Uniform distribution width estimation from data observed with Laplace additive error

M. Benšić[a], K. Sabo[a,*]

[a]*Department of Mathematics, University of Osijek, Trg Ljudevita Gaja 6, HR-31000 Osijek, Croatia*

Abstract

A one-dimensional problem of a uniform distribution width estimation from data observed with a Laplace additive error is analyzed. The error variance is considered as a nuisance parameter and it is supposed to be known or consistently estimated before. It is proved that the maximum likelihood estimator in the described model is consistent and asymptotically efficient and sufficient conditions for its existence are given. The method of moment estimator is also analyzed in this model and compared with the maximum likelihood estimator theoretically and in simulations. Finally, one real-world example illustrates the possibility for applications in two-dimensional problems.
Keywords: Maximum likelihood estimator, Method of moments estimator, Measurement error, Laplace additive error, Uniform distribution

1. Introduction

How to estimate the support of a uniform distribution from the data measured with additive errors is the problem that comes from different appli-

*Corresponding author
Email addresses: mirta@mathos.hr (M. Benšić), ksabo@mathos.hr (K. Sabo)

cations. Generally, if the object is measured with errors, it is complicated to determine its edges by using the well known edge detection methods (see e.g. [18]) and consequently, it is not easy to accurately estimate its dimensions either. Such problems can arise, for example, when the object is observed with a fluorescent microscope ([19]), a ground penetrating radar, ultrasound, etc. The same type of model can be used in the problem of protein secondary structure assignment ([9, 13]), detection of shapes in image analysis ([10]), etc.

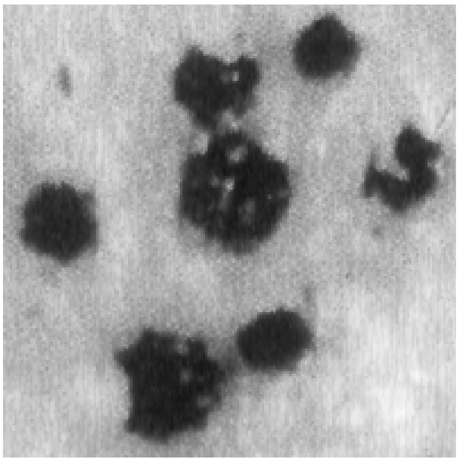


Figure 1: Real world problem: Estimation the size of colonies of black fungi ([10], [25])

To illustrate applications more precisely, let us mention the problem of estimating the size of black fungi colonies on the basis of a monochromatic image (see Fig. 1). Here we have seven colonies captured with errors and we focus our attention on a diameter of each colony (see [10]). In Subsection 5.3 we give a complete analysis of this problem and present the results obtained by the method presented in this paper.

The basic statistical model that we are going to use for our purpose is a parametric one and it is built for one-dimensional data in the data set. It is a simple random sample $(X_1, \ldots, X_n)$ from the distribution that is a mixture of two independent random variables U and Y

$$X = U + Y,$$

where U is uniformly distributed on the segment $[-a, a]$, $a > 0$, and Y is the error variable. In this setting, the estimation of the parameter a means also the estimation of the boundary points.

Boundary estimation in the presence of measurement errors is a problem extensively treated in literature in different contexts (see e.g. [6, 15, 16, 17]). Although this problem generally can be treated as a classical deconvolution density estimation problem (see e.g. [4, 7, 8, 24, 26]), these methods usually face problems at density discontinuity points. That is why the modified kernel estimator has been proposed in [27] if the boundary points are of interest. Also, in [6] a diagnostic function which is proportional to the derivative of the deconvolution kernel density estimator has been optimized in order to estimate the boundary points. However, some computational problems for an easy application remain, for instance the choice of bandwidth which is very important for a good performance of any kernel estimator.

Here we discuss a completely different approach which assumes the error distribution type to be known. Similar models have been analyzed in [1, 2, 3, 23]. In these papers, the basic model assumes that the distribution of the part U is uniform on an interval and the distribution of the part Y, which is considered an error, is Gaussian. The maximum likelihood approach applied in the aforementioned papers shows that the length of the uniform support can be estimated consistently and in an asymptotically efficient way even if we have the added Gaussian error part in the data. Also, the asymptotic variance of the estimator is calculated so we can construct confidence intervals and statistical tests about the length of the uniform support in a classical parametric way.

However, the estimation procedures based on the ML estimator from the mentioned papers are shown to be very sensitive to outliers in applications. Unfortunately, the outliers are often present in images, so we tried to consider the estimation procedure which is less sensitive to outliers and allows a parametric approach at the same time. As expected, the ML estimator derived from a similar model but with a Laplace error distribution fulfilled our expectation.

In this paper, we present results which are based on the assumptions that U is uniformly distributed on the interval $[-a, a]$ for some $a > 0$, which is to be estimated, and Y is a Laplace random variable with a location parameter $\mu = 0$ and a scale parameter $\lambda > 0$.

If U is uniform with a density function

$$f_U(x|a) = \begin{cases} \frac{1}{2a}, & x \in [-a, a] \\ 0, & \text{else,} \end{cases} \tag{1}$$

and Y has a Laplace distribution with a location parameter $\mu = 0$ and a scale parameter $\lambda > 0$ i.e. a density function

$$f_Y(x|\lambda) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}, \tag{2}$$

then a density function for X has a form

$$f_X(x|a, \lambda) = \frac{1}{2a} \left(G\left(\frac{x+a}{\lambda}\right) - G\left(\frac{x-a}{\lambda}\right) \right), \tag{3}$$

where

$$G(x) = \begin{cases} \frac{1}{2}e^{x}, & x < 0, \\ 1 - \frac{1}{2}e^{-x}, & x \geq 0. \end{cases}$$

It can be easily shown that the limiting density for λ tending to zero is uniform on $[-a, a]$ and for a tending to zero it is Laplace with a location parameter $\mu = 0$ and a scale parameter $\lambda > 0$.

Here we consider the estimation problem of the uniform support $[-a, a]$ in case when the scale parameter $\lambda > 0$ is known or consistently estimated before.

