Estimating the size of an object captured with error

Safet, Hamedović; Mirta Benšić; Kristian Sabo; Petar Taler

Abstract

In many applications we are faced with the problem of estimating object dimensions from a noisy image. Some devices like a fluorescent microscope, X-ray or ultrasound machines, etc., produce imperfect images. Image noise comes from a variety of sources. It can be produced by the physical processes of imaging, or may be caused by the presence of some unwanted structures (e.g. soft tissue captured in images of bones). In the proposed models we suppose that the data are drawn from uniform distribution on the object of interest, but contaminated by an additive error. Here we use two one-dimensional parametric models to construct confidence intervals and statistical tests pertaining to the object size and suggest the possibility of application in two-dimensional problems. Normal and Laplace distributions are used as error distributions. Finally, we illustrate ability of the R-programs we created for these problems on a real-world example.

Key words: noisy image, additive error, maximum likelihood estimator, uniform distribution, normal distribution, Laplace distribution

1 Introduction

Suppose we want to estimate the size of an object from a noisy image. This problem occurs, for example, when the object is captured by a fluorescent microscope [23], a ground penetrating radar, X-ray or ultrasound machines, etc. The dimensions of an object can be calculated from its edges, using some edge detection techniques ([5, 21]), but it is not an easy task. In our approach we don't perform preliminary image restoration or some other image processing techniques. In fact, the model we propose might be used in image analysis.

We suppose that the data taken from the picture are from random variable Y, $Y = U + \varepsilon$, where U is uniformly distributed over the object of interest and ε represents measurement (or some other source of) error (see [1–4, 25]). Although this model is not universal in image describing, we find it is appropriate in some cases. In particularly, it may be helpful in the reconstruction of an object size from a grayscale image.

Here we use one-dimensional case of the model in order to simplify inference. A medical image in Figure 1 gives us some clue how one-dimensional models can help in two-dimensional problems.

Assuming that support of U has circular or elliptical shape it is possible to reconstruct its boundary employing one-dimensional models and parametric curve fitting. Two-phased procedure dealing with this problem can be seen in [2, 25].

In section 2, we briefly describe general one-dimensional model we use. Two types of error distribution, namely normal and Laplace distributions, are analized in section 3. We recall some properties of estimators found by method of moments (MM) and maximum likelihood method (ML) described in [1, 3, 4]. In section 4 we give two classic methods of constructing a confidence intervals and hypothesis testing. In the last section we demonstrate R-program created on the basis of this model.

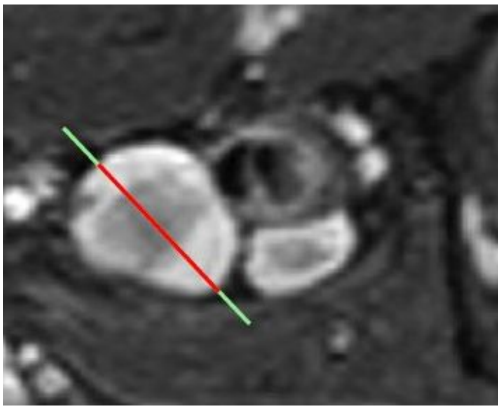

Figure 1: The line-object intersection

2 The Model

Suppose that random variable U has the uniform distribution on the interval $[-a, a]$, $a > 0$, but observations from U are contaminated by an additive error ε, so that we can only observe

$$Y = U + \varepsilon.$$

Furthermore, assume that the random variables U and ε are independent. Our goal is to estimate the parameter a, based on an i.i.d. sample $Y_1, Y_2, \dots, Y_n$ (see [1–4, 25])

This is the special case of general additive error model $Y = X + \varepsilon$, where X and ε are assumed to be independent continuous random variables, but only Y is observable. Usually, density f_ε is supposed to be completely known, but some papers deal with only partially known error densities. Recovering the unknown density f_X from i.i.d. sample $Y_1, Y_2, \dots, Y_n, Y_1 \sim Y$, is known as deconvolution problem. Several nonparametric methods are developed to estimate f_X (see [20]), but the most popular and studied is the deconvolution kernel density estimator ([6, 29]). It is known that although this estimator is optimal, it converges at low rates, particularly in the case of so-called supersmooth error densities. For example, with normal error the rate of convergence is only logarithmic ([10]). Some papers deal with the problem of estimating only the support of f_X ([19]). Here we suppose that the X part is from simple parametric model (the uniform one) and discuss parametric approach in support estimation.

