

Nouvelles fonctionnalités du package `fitdistrplus`

ML. Delignette-Muller^a and C. Dutang^b

^aUniversité de Lyon

Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive
VetAgro Sup, Campus vétérinaire de Lyon
1 avenue Bourgelat, 69280 Marcy l'Etoile
marielaure.delignettemuller@vetagro-sup.fr

^bUniversité de Strasbourg

CNRS, UMR7501, Institut de Recherche Mathématiques Avancée
7 rue René Descartes, 67084 Strasbourg Cedex
dutang@math.unistra.fr

Mots clefs : ajustement de distributions, bootstrap, données censurées.

`fitdistrplus` est un package **R** dédié à l'ajustement de distributions paramétriques à des données univariées. Il propose diverses fonctions visant à faciliter le processus global de description d'une distribution empirique par une distribution paramétrique, incluant :

1. le choix de distributions candidates pour décrire les données,
2. l'ajustement de chacune des distributions candidates aux données,
3. la comparaison des ajustements en vue de choisir la distribution la plus adaptée,
4. le calcul, par bootstrap, de l'incertitude sur les paramètres estimés de la distribution choisie.

Plusieurs méthodes d'estimation des paramètres sont proposées dans le package : les plus courantes que sont la méthode du maximum de vraisemblance et la méthode des moments [1], mais aussi la méthode des quantiles [2] et la méthode de minimisation d'une statistique d'ajustement [3].

Une spécificité importante du package est la prise en compte de données de types variés. En effet, les fonctions relatives à l'ajustement de distributions par maximum de vraisemblance ont été adaptées pour permettre la prise en compte d'une part des données discrètes, et d'autre part des données censurées [4], quel que soit le type de censures (à droite, à gauche ou par intervalle).

Ce package a tout d'abord été développé pour une utilisation dans le cadre de l'appréciation quantitative des risques [5], mais les outils qu'il propose sont de nature à aider tout scientifique dans l'ajustement de distributions à des données observées. Depuis sa publication sur le CRAN en 2009, il a d'ailleurs été utilisé dans des domaines d'application très variés : appréciation quantitative des risques, épidémiologie, biologie moléculaire, génomique, bioinformatique, mathématiques financières et actuarielles, . . .

Les divers retours des utilisateurs nous ont conduit récemment à développer de nouvelles fonctionnalités que nous souhaitons présenter. Sont en particulier maintenant disponibles de nou-

velles fonctions facilitant la comparaison des ajustements de plusieurs distributions sur un même jeu de données, tant au niveau graphique [1] (fonctions `denscomp`, `cdfcomp`, `ppcomp` et `qqcomp` respectivement pour les diagrammes en densité, en fréquences cumulées et les P-P plot et Q-Q plot) qu'au niveau numérique [6] (amélioration de la fonction `gofstat` pour le calcul des statistiques d'ajustement et des critères d'information). La fonction générique `quantile` a aussi été créée pour diverses classes **S3** d'objets définies dans le package : elle permet le calcul de quantiles à partir d'une distribution ajustée sur des données censurées ou non, ainsi que le calcul par bootstrap de l'incertitude sur ces quantiles estimés.

Références

- [1] Cullen, A., Frey, H. (1999). *Probabilistic techniques in exposure assessment*. First edition. Plenum Publishing Co.
- [2] Tse, Y. (2009). *Nonlife Actuarial Models: Theory, Methods and Evaluation*. International Series on Actuarial Science. Cambridge University Press.
- [3] Luceno, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics and data analysis*, **51**(2), 904-917.
- [4] Helsel, D. (2005). *Nondetects and data analysis: statistics for censored environmental data*. First edition. Wiley.
- [5] Pouillot, R., Delignette-Muller, M.L. (2010). Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. *International Journal of Food Microbiology*, **142**[3], 330-340.
- [6] D'Agostino, R., Stephens, M. (1986). *Googness-of-fit techniques*. First edition. Dekker.