The paper is organized as follows. In Section 2, we give some useful properties of the model density function that we use in Sections 3 and 4 where we discuss properties of estimators obtained by the method of moment (MM) and the maximum likelihood (ML) method, respectively. We prove consistency, derive asymptotic variances of estimators and give conditions on the data that guarantee the existence of the MM and the ML estimators. In Section 5, we give different numerical examples. First of all, MM and ML estimators are compared in simulations. Further, we give the simulation results for the ML estimator which include bias, MSE, the length of the confidence intervals and the cover rate for small and large data sets. In order to discuss robustness we simulate data from the model with Gaussian error (see [1]), but we replace some data with outliers. It is shown in this example that, in the presence of outliers, the ML estimator based on Laplace errors behaves much better than the ML estimator based on Gaussian errors. Finally, we apply our approach to the problem of detecting circular shapes of black fungi colonies, which was presented in [10]. The last Section is devoted to concluding remarks.

2. Properties of the log-likelihood function

In this section, we derive several useful technical properties of the density $f_X(x|a,\lambda)$, $a, \lambda > 0$ and the corresponding log-likelihood that will be used in the sequel. As usual, we use upper and lower case to denote a random variable and its realizations, respectively.

First of all, let us notice that the density function (1) is even and continuous on $\mathbb{R}$. We can also represent it in the form

$$f_X(x|a,\lambda) = \frac{1}{2a}\begin{cases} e^{-\frac{|x|}{\lambda}} \sinh \frac{a}{\lambda}, & |x| \geq a, \\ \left(1 - e^{-\frac{a}{\lambda}} \cosh \frac{x}{\lambda}\right), & |x| < a. \end{cases} \tag{4}$$

As a consequence, the function

$$\begin{aligned} x \mapsto \varrho(x, a, \lambda) :&= \log f_X(x|a,\lambda) \\ &= -\log(2a) + \begin{cases} -\frac{|x|}{\lambda} + \log \sinh \frac{a}{\lambda}, & |x| \geq a, \\ \log\left(1 - e^{-\frac{a}{\lambda}} \cosh \frac{x}{\lambda}\right), & |x| < a, \end{cases} \end{aligned} \tag{5}$$

is also even and continuous. The following property is obvious.

Property 1. Let the values $a, \lambda > 0$ be fixed.

$$\lim_{x \to a} \varrho(x, a, \lambda) = -\log(2a) + \log\left(\sinh \frac{a}{\lambda}\right) - \frac{a}{\lambda} = -\log(2a) + \log\left[1 - e^{-\frac{a}{\lambda}} \cosh \frac{a}{\lambda}\right].$$

In what follows we summarize some additional important properties of the model log-likelihood function that we need in the next sections.

Property 2. Let the values $x \in \mathbb{R}, \lambda > 0$ be fixed.

a) For all $x \in \mathbb{R}$, the function $a \mapsto \varrho(x, a, \lambda)$ is continuous on $(0, \infty)$.
   If $x \neq 0$, then

$$\lim_{a \to |x|} \varrho(x, a, \lambda) = -\log(2|x|) + \log\left(\sinh \frac{|x|}{\lambda}\right) - \frac{|x|}{\lambda}.$$

If $x = 0$, then

$$\lim_{a \to 0^+} \varrho(0, a, \lambda) = -\log(2\lambda).$$

b) For all $x \in \mathbb{R}$,

$$\varrho(x, a, \lambda) \leq \lim_{a \to 0^+} \varrho(x, a, \lambda)$$

$$\lim_{a \to 0^+} \varrho(x, a, \lambda) = -\log(2\lambda) - \frac{|x|}{\lambda} \leq \lim_{a \to 0^+} \varrho(0, a, \lambda).$$

c) The function $a \mapsto \varrho(x, a, \lambda)$ is continuously differentiable on $(0, \infty)$,

$$c(x, a, \lambda) := \frac{\partial \varrho(x, a, \lambda)}{\partial a} = \begin{cases} -\frac{1}{a} + \frac{1}{\lambda} \coth \frac{a}{\lambda}, & 0 < a \leq |x| \\ -\frac{1}{a} + \frac{1}{\lambda} \frac{\cosh \frac{x}{\lambda}}{e^{\frac{a}{\lambda}} - \cosh \frac{x}{\lambda}}, & a > |x|, \end{cases}$$

and

$$\lim_{a \to |x|} c(x, a, \lambda) = -\frac{1}{a} + \frac{1}{\lambda} \coth \frac{a}{\lambda}.$$

Property 3. For fixed $\lambda > 0$, the function $(x, a) \mapsto c(x, a, \lambda)$ is bounded, i.e.,

$$|c(x, a, \lambda)| < \frac{1}{\lambda}, \text{ for all } x \in \mathbb{R}, a > 0.$$

Since Properties 1 and 2 are obvious, we shall prove only Property 3.

Proof of Property 3. The statement for $0 < a \leq |x|$ follows immediately from the inequality $\left|-\frac{1}{u} + \coth u\right| < 1$ for $u \neq 0$ (Fig. 2)

For $a > |x|$, we have

$$|c(x, a, \lambda)| = \frac{1}{\lambda} \left| \frac{1}{\frac{e^{\frac{a}{\lambda}}}{\cosh \frac{x}{\lambda}} - 1} - \frac{\lambda}{a} \right|.$$

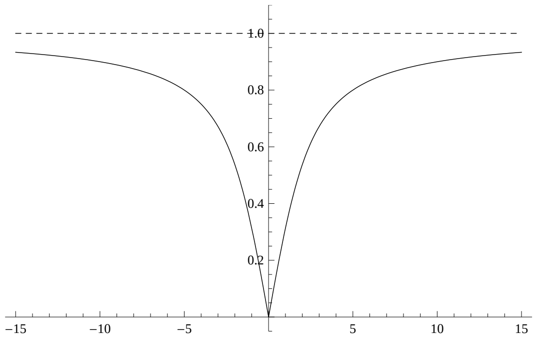


Figure 2: Graph of the function $x \mapsto \left|-\frac{1}{x} + \coth x\right|$

Classical analysis of the function $u \mapsto \frac{1}{\frac{e^b}{\cosh u} - 1} - \frac{1}{b}, \quad u \in [0, b], \quad b > 0,$ shows that it is strictly increasing for $u \in [0, b]$. So the function

$$u \mapsto h(u, b) = \left| \frac{1}{\frac{e^b}{\cosh u} - 1} - \frac{1}{b} \right|, \;\; u \in [0, b], \;\; b > 0$$

is bounded by numbers $h(0, b)$ and $h(b, b)$.