If U is uniformly distributed on $[-a, a]$, i.e. has density function

$$f_U(x) = \frac{1}{2a} I_{[-a,a]}(x),$$

and ε is continuous random variable with density f_ε and distribution function F_ε, then the density of $Y = U + \varepsilon$ is

$$f_Y(x; a) = \int_{-\infty}^{\infty} f_U(t) f_\varepsilon(x - t)dt = \frac{1}{2a} \left(F_\varepsilon(x + a) - F_\varepsilon(x - a) \right).$$

3 Two types of error

3.1 Normal error

The most common assumption is that error part ε has a normal distribution with zero mean. The case $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ with known σ^2 is analysed in [1]. In that paper the MM and ML estimators are suggested and compared.

Let $\mathbf{y} = (y_1, \ldots, y_n)$ denotes the realization of the sample $\mathbf{Y} = (Y_1, \ldots, Y_n)$. The likelihood function has the form

$$L(a; \mathbf{y}) = \prod_{i=1}^{n} f_Y(y_i; a) = \frac{1}{(2a)^n} \prod_{i=1}^{n} \left(\Phi\left(\frac{y_i + a}{\sigma}\right) - \Phi\left(\frac{y_i - a}{\sigma}\right) \right) \tag{1}$$

and the log-likelihood function is given by

$$l(a) = -n \log(2a) + \sum_{i=1}^{n} \log\left(\Phi\left(\frac{y_i + a}{\sigma}\right) - \Phi\left(\frac{y_i - a}{\sigma}\right) \right) \tag{2}$$

where $\Phi(x)$ is the standard normal cdf. In [1] the regularity conditions are checked, thus

$$\sqrt{n}\,(\hat{a}_{ML} - a_0) \longrightarrow \mathcal{N}\left(0, \frac{1}{I(a_0)}\right), \tag{3}$$

where

$$I(a) = \frac{-1}{a^2} + \frac{1}{a\sigma^2} \int_0^\infty \frac{\left(\varphi\left(\frac{x+a}{\sigma}\right) + \varphi\left(\frac{x-a}{\sigma}\right)\right)^2}{\Phi\left(\frac{x+a}{\sigma}\right) - \Phi\left(\frac{x-a}{\sigma}\right)}\, dx \tag{4}$$

with $\varphi(x)$ being the standard normal pdf. Also, in ([1]) is shown that the ML estimator is more efficient than the MM estimator. The case with unknown σ^2 is considered in [3], where regularity of corresponding two-parametric model is proved.

3.2 Laplace error

However, estimators based on normal error model are shown to be very sensitive to the presence of outliers in the data. One possible choice is to use Laplace error model instead, since the Laplace distribution has heavier tails than the normal distribution (see [4]). Laplace distribution with the location parameter $\mu = 0$ and a scale parameter $\lambda > 0$ has the density function

$$f_\lambda(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}.$$

Similarly to the normal error case, the likelihood function and the log-likelihood function have the form

$$L(a; \mathbf{y}) = \prod_{i=1}^{n} f_Y(y_i; a) = \frac{1}{(2a)^n} \prod_{i=1}^{n} \left(F\left(\frac{y_i + a}{\lambda}\right) - F\left(\frac{y_i - a}{\lambda}\right) \right) \tag{5}$$

$$l(a) = -n \log(2a) + \sum_{i=1}^{n} \log\left(F\left(\frac{y_i + a}{\lambda}\right) - F\left(\frac{y_i - a}{\lambda}\right) \right) \tag{6}$$

where $F(x)$ is the standard Laplace ($\lambda = 1$) cdf. Classical regularity conditions are not satisfied in the model with Laplace error. Nevertheless, consistency and asymptotic efficiency of ML estimator are proven in [4] using some nonstandard conditions. Thus we have

$$\sqrt{n}\,(\hat{a}_{ML} - a_0) \longrightarrow \mathcal{N}\left(0, \frac{1}{I(a_0)}\right), \tag{7}$$

where

$$I(a) = \frac{-1}{a^2} + \frac{1}{a\lambda^2}\int_0^\infty \frac{\left(f\left(\frac{x+a}{\lambda}\right) + f\left(\frac{x-a}{\lambda}\right)\right)^2}{F\left(\frac{x+a}{\lambda}\right) - F\left(\frac{x-a}{\lambda}\right)}\,dx \tag{8}$$

Here, $f(x)$ is the standard Laplace ($\lambda = 1$) pdf. The ML estimator in the model with Laplace error is shown to be more robust than in the model with normal error, as expected (see [4]).

