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.

# 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.

**fitdistrplus** extends the `fitdistr()` function (of the MASS package)

- Censored data may contain left censored, right censored and interval censored values, with several lower and upper bounds.
- the package also provides moment matching (MME), quantile matching (QME) and maximum goodness-of-fit estimation (MGE) methods.

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

- This method was introduced by Cheng and Amin (1983) and Ranneby (1984) independently.
- Currently, the **BMT** package provides MSE for the Bezier-Montenegro-Torres distribution.
- the **MPS** package provides MSE only for a selected number of distribution.

# Maximum spacing estimation (MSE)

Consider a sample of observations $(x_1, \dots, x_n)$ (assuming real-valued observations).

- We can compute order statistics as $x_{(1)} < \dots < x_{(j)} < \dots < x_{(n)}$.
- Spacings on the distribution function $F(;\theta)$ are defined as

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

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

- we know $D_i(\theta) > 0$ for any continuous cdf.

# Maximum spacing estimation (MSE)

Consider a sample of observations $(x_1, \ldots, x_n)$ (assuming real-valued observations).

- We can compute order statistics as $x_{(1)} < \cdots < x_{(j)} < \cdots < x_{(n)}$.
- Spacings on the distribution function $F(;\theta)$ are defined as

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

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

- we know $D_i(\theta) > 0$ for any continuous cdf.

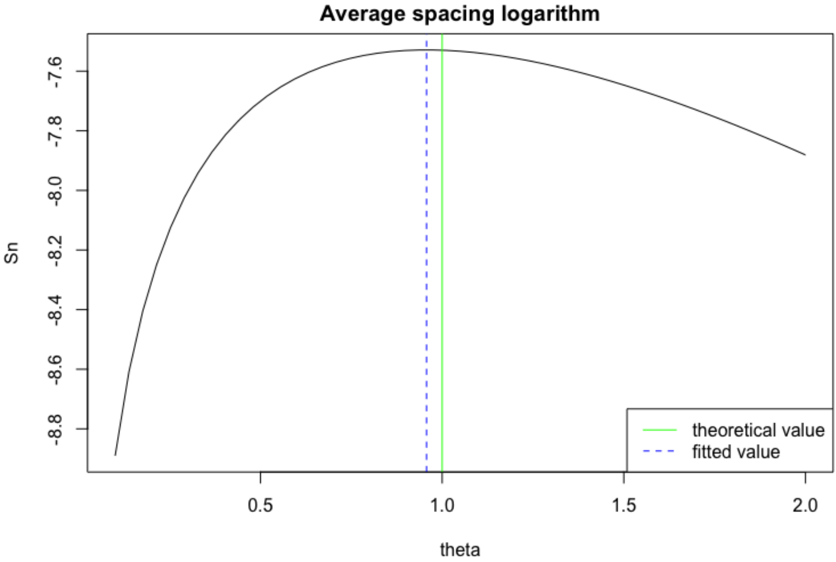
MSE consists in maximizing the average of the spacing logarithm

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

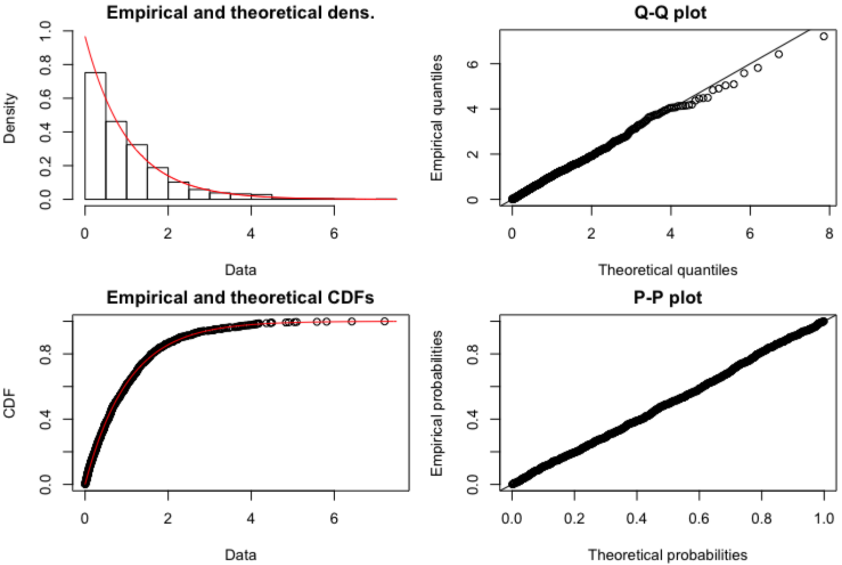
Under certain regularity conditions, MSE has asymptotically a normal distribution as MLE, see Ekstrom (2015).

