

Dang, Utkarsh J.; Browne, Ryan P.; McNicholas, Paul D.

Mixtures of multivariate power exponential distributions. (English) Zbl 1419.62330

Biometrics 71, No. 4, 1081-1089 (2015).

Summary: An expanded family of mixtures of multivariate power exponential distributions is introduced. While fitting heavy-tails and skewness have received much attention in the model-based clustering literature recently, we investigate the use of a distribution that can deal with both varying tail-weight and peakedness of data. A family of parsimonious models is proposed using an eigen-decomposition of the scale matrix. A generalized expectation-maximization algorithm is presented that combines convex optimization via a minorization-maximization approach and optimization based on accelerated line search algorithms on the Stiefel manifold. Lastly, the utility of this family of models is illustrated using both toy and benchmark data.

#### MSC:

62P10 Applications of statistics to biology and medical sciences; meta analysis

Cited in 18 Documents

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

#### Keywords:

GEM-algorithm; mixture models; model-based clustering; multivariate power exponential; Stiefel manifold

#### Software:

Rmixmod; PARVUS; MNM; R; mixture

Full Text: [DOI](#) [arXiv](#)

#### References:

- [1] Absil , P.-A. Mahony , R. Sepulchre , R. 2009
- [2] Airoldi, Age variation in voles (*Microtus californicus*, *M. ochrogaster*) and its significance for systematic studies, Occasional Papers of the Museum of Natural History. University of Kansas (1984)
- [3] Aitken , A. C. 1926 On Bernoulli's numerical solution of algebraic equations 289 305 · [Zbl 52.0098.05](#)
- [4] Anderson, The irises of the Gaspe peninsula, Bulletin of the American Iris Society 59 pp 2– (1935)
- [5] Andrews, Extending mixtures of multivariate t-factor analyzers, Statistics and Computing 21 pp 361– (2011) · [Zbl 1255.62171](#) · [doi:10.1007/s11222-010-9175-2](#)
- [6] Andrews, Model-based clustering, classification, and discriminant analysis via mixtures of multivariate t-distributions, Statistics and Computing 22 pp 1021– (2012) · [Zbl 1252.62062](#) · [doi:10.1007/s11222-011-9272-x](#)
- [7] Andrews , J. L. McNicholas , P. D. 2015
- [8] Banfield, Model-based Gaussian and non-Gaussian clustering, Biometrics 49 pp 803– (1993) · [Zbl 0794.62034](#) · [doi:10.2307/2532201](#)
- [9] Biernacki, Assessing a mixture model for clustering with the integrated completed likelihood, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 pp 719– (2000) · [doi:10.1109/34.865189](#)
- [10] Bombrun , L. Pascal , F. Tourneret , J.-Y. Berthoumieu , Y. 2012 3525 3528
- [11] Boulesteix , A.-L. Lambert-Lacroix , S. Peyre , J. Strimmer , K. 2014
- [12] Browne , R. P. ElSherbiny , A. McNicholas , P. D. 2014 mixture
- [13] Browne, Estimating common principal components in high dimensions (in press), Advances in Data Analysis and Classification 8 pp 217– (2014a) · [doi:10.1007/s11634-013-0139-1](#)
- [14] Browne, Orthogonal Stiefel manifold optimization for eigen-decomposed covariance parameter estimation in mixture models, Statistics and Computing 24 pp 203– (2014b) · [Zbl 1325.62008](#) · [doi:10.1007/s11222-012-9364-2](#)
- [15] Browne, A mixture of generalized hyperbolic distributions, Canadian Journal of Statistics 43 pp 176– (2015) · [Zbl 1320.62144](#) · [doi:10.1002/cjs.11246](#)
- [16] Campbell, A multivariate study of variation in two species of rock crab of the genus *Leptograpsus*, Australian Journal of Zoology 22 pp 417– (1974) · [doi:10.1071/ZO9740417](#)

