

Palarea-Albaladejo, Javier; Martín-Fernández, Josep A.; Gómez-García, Juan
A parametric approach for dealing with compositional rounded zeros. (English)

Zbl 1130.86001

Math. Geol. 39, No. 7, 625-645 (2007).

Summary: In this work, a parametric approach for replacing data below the detection limit, also known as rounded zeros, in compositional data sets is proposed. Compositional rounded zeros correspond to small proportions of some whole that cannot be reliably detected by the analytical instruments under given operating conditions. This kind of zeros appear frequently in the data collection process in geosciences. They must be treated in an adequate way before some multivariate analysis can be applied. Our procedure results from a modification of the Expectation-Maximization (EM) algorithm and is based on the additive log-ratio transformation. Its coherence with the nature of compositional data and with basic operations in the simplex sample space is checked. Using real data sets, we find that this approach improves other parametric and non-parametric techniques for compositional rounded zeros.

MSC:

- 86-08 Computational methods for problems pertaining to geophysics
- 86A32 Geostatistics
- 62H30 Classification and discrimination; cluster analysis (statistical aspects)
- 65C60 Computational problems in statistics (MSC2010)

Cited in 10 Documents

Keywords:

Additive log-ratio transformation; Compositional data; EM algorithm; Simplex

Full Text: DOI

References:

- [1] Aitchison J (1986) The statistical analysis of compositional data. Chapman and Hall, London. Reprinted in 2003 by Blackburn Press, 416 p · Zbl 0688.62004
- [2] Aitchison J, Greenacre M (2002) Biplots of compositional data. Appl Stat 51(4):375–392 · Zbl 1111.62300
- [3] Aitchison J, Kay JW (2004) Possible solutions of some essential zero problems in compositional data analysis. In: Thió-Henestrosa S, Martín-Fernández JA (eds) Compositional data analysis workshop, Girona, Spain. <http://ima.udg.es/Activitats/CoDaWork03/>
- [4] Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V (2000) Logratio analysis and compositional distance. Math Geol 32(3):271–275 · Zbl 1101.86309 · doi:10.1023/A:1007529726302
- [5] Amemiya T (1984) Tobit models: a survey. J Econom. 24:3–61 · Zbl 0539.62121 · doi:10.1016/0304-4076(84)90074-5
- [6] Bacon-Shone J (2003) Modelling structural zeros in compositional data. In: Thió-Henestrosa S, Martín-Fernández JA (eds) Compositional data analysis workshop, Girona, Spain. <http://ima.udg.es/Activitats/CoDaWork03/>
- [7] Buccianti A, Rosso F (1999) A new approach to the statistical analysis of compositional (closed) data with observations below the "detection limit". Geoinformatica 3:17–31
- [8] Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). J Roy Stat Soc Ser B 39:1–38 · Zbl 0364.62022
- [9] Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal (2003) Isometric logratio transformation for compositional data analysis. Math Geol 35(3):279–300 · Zbl 1302.86024 · doi:10.1023/A:1023818214614
- [10] Fry JM, Fry TRL, McLaren KR (2000) Compositional data analysis and zeros in micro data. Appl Econom 32:953–959 · doi:10.1080/000368400322002
- [11] Gómez-García J, Palarea-Albaladejo J, Martín-Fernández JA (2006) Métodos de inferencia estadística con datos faltantes. Estudio de simulación sobre los efectos en las estimaciones. Revista Estadística Española 48(162):241–270
- [12] Heckman J (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. Ann Econom Soc Meas 5:475–492
- [13] Honaker J, Katz JN, King G (2002) A fast, easy, and efficient estimator for multiparty electoral data. Political Anal 10(1):84–100 · doi:10.1093/pan/10.1.84
- [14] King G, Honaker J, Joseph A, Scheve K (2001) Analyzing incomplete political science data: an alternative algorithm for multiple imputation. Am Political Sci Rev 95(1):49–69

