The Journal of Open Source Software

univariateML: An R package for maximum likelihood estimation of univariate densities

Jonas Moss[1]

1 University of Oslo

DOI: 10.21105/joss.01863

Software

- Review
- Repository
- Archive

Editor: Arfon Smith

Reviewers:

- @MaaniBeigy
- @vbaliga

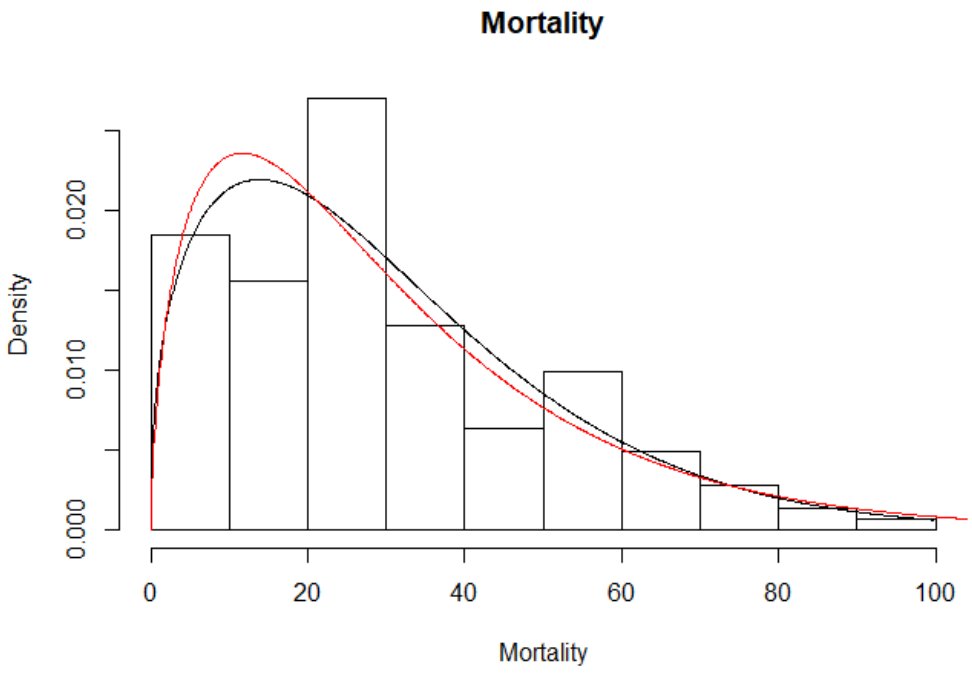
Submitted: 01 November 2019
Published: 04 December 2019

Summary

`univariateML` is an R (R Core Team, 2019) package for user-friendly univariate maximum likelihood estimation (Cam, 1990). It supports more than 20 densities, the most popular generic functions such as `plot`, `AIC`, and `confint`, and a simple parametric bootstrap (Efron & Tibshirani, 1994) interface.

When looking at univariate data it is natural to ask if there is a known parametric density that fits the data well. The following example uses the egypt (Pearson, 1902) data set included in the package and a plot of the Weibull and Gamma densities (Johnson, Kotz, & Balakrishnan, 1995, Chapter 17 & 21).

```
# install.packages("univariateML")
library("univariateML")
hist(egypt$age, freq = FALSE, main = "Mortality", xlab = "Mortality")
lines(mlweibull(egypt$age)) # Plots a Weibull fit.
lines(mlgamma(egypt$age), col = "red")  # Plots a Gamma fit.
```



A natural question to ask is which among several models fits the data best. This can be done using tools of model selection such as the `AIC` (Akaike, 1998).

```
AIC(mlweibull(egypt$age),
    mlgamma(egypt$age))
```

```
##                        df      AIC
## mlweibull(egypt$age)  2 1230.229
## mlgamma(egypt$age)    2 1234.772
```

Problems involving estimation of univariate densities are common in statistics. Estimation of univariate densities is used in for instance exploratory data analysis, in the estimation of copulas (Ko, Hjort, & Hobæk Haff, 2019), as parametric starts in density estimation (Hjort & Glad, 1995; Moss & Tveten, 2019), and is of interest in and of itself.

Analytic formulas for the maximum likelihood estimates are used whenever they exist. Most estimators without analytic solutions have a custom made Newton-Raphson solver. This is in contrast to the `mle` function in the built-in `R` package `stats4`, which supports more general maximum likelihood estimation through numerical optimization on a supplied negative log-likelihood function.

`Rfast` (Papadakis et al., 2019) is an `R` package with fast Newton-Raphson implementations of many univariate density estimators. `univariateML` differs from `Rfast` mainly in focus: While `univariateML`is focused on user-friendly univariate density estimation, `Rfast` aims to have the fastest possible implementations of many kinds of functions.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer.

Cam, L. L. (1990). Maximum likelihood: An introduction. *International Statistical Review / Revue Internationale de Statistique*, *58*(2), 153–171. doi:10.2307/1403464

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Hjort, N. L., & Glad, I. K. (1995). Nonparametric density estimation with a parametric start. *The Annals of Statistics*, 882–904. doi:10.1214/aos/1176324627

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd ed., Vol. 1). Wiley.

Ko, V., Hjort, N. L., & Hobæk Haff, I. (2019). Focused information criteria for copulas. *Scandinavian Journal of Statistics*. doi:10.1111/sjos.12387

Moss, J., & Tveten, M. (2019). Kdensity: An r package for kernel density estimation with parametric starts and asymmetric kernels. *The Journal of Open Source Software*, *4*. doi:10.21105/joss.01566

Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., Borboudakis, G., et al. (2019). *Rfast: A collection of efficient and extremely fast r functions*. Retrieved from https://CRAN.R-project.org/package=Rfast

Pearson, K. (1902). On the change in expectation of life in man during a period of circa 2000 years. *Biometrika*, *2*, 261–264. doi:10.2307/2331493

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/