36 GM078000-01 from the National Institutes of Health. The work of N. Buntao was supported by a grant from the Office of the Higher Education Commission, Thailand, under the Strategic Scholarships for Frontier Research Network, and by the Graduate College, King Mongkut's University of Technology North Bangkok.

The authors are thankful to Hung T. Nguyen for valuable discussions.

References

[1] J. Aitchison and J. Brown, *The Lognormal Distribution*, Cambridge University Press, Cambridge, UK, 1969.

[2] N. Buntao, "Estimating Parameters of Pareto Distribution under Interval and Fuzzy Uncertainty", *Proceedings of the 30th Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2011*, El Paso, Texas, March 18–20, 2011.

[3] N. Buntao and V. Kreinovich, "Measures of Deviation (and Dependence) for Heavy-Tailed Distributions and their Estimation under Interval and Fuzzy Uncertainty", In: R. R. Yager, M. Z. Reformat, S. N. Shahbazova, and S. Ovchinnikov (eds.), *Proceedings of the World Conference on Soft Computing*, San Francisco, CA, May 23–26, 2011.

[4] M. A. Cantos Rosales, *The Robustness of Confidence Intervals for the Mean of Delta Distribution*, PhD Dissertation, Department of Statistics, Western Michigan University, Kalamazoo, Michigan, 2009.

[5] E. Dantsin, V. Kreinovich, A. Wolpert, and G. Xiang, "Population Variance under Interval Uncertainty: A New Algorithm", *Reliable Computing*, 2006, Vol. 12, No. 4, pp. 273–280.

[6] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg, *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*, Sandia National Laboratories, Report SAND2007-0939, May 2007.

[7] V. Kreinovich, L. Longpré, S. A. Starks, G. Xiang, J. Beck, R. Kandathi, A. Nayak, S. Ferson, and J. Hajagos, "Interval Versions of Statistical Techniques, with Applications to Environmental Analysis, Bioinformatics, and Privacy in Statistical Databases", *Journal of Computational and Applied Mathematics*, 2007, Vol. 199, No. 2, pp. 418–423.

[8] V. Kreinovich, H. T. Nguyen, and S. Niwitpong, "Statistical Hypothesis Testing Under Interval Uncertainty: An Overview", *International Journal of Intelligent Technologies and Applied Statistics*, 2008, Vol. 1, No. 1, pp. 1–32.

[9] V. Kreinovich and G. Xiang, "Fast Algorithms for Computing Statistics under Interval Uncertainty: An Overview", In: V.-N. Huynh, Y. Nakamori, H. Ono, J. Lawry, V. Kreinovich, and H. T. Nguyen (eds.), *Interval/Probabilistic Uncertainty and Non-Classical Logics*, Springer-Verlag, Berlin-Heidelberg-New York, 2008, pp. 19–31.

[10] V. Kreinovich, G. Xiang, S. A. Starks, L. Longpré, M. Ceberio, R. Araiza, J. Beck, R. Kandathi, A. Nayak, R. Torres, and J. Hajagos, "Towards combining probabilistic and interval uncertainty in engineering calculations: algorithms for computing statistics under interval uncertainty, and their computational complexity", *Reliable Computing*, 2006, Vol. 12, No. 6, pp. 471–501.

[11] S. Niwitpong, H. T. Nguyen, I. Neumann, and V. Kreinovich, "Hypothesis testing with interval data: case of regulatory constraints", *International Journal of Intelligent Technologies and Applied Statistics*, 2008, Vol. 1, No. 2, pp. 19-41.

[12] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, American Institute of Physics, New York, 2005.

[13] D. J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, Chapman & Hall/CRC, Boca Raton, Florida, 2007.

[14] L. Tian and J. Wu, "Confidence Intervals for the Mean of Lognormal Data with Excess Zeros", *Biometrical Journal*, 2006, Vol. 48, No. 1, pp. 149–156.

[15] G. Xiang, M. Ceberio, and V. Kreinovich, "Computing Population Variance and Entropy under Interval Uncertainty: Linear-Time Algorithms", *Reliable Computing*, 2007, Vol. 13, No. 6, pp. 467–488.