4 Confidence intervals and tests

Asymptotic distributions of ML estimators in (3) and (7) can be used to construct approximate confidence intervals for a. For a specified $0 < \alpha < 1$, regardless of error type, an asymptotic $(1-\alpha)100\%$ confidence interval (AD.CI) for a is[1]

$$\left(\hat{a}_{ML} - \frac{z_{\alpha/2}}{\sqrt{nI(\hat{a}_{ML})}}, \hat{a}_{ML} + \frac{z_{\alpha/2}}{\sqrt{nI(\hat{a}_{ML})}}\right).$$

Note that we use $I(\hat{a}_{ML})$ as a consistent estimator of $I(a_0)$.

Another way to compute a confidence interval is based on the likelihood ratio statistic

$$\lambda(\mathbf{Y}) = \frac{\sup\limits_{a_0} L\,(a;\mathbf{Y})}{\sup\limits_{(0,\infty)} L\,(a;\mathbf{Y})} = \frac{L\,(a_0;\mathbf{Y})}{L\,(\hat{a}_{ML};\mathbf{Y})}$$

We know that (recall that models with both type of errors are regular)

$$-2\log\lambda(\mathbf{Y}) \xrightarrow{D} \chi_1^2$$

where χ_1^2 is a χ^2 random variable with 1 degree of freedom, provided a_0 is the true value of a. An approximate $(1-\alpha)100\%$ confidence interval (LR.CI) for a is

$$\left\{a \middle|\, l(\hat{a}_{ML}) - l(a) \leq 0.5\chi_1^2(1-\alpha)\right\},$$

where $\chi_1^2(1-\alpha)$ is the $1-\alpha$ quantile of χ_1^2 distribution.

We performed a simulation study to compare the AD and LR confidence intervals. The parameter a is fixed at $a = 1$ and the scale parameters are varied. For each of the three sample sizes ($n \in \{30, 300, 1000\}$), $N = 1000$ samples of size n were generated. The

[1] as usual, z_α is the $1-\alpha$ quantile of standard normal distribution

average length and cover rate for confidence coefficient 0.95 are presented in Tables 1-6. In small samples some difficulties in computing confidence intervals may occur. Namely, in the AD approach it may happen that estimated asymptotic variance of $\hat{a}_{ML}$ does not exist. On the other hand, in the LR approach it happens that equation $l(\hat{a}_{ML}) - l(a) = 0.5\chi_1^2(1-\alpha)$ has only one solution. Therefore, the number of successfully computed confidence intervals is also presented in these tables. As expected, smaller values of scale parameter lead to shorter intervals. For moderate and large samples, the AD.CI and LR.CI are similar in length and cover rate. In small samples AD.CI are slightly shorter than LR.CI (except for large values of scale parameter), but have worse cover rate. The AD.CI are computationally more acceptable, but the LR.CI have the advantage of being exact invariant under reparametrization.

The two approaches above can be used to testing hypothesis about a parameter a. The asymptotic tests are based on limiting distributions of the test statistics $W = \sqrt{nI(a_0)}(\hat{a}_{ML} - a_0)$ (AD approach) and $\chi_L^2 = -2\log\lambda(\mathbf{Y})$ (LR approach).

