

# Nonparametric estimation for probability mass function with Disake

## an R package for discrete associated kernel estimators

W. E. Wansouwé\* — C. C. Kokonendji\*\* — D. T. Kolyang\*

\* Department of Computer Science  
The University of Maroua  
P.O. box 55 MAROUA  
CAMEROON  
ericwansouwe@gmail.com, dtaiwe@yahoo.fr

\*\* Laboratoire de Mathématiques de Besançon-UMR 6623 CNRS-UFC  
Université de Franche-Comté  
16 route de Gray, 25030 BESANÇON Cedex  
FRANCE  
celestin.kokonendji@univ-fcomte.fr

.....  
**RÉSUMÉ.** La méthode des noyaux est l'une des techniques d'estimation les plus répandues en statistique non paramétrique. Nous introduisons un module en R, **Disake**, pour l'estimation d'une distribution de probabilité par noyaux associés discrets. Dans l'estimation par noyau, deux choix importants sont à faire : le choix du noyau et celui de la fenêtre de lissage. Le module **Disake** implémente principalement les noyaux associés discrets ainsi que la validation croisée et l'approche bayésienne locale pour la sélection du paramètre de lissage. Des applications sur des données simulées et réelles montrent que le noyau binomial est approprié pour les échantillons de petite ou moyenne taille et, l'estimateur naïf ou le noyau triangulaire discret, indiqué pour les échantillons de grande taille.

**ABSTRACT.** Kernel smoothing is one of the most widely used nonparametric data smoothing techniques. We introduce a new R package, **Disake**, for computing discrete associated kernel estimators for probability mass function. When working with a kernel estimator, two choices must be made: the kernel function and the smoothing parameter. The **Disake** package focuses on discrete associated kernels and also on cross-validation and local Bayesian techniques to select the appropriate bandwidth. Applications on simulated data and real data show that the binomial kernel is appropriate for small or moderate count data while the empirical estimator or the discrete triangular kernel is indicated for large samples.

**MOTS-CLÉS :** Module R, noyau associé discret standard, validation croisée.

**KEYWORDS :** R package, standard discrete associated kernel, cross-validation.

.....















































- [17] PARZEN, E. « On estimation of a probability density function and mode », *Annals of Mathematical Statistics*, vol. 33, n° 3, pp. 1065 - 1076, 1962.
- [18] R DEVELOPMENT CORE TEAM « R : A Language and Environment for Statistical Computing, R Foundation for Statistical Computing », *Vienna, Austria, ISBN 3-900051-07-0, URL <http://www.R-project.org>*, 2008.
- [19] ROSENBLATT, M. « Remarks on some nonparametric estimates of a density function », *Annals of Mathematical Statistics*, vol. 27, n° 13, pp. 832 - 837, 1956.
- [20] SCAILLET, O. « Density estimation using inverse and reciprocal inverse Gaussian kernels », *Journal of Nonparametric Statistics*, vol. 16, n° 1-2, pp. 217 - 226, 2004.
- [21] SCOTT, D.W. « Multivariate Density Estimation : Theory, Practice, and Visualization », *Wiley, New York.*, 1992.
- [22] SENG A KIESSÉ, T. « Nonparametric Approach by Discrete Associated-Kernel for Count Data », *Ph.D. manuscript*, University of Pau, URL <http://tel.archives-ouvertes.fr/tel-00372180/fr/> (in French), 2008.
- [23] SENG A KIESSÉ, T., LIBENGUÉ, F.G., ZOCCHI, S.S. & KOKONENDJI, C.C. « The R package for general discrete triangular distributions. ». R package Version 1.0, URL <http://www.CRAN.R-project.org/package=TRIANGG>, 2010.
- [24] SILVERMAN, B.W. « Density Estimation for Statistics and Data Analysis », *Chapman & Hall, London.*, 1986.
- [25] SIMONOFF, J.S. « Smoothing Methods in Statistics », *Springer, New York*, 1996.
- [26] SIMONOFF, J.S. & TUTZ, G. « Smoothing methods for discrete data », *In : M.G. Schimek (Ed.), Smoothing and Regression : Approaches, Computation, and Application. Wiley, New York*, pp. 193 - 228, 2000.
- [27] TSYBAKOV, A.B. « Introduction à l'estimation non-paramétrique », *Springer, Paris*, 2004.
- [28] WANSOUWÉ, W.E., KOKONENDJI, C.C. KOLYANG, D.T. « Disake : an R package for discrete associated kernel estimators ». R package Version 1.5, URL <http://www.CRAN.R-project.org/package=Disake>, 2015.
- [29] ZOUGAB, N., ADJABI, S. & KOKONENDJI, C.C. « Binomial kernel and Bayes local bandwidth in discrete functions estimation », *Journal of Nonparametric Statistics*, vol. 24, n° 3, pp. 783 - 795, 2012.