CEREMADE
UMR CNRS 7534

Dauphine | PSL
UNIVERSITÉ PARIS

MAXIMUM SPACING ESTIMATION

A NEW METHOD IN FITDISTRPLUS

Christophe Dutang[1]

[1] CEREMADE, Université Paris Dauphine

useR 2019 conference, July 10

Introduction

The **fitdistrplus** project

- started in 2009: stable version 1.0-9 on CRAN (first release),
- extensively enhanced between 2009-2018: 17 versions on CRAN,
- published 2015: publication in JSS [**?**],
- currently, 2019: last stable version 1.0-14.

Presented at

- useR 2009 in Rennes, useR 2011 in Warwick,
- Rencontres R 2013 in Lyon, Rencontres R 2018 in Rennes.

Today, we present the implementation of a new estimation method: maximum spacing estimation (MSE)

- This method was introduced by Cheng and Amin (1986) and Ranneby (1984) independently.
- Currently, only the **BMT** package provides MSE for the Bezier-Montenegro-Torres distribution.

Maximum spacing estimation (MSE)

Consider a sample of observations $(x_1, \ldots, x_n)$.

Order statistics are denoted by $(x_{(1)} < \cdots < x_{(j)} < \cdots < x_{(n)})$.

Spacings on the distribution function $F(;\theta)$ are defined as

$$D_i(\theta) = F(x_{(i)};\theta) - F(x_{(i-1)};\theta),\ i = 1,;n+1$$

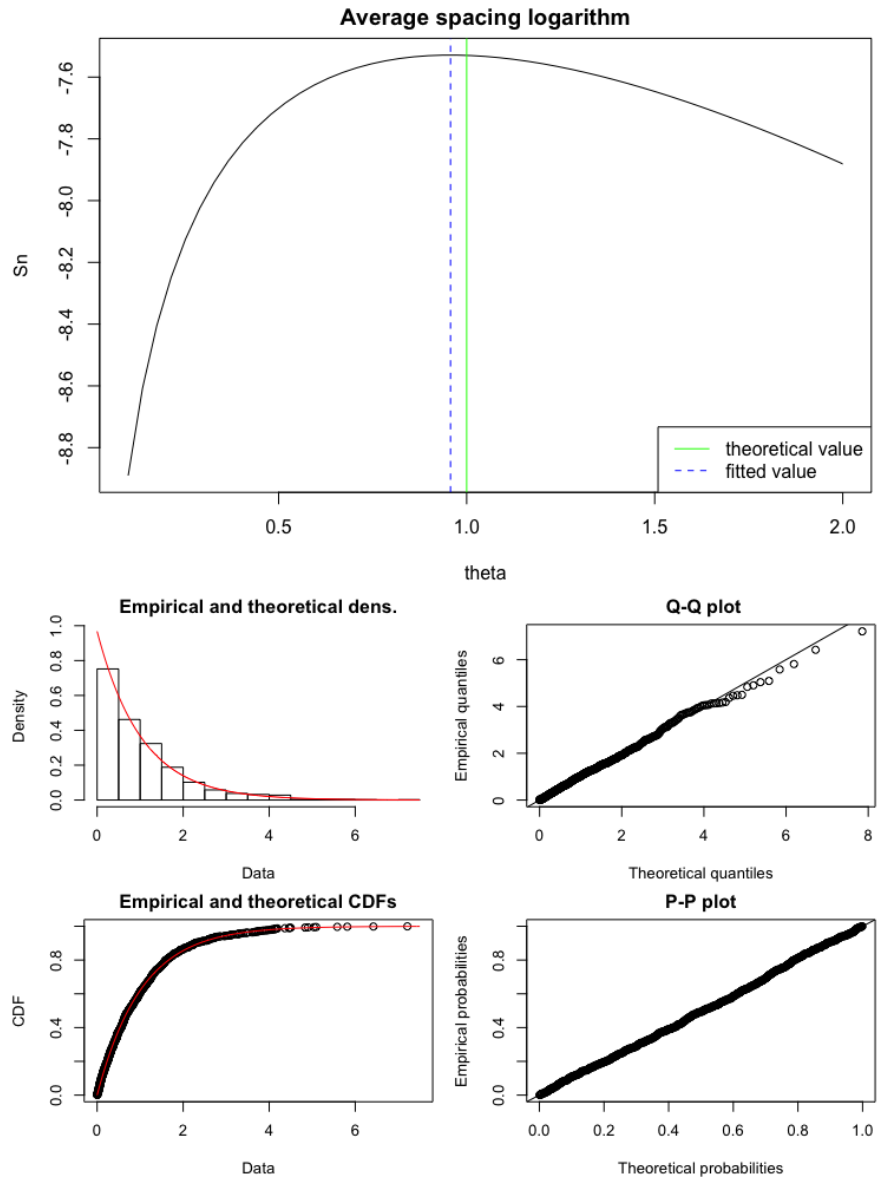
where $x_{(0)} = -\infty$ and $x_{(n+1)} = +\infty$.

