The width of a uniform distribution: estimation in additive error models

M. Benšić[1] K. Sabo[1] S. Hamedović[2]

[1]Department of Mathematics
University of Osijek

[2]Faculty of Metallurgy and Technology
University of Zenica

This work is supported by the Croatian Science Foundation through the research grant IP-2016-06-6545.

ISCCRO'18, Opatija, May 2018

Additive error model

$$X = Y + \varepsilon$$

- Typical assumptions
  - Y and ε are independent
  - error variable ε has a known PDF f_ε
  - an i.i.d. sample $(X_1, X_2, \ldots, X_n)$ on X is available
- The goal: to estimate the unknown density f_Y
- Deconvolution: non-parametric methods

Deconvolution scheme

$$X = Y + \varepsilon$$

- Technique of characteristic functions

$$\phi_X(t) = \phi_Y(t)\phi_\varepsilon(t)$$

- Rough scheme
  - estimate $\phi_X(t)$ by an empirical version $\hat{\phi}_X(t)$
  - get estimate of $\phi_Y(t)$ as $\hat{\phi}_Y(t) = \frac{\hat{\phi}_X(t)}{\phi_\varepsilon(t)}$
  - regularize $\hat{\phi}_Y(t)$ before applying inversion and get $\hat{f}_Y(x)$
- Deconvolution kernel density estimator

Model

$$X = U + \sigma\varepsilon,\ U \sim \mathcal{U}\left[-a, a\right],\ a > 0, \sigma > 0$$

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

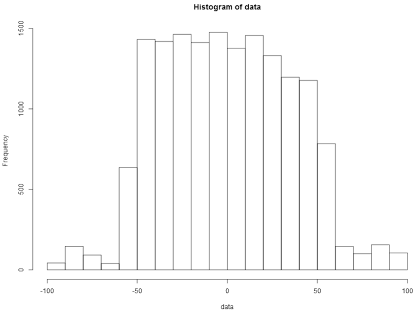
- ε symmetric absolutely continuous with CDF F and PDF f
- PDF of X (convolution of f_U and $f_{\sigma\varepsilon}$):

$$f_X(x; a, \sigma) = \frac{1}{2a}\left(F\left(\frac{x+a}{\sigma}\right) - F\left(\frac{x-a}{\sigma}\right)\right)$$

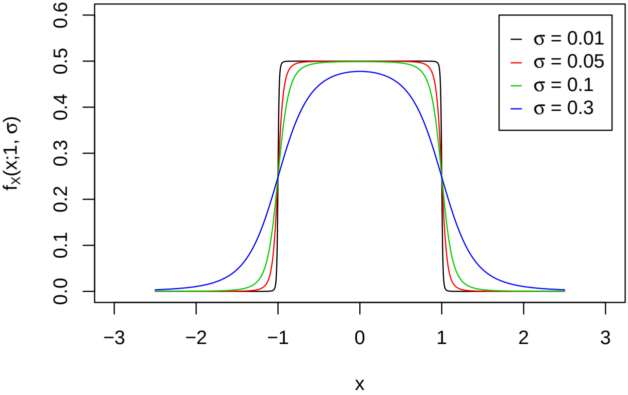
- To estimate:
  - a (half-length of the uniform support)
  - σ

Motivation

- Estimating the size of an object from a noisy image
- Example, a medical (RTG) image



- PDF of $X = U + \varepsilon$, $U \sim \mathcal{U}[-1, 1]$, $\varepsilon \sim t_3(0, \sigma)$



Some references (1)

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

Some references (2)

- 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
- Benšić, M., Sabo, K. (2016). Uniform distribution width estimation from data observed with Laplace additive error, Journal of the Korean Statistical Society
- Benšić, M., Taler, P., Hamedović, S., Nyarko, E.K., Sabo, K. (2017). LeArEst: Length and Area Estimation from Data Measured with Additive Error, The R Journal, 9:2, pages 461-473.
- Hamedović, S., Benšić, M., Sabo, K., Taler, P. (2017). Estimating the size of an object captured with error, CEJOR,

ML estimator

- Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a realization of $\mathbf{X} = (X_1, \ldots, X_n)$
- log-likelihood function:

$$l(a; \mathbf{x}) = -n \log(2a) + \sum_{i=1}^{n} \log\left(F\left(\frac{x_i + a}{\sigma}\right) - F\left(\frac{x_i - a}{\sigma}\right)\right)$$

- ML estimator of a

$$\hat{a}_{ML} = \underset{a \in (0,\infty)}{\arg\max}\, l(a; \mathbf{x})$$

Some examples of error

- Normal error; $\varepsilon \sim \mathcal{N}(0,1)$; $f_\varepsilon(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
  - corresponding ML estimator of a is CAE
  - classical regularity conditions
  - sensitive to outliers
- Laplace error; $\varepsilon \sim \mathcal{L}\left(0, \sqrt{2}/2\right)$; $f_\varepsilon(x) = \frac{\sqrt{2}}{2} e^{-2|x|/\sqrt{2}}$
  - corresponding ML estimator of a is CAE
  - nonstandard regularity conditions
  - more robust

Asymptotic distribution

- So, in both cases

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

- where

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

- We can construct confidence intervals and tests

Confidence intervals

- Approximate confidence intervals (CI) derived from asymptotic distribution of $\hat{a}_{ML}$:

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

- Likelihood ratio (LR) based confidence intervals:

$$\lambda(\mathbf{X}) = \frac{\sup_{a_0} L\left(a; \mathbf{X}\right)}{\sup_{(0,\infty)} L\left(a; \mathbf{X}\right)} = \frac{L\left(a_0; \mathbf{X}\right)}{L\left(\hat{a}_{ML}; \mathbf{X}\right)}$$

- Under a_0 (the true value of a):

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

- LR.CI is the set:

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

Testing

- For testing $H_0 : a = a_0$ against $H_1 : a \neq a_0$, critical regions:
  - AD (Wald) approach $\left\{\mathbf{x} | \sqrt{nI(a_0)} |\hat{a}_{ML} - a_0| \geq z_{\alpha/2}\right\}$ or

$$\left\{\mathbf{x} | \chi^2_W := nI(a_0) \left(\hat{a}_{ML} - a_0\right)^2 \geq \chi^2_1(1-\alpha)\right\}$$

  - LR approach

$$\{\mathbf{x} | \chi^2_L := -2 \log \lambda(\mathbf{x}) \geq \chi^2_1(1-\alpha)\})$$

- Approximate power

$$\beta(a) = 1 - F\left(\chi^2_1(1-\alpha)\right)$$

- $F(x)$ is the cdf of non-central χ^2 distribution with 1 df and ncp $nI(a_0)(a - a_0)^2$

General model

$$X = U + \sigma\varepsilon,\ U \sim \mathcal{U}\left[-a, a\right],\ a > 0, \sigma > 0$$

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

- We have conditions on error PDF for the MLE to be CAE
- Examples of errors with which the model is regular
    - normal error
    - logistic error
    - student t error
- student t error - more robust model

The LeArEst R package

- LeArEst: Border and Area Estimation of Data Measured with Additive Error
- available on CRAN: https://cran.r-project.org/package=LeArEst
- estimation of the uniform support from the contaminated data
- constructing CI and tests
- area estimation of an elliptical object