

Mazo, Gildas

A semiparametric and location-shift copula-based mixture model. (English) Zbl 1381.62186
J. Classif. 34, No. 3, 444-464 (2017).

Summary: Modeling mixtures of distributions has rested on Gaussian distributions and/or a conditional independence hypothesis for a long time. Only recently have researchers begun to construct and study broader generic models without appealing to such hypotheses. Some of these extensions use copulas as a tool to build flexible models, as they permit modeling the dependence and the marginal distributions separately. But this approach also has drawbacks. First, the practitioner has to make more arbitrary choices, and second, marginal misspecification may loom on the horizon. This paper aims at overcoming these limitations by presenting a copulabased mixture model which is semiparametric. Thanks to a location-shift hypothesis, semiparametric estimation, also, is feasible, allowing for data adaptation without any modeling effort.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)
62G05 Nonparametric estimation

Cited in **3** Documents

Keywords:

location; shift; copula; mixture; clustering; semiparametric; nonparametric

Software:

ks; mclust; AS 136

Full Text: [DOI](#)

References:

- [1] AKAIKE, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle", in *\textit{2nd International Symposium on Information Theory}*, eds. B.N. Petrov and F. Csaki, Hungary: Akadémiai Kiadó. · [Zbl 0283.62006](#)
- [2] AKAIKE, H, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19, 716-723, (1974) · [Zbl 0314.62039](#) · [doi:10.1109/TAC.1974.1100705](#)
- [3] BENAGLIA, T; CHAUVEAU, D; HUNTER, DR, An EM-like algorithm for semi- and nonparametric estimation in multivariate mixtures, *Journal of Computational and Graphical Statistics*, 18, 505-526, (2009) · [doi:10.1198/jcgs.2009.07175](#)
- [4] BORDES, L; CHAUVEAU, D; VANDEKERKHOVE, P, A stochastic EM algorithm for a semiparametric mixture model, *Computational Statistics and Data Analysis*, 51, 5429-5443, (2007) · [Zbl 1445.62056](#) · [doi:10.1016/j.csda.2006.08.015](#)
- [5] DEMPSTER, AP; LAIRD, NM; RUBIN, DB, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B*, 39, 1-38, (1977) · [Zbl 0364.62022](#)
- [6] DUONG, T. (2015), "ks: Kernel Smoothing", R package, <https://cran.r-project.org/package=ks/> · [Zbl 0536.62021](#)
- [7] FORBES, F; WRAITH, D, A new family of multivariate heavy-tailed distributions with variable marginal amounts of tail-weight: application to robust clustering, *Statistics and Computing*, 24, 971-984, (2014) · [Zbl 1332.62204](#) · [doi:10.1007/s11222-013-9414-4](#)
- [8] FRALEY, C., RAFTERY, A.E., BRENDAN, T.M., and SCRUCCA, L. (2012), *\textit{mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation}*, Technical Report 597, Department of Statistics, University of Washington, <https://cran.r-project.org/package=mclust>.
- [9] GENEST, C; FAVRE, A-C, Everything you always wanted to know about copula modeling but were afraid to ask, *Journal of Hydrologic Engineering*, 12, 347-368, (2007) · [doi:10.1061/\(ASCE\)1084-0699\(2007\)12:4\(347\)](#)
- [10] GENEST, C; GHOUDI, K; RIVEST, L-P, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, 82, 543-552, (1995) · [Zbl 0831.62030](#) · [doi:10.1093/biomet/82.3.543](#)
- [11] GOOD, IJ; GASKINS, RA, Nonparametric roughness penalties for probability densities, *Biometrika*, 58, 255-277, (1971) · [Zbl 0221.62012](#) · [doi:10.2307/2334515](#)
- [12] HARTIGAN, JA; WONG, MA, "A K-means clustering algorithm", *\textit{journal of the royal statistical society}*, Series C, 28, 100-108, (1979) · [Zbl 0447.62062](#)
- [13] HUNTER, DR; WANG, S; HETTMANSPERGER, TP, Inference formixtures of symmetric distributions, *The Annals of Statistics*, 35, 224-251, (2007) · [Zbl 1114.62035](#) · [doi:10.1214/009053606000001118](#)
- [14] JOE, H. (2001), *\textit{Multivariate Models and Dependence Concepts}*, Boca Raton FL: Chapman and Hall/CRC. · [Zbl](#)

- [15] JOE, H. (2014), \textit{Dependence Modeling with Copulas}, Chapman and Hall/CRC. · Zbl 1346.62001
- [16] JONES, MC; MARRON, JS; SHEATHERM, SJ, A brief survey of bandwidth selection for density estimation, *Journal of the American Statistical Association*, 91, 401-407, (1996) · Zbl 0873.62040 · doi:10.1080/01621459.1996.10476701
- [17] KOSMIDIS, I; KARLIS, D, Model-based clustering using copulas with applications, *Statistics and Computing*, 26, 1079-1099, (2016) · Zbl 06652996 · doi:10.1007/s11222-015-9590-5
- [18] LEE, S; and MCLACHLAN, GJ, Finitemixtures of multivariate skew t-distributions: some recent and new results, *Statistics and Computing*, 24, 181-202, (2014) · Zbl 1325.62107 · doi:10.1007/s11222-012-9362-4
- [19] LEVINE, M; HUNTER, DR; CHAUVEAU, D, Maximum smoothed likelihood for multivariate mixtures, *Biometrika*, 98, 403-416, (2011) · Zbl 1215.62055 · doi:10.1093/biomet/asq079
- [20] MARBAC, M., BIERNACKI, C., and VANDEWALLE, V. (2014), "Model-Based Clustering of Gaussian Copulas for Mixed Data", \textit{arXiv preprint arXiv:1405.1299}. · Zbl 1384.62198
- [21] MARBAC, M; BIERNACKI, C; VANDEWALLE, V, Model-based clustering for conditionally correlated categorical data, *Journal of Classification*, 32, 145-175, (2015) · Zbl 1335.62103 · doi:10.1007/s00357-015-9180-4
- [22] MCLACHLAN, G., and KRISHNAN, T. (2007), \textit{The EM Algorithm and Extensions} (Vol. 382), New York: John Wiley and Sons.
- [23] MCLACHLAN, G., and PEEL, D. (2004), \textit{Finite Mixture Models}, New York: John Wiley and Sons. · Zbl 0963.62061
- [24] NELSEN, R.B. (2006), \textit{An Introduction to Copulas}, New York: Springer. · Zbl 1152.62030
- [25] REDNER, RA; WALKER, HF, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, 26, 195-239, (1984) · Zbl 0536.62021 · doi:10.1137/1026034
- [26] SAKAMOTO, Y., ISHIGURO, M., and KITAGAWA, G. (1986), \textit{Akaike Information Criterion Statistics}, Tokyo: KTK Scientific Publishers. · Zbl 0608.62006
- [27] SILVERMAN, B.W. (1998), \textit{Density Estimation for Statistics and Data Analysis}, London: Chapman and Hall. · Zbl 0617.62042
- [28] SKLAR, A, Fonction de Répartition dont LES marges sont données, *Institute de Statistique de l'Université de Paris*, 8, 229-231, (1959)
- [29] VRAC, M; BILLARD, L; DIDAY, E; CHÉDIN, A, Copula analysis of mixture models, *Computational Statistics*, 27, 427-457, (2012) · Zbl 1304.65087 · doi:10.1007/s00180-011-0266-0
- [30] WAND, MP; JONES, MC, Multivariate plug-in bandwidth selection, *Computational Statistics*, 9, 97-116, (1994) · Zbl 0937.62055

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.