

**Arevalillo, Jorge M.; Navarro, Hilario**

**Data projections by skewness maximization under scale mixtures of skew-normal vectors.**  
(English) [Zbl 1459.62075](#)  
*Adv. Data Anal. Classif., ADAC 14, No. 2, 435-461 (2020).*

Summary: Multivariate scale mixtures of skew-normal distributions are flexible models that account for the non-normality of data by means of a tail weight parameter and a shape vector representing the asymmetry of the model in a directional fashion. Its stochastic representation involves a skew-normal vector and a non negative mixing scalar variable, independent of the skew-normal vector, that injects tail weight behavior into the model. In this paper we look into the problem of finding the projection that maximizes skewness for vectors that follow a scale mixture of skew-normal distribution; when a simple condition on the moments of the mixing variable is fulfilled, it can be shown that the direction yielding the maximal skewness is proportional to the shape vector. This finding stresses the directional nature of the shape vector to regulate the asymmetry; it also provides the theoretical foundations motivating the skewness based projection pursuit problem in this class of distributions. Some examples that illustrate the application of our results are also given; they include a simulation experiment with artificial data, which sheds light on the usefulness and implications of our results, and the application to real data.

**MSC:**

- [62H05](#) Characterization and structure theory for multivariate probability distributions; copulas
- [62H10](#) Multivariate distribution of statistics
- [62H30](#) Classification and discrimination; cluster analysis (statistical aspects)
- [60E05](#) Probability distributions: general theory

**Keywords:**

[skew-normal](#); [scale mixtures of skew-normal distributions](#); [maximal skewness projection](#)

**Software:**

[sn](#); [MaxSkew](#)

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

**References:**

- [1] Arevalillo, JM; Navarro, H., A study of the effect of kurtosis on discriminant analysis under elliptical populations, *J Multivar Anal*, 107, 53-63 (2012) · [Zbl 1236.62066](#)
- [2] Arevalillo, JM; Navarro, H., A note on the direction maximizing skewness in multivariate skew-t vectors, *Stat Probab Lett*, 96, 328-332 (2015) · [Zbl 1396.60015](#)
- [3] Arevalillo, JM; Navarro, H., A stochastic ordering based on the canonical transformation of skew-normal vectors, *TEST*, 28, 2, 475-498 (2019) · [Zbl 1456.60046](#)
- [4] Azzalini, A., The skew-normal distribution and related multivariate families, *Scand J Stat*, 32, 2, 159-188 (2005) · [Zbl 1091.62046](#)
- [5] Azzalini, A.; Capitanio, A., Statistical applications of the multivariate skew normal distribution, *J R Stat Soc Ser B*, 61, 3, 579-602 (1999) · [Zbl 0924.62050](#)
- [6] Azzalini, A.; Capitanio, A., Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *J R Stat Soc Ser B*, 65, 2, 367-389 (2003) · [Zbl 1065.62094](#)
- [7] Azzalini, A.; Capitanio, A., *The skew-normal and related families (2014)*, Cambridge: Cambridge University Press, Cambridge · [Zbl 0924.62050](#)
- [8] Azzalini, A.; Dalla Valle, A., The multivariate skew-normal distribution, *Biometrika*, 83, 4, 715-726 (1996) · [Zbl 0885.62062](#)
- [9] Balakrishnan, N.; Scarpa, B., Multivariate measures of skewness for the skew-normal distribution, *J Multivar Anal*, 104, 1, 73-87 (2012) · [Zbl 1226.60020](#)
- [10] Balakrishnan, N.; Capitanio, A.; Scarpa, B., A test for multivariate skew-normality based on its canonical form, *J Multivar Anal*, 128, 19-32 (2014) · [Zbl 1352.62084](#)
- [11] Bickel, PJ; Kur, G.; Nadler, B., Projection pursuit in high dimensions, *Proc Natl Acad Sci*, 115, 37, 9151-9156 (2018) · [Zbl](#)