# EXAMPLE 1 – STD. EXPONENTIAL DISTRIBUTION – AVG. SPACING

Simulated dataset from $\mathcal{E}(1)$ and $n = 1000$.

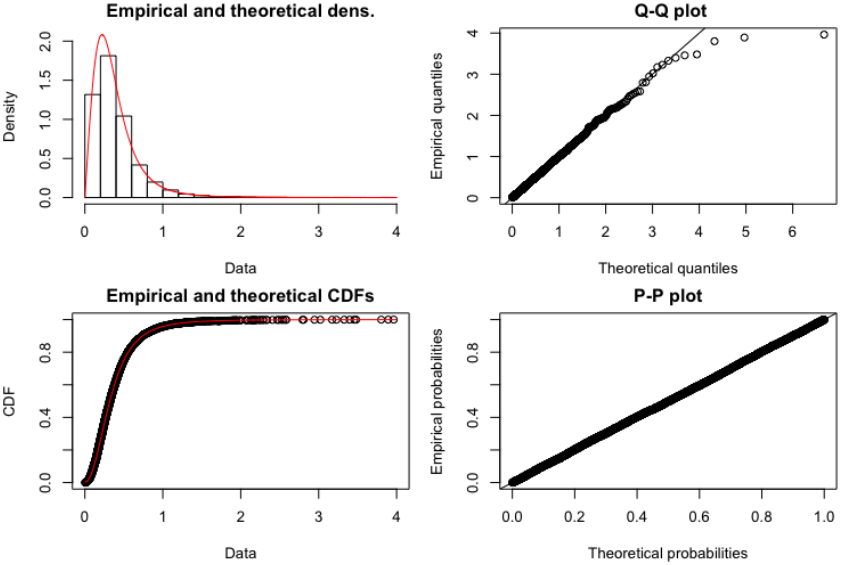


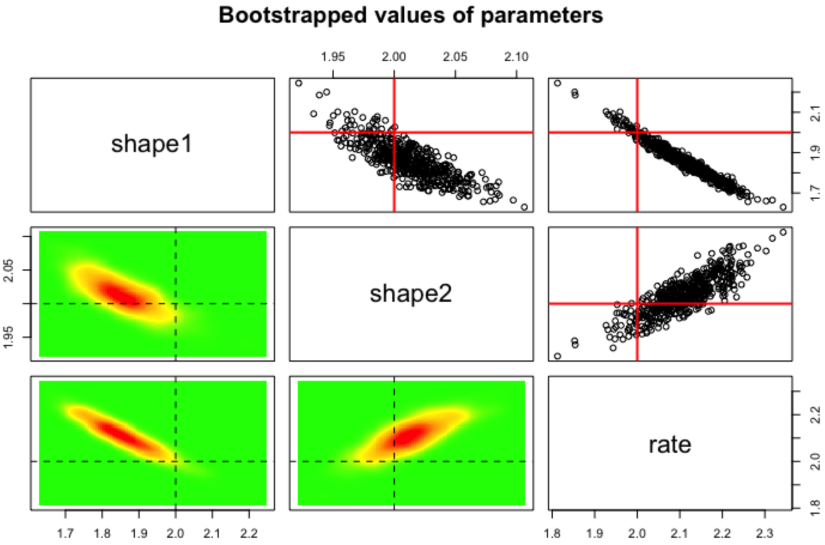
# EXAMPLE 1 – STD. EXPONENTIAL DISTRIBUTION – GOF. PLOTS



# EXAMPLE 2 – BURR DISTRIBUTION – GOF. PLOTS

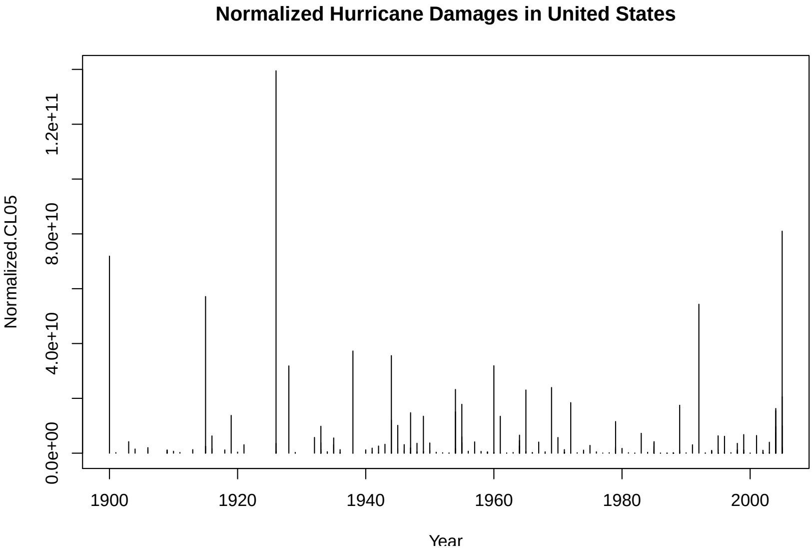
Simulated dataset from $\mathcal{B}urr(2,2,2)$ and $n = 10000$.