By analyzing functions $b \mapsto h(0, b)$ and $b \mapsto h(b, b)$ for $b > 0$ we see that

- $h(0, b)$ and $h(b, b)$ are strictly monotone for $b > 0$;
- $\lim\limits_{b \to 0^+} h(0, b) = \frac{1}{2}$, $\lim\limits_{b \to \infty} h(0, b) = 0$, $\lim\limits_{b \to 0^+} h(b, b) = 0$,$\lim\limits_{b \to \infty} h(b, b) = 1$.

Thus, we see that $|c(x, a, \lambda)| < \frac{1}{\lambda}$ for $a > |x|$, too.

3. The method of moments estimator

Let us suppose that λ is known. The second order MM estimator can be calculated from the equation

$$m_2 = \frac{a^2}{3} + 2\lambda^2,$$

where $m_2 = \frac{1}{n}\sum_{i=1}^n x_i^2$ is the second sample moment. Thus, the MM method estimator has an explicit form

$$\hat{a}_{MM} = \sqrt{3m_2 - 6\lambda^2},$$

and can be calculated only if data satisfy the condition $m_2 \geq 2\lambda^2$. The asymptotic variance for $\hat{a}_{MM}$ is then

$$\operatorname{Var}\hat{a}_{MM} \approx \frac{1}{5\,n}\left(a + 15\frac{\lambda^2}{a}\right)^2.$$

If the parameter λ is also supposed to be unknown, then we can calculate the MM estimator for a and λ from the expressions:

$$m_2 = \frac{a^2}{3} + 2\lambda^2, \tag{6}$$

$$m_4 = \frac{a^4}{5} + 24\lambda^4 + 4a^2\lambda^2, \tag{7}$$

where $m_4 = \frac{1}{n}\sum_{i=1}^n x_i^4$. Equations (6)-(7) have a positive solution if and only if

$$\frac{9}{5} \leq \frac{m_4}{m_2^2} < 6,$$

and the solution has an explicit form:

$$\hat{a}_{MM} = \sqrt{5m_2 - \sqrt{5}\sqrt{m_4 - m_2^2}}, \quad \hat{\lambda}_{MM} = \frac{1}{\sqrt{6}}\sqrt{-2m_2 + \sqrt{5}\sqrt{m_4 - m_2^2}}.$$

4. The ML estimator

Let us denote the realization of the sample with $\mathbf{x} = (x_1, \ldots, x_n)$. According to (3), the likelihood function of the random sample model has the following form

$$\mathcal{L}(a,\lambda) := \mathcal{L}(\mathbf{x}|a,\lambda) = \prod_{i=1}^n f_X(x_i|a,\lambda) = \frac{1}{(2a)^n}\prod_{i=1}^n \left(G\left(\frac{x_i + a}{\lambda}\right) - G\left(\frac{x_i - a}{\lambda}\right)\right)$$

and the log-likelihood function for the random sample model is as follows:

$$\ell(a,\lambda) := \sum_{i=1}^{n} \varrho(x_i, a, \lambda).$$

For fixed $\lambda > 0$, let us shorten the notation and define $\varrho(x,a) := \varrho(x,a,\lambda)$ and $\ell(a) := \ell(a,\lambda)$.

4.1. The existence of the ML estimator

In the following theorem we give a simple sufficient condition on the data that guarantee the existence of the ML estimator.

Theorem 1. *Let $x_i \in \mathbb{R}$, $i = 1,\dots,n$ be the data, and let $\lambda > 0$ be given. The function $a \mapsto \ell(a)$ is bounded from above on $(0,\infty)$. Particularly, if $x_i \neq 0$, $i = 1,\dots,n$, there exists $a^\star \in (0,\infty)$ satisfying*

$$\ell(a^\star) = \sup_{a\in(0,\infty)} \ell(a). \tag{8}$$

Proof. Let us first prove that the function $a \mapsto \ell(a)$ is bounded from above on $(0,\infty)$. From the continuity of $a \mapsto \ell(a)$, it is obvious that it is bounded from above on every compact subset of $(0,\infty)$. Problems can occur only at the domain boundaries, i.e., $a \to \infty$ or $a \to 0^+$.

Since

$$\lim_{a\to\infty} \ell(a) = \lim_{a\to\infty} \sum_{i=1}^{m} \log\left(\frac{1}{2a}\left(1 - e^{-\frac{a}{\lambda}} \cosh\frac{x_i}{\lambda}\right)\right) = -\infty,$$

and

$$\lim_{a\to 0^+} \ell(a) = -n\log(2\lambda) - \frac{1}{\lambda}\sum_{i=1}^{n} |x_i|,$$

the function $a \mapsto \ell(a)$ is bounded from above on $(0,\infty)$.

It remains to show that the assumption $x_i \neq 0$ for all $i = 1, \ldots, n$, guarantees the existence of $a > 0$ such that

$$\ell(a) > \lim_{a \to 0^+} l(a) = -n \log(2\lambda) - \frac{1}{\lambda} \sum_{i=1}^{n} |x_i|.$$

In order to do this, let us note that

$$\ell(a) = \sum_{i=1}^{n} \left(-\log(2a) + \log\left(\sinh \frac{a}{\lambda}\right) - \frac{|x_i|}{\lambda} \right),$$

for all $a \in (0, \min\{|x_i| : i = 1, \ldots, n\})$ and $\min\{|x_i| : i = 1, \ldots, n\} > 0$.

