



A goodness-of-fit test for parametric models based on dependently truncated data

Takeshi Emura^a, Yoshihiko Konno^{b,*}

^a Graduate Institute of Statistics, National Central University, Zhongda Road, Taoyuan, Taiwan

^b Department of Mathematical and Physical Sciences, Japan Women's University, 2-8-1 Mejirodai, Bunkyo-ku, Tokyo 112-8681, Japan

ARTICLE INFO

Article history:

Received 24 January 2011

Received in revised form 24 December 2011

Accepted 29 December 2011

Available online 18 January 2012

Keywords:

Central limit theorem

Empirical process

Truncation

Maximum likelihood

Parametric bootstrap

Shrinkage estimator

ABSTRACT

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 null distribution, may not be computationally feasible. In this paper, 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.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Truncated data are those from which part of them are entirely excluded. For instance, in the study of aptitude test scores in experimental education, only those individuals whose test scores are above (or below) a threshold may appear in the sample (Schiel, 1998; Schiel and Harmston, 2000). Many different types of truncation are possible depending on how to determine the truncation criteria. A classical parametric method for analyzing truncated data is based on a fixed truncation. That is, a variable X^0 of interest can be included in the sample if it exceeds a fixed value l , where l is known. Parametric estimation for the normal distribution of X^0 has been given by Cohen (1991). Other examples of the fixed truncation include the zero-truncated Poisson model in which X^0 is a Poisson random variable and $l = 1$.

A more general truncation scheme is the so-called “left-truncation” in which the sample is observed when a variable X^0 exceeds another random variable L^0 . The left-truncated data is commonly seen in studies of biomedicine, epidemiology and astronomy (Klein and Moeschberger, 2003). Construction of nonparametric estimators for $\Pr(X^0 \leq x)$ under the left-truncation has been extensively studied (e.g., Woodroffe, 1985; Wang et al., 1986). It is well known that the nonparametric methods rely on the independence assumption between L^0 and X^0 . Accordingly, Chen et al. (1996), Emura and Wang (2010),

* Corresponding author. Tel.: +81 3 5981 3642; fax: +81 3 5981 3637.

E-mail addresses: emura@stat.ncu.edu.tw (T. Emura), konno@fc.jwu.ac.jp (Y. Konno).

Martin and Betensky (2005) and Tsai (1990) have presented methods for testing the independence assumption. For positive random variables L^0 and X^0 , semiparametric approaches proposed by Emura et al. (2011) and Lakhali-Chaieb et al. (2006) are alternatives in the absence of independence assumption, where the association structure between L^0 and X^0 is modeled via an Archimedean copula.

Compared with the nonparametric and semiparametric inferences, there is not much in the literature on the analysis of left-truncated data based on parametric modeling. Although parametric modeling easily incorporates the dependence structure between L^0 and X^0 , it involves strong distributional assumptions, and the inference procedure may not be robust to departures from these assumptions (Emura and Konno, 2012). Nevertheless, parametric modeling is still useful in many applications where parameters in the model provide useful interpretation or a particular parametric form is supported by the subject matter knowledge. For instance, in the study of aptitude test scores in educational research, researchers may be interested in estimating the mean and standard deviation of the test score X^0 rather than $\Pr(X^0 \leq x)$ (Schiel and Harmston, 2000; Emura and Konno, 2009). Hence, parameters of the normal distribution usually provide useful summary information (see Section 5 for details). For another example, the study of count data in epidemiological research often encounters the zero-modified Poisson model (Dietz and Böhning, 2000) for X^0 (see Example 3 in Appendix A for details). For count data, the main focus is to estimate the intensity parameter of the Poisson distribution rather than $\Pr(X^0 \leq x)$. In the preceding two examples, one needs to specify the parametric forms of (L^0, X^0) . If the goodness-of-fit tests are used appropriately, the robustness concern about the parametric analysis can be circumvented.

In this paper, we develop a computationally attractive and reliable algorithm for the goodness-of-fit test by utilizing the multiplier central limit theorem. The basic idea behind the proposed approach follows the goodness-of-fit procedure for copula models (Kojadinovic and Yan, 2011; Kojadinovic et al., 2011), though the technical details and the computational advantages in the present setting are different. The rest of the paper is organized as follows: Section 2 briefly reviews the parametric formulation given in Emura and Konno (2012). Section 3 presents the theory and algorithm of the proposed goodness-of-fit test based on the multiplier central limit theorem. Simulations and data analysis are presented in Sections 4 and 5, respectively. Section 6 concludes this paper.