- [17] Celeux, Gaussian parsimonious clustering models, *Pattern Recognition* 28 pp 781– (1995) · [Zbl 05480211](#) · [doi:10.1016/0031-3203\(94\)00125-6](#)
- [18] Cho, Multivariate statistical modeling for image denoising using wavelet transforms, *Signal Processing: Image Communication* 20 pp 77– (2005)
- [19] Coretto, A simulation study to compare robust clustering methods based on mixtures, *Advances in Data Analysis and Classification* 4 pp 111– (2010) · [Zbl 1284.62366](#) · [doi:10.1007/s11634-010-0065-4](#)
- [20] Dempster, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B* 39 pp 1– (1977) · [Zbl 0364.62022](#)
- [21] Fisher, The use of multiple measurements in taxonomic problems, *Annals of Eugenics* 7 pp 179– (1936) · [doi:10.1111/j.1469-1809.1936.tb02137.x](#)
- [22] Flury , B. 2012 *Flury R*
- [23] Forbes, A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweights: Application to robust clustering (in press), *Statistics and Computing* 24 pp 971– (2014) · [Zbl 1332.62204](#) · [doi:10.1007/s11222-013-9414-4](#)
- [24] Forina, Parvus: An extendable package of programs for data exploration, classification and correlation, *Journal of Chemometrics* 4 pp 191– (1988)
- [25] Fraley , C. Raftery , A. E. Murphy , T. B. Scrucca , L. 2012
- [26] Franczak, Mixtures of shifted asymmetric Laplace distributions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 pp 1149– (2014) · [doi:10.1109/TPAMI.2013.216](#)
- [27] Ghahramani , Z. Hinton , G. E. 1997
- [28] Golub, Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science* 286 pp 531– (1999) · [doi:10.1126/science.286.5439.531](#)
- [29] Gómez, A multivariate generalization of the power exponential family of distributions, *Communications in Statistics-Theory and Methods* 27 pp 589– (1998) · [Zbl 0895.62053](#) · [doi:10.1080/03610929808832115](#)
- [30] Gómez-Sánchez-Manzano, Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications, *Communications in Statistics-Theory and Methods* 37 pp 972– (2008) · [Zbl 1135.62041](#) · [doi:10.1080/03610920701762754](#)
- [31] Hartigan, A k-means clustering algorithm, *Journal of the Royal Statistical Society: Series C* 28 pp 100– (1979) · [Zbl 0447.62062](#)
- [32] Hennig , C. Coretto , P. 2008 127 138
- [33] Hubert, Comparing partitions, *Journal of Classification* 2 pp 193– (1985) · [Zbl 0587.62128](#) · [doi:10.1007/BF01908075](#)
- [34] Hunter, Rejoinder to discussion of Optimization transfer using surrogate objective functions, *Journal of Computational and Graphical Statistics* 9 pp 52– (2000)
- [35] Hurley , C. 2012
- [36] Karlis, Model-based clustering with non-elliptically contoured distributions, *Statistics and Computing* 19 pp 73– (2009) · [doi:10.1007/s11222-008-9072-0](#)
- [37] Khan, Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, *Nature Medicine* 7 pp 673– (2001) · [doi:10.1038/89044](#)
- [38] Landsman, Tail conditional expectations for elliptical distributions, *North American Actuarial Journal* 7 pp 55– (2003) · [Zbl 1084.62512](#) · [doi:10.1080/10920277.2003.10596118](#)
- [39] Lebet , R. Iovleff , S. Longeville , A. 2012 *Rmixmod R*
- [40] Lin, Finite mixture modelling using the skew normal distribution, *Statistica Sinica* 17 pp 909– (2007) · [Zbl 1133.62012](#)
- [41] Lindsey, Multivariate elliptically contoured distributions for repeated measurements, *Biometrics* 55 pp 1277– (1999) · [Zbl 1059.62543](#) · [doi:10.1111/j.0006-341X.1999.01277.x](#)
- [42] Liu, Multivariate regression models with power exponential random errors and subset selection using genetic algorithms with information complexity, *European Journal of Pure and Applied Mathematics* 1 pp 4– (2008) · [Zbl 1132.62042](#)
- [43] Mardia , K. V. Kent , J. T. Bibby , J. M. 1980
- [44] McLachlan , G. Peel , D. 2000a 599 606
- [45] McLachlan , G. J. Peel , D. 2000b
- [46] McNicholas, Parsimonious Gaussian mixture models, *Statistics and Computing* 18 pp 285– (2008) · [doi:10.1007/s11222-008-9056-0](#)
- [47] McNicholas, Model-based clustering of microarray expression data via latent Gaussian mixture models, *Bioinformatics* 26 pp 2705– (2010) · [doi:10.1093/bioinformatics/btq498](#)
- [48] Meng, Maximum likelihood estimation via the ECM algorithm: A general framework, *Biometrika* 80 pp 267– (1993) · [Zbl 0778.62022](#) · [doi:10.1093/biomet/80.2.267](#)
- [49] Murray, Mixtures of skew-factor analyzers, *Computational Statistics and Data Analysis* 77 pp 326– (2014) · [Zbl 06984029](#) · [doi:10.1016/j.csda.2014.03.012](#)
- [50] Nordhausen, Multivariate l1 methods: The package MNM, *Journal of Statistical Software* 43 pp 1– (2011) · [doi:10.18637/jss.v043.i05](#)
- [51] Pascal, Parameter estimation for multivariate generalized Gaussian distributions, *IEEE Transactions on Signal Processing* 61 pp 5960– (2013) · [Zbl 1394.62071](#) · [doi:10.1109/TSP.2013.2282909](#)
- [52] Reaven, An attempt to define the nature of chemical diabetes using a multidimensional analysis, *Diabetologia* 16 pp 17–