- [12] Branco, MD; Dey, DK, A general class of multivariate skew-elliptical distributions, *J Multivar Anal*, 79, 1, 99-113 (2001) · [Zbl 0992.62047](#)
- [13] Capitanio A (2012) On the canonical form of scale mixtures of skew-normal distributions. [arXiv:1207.0797](#)
- [14] Capitanio, A.; Azzalini, A.; Stanghellini, E., Graphical models for skew-normal variates, *Scand J Stat*, 30, 1, 129-144 (2003) · [Zbl 1035.60008](#)
- [15] Caussinus, H.; Ruiz-Gazen, A., *Exploratory projection pursuit*, chap 3, 67-92 (2010), Hoboken: Wiley, Hoboken
- [16] Chow, CK; Liu, CN, Approximating discrete probability distributions with dependence trees, *IEEE Trans Inf Theory*, 14, 3, 462-467 (1968) · [Zbl 0165.22305](#)
- [17] Contreras-Reyes, JE; Arellano-Valle, RB, Kullback-Leibler divergence measure for multivariate skew-normal distributions, *Entropy*, 14, 9, 1606-1626 (2012) · [Zbl 1306.62040](#)
- [18] Cook, R.; Weisberg, S., *An introduction to regression graphics* (2009), Hoboken: Wiley, Hoboken
- [19] Cook, D.; Buja, A.; Cabrera, J., Projection pursuit indexes based on orthonormal function expansions, *J Comput Gr Stat*, 2, 3, 225-250 (1993)
- [20] Edwards, D.; de Abreu, GC; Labouriau, R., Selecting high-dimensional mixed graphical models using minimal AIC or BIC forests, *BMC Bioinform*, 11, 1, 18 (2010)
- [21] Franceschini, C.; Loperfido, N., Testing for normality when the sampled distribution is extended skew-normal, 159-169 (2014), Cham: Springer International Publishing, Cham · [Zbl 1407.62198](#)
- [22] Franceschini C, Loperfido N (2016) MaxSkew: orthogonal data projections with maximal skewness. R package version 1.0, <https://CRAN.R-project.org/package=MaxSkew>
- [23] Friedman, JH, Exploratory projection pursuit, *J Am Stat Assoc*, 82, 397, 249-266 (1987) · [Zbl 0664.62060](#)
- [24] Friedman, JH; Tukey, JW, A projection pursuit algorithm for exploratory data analysis, *IEEE Trans Comput*, 23, 9, 881-890 (1974) · [Zbl 0284.68079](#)
- [25] Gamez-Pozo, A.; Berges-Soria, J.; Arevalillo, JM; Nanni, P.; Lopez-Vacas, R.; Navarro, H.; Grossmann, J.; Castaneda, CA; Main, P.; Diaz-Almiron, M.; Espinosa, E.; Ciruelos, E.; Vara, JAF, Combined label-free quantitative proteomics and microRNA expression analysis of breast cancer unravel molecular differences with clinical implications, *Cancer Res*, 75, 11, 2243-2253 (2015)
- [26] Gómez-Sánchez-Manzano, E.; Gómez-Villegas, M.; Marín, J., Multivariate exponential power distributions as mixtures of normal distributions with bayesian applications, *Commun Stat Theory Methods*, 37, 6, 972-985 (2008) · [Zbl 1135.62041](#)
- [27] Huber, PJ, Projection pursuit, *Ann Stat*, 13, 2, 435-475 (1985) · [Zbl 0595.62059](#)
- [28] Jones, MC; Sibson, R., What is projection pursuit?, *J R Stat Soc Ser A (Gen)*, 150, 1, 1-37 (1987) · [Zbl 0632.62059](#)
- [29] Kim, HM, A note on scale mixtures of skew normal distribution, *Stat Probab Lett*, 78, 13, 1694-1701 (2008) · [Zbl 1152.62032](#)
- [30] Kim, HM; Kim, C., Moments of scale mixtures of skew-normal distributions and their quadratic forms, *Commun Stat Theory Methods*, 46, 3, 1117-1126 (2017) · [Zbl 1364.62119](#)