2. Parametric inference for dependently truncated data

In this section, we introduce the parametric approach to dependent truncation data based on Emura and Konno (2012) and derive the asymptotic results of the maximum likelihood estimator (MLE) that are the basis for the subsequent developments.

Let $f_{\theta}(l, x)$ be a density or probability function of a bivariate random variable (L^0, X^0) , where $\theta \in \Theta$ is a p -variate vector of parameters and where $\Theta \subset R^p$ is a parameter space. In a truncated sample, a pair (L^0, X^0) is observed when $L^0 \leq X^0$ holds. For observed data $\{(L_j, X_j); j = 1, 2, \dots, n\}$ subject to $L_j \leq X_j$, the likelihood function has the form

$$L(\theta) = c(\theta)^{-n} \prod_j f_{\theta}(L_j, X_j), \quad (1)$$

where $c(\theta) = \Pr(L^0 \leq X^0)$. Let $\dot{l}_j(\theta)$ be the column vector of partial derivatives (with respect to the component of θ) of $l_j(\theta) = \log\{c(\theta)^{-1} f_{\theta}(L_j, X_j)\}$, i.e., $\dot{l}_j(\theta) = \partial l_j(\theta) / \partial \theta$ for $j = 1, 2, \dots, n$, and let $\hat{\theta}$ be the maximum likelihood estimator (MLE) that maximizes (1) in Θ . Emura and Konno (2012) noted that for computing the MLE, it is crucial that the simple formula of $c(\theta)$ is available. This also has a crucial role in the subsequent developments for the proposed goodness-of-fit test procedure. For easy reference, Appendix A lists three examples of the parametric forms that permit a tractable form in $c(\theta)$. The following Lemma is a basis for deriving the asymptotic expression for the goodness-of-fit statistics.

Lemma 1. Suppose that (R1) through (R7) listed in Appendix B hold. Then,

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_j I^{-1}(\theta) \dot{l}_j(\theta) + o_p(1), \quad (2)$$

where $I(\theta) = E\{\dot{l}_j(\theta) \dot{l}_j^T(\theta)\}$ is the Fisher information matrix and $\dot{l}_j^T(\theta)$ is the transposed vector of $\dot{l}_j(\theta)$.

3. Goodness-of-fit procedure under truncation

3.1. Asymptotic linear expression of the goodness-of-fit process

Let $\mathfrak{S} = \{f_{\theta} \mid \theta \in \Theta\}$ be a given parametric family. Also, let f be the underlying (true) density or probability function of a bivariate random variable (L^0, X^0) . Given the observed data $\{(L_j, X_j); j = 1, 2, \dots, n\}$, we wish to test the null hypothesis

$$H_0 : f \in \mathfrak{S} \quad \text{against} \quad H_1 : f \notin \mathfrak{S}.$$

One of the popular classes of goodness-of-fit tests consists of comparing the distance between $\hat{F}(l, x) = \sum_j \mathbf{I}(L_j \leq l, X_j \leq x) / n$ and $F_{\hat{\theta}}(l, x) / c(\hat{\theta})$, where $\mathbf{I}(\cdot)$ is the indicator function and

$$F_{\theta}(l, x) = \begin{cases} \iint_{u \leq l, u \leq v \leq x} f_{\theta}(u, v) du dv & \text{for a continuous distribution;} \\ \sum_{u \leq l} \sum_{u \leq v \leq x} f_{\theta}(u, v) & \text{for a discrete distribution.} \end{cases}$$

The Kolmogorov–Smirnov type test is based on

$$K = \sup_{l \leq x} |\hat{F}(l, x) - F_{\hat{\theta}}(l, x)/c(\hat{\theta})|. \tag{3}$$

The calculation of K requires the numerical integrations (or summations) for $F_{\theta}(l, x)$ at $O(n^2)$ different points in $\{(l, x); l \leq x\}$. A computationally attractive alternative is the Cramér–von Mises type statistic

$$C = \iint_{l \leq x} \{\hat{F}(l, x) - F_{\hat{\theta}}(l, x)/c(\hat{\theta})\}^2 d\hat{F}(l, x) = \sum_j \{\hat{F}(L_j, X_j) - F_{\hat{\theta}}(L_j, X_j)/c(\hat{\theta})\}^2. \tag{4}$$

This requires exactly n evaluations of $F_{\theta}(l, x)$. The null distributions for K and C have not been derived and depend on the true value of θ .