MSE consists in maximizing the average of the spacing logarithm

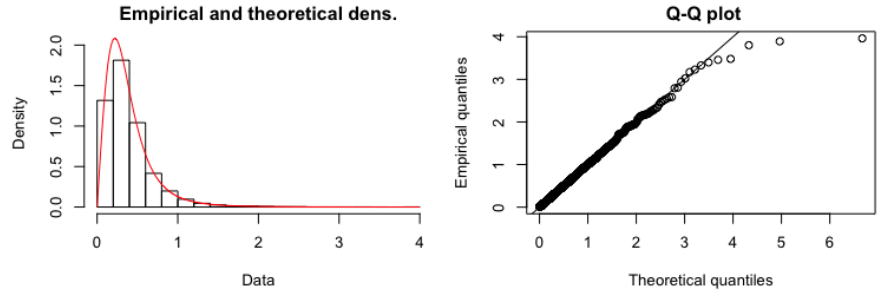
$$S_n(\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i(\theta).$$

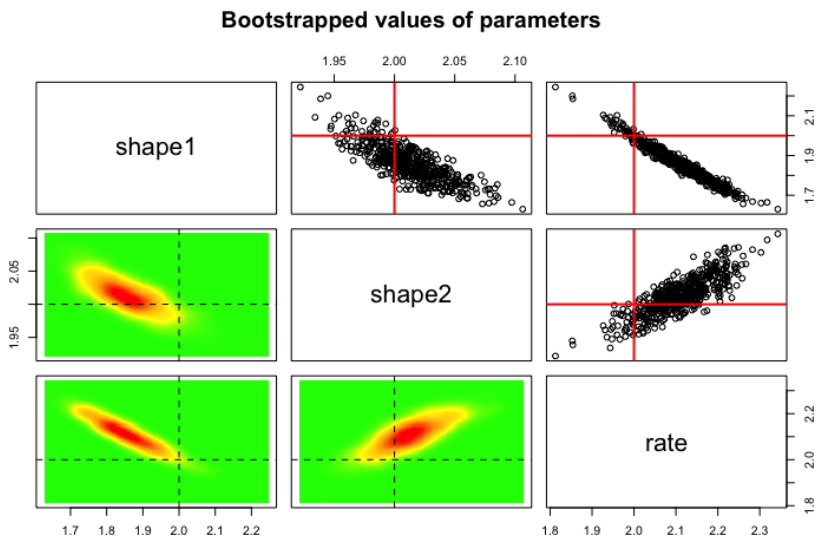
Under certain conditions, MSE has asymptotically a normal distribution as MLE.

Example 1 – Light-tailed continuous distribution

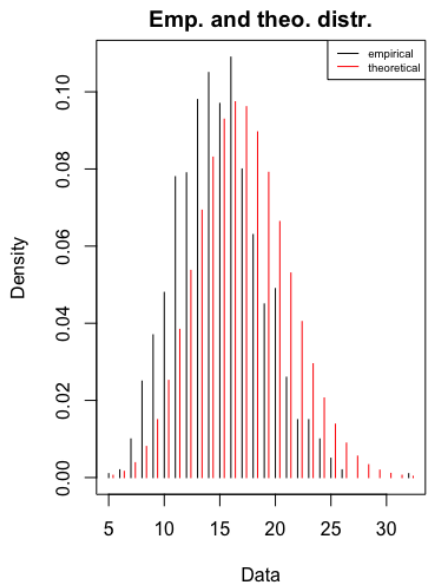


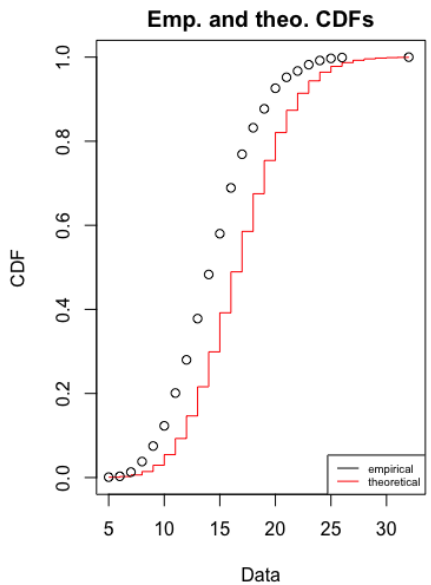
EXAMPLE 2 – HEAVY-TAILED CONTINUOUS DISTRIBUTION