- [15] Little RJA, Rubin DB (2002) *Statistical analysis with missing data*. Wiley, New York, 381 p
- [16] Martín-Fernández JA, Thió-Henestrosa S (2006) Rounded zeros: some practical aspects for compositional data. In: Buccianti A, Mateu-Figueras G, Pawlowsky-Glahn V (eds) *Compositional data analysis: from theory to practice*, vol 264. The Geological Society, London, pp 191–201
- [17] Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V (2000) Zero replacement in compositional data sets. In: Kiers H, Rasson J, Groenen P, Shader M (eds) *Studies in classification, data analysis, and knowledge organization*. Springer, Berlin, pp 155–160 · Zbl 1101.62352
- [18] Martín-Fernández JA, Olea-Mensese R, Pawlowsky-Glahn V (2001) Criteria to compare estimation methods of regionalized compositions. *Math Geol* 33(8):889–909 · Zbl 1011.86505 · doi:10.1023/A:1012293922142
- [19] Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V (2003a) Dealing with zeros and missing values in compositional data sets. *Math Geol* 35(3):253–278 · Zbl 1302.86027 · doi:10.1023/A:1023866030544
- [20] Martín-Fernández JA, Palarea-Albaladejo J, Gómez-García J (2003b) Markov chain Monte Carlo method applied to rounding zeros of compositional data: first approach. In: Thió-Henestrosa S, Martín-Fernández JA (eds) *Compositional data analysis workshop*, Girona, Spain. <http://ima.udg.es/Activitats/CoDaWork03/>
- [21] Mateu-Figueras G, Barceló-Vidal C (eds) (2005) *Second compositional data analysis workshop–CoDaWork’05*, Proceedings, Universitat de Girona, CD-ROM, ISBN: 84-8458-222-1; available at <http://ima.udg.es/Activitats/CoDaWork05/>
- [22] Mateu-Figueras G, Pawlowsky-Glahn V (2007) The skew-normal distribution on SD. Special issue: Skew-elliptical distributions and their application. *Commun Stat Theory Methods* 36(9):1787–1802 · Zbl 1315.60023 · doi:10.1080/03610920601126258
- [23] McLachlan GJ, Krishnan T (1997) *The EM algorithm and extensions*. Wiley, New York, 274 p · Zbl 0882.62012
- [24] Palarea-Albaladejo J, Martín-Fernández JA, Gómez-García J (2005) ALR approach for replacing values below the detection limit. In: Mateu-Figueras G, Barceló-Vidal C (eds) *Compositional data analysis workshop*, Girona, Spain, 2005. <http://ima.udg.es/Activitats/CoDaWork05/>
- [25] Palarea-Albaladejo J, Martín-Fernández JA (2007) A modified EM alr-algorithm for replacing rounded zeros in compositional data sets. *Comput Geosci* (submitted)
- [26] Pawlowsky-Glahn V (guest ed) (2005) Special issue: Advances in compositional data. *Math Geol* 37(7): 671–850 · Zbl 1109.86300
- [27] Rubin DB (1987) *Multiple imputation for nonresponse in survey*. Wiley, New York, 258 p
- [28] Sandford RF, Pierson CT, Crovelli RA (1993) An objective replacement method for censored geochemical data. *Math Geol* 25(1):59–80 · doi:10.1007/BF00890676
- [29] Schafer JL (1997) *Analysis of incomplete multivariate data*. Chapman and Hall, London, 430 p · Zbl 0997.62510
- [30] Thió-Henestrosa S, Martín-Fernández JA (eds) (2003) *Compositional data analysis workshop–CoDaWork’03*, Proceedings, Universitat de Girona, CD-ROM, ISBN: 84-8458-111-X; available at <http://ima.udg.es/Activitats/CoDaWork03/>
- [31] Wu CFJ (1983) On the convergence properties of the EM algorithm. *Ann Stat* 11:95–103 · Zbl 0517.62035 · doi:10.1214/aos/1176346060

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.