For testing the hypothesis $H_0 : a = a_0$ against $H_1 : a \neq a_0$ the critical regions of asymptotic size α are

$$\left\{\mathbf{y}|\ \sqrt{nI(a_0)}|\hat{a}_{ML} - a_0| \geq z_{\alpha/2}\right\}, \text{(AD approach) and} \tag{9}$$

$$\left\{\mathbf{y}|\ -2\log\lambda(\mathbf{y}) \geq \chi_1^2(1-\alpha)\right\}, \text{(LR approach)}.$$

In order to make meaningful comparison of tests we should compute the power in a sequence of so-called local alternatives $a_n = a_0 + \delta/\sqrt{n}$, $\delta \in \mathbb{R}$. For the AD approach it is more convenient to use statistic $\chi_W^2 := W^2 = nI(a_0)\left(\hat{a}_{ML} - a_0\right)^2$ (known as Wald statistic). In terms of this statistic (note that $\chi_1^2(1-\alpha) = z_{\alpha/2}^2$), the critical region (9) has the form $\{\mathbf{y}|\ \chi_W^2 \geq \chi_1^2(1-\alpha)\}$. Under the local alternatives, both χ_W^2 and χ_L^2 statistics have an asymptotic $\chi^2_{1,\delta^2 I(a_0)}$ distribution[2] (see [28]).

For a given alternative a and sample size n, this result can be used to get an approximate power at a by equating a and a_n. In this way, an approximate power function of both tests is

$$\beta(a) = 1 - F\left(\chi_1^2(1-\alpha)\right),$$

where $F(x)$ is the cdf of non-central χ^2 distribution with 1 degree of freedom and non-centrality parameter $nI(a_0)(a-a_0)^2$.

For moderate values of scale parameter nominal (theoretical) power is close to the empirical ones of both tests, even in relatively small samples. In any case, moderate or large samples ensure very close power functions. Also, the power increases as the value of scale parameter decreases. As an example, simulation results with normal error $\mathcal{N}(0,(0.25)^2)$ are shown in Figure 2. The null value is set at $a_0 = 1$. For each of the alternatives $a \in \{0.60, 0.64, \ldots, 1.40\}$, $N = 1000$ samples of size n ($n \in \{30, 100\}$) are generated. We computed the proportion of rejected null hypotheses under a (at 0.05 level), which we refer to as empirical power at a.

[2] $\chi^2_{k,p}$ is the non-central χ^2 distribution with k degree of freedom and non-centrality parameter p

σ	AD.CI Success	AD.CI Length	AD.CI Cover rate	LB.CI Success	LB.CI Length	LB.CI Cover rate
1.00	880	2.5608	0.939	214	1.7958	0.855
0.50	999	0.8436	0.967	937	0.8314	0.963
0.25	1000	0.4412	0.933	1000	0.4471	0.938
0.10	1000	0.2508	0.935	1000	0.2562	0.952
0.05	1000	0.1719	0.921	1000	0.1788	0.946
0.01	1000	0.0749	0.748	1000	0.0944	0.914

Table 1: Simulation results for the CI, $a = 1, n = 30, N = 1000$ (normal error)

σ	AD.CI Success	AD.CI Length	AD.CI Cover rate	LB.CI Success	LB.CI Length	LB.CI Cover rate
1.00	1000	0.6451	0.966	942	0.6548	0.968
0.50	1000	0.2506	0.940	1000	0.2514	0.939
0.25	1000	0.1400	0.956	1000	0.1404	0.958
0.10	1000	0.0799	0.959	1000	0.0801	0.958
0.05	1000	0.0548	0.945	1000	0.0551	0.951
0.01	1000	0.0239	0.938	1000	0.0243	0.955

Table 2: Simulation results for the CI, $a = 1, n = 300, N = 1000$ (normal error)

For testing $H_0 : a = a_0$ against one-sided alternatives, it is straightforward to create tests based on W statistic. For example, in the case of alternative hypothesis $H_1 : a > a_0$ the critical region of size α is

$$\left\{ \mathbf{y} | \sqrt{nI(a_0)}\,(\hat{a}_{ML} - a_0) \geq z_\alpha \right\}$$

with approximate power function

$$\beta(a) = \Phi\left(\sqrt{nI(a_0)}(a - a_0) - z_\alpha\right)$$

However, it is more difficult to adapt the likelihood ratio test to the one-sided alternatives since the null value is on the boundary of the parameter space.