(1979) · [doi:10.1007/BF00423145](https://doi.org/10.1007/BF00423145)

- [53] Schwarz, Estimating the dimension of a model, *Annals of Statistics* 6 pp 461– (1978) · [Zbl 0379.62005](https://zbmath.org/journals/aos/1176344136) · [doi:10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)
- [54] R 2013 R: A Language and Environment for Statistical Computing R
- [55] Subedi, Variational Bayes approximations for clustering via mixtures of normal inverse Gaussian distributions, *Advances in Data Analysis and Classification* 8 pp 167– (2014) · [doi:10.1007/s11634-014-0165-7](https://doi.org/10.1007/s11634-014-0165-7)
- [56] Titterton , D. M. Smith , A. F. M. Makov , U. E. 1985
- [57] Venables , W. N. Ripley , B. D. 2002
- [58] Verdoolaege , G. De Backer , S. Scheunders , P. 2008 169 172
- [59] Vrbik, Parsimonious skew mixture models for model-based clustering and classification, *Computational Statistics and Data Analysis* 71 pp 196– (2014) · [Zbl 1471.62202](https://zbmath.org/journals/csda/2013.07.008) · [doi:10.1016/j.csda.2013.07.008](https://doi.org/10.1016/j.csda.2013.07.008)
- [60] Zhang, Robust clustering using exponential power mixtures, *Biometrics* 66 pp 1078– (2010) · [Zbl 1233.62192](https://zbmath.org/journals/j.1541-0420.2010.01389.x) · [doi:10.1111/j.1541-0420.2010.01389.x](https://doi.org/10.1111/j.1541-0420.2010.01389.x)
- [61] Zhang, Multivariate generalized gaussian distribution: Convexity and graphical models, *IEEE Transactions on Signal Processing* 61 pp 4141– (2013) · [Zbl 1394.62072](https://zbmath.org/journals/TSP/2013.2267740) · [doi:10.1109/TSP.2013.2267740](https://doi.org/10.1109/TSP.2013.2267740)

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.