

**Emura, Takeshi; Konno, Yoshihiko**

**A goodness-of-fit test for parametric models based on dependently truncated data.** (English)

Zbl 1252.62052

Comput. Stat. Data Anal. 56, No. 7, 2237-2250 (2012).

Summary: Suppose that one can observe bivariate random variables  $(L, X)$  only when  $L \leq X$  holds. Such data are called left-truncated data and found in many fields, such as experimental education and epidemiology. Recently, a method of fitting a parametric model on  $(L, X)$  has been considered, which can easily incorporate the dependent structure between the two variables. A primary concern for the parametric analysis is the goodness-of-fit for the imposed parametric forms. Due to the complexity of dependent truncation models, the traditional goodness-of-fit procedures, such as Kolmogorov-Smirnov type tests based on the bootstrap approximation to the null distribution, may not be computationally feasible. We develop a computationally attractive and reliable algorithm for the goodness-of-fit test based on the asymptotic linear expression. By applying the multiplier central limit theorem to the asymptotic linear expression, we obtain an asymptotically valid goodness-of-fit test. Monte Carlo simulations show that the proposed test has correct type I error rates and desirable empirical power. It is also shown that the method significantly reduces the computational time compared with the commonly used parametric bootstrap method. Analysis on law school data is provided for illustration. R codes for implementing the proposed procedure are available in the supplementary material.

**MSC:**

- 62G10 Nonparametric hypothesis testing
- 62F40 Bootstrap, jackknife and other resampling methods
- 62H12 Estimation in multivariate analysis
- 60F05 Central limit and other weak theorems
- 62F03 Parametric hypothesis testing
- 62N01 Censored data models
- 65C05 Monte Carlo methods
- 65C60 Computational problems in statistics (MSC2010)

Cited in 5 Documents

**Keywords:**

central limit theorem; empirical process; truncation; maximum likelihood; parametric bootstrap; shrinkage estimator, R

**Software:**

R; MASS (R); bootstrap; numDeriv

**Full Text:** DOI

**References:**

- [1] Bücher, A.; Dette, H., A note on bootstrap approximations for the empirical copula process, *Statistics & probability letters*, 80, 1925-1932, (2010) · Zbl 1202.62055
- [2] Chen, C.-H.; Tsai, W.-Y.; Chao, W.-H., The product-moment correlation coefficient and linear regression for truncated data, *Journal of American statistical association*, 91, 1181-1186, (1996) · Zbl 0882.62054
- [3] Cohen, A.C., *Truncated and censored samples: theory and applications*, (1991), Marcel Dekker, Inc New York · Zbl 0742.62027
- [4] Dietz, E.; Böhning, D., On estimation of the Poisson parameter in zero-modified Poisson models, *Computational statistics & data analysis*, 34, 441-459, (2000) · Zbl 1046.62085
- [5] Efron, B.; Tibshirani, R.J., *An introduction to the bootstrap*, (1993), Chapman & Hall London · Zbl 0835.62038
- [6] Emura, T., Konno, Y., 2009. Multivariate parametric approaches for dependently left-truncated data. *Technical Reports of Mathematical Sciences, Chiba University* 25, No. 2. · Zbl 1241.62094
- [7] Emura, T.; Konno, Y., Multivariate normal distribution approaches for dependently truncated data, *Statistical papers*, 53, 1, 133-149, (2012) · Zbl 1241.62094
- [8] Emura, T.; Wang, W., Testing quasi-independence for truncation data, *Journal of multivariate analysis*, 101, 223-239, (2010) · Zbl 1352.62148

- [9] Emura, T.; Wang, W.; Hung, H., Semi-parametric inference for copula models for truncated data, *Statistica sinica*, 21, 349-367, (2011) · [Zbl 05849522](#)
- [10] Fang, K.-T.; Kotz, S.; Ng, K.W., *Symmetric multivariate and related distributions*, (1990), Chapman & Hall New York
- [11] Gilberet, P., 2010. *Accurate Numerical Derivatives*. R package version 2010.11-1.
- [12] Holgate, Estimation of the bivariate Poisson distribution, *Biometrika*, 51, 241-245, (1964) · [Zbl 0133.11802](#)
- [13] Jin, Z.; Lin, D.Y.; Wei, L.J., Rank-based inference for the accelerated failure time model, *Biometrika*, 90, 341-353, (2003) · [Zbl 1034.62103](#)
- [14] Klein, J.P.; Moeschberger, M.L., *Survival analysis: techniques for censored and truncated data*, (2003), Springer New York · [Zbl 1011.62106](#)
- [15] Knight, K., *Mathematical statistics*, (2000), Chapman & Hall · [Zbl 0935.62002](#)
- [16] Kojadinovic, I.; Yan, J., A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems, *Statistics and computing*, 21, 17-30, (2011) · [Zbl 1274.62400](#)
- [17] Kojadinovic, I.; Yan, J.; Holmes, M., Fast large-sample goodness-of-fit tests for copulas, *Statistica sinica*, 21, 841-871, (2011) · [Zbl 1214.62049](#)
- [18] Lakhali-Chaieb, L.; Rivest, L.-P.; Abdous, B., Estimating survival under a dependent truncation, *Biometrika*, 93, 665-669, (2006) · [Zbl 1109.62084](#)
- [19] Lang, K.L.; Little, J.A.; Taylor, J.M.G., Robust statistical modeling using the t distribution, *Journal of the American statistical association*, 84, 881-896, (1989)
- [20] Lehmann, E.L.; Casella, G., *Theory of point estimation*, (1993), Springer-Verlag New York
- [21] Lynden-Bell, D., A method of allowing for known observational selection in small samples applied to 3RC quasars, *Monthly notices of the royal astronomical society*, 155, 95-118, (1971)
- [22] Martin, E.C.; Betensky, R.A., Testing quasi-independence of failure and truncation via conditional kendall's tau, *Journal of the American statistical association*, 100, 484-492, (2005) · [Zbl 1117.62397](#)
- [23] Ripley, P., Hornik, K., Gebhardt, A., 2011. *Support Functions and Datasets for Venables and Ripley's MASS*, R package version 7.3-13.
- [24] Schiel, J.L., 1998. *Estimating conditional probabilities of success and other course placement validity statistics under soft truncation*. ACT Research Report Series 1998-2, Iowa City, IA: ACT.
- [25] Schiel, J.L., Harmston, M., 2000. *Validating two-stage course placement systems when data are truncated*. ACT Research Report Series 2000-3, Iowa City, IA: ACT.
- [26] Spiekerman, C.F.; Lin, D.Y., Marginal regression models for multivariate failure time data, *Journal of the American statistical association*, 93, 1164-1175, (1998) · [Zbl 1064.62572](#)
- [27] Tsai, W.-Y., Testing the assumption of independence of truncation time and failure time, *Biometrika*, 77, 169-177, (1990) · [Zbl 0692.62045](#)
- [28] van der Vaart, A.W., ()
- [29] van der Vaart, A.W.; Wellner, J.A., ()
- [30] Waiker, V.B.; Schuurmann, F.J.; Raghunathan, T.E., On a two-stage shrinkage estimator of the Mean of a normal distribution, *Communications in statistics — theory and methods*, 13, 1901-1913, (1984)
- [31] Wang, M.C.; Jewell, N.P.; Tsai, W.Y., Asymptotic properties of the product-limit estimate and right censored data, *Annals of statistic*, 13, 1597-1605, (1986) · [Zbl 0656.62048](#)
- [32] Woodroffe, M., Estimating a distribution function with truncated data, *Annals of statistics*, 13, 163-177, (1985) · [Zbl 0574.62040](#)

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.