For $a < x$, we have

$$\ell'(a) = \sum_{i=1}^{n} \left(\frac{\coth \frac{a}{\lambda}}{\lambda} - \frac{1}{a} \right).$$

Since

$$\lim_{a \to 0^+} \frac{1}{a} \ell'(a) = \frac{1}{3\lambda^2} > 0,$$

there is a real number $\delta \in (0, \min\{|x_i| : i = 1, \ldots, n\})$ such that the function $a \mapsto \ell(a)$ is strictly increasing on $(0, \delta)$. Hence, for every $a \in (0, \delta)$ we have

$$\ell(a) > \lim_{a \to 0^+} \ell(a).$$

□

From the computational aspect, it can be useful to mention that we can continuously extend the log-likelihood function to $a = 0$. Thus, we can consider the continuous function $\tilde{\ell} : [0, \infty) \to \mathbb{R}$ given by the formula:

$$\tilde{\ell}(a) = \begin{cases} \ell(a), & a > 0 \\ -n \log(2\lambda) - \frac{1}{\lambda} \sum_{i=1}^{n} |x_i|, & a = 0. \end{cases}$$

Since the function $a \mapsto \tilde{\ell}(a)$ is continuous and bounded from above on $[0, \infty)$, obviously there exists $\tilde{a}^\star \in [0, \infty)$ satisfying $\tilde{\ell}(\tilde{a}^\star) = \sup_{a \in [0,\infty)} \tilde{\ell}(a)$. Also note

$$\tilde{a}^\star = \begin{cases} a^\star > 0, & \text{if there exists } a^\star > 0 \text{ such that } \ell(a^\star) = \sup\limits_{a \in (0,\infty)} \ell(a), \\ 0, & \text{else.} \end{cases}$$

The next corollary follows immediately from the previous consideration and from Theorem 1.

Corollary 1. *Let $x_i \in \mathbb{R}$, $i = 1, \ldots, n$ be the data, and let $\lambda > 0$ be given. Then the function $a \mapsto \tilde{\ell}(a)$ is bounded from above on $[0, \infty)$ and consequently there exists $\tilde{a}^\star \in [0, \infty)$ satisfying*

$$\tilde{\ell}(\tilde{a}^\star) = \sup_{a \in [0,\infty)} \tilde{\ell}(a). \tag{9}$$

4.2. Asymptotic efficiency of the ML estimator

In order to analyze asymptotic efficiency we will use conditions suggested and partially proved in [5], which are also discussed and finally proved in [28]. In fact, we will use the second set of sufficient conditions from [5] which we quote in the sequel in terms of our notation. Also, we drop the parameter λ from the function variable list to shorten the notation.

Conditions for asymptotic efficiency

1. $\varrho(x,a)$ is continuous in a through $(0,\infty)$. At every a_0, there is a neighborhood such that for all a, a' therein
$$|\varrho(x,a) - \varrho(x,a')| < A(x,a_0)|a - a'|,$$
where $E_{a_0}(A^3) < \infty$.
2. At every a, there exists $\partial\varrho(x,a)/\partial a$ for almost all x and it is not zero almost everywhere. It is continuous in a, except at a finite number of discontinuity points at which it has finite jumps of either sign.
3. The probability that the interval (a,a') contains a discontinuity point of $\partial\varrho(x,a)/\partial a$ is $\mathcal{O}(a'-a)$ for any true value a_0.
4. $\partial\varrho(x,a)/\partial a$ is assumed to be continuous on the right. For the representation
$$\partial\varrho(x,a)/\partial a = c(x,a) + h(x,a),$$
where $c(x,a)$ is continuous at every a and $h(x,a)$ is a step function, it is required that
$$|c(x,a') - c(x,a)| < B(x,a_0)|a' - a|, \quad E_{a_0}(B^2) < \infty.$$

In addition to the above conditions, consistency should be confirmed. In our model this can be done following Theorem 1 in [11], p. 223 (see also [12]).

To prove consistency, let us spread the parameter space $(0,\infty)$ and consider the whole $\mathbb{R}$. Now we extend the log-likelihood function such that we use the function:

$$\varrho^\star(x,a) = \begin{cases} \varrho(x,|a|), & a \neq 0 \\ -\log(2\lambda) - \frac{|x|}{\lambda}, & a = 0. \end{cases}$$

Also, let us adjust the notation with paper [11]. So, we define

$$\rho(x,a) := -\log(\varrho^\star(x,a)).$$

The maximization problem in terms of $\varrho^\star$ can now be considered as a minimization problem in terms of ρ. The function $a(\cdot)$ from (A-3) in [11] is denoted by $\alpha(\cdot)$ to avoid confusion with the parameter.

The first two assumptions, (A-1) and (A-2), are obviously fulfilled in our model. To consider other assumptions, let us denote by a_0 the true value of the parameter and note that

$$E_{a_0}|\rho(X,a_0)| < \infty.$$

This can be checked by analyzing the integral which appears in the expectation. Namely, this integral consists of two parts. The first one, on the interval $[-a,a]$, is finite as the integrand is continuous on this segment. The second part, on the set $(-\infty,a)\cup(a,\infty)$, can be easily solved and is also finite.

Assumption (A-3) from [11] requires the existence of a measurable function $\alpha(\cdot)$ such that the quantity

$$\gamma(a) = E_{a_0}[\rho(X,a) - a(X)]$$

is well defined for all a. We choose $\alpha(x) = \rho(x,a_0)$. To substantiate the claim that $\gamma(a)$ is well defined we note that

$$E_{a_0}|\rho(X,a) - \alpha(X)| < \infty, \quad \forall a \in \mathbb{R}.$$

This is an immediate consequence of the mean value theorem and Property 3 from Section 2.

According to the assumption (A-4) for the function $\gamma(a)$ we need to show the existence of the number a' such that $\gamma(a) > \gamma(a')$ for all $a \neq a'$. We see that this condition is fulfilled by choosing $a' = a_0$ and the fact that we consider a maximum likelihood estimator.