- [31] Lachos, VH; Ghosh, P.; Arellano-Valle, RB, Likelihood based inference for skew-normal independent linear mixed models, *Stat Sin*, 20, 1, 303-322 (2010) · [Zbl 1186.62071](#)
- [32] Lachos, VH; Labra, FV; Bolfarine, H.; Ghosh, P., Multivariate measurement error models based on scale mixtures of the skew-normal distribution, *Statistics*, 44, 6, 541-556 (2010) · [Zbl 1291.62120](#)
- [33] Lee, SX; McLachlan, GJ, Finite mixtures of canonical fundamental skew t-distributions: the unification of the restricted and unrestricted skew t-mixture models, *Stat Comput*, 26, 573-589 (2016) · [Zbl 1420.60020](#)
- [34] Lin, TI, Robust mixture modeling using multivariate skew t distributions, *Stat Comput*, 20, 343-356 (2010)
- [35] Lin, TI; Ho, HJ; Lee, CR, Flexible mixture modelling using the multivariate skew-t-normal distribution, *Stat Comput*, 24, 531-546 (2014) · [Zbl 1325.62113](#)
- [36] Loperfido, N., *Generalized skew-normal distributions*, Chap 4, 65-80 (2004), Boca Raton: CRC/Chapman & Hall, Boca Raton
- [37] Loperfido, N., Canonical transformations of skew-normal variates., *TEST*, 19, 1, 146-165 (2010) · [Zbl 1203.62102](#)
- [38] Loperfido, N., Skewness-based projection pursuit: a computational approach, *Comput Stat Data Anal*, 120, 42-57 (2018) · [Zbl 1469.62111](#)
- [39] Loperfido, N., Finite mixtures, projection pursuit and tensor rank: a triangulation, *Adv Data Anal Classif*, 13, 1, 145-173 (2019) · [Zbl 1466.62355](#)
- [40] Malkovich, JF; Afifi, AA, On tests for multivariate normality, *J Am Stat Assoc*, 68, 341, 176-179 (1973)
- [41] Merkle, M., Conditions for convexity of a derivative and some applications to the gamma function, *Aequ Math*, 55, 3, 273-280 (1998) · [Zbl 0922.26005](#)
- [42] Prado-Vázquez, G.; Gámez-Pozo, A.; Trilla-Fuertes, L.; Arevalillo, JM; Zapater-Moros, A.; Ferrer-Gómez, M.; Díaz-Almirón, M.; López-Vacas, R.; Navarro, H.; Mañá, P.; Feliú, J.; Zamora, P.; Espinosa, E.; Fresno Vara, JÁ, A novel approach to triple-negative breast cancer molecular classification reveals a luminal immune-positive subgroup with good prognoses, *Sci Rep*, 9, 1, 1538 (2019)
- [43] Rody, A.; Karn, T.; Liedtke, C.; Pusztai, L.; Ruckhaeberle, E.; Hanker, L.; Gaetje, R.; Solbach, C.; Ahr, A.; Metzler, D.; Schmidt, M.; Müller, V.; Holtrich, U.; Kaufmann, M., A clinically relevant gene signature in triple negative and basal-like

breast cancer, *Breast Cancer Res*, 13, 5, R97 (2011)

- [44] Wang, J., A family of kurtosis orderings for multivariate distributions, *J Multivar Anal*, 100, 3, 509-517 (2009) · [Zbl 1154.62043](#)
- [45] Wang, J.; Genton, MG, The multivariate skew-slash distribution, *J Stat Plan Inference*, 136, 1, 209-220 (2006) · [Zbl 1081.60013](#)
- [46] Zapater-Moros, A.; Gámez-Pozo, A.; Prado-Vázquez, G.; Trilla-Fuertes, L.; Arevalillo, JM; Díaz-Almirón, M.; Navarro, H.; Maín, P.; Feliú, J.; Zamora, P.; Espinosa, E.; Fresno Vara, JÁ, Probabilistic graphical models relate immune status with response to neoadjuvant chemotherapy in breast cancer, *Oncotarget*, 9, 45, 27586-27594 (2018)

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.