5 A real world example

To enable the use of these models in the real world, we are building an R-package (the work is still in progress). In the loaded image we choose the line in desired direction and extract the data along this line. We can estimate the length of the line-object intersection (line profile) using our one-dimensional models (see screenshot at Figure 3). The arguments of the main R function are: the data extracted from the line, the type of the error distribution, standard error of the error distribution (known, ML estimate or MM estimate), confidence level (for the CI), the null value and the type of alternative hypothesis

σ	AD.CI Success	AD.CI Length	AD.CI Cover rate	LB.CI Success	LB.CI Length	LB.CI Cover rate
1.00	1000	0.3446	0.955	1000	0.3477	0.955
0.50	1000	0.1371	0.956	1000	0.1372	0.957
0.25	1000	0.0767	0.958	1000	0.0767	0.958
0.10	1000	0.0437	0.952	1000	0.0438	0.953
0.05	1000	0.0300	0.951	1000	0.0300	0.951
0.01	1000	0.0131	0.947	1000	0.0132	0.952

Table 3: Simulation results for the CI, $a = 1, n = 1000, N = 1000$ (normal error)

λ	AD.CI Success	AD.CI Length	AD.CI Cover rate	LB.CI Success	LB.CI Length	LB.CI Cover rate
1.00	1000	4.0450	0.951	206	2.1543	0.815
0.50	1000	1.0691	0.957	766	1.0628	0.953
0.25	1000	0.5375	0.920	999	0.5622	0.932
0.10	1000	0.2926	0.917	1000	0.3075	0.930
0.05	1000	0.1984	0.900	1000	0.2133	0.935
0.01	1000	0.0857	0.772	1000	0.1016	0.898

Table 4: Simulation results for the CI, $a = 1, n = 30, N = 1000$ (Laplace error)

λ	AD.CI Success	AD.CI Length	AD.CI Cover rate	LB.CI Success	LB.CI Length	LB.CI Cover rate
1.00	1000	0.7569	0.957	924	0.7619	0.976
0.50	1000	0.3201	0.952	1000	0.3240	0.949
0.25	1000	0.1697	0.960	1000	0.1706	0.961
0.10	1000	0.0929	0.948	1000	0.0934	0.952
0.05	1000	0.0631	0.949	1000	0.0637	0.949
0.01	1000	0.0274	0.927	1000	0.0285	0.942

Table 5: Simulation results for the CI, $a = 1, n = 300, N = 1000$ (Laplace error)

λ	AD.CI Success	AD.CI Length	AD.CI Cover rate	LB.CI Success	LB.CI Length	LB.CI Cover rate
1.00	1000	0.4030	0.953	999	0.4077	0.950
0.50	1000	0.1751	0.951	1000	0.1759	0.949
0.25	1000	0.0930	0.951	1000	0.0932	0.948
0.10	1000	0.0509	0.945	1000	0.0509	0.950
0.05	1000	0.0346	0.947	1000	0.0348	0.947
0.01	1000	0.0150	0.938	1000	0.0152	0.945

Table 6: Simulation results for the CI, $a = 1, n = 1000, N = 1000$ (Laplace error)

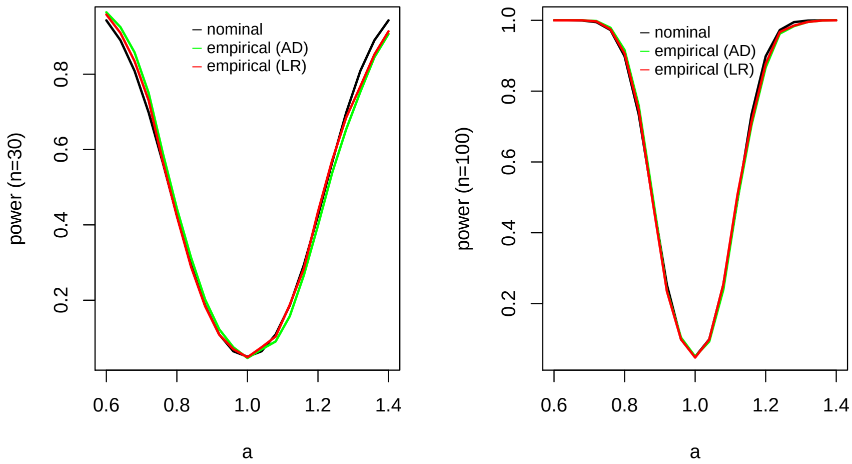

Figure 2: Power functions (empirical and nominal) with $a_0 = 1$ and error $\varepsilon \sim \mathcal{N}(0, (0.25)^2)$