Assumption (A-5) consists of three parts. It demands the existence of a continuous function $b(a) > 0$ such that

(i) $\inf\limits_{a\in\mathbb{R}} \frac{\rho(x,a)-a(x)}{b(a)} \geq h(x)$ for some integrable h;

(ii) $\lim\limits_{a\to\pm\infty} \inf b(a) > \gamma(a_0)$;

(iii) $E\left[\lim\limits_{a\to\pm\infty} \inf \frac{\rho(X,a)-a(X)}{b(a)}\right] \geq 1$.

If we choose $b(a) = |\log(2\lambda)|$, fulfillment of this assumption is a consequence of the fact that $\gamma(a_0) = 0$, $\gamma(a) < \infty$, that the function $(x, a) \mapsto \rho(x, a)$ has a lower bound $\log(2\lambda)$ (Property 2, Section (2)) and $\lim\limits_{a\to\pm\infty} \rho(x, a) = \infty$ for all $x \in \mathbb{R}$.

Thus, the ML estimator for the parameter a in our model is consistent.

Regarding conditions for asymptotic efficiency quoted at the beginning of this section, we see that the second and the third condition are obviously fulfilled. The fourth condition is a consequence of Property 3 from Section 2. Fulfillment of the first condition follows from Property 3 combined with the mean value theorem. Thus, the ML estimator for the parameter a in our model is asymptotically efficient which means that for any consistent sequence $\hat{a}_n = \hat{a}_n(X_1, \ldots, X_n)$ of roots of the likelihood equation $(\hat{a}_n)$, the sequence $\sqrt{n}(\hat{a}_n - a_0)$ converges in law to the random variable with $\mathcal{N}\left(0, \frac{1}{I(a_0)}\right)$ distribution.

Here

$$I(a) = -\frac{1}{a^2} + \frac{1}{a\lambda^2} \int\limits_0^\infty \frac{(\varphi(\frac{x+a}{\lambda}) + \varphi(\frac{x-a}{\lambda}))^2}{\Phi(\frac{x+a}{\lambda}) - \Phi(\frac{x-a}{\lambda})} dx \tag{10}$$

and the asymptotic variance of the ML estimator $\hat{a}_{ml}$ has the form

$$\text{Var}\, \hat{a}_{ML} = \frac{1}{n} \left[-\frac{1}{a_0^2} + \frac{1}{a_0\lambda^2} \int\limits_0^\infty \frac{(\varphi(\frac{x+a_0}{\lambda}) + \varphi(\frac{x-a_0}{\lambda}))^2}{\Phi(\frac{x+a_0}{\lambda}) - \Phi(\frac{x-a_0}{\lambda})} dx \right]^{-1}. \tag{11}$$

Figure 3 illustrates rates of asymptotic variances for the MM and ML estimators in our model.

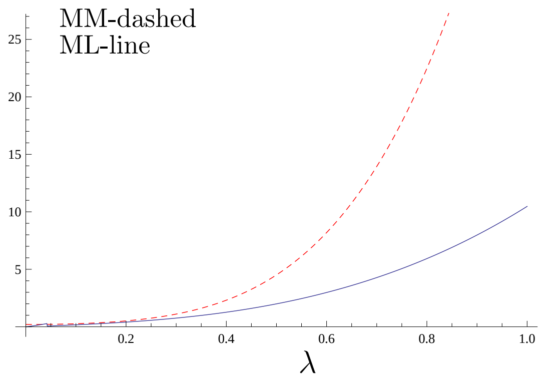


Figure 3: Asymptotic variance multiplied by n for the MM and the ML estimator and $a = 1$.

5. Numerical experiments and examples

5.1. Simulation study

A simulation study was conducted to analyze the quality of the ML estimator in small ($n = 30$), relatively large ($n = 300$) and large (n=1000) samples. $N = 1000$ replications were carried out for each sample size and different λ.

Here we present results of the study where we kept a fixed at $a = 1$. All simulations were completed in `R` with the `optim(method="Brent")` optimization procedure.

For discussion purposes we present a bias and an MSE of the ML estimator as well as the average length of the likelihood based approximate 95% confidence interval (LB.CI) and the cover rate (Tables 1, 2 and 3). Also, in these tables, we add a column with the number of successfully computed LB.CI in one thousand replications. Namely, an approximate $100(1-\alpha)\%$ confidence interval for a consists of all the possible values of a for which the log-likelihood function drops off by no more than $0.5\chi_1^2(1-\alpha)$[1] units. It means that the intersections of the straight line $y = \ell(\hat{a}_{ML}) - \frac{1}{2}0.5\chi_1^2(1-\alpha)$ and the graph of the log-likelihood function should be computed. In small samples it often happens that these two lines intersect only once, in which case the LB.CI does not exist.

An approximate confidence interval can also be computed from the asymptotic distribution of $\hat{a}_{ML}$. It was proved in Section 4 that $\hat{a}_{ML}$ is consistent and asymptotically normal with the asymptotic variance (11). Also, we can put $\hat{a}_{ML}$ in the variance formula without affecting the asymptotic distribution. If the LB.CI does not exist we suggest computing confidence intervals in this way. Both methods produce approximate confidence intervals with similar average lengths and cover rates for large samples but the LB.CI is computationally more convenient.

Simulations confirm our expectations and theoretical considerations. In

[1] $\chi_1^2(1-\alpha)$ is the $1-\alpha$ quantile of the χ^2 distribution with one degree of freedom.

the presented model, the ML estimator is always better than the MM estimator by the MSE (see Figure 4). It is also worth mentioning that we had some cases where the MM estimator did not exist in simulations with huge variances but we did not have existence problems with the ML estimation procedure.