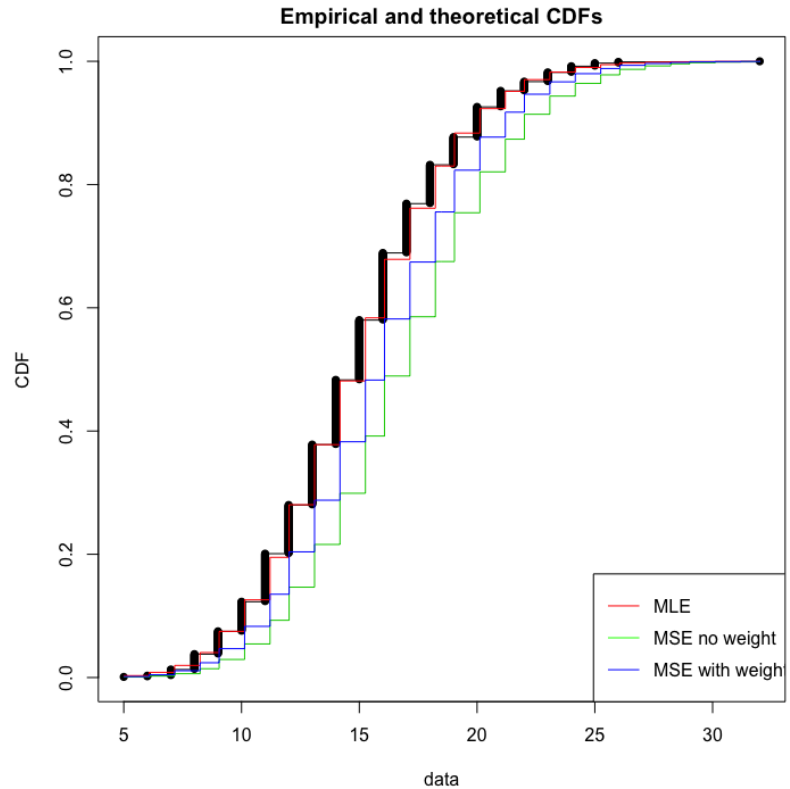




EXAMPLE 3 – DISCRETE DISTRIBUTION







Conclusion and perspectives

We investigate in details **survival**, **fitdistrplus**, **flexsurv** packages for

- fitting parametric models,
- assessing goodness of fits,
- using bootstrap to quantify uncertainty.

Similar results obtained on a Canadian dataset.

Competitors

`https://cran.r-project.org/web/packages/MPS/index.html`
`https://cran.r-project.org/web/packages/BMT/index.html`

REFERENCES

Kaplan, E. L. & Meier, P. (1958),

Nonparametric estimation from incomplete observations,
Journal of the American Statistical Association, 53(282), 457-481

Hagar, Y., & Dukic, V. (2015),

Comparison of hazard rate estimation in R,
arxiv working paper.

Kalbfleisch, J. D. and Prentice, R. L. (2002),
The statistical analysis of failure time data,
Wiley.

McCullagh, P., Nelder, J.A. (1989),
Generalized linear models, 2nd edn,
Springer.

R Core Team (2019).

R: A Language and Environment for Statistical Computing,
https://www.R-project.org/

Allignol A. (2019).

Survival Task view,
https://CRAN.R-project.org/view=Survival

Therneau T (2019),

A Package for Survival Analysis in S,
https://CRAN.R-project.org/package=survival

Jackson C (2019),

flexsurv: A Platform for Parametric Survival,
https://CRAN.R-project.org/package=flexsurv

Hess K & Gentleman R (2019),

muhaz: Hazard Function Estimation in Survival Analysis,
https://CRAN.R-project.org/package=muhaz

Spedicato GA et al. (2019),

The lifecontingencies Package: Performing Financial and Actuarial Mathematics Calculations in R,
https://CRAN.R-project.org/package=lifecontingencies

Delignette-Muller ML & Dutang C (2019),

fitdistrplus: An R Package for Fitting Distributions,
https://CRAN.R-project.org/package=fitdistrplus