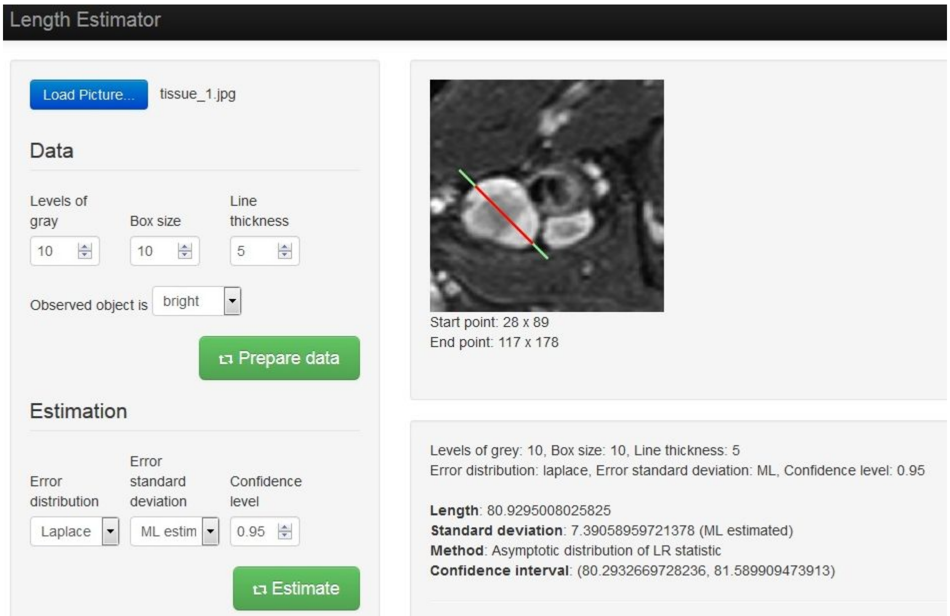

Figure 3: Screenshot of a line profile length estimation.

(for the tests), the number of gray levels, box size, line thickness (the "line" is not a line but a thin rectangle). The values of R function, among others, are: estimated line profile, estimated standard error of the error distribution, confidence interval, p-value. We are going to enrich the program with some more features. For example, we plan to include: more error distributions, computing the maximum width and height of the object, testing if there is a change in size of the same object but captured at different times, etc..

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 - A Journal of Theoretical and Applied 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 - A Journal of Theoretical and Applied Statistics, 44, 235–246

[4] Benšić, M., Sabo, K. (2016). Uniform distribution width estimation from data observed with Laplace additive error, Journal of the Korean Statistical Society

[5] Canny, J. (1986) A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. 8, 679â“-698

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

[7] Carroll, R. J., Ruppert, D., Stefanski, L. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models.* Second Edition. Chapman and Hall/CRC, Boca Raton.

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

[9] Delaigle, A., Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems, Statistica Sinica 16, 773–788

[10] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist.,19, 1257–1272.

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

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

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

[14] 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.

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

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

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

[18] Lehmann, E.L., Casella, G. (1998). *Theory of Point Estimation,* Springer, Berlin, Germany

[19] Meister, A., (2006). Estimating the support of multivariate densities under measurement error, Journal of Multivariate Analysis 97 (2006) 1702 â"- 1717

[20] Meister, A., (2009). *Deconvolution Problems in Nonparametric Statistics.* Springer-Verlag, Berlin Heidelberg.

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

[22] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[23] Ruzin, S.E. (1999). *Plant Microtechnique and Microscopy.* Oxford University Press, Oxford.

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

[25] 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.

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

[27] Schneeweiss, H. (2004). Estimating the endpoint of a uniform distribution under measurement errors. CEJOR 12, 221–231.

[28] Sen, P.B., Singer, J.M., de Lima, A.C.P. (2010), *From Finite Sample to Asymptotic Methods in Statistics*, Cambridge University Press, New York.

[29] Stefanski, L., Carroll, R.J. (1990). Deconvoluting kernel density estimators. Statistics, 21, 169–184

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

[31] Zhang, C.H. (1990). Fourier methods for estimating mixing densities and distributions. Ann. Statist., 18, 806–830.

[32] Williamson, J.A. (1984), A note on the proof by H.E. Daniels of the asymptotic efficiency of a maximum likelihood estimator, Biometrika, 71,3, 651–653.