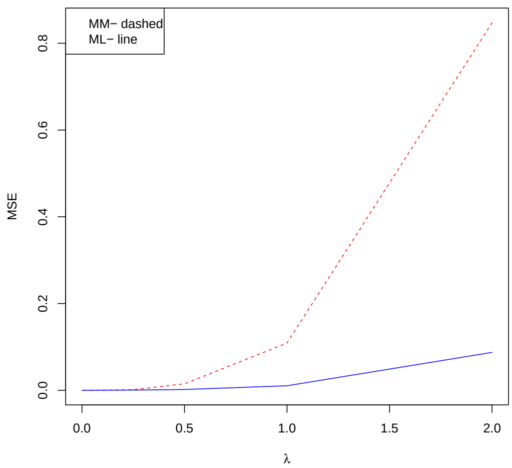


Figure 4: MSE for $n = 1000$ data, $a = 1$.

The results presented in Tables 1, 2 and 3 confirm the expected behaviour of the ML estimator. Although the LB.CI cover rate is acceptable for all cases, we face rather long confidence intervals for small sample sizes. It is worth mentioning that, with small sample sizes, the alternative way of confidence interval calculation (by using $\hat{a}_{ML}$ asymptotic distribution) assures shorter intervals but with unsatisfactory cover rates. In our opinion, it is important to assure a large sample in order for the estimation procedure with this approach to be reliable.

Table 1: Simulation results for the ML estimator, $a = 1$, $n = 30$ data and $N = 1000$ replications.

λ	Bias	MSE	LB.CI No. of success	LB.CI Length	LB.CI Cover rate
1.000	0.025716	0.366073	206	2.154386	0.815
0.500	-0.009745	0.080511	766	1.064041	0.953
0.250	-0.011702	0.023240	999	0.566858	0.933
0.100	-0.034288	0.009456	1000	0.333181	0.952
0.050	-0.048648	0.007786	1000	0.272064	0.953
0.010	-0.066304	0.008849	986	0.479989	0.963
0.001	-0.061918	0.007250	780	0.894387	0.959

Table 2: Simulation results for the ML estimator, $a = 1$, $n = 300$ data and $N = 1000$ replications.

λ	Bias	MSE	LB.CI No. of success	LB.CI Length	LB.CI Cover rate
1.000	-0.014906	0.041377	924	0.761894	0.976
0.500	0.001246	0.007006	1000	0.322907	0.942
0.250	-0.000026	0.001846	1000	0.170317	0.952
0.100	-0.000077	0.000624	1000	0.093516	0.937
0.050	-0.000963	0.000261	1000	0.063908	0.951
0.010	-0.003812	0.000094	1000	0.031168	0.953
0.001	-0.006398	0.000080	1000	0.033283	0.945

Table 3: Simulation results for the ML estimator, $a = 1$, $n = 1000$ data and $N = 1000$ replications.

λ	Bias	MSE	LB.CI No. of success	LB.CI Length	LB.CI Cover rate
1.000	-0.000610	0.010709	1000	0.407844	0.951
0.500	0.001185	0.002050	1000	0.175497	0.952
0.250	-0.000603	0.000544	1000	0.093249	0.953
0.100	0.000109	0.000175	1000	0.050940	0.949
0.050	0.000116	0.000075	1000	0.034782	0.958
0.010	-0.000314	0.000017	1000	0.015308	0.949
0.001	-0.001548	0.000008	1000	0.007442	0.964

5.2. Robustness

The fundamental reason why we conducted the Laplace distribution as an error distribution in our model is a reasonable expectation that this model is better for application in the presence of outliers than the model with Gaussian errors. To confirm this expectation, we simulated 300 data from the model of the same type but with Gaussian errors (see eg. [1]) where we changed some data with outliers by the rules described in Table 4. We applied two ML estimation procedures to this data set. The first was based on the model with Gaussian errors (see e.g. [1]) and the second one with errors following the Laplace distribution. Also, we applied these two estimation procedures on the same type of data but with the error distribution that has long tails. For discussion purposes in Table 4 we include results for data simulated from the model of the same type but with scaled student errors with four degrees of freedom and variance 1.

Table 4: Simulation results in the presence of outliers. Error variance was kept fixed at $\sigma = 1$, $a = 2$, $n = 300$ and N=1000 replications.

Simulatons	Model with normal error		Model with Laplace error	
	Bias	MSE	Bias	MSE
normal errors no outliers	-0.007686	0.017691	0.092710	0.026495
normal errors 5 one-sided outliers $P(X > o) < 10^{-8}$	0.499169	0.267460	0.172122	0.051828
normal errors 10 two-sided outliers $P(X > o) < 10^{-8}$ $P(X < o) < 10^{-8}$	1.10706	1.241221	0.238137	0.073474
scaled student errors df=4 Var$Y = 1$	2.463928	15.522259	1.153521	2.161191

For all simulations we chose $a = 2$ and the error variance $\sigma = 1$, which means that we set the parameter λ in the Laplace error model to $\frac{1}{\sqrt{2}}$. The empirical MSE and bias for $N = 1000$ replications are shown in Table 4. The difference between estimation procedures is also illustrated in Figure 5. Our results confirmed that the ML estimator which comes from the Laplace error model gives much better results in the presence of outliers. In the case of incorrectly specified error distribution, if it has heavy tails, both estimators included in Table 4 seriously overestimate the targeted parameter a although the model with Laplace errors gives better results.

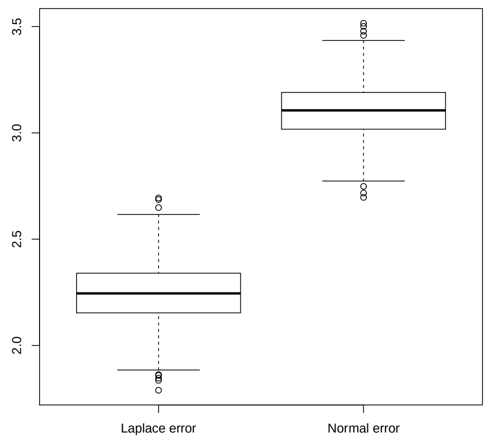


Figure 5: Distribution of estimators for $n = 300$ and 10 two-sided outlier.