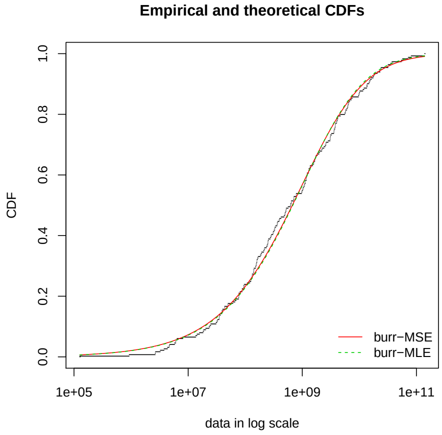

Bootstrapped values of parameters
shape1
shape2
rate
1.95
2.00
2.05
2.10
2.1
1.9
1.7
2.05
1.95
2.2
2.0
1.8
1.7
1.8
1.9
2.0
2.1
2.2
1.8
1.9
2.0
2.1
2.2
2.3

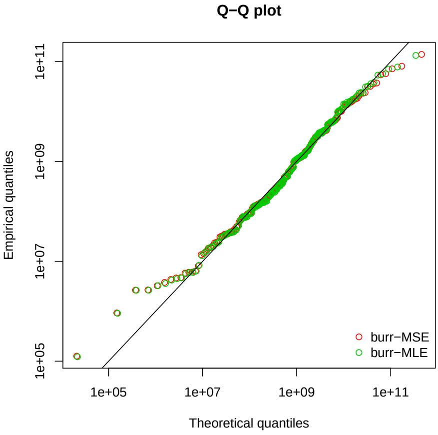
# EXAMPLE 3 – REAL DATASET

Consider Normalized Hurricane Damages in the United States: 1900-2005 used in Pielke et al. (2008).



# EXAMPLE 3 – REAL DATASET – GRAPHICS



# Example 3 – real dataset – GOF. statistics for Burr distribution

Coefficients values

| | MSE | MLE |
|---|---|---|
| shape1 | 2.2075 | 2.4563 |
| shape2 | 0.5611 | 0.5604 |
| rate | 2.5318e-10 | 1.9825e-10 |

Goodness-of-fit statistics

| | MSE | MLE |
|---|---|---|
| Kolmogorov-Smirnov statistic | 0.04421733 | 0.04867166 |
| Cramer-von Mises statistic | 0.06798511 | 0.08520099 |
| Anderson-Darling statistic | 0.43766768 | 0.51206510 |

Goodness-of-fit criteria

| | MSE | MLE |
|---|---|---|
| Akaike's Information Criterion | 9301.882 | 9301.723 |
| Bayesian Information Criterion | 9311.880 | 9311.721 |

We implements (very quickly) in **fitdistrplus**

- a new statistical procedure Maximum Spacing Estimation.
- as well as the generalized Maximum Spacing Estimation considering $\phi$ divergence function.
- automatically all generic functions (`plot`, `summary`, `coef`,...) are available.

# References


Cheng, R.C.H., Amin, N.A.K. (1983),
Estimating parameters in continuous univariate distributions with a shifted origin,
*Journal of the Royal Statistical Society Ser. B*, 45, 394-403.

Ranneby B (1984),
The Maximum Spacing Method. An Estimation Method Related to the Maximum Likelihood Method,
*Scandinavian Journal of Statistics*, 11(2), 93-112

Ranneby B & Ekstrom M (1997),
Maximum spacing estimates based on different metrics,
*Umea universitet*, research paper.

Ekstrom M (2015),
On the consistency of the maximum spacing method,
*Journal of Statistical Planning and Inference*, 70, 209-224.

Delignette-Muller ML & Dutang C (2019),
fitdistrplus: An R Package for Fitting Distributions,
`https://CRAN.R-project.org/package=fitdistrplus`

Teimouri M. (2018),
MPS: Estimating Through the Maximum Product Spacing Approach,
`https://CRAN.R-project.org/package=MPS`

Torres-Jimenez C.J. (2017),
BMT: The BMT Distribution,
`https://CRAN.R-project.org/package=BMT`