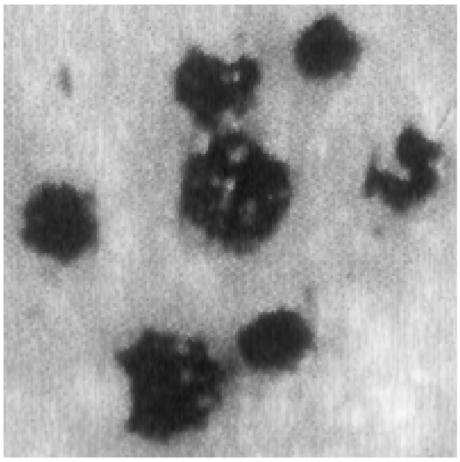
5.3. Size estimation of the black fungi colonies

This real world problem illustrates the possibility of extension to a two-dimensional case. It is motivated by biological research aimed at estimating the size of colonies of black fungi that grow under different conditions ([25]). Since these colonies grow approximately with the same rate in each direction, it is reasonable to consider the problem of finding circles within an image (see Fig. 6 (a)) of the surface and to estimate their diameter ([10]).

In order to separate data points into clusters, we use the method for searching for a nearly optimal partition from [22], combined with the center-based L_1- clustering method (see [20]). An appropriate number of clusters is seven, and it is estimated by the Silhouette Width Criterion index ([14]). The corresponding nearly optimal partition is shown in Fig 6(b).

In order to determine the discrete approximation for the border of circles, for each cluster we apply Algorithm 1.

(a) original data



(b) clustered data

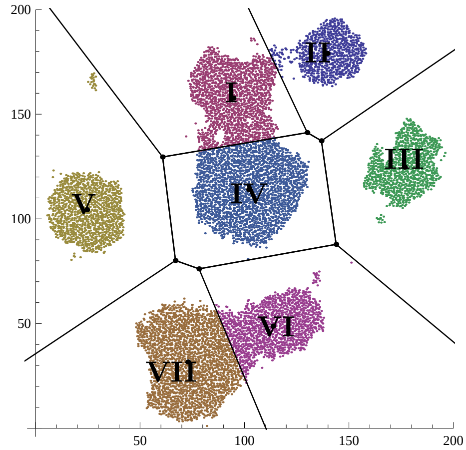


Figure 6: Black fungi

Algorithm 1 Approximation of the circle border

Input: cluster π, the number of tiny strips n_S.

1: Cut the data into n_S tiny strips parallel to the $x-$axis.

2: Keep only the first coordinate of the data and make one data set from each slide.

3: Centralize the data from the slide with the mean for this slide.

4: Apply the model described in this paper to each slide. Estimate the parameter λ by the MM method. Calculate the average value of $\hat{\lambda}_{MM}$ through slides. Estimate the width of the uniform support with the ML method using this value for λ.

5: Calculate the edges of the slide from this estimation, the slide mean and the $y-$coordinate of the slide $y-$center.

6: Repeat the procedure by cutting the data into tiny strips parallel to the $y-$axis and changing all steps in accordance to that.

Output: Approximation of the border of a circle.

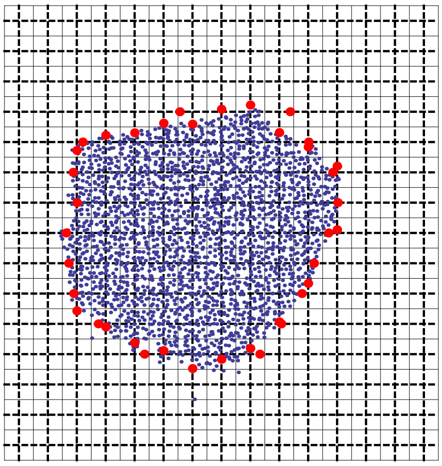


Figure 7: Approximation for the colony edge

Fig. 7 shows $n_S = 15$ tiny strips parallel to the $x-$axis and $n_S = 15$ tiny strips parallel to the $y-$axis. Large points represent the corresponding approximation of the edge points for cluster (IV) (see Fig 6 b)) in the model with Laplace errors obtained by Algorithm 1. From these edge points we can estimate the circle radius by the least squares method (for more details, see [1] and [21]). The results for all clusters are given in Table 5.

Figures 8 a) and b) show estimated circles for Gaussian and Laplace model errors, respectively. The method described in [1] in Steps 4 and 5 in Algorithm 1 has been used for Gaussian model errors.

Table 5: Estimated circle radii (in the same measurement units).

Radius/cluster	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Laplace form error	22.9951	14.9327	25.5602	18.7622	25.5602	18.0291	26.6106
Gaussian form error	24.5689	16.2531	26.4455	19.54321	26.4455	23.3707	26.9301

Note that by applying the model with Gaussian errors, the resulting circle radii are less precise than in [10] due to the presence of outliers. At the same time, the resulting circles obtained by applying the model with Laplace errors obviously give good results.

(a) (b)

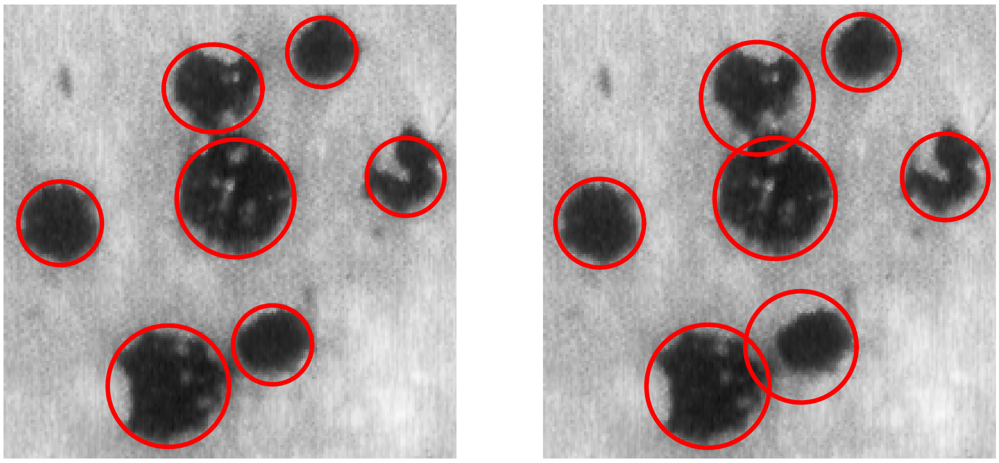


Figure 8: Size estimation of black fungi colonies. (a) Estimated circles (Laplace form error model) (b) Estimated circles (Gaussian form error model)

6. Concluding remarks

Once we have data that is known to be collected from disjoint bounded regions but measured with an additive error, the estimation of these region's boundaries becomes a problem. This is a typical measurement error problem that can be treated in different contexts (e.g. deconvolution problem, mixed effect problem). We are considering a combination of a one-dimensional parametric model and a computational procedure that was suggested in [1], [2], [3] and [21].

Previous researches that have been described in [1], [2], [3] and [21] guarantee that a simple symmetric one-dimensional model can serve as a base for the boundary reconstruction even if the region is multidimensional. However, the procedures presented in these papers are very sensitive to the outliers. Here we have presented an approach similar to the procedures described in the mentioned papers but less sensitive to the outliers. We have improved robustness by changing the error distribution in the basic model from the normal distribution to the Laplace distribution.

The results confirm our expectations that came from the classical M-estimator theory. As it is well-known, the robustness of the specific M-estimator depends on the form of the ρ-function that is used as a criterion function in the definition of this M-estimator. The maximum likelihood estimator is also an M-estimator. If the ρ-function is defined as a log-likelihood from the model with a normal error we create an M-estimator which is more sensitive to outliers than if the ρ-function is defined as a log-likelihood from the model with a Laplace error, by nature of these distributions. Obviously, other ρ-functions in the definition of an M-estimator can be considered in order to control robustness, but this approach requires further detailed study in future research.

Acknowledgements.

We are very grateful to the anonymous referees for the valuable comments which have improved considerably the first version of the manuscript.

The authors gratefully acknowledges the financial support of the Ministry of Science, Education and Sports, Republic of Croatia.

References

[1] Benšić, M., Sabo, K., 2007. Estimating the width of a uniform distribution when data are measured with additive normal errors with known variance, Computational Statistics and Data Analysis, 51, 4731–4741

[2] Benšić, M., Sabo, K., 2007. Border Estimation of a Two-dimensional Uniform Distribution if Data are Measured with Additive Error, Statistics, 41, 311–319

[3] Benšić, M., Sabo, K., 2010. Estimating a uniform distribution when data are measured with a normal additive error with unknown variance, Statistics, 44, 235–246

[4] Carroll, R.J., Hall, P., 1988. Optimal rates of convergence for deconvoluting a density. J. Amer. Statist. Assoc., 83, 1184-1186

[5] Daniels, H.E., 1961. The asymptotic efficiency of a maximum likelihood estimator, Proc. 4th Berkeley Symp. 1, 151–164

[6] A. Delaigle, I. Gijbels, 2006. Estimation of boundary and discontinuty points in deconvolution problems, Statistica Sinica 16, 773-788

[7] Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems, Ann. Statist., 19, 1257-1272

[8] Feuerverger, A., Kim P.T., Sun J., 2008. On Optimal Uniform Deconvolution, Journal of Statistical Theory and Practice 2, 433–451

[9] Frishman, D., Argos, P., 1995. Knowledge-based secondary structure assignment. Proteins: Structure Function and Genetics 23, 566-579

[10] Garlipp, T., Müller, C. H., 2006. Detection of linear and circular shapes in image analysis, Computational Statistics & Data Analysis 51, 1479–1490

[11] Huber, P.J., 1967., The behavior of maximum likelihood estimates under nonstandard conditions, Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 221–233

[12] Huber, P.J., Ronchetti, E.M., 2009. Robust statistics, John Wiley & Sons, Inc., Hoboken, New Jersey.

[13] Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577-2637

[14] Kaufman, L., Rousseeuw, P., 2005. Finding groups in data: An introduction to cluster analysis, John Wiley & Son, Hoboken, USA

[15] Kneip, A., Simar, L., 1996. A general framework for frontier estimation with panel data. J. Productivity Anal. 7, 187–212

[16] Meister, A., 2006, Estimating the support of multivariate densities under measurement error, Journal of Multivariate Analysis 97, 1702–1717

[17] Neumann, M. H., 1997. Optimal change-point estimation in inverse problems. Scand. J. Statist. 24, 503–521

[18] Qiu, P., 2005. Image Processing and Jump Regression Analysis, John Wiley&Sons, Inc., Hoboken, New Jersey

[19] Ruzin, S.E., 1999. Plant Microtechnique and Microscopy. Oxford University Press, Oxford.

[20] Sabo, K., 2014. Center-based l_1-clustering method, International Journal of Applied Mathematics and Computer Science, 24, 151–163

[21] Sabo K., Benšić, M., 2009. Border estimation of a disc observed with random errors solved in two steps, Journal of Computational and Applied Mathematics, 229, 16–26

[22] Scitovski, R., Scitovski, S., 2013. A fast partitioning algorithm and its application to earthquake investigation, Computers & Geosciences 59, 124–131

[23] Schneeweiss, H., 2004. Estimating the endpoint of a uniform distribution under measurement errors. CEJOR 12, 221-231

[24] Stefanski, L., Carroll, R.J., 1990. Deconvoluting kernel density estimators. Statistics, 21, 169-184

[25] Sterflinger, K., Krumbein, W. E., 1995. Multiple stress factors affecting growth of rock-inhabiting black fungi, Botanica Acta 108, 490-496

[26] Zhang, C.H., 1990. Fourier methods for estimating mixing densities and distributions. Ann. Statist., 18, 806-830

[27] Zhang, S., Karunamuni, R., 2000. Boundary bias correction for nonparametric deconvolution. Ann. Inst. Statist. Math., 52, 612–629

[28] Williamson, J.A., 1984, A note on the proof by H.E. Daniels of the asymptotic efficiency of a maximum likelihood estimator, Biometrika, 71, 651–653