Empirical process techniques are useful for analyzing the goodness-of-fit process $\hat{F} - F_{\hat{\theta}}/c(\hat{\theta})$ defined on $(-\infty, \infty)^2$. Let $\varphi : \Theta \mapsto D\{(-\infty, \infty)^2\}$ be $\varphi(\hat{\theta}) = F_{\hat{\theta}}/c(\hat{\theta})$, where $D\{(-\infty, \infty)^2\}$ is a collection of bounded functions on $(-\infty, \infty)^2$ that are continuous from above, equipped with the uniform norm $\|F\|_{\infty} = \sup_{-\infty < l, x < \infty} |F(l, x)|$. Under the assumptions (R1), (R3), (R4) and (R8) listed in [Appendix B](#), the map φ is shown to be the Hadamard differentiable with the derivative given by

$$\mathbf{g}_{\theta} = \mathbf{g}_{\theta}(l, x) = c(\theta)^{-2} \{\dot{F}_{\theta}(l, x)c(\theta) - F_{\theta}(l, x)\dot{c}(\theta)\},$$

where $\dot{F}_{\theta}(l, x) = \partial F_{\theta}(l, x)/\partial \theta$ and $\dot{c}(\theta) = \partial c(\theta)/\partial \theta$, the column vectors of partial derivatives (with respect to the component of θ) of $F_{\theta}(l, x)$ and $c(\theta)$, respectively. Applying the functional delta method under the regularity conditions for Theorem 20.8 of [van der Vaart \(1998\)](#), we have

$$\sqrt{n}\{F_{\hat{\theta}}(l, x)/c(\hat{\theta}) - F_{\theta}(l, x)/c(\theta)\} = \frac{1}{\sqrt{n}} \sum_j \mathbf{g}'_{\theta}(l, x)I^{-1}(\theta)\dot{l}_j(\theta) + o_p(1),$$

where $o_p(1)$ is uniform in (l, x) . Making the above arguments rigorous, we obtain the following theorem whose proof is given in [Appendix C](#):

Theorem 1. *Suppose that (R1) through (R8) listed in [Appendix B](#) hold. Under H_0 ,*

$$\sqrt{n}\{\hat{F}(l, x) - F_{\hat{\theta}}(l, x)/c(\hat{\theta})\} = \frac{1}{\sqrt{n}} \sum_j V_j(l, x; \theta) + o_p(1), \tag{5}$$

where $V_j(l, x; \theta) = \mathbf{I}(L_j \leq l, X_j \leq x) - F_{\theta}(l, x)/c(\theta) - \mathbf{g}'_{\theta}(l, x)I^{-1}(\theta)\dot{l}_j(\theta)$.

3.2. Algorithm based on the multiplier central limit theorem

Eq. (5) is the basis for developing a resampling scheme based on the multiplier central limit theorem ([van der Vaart and Wellner, 1996](#)). Let $\{Z_j; j = 1, 2, \dots, n\}$ be independent random variables with $E(Z_j) = 0$ and $\text{Var}(Z_j) = 1$. Consider

$$\frac{1}{\sqrt{n}} \sum_j Z_j V_j(l, x; \theta). \tag{6}$$

Conditional on $\{(L_j, X_j); j = 1, 2, \dots, n\}$, the only random quantities in Eq. (6) are $\{Z_j; j = 1, 2, \dots, n\}$, which will be called the “multiplier”. The conditional distribution of (6) can approximate the asymptotic distribution of $\sqrt{n}\{\hat{F}(l, x) - F_{\hat{\theta}}(l, x)/c(\hat{\theta})\}$.

The random variable $V_j(l, x; \theta)$ contains $I(\theta)$, whose analytical expression is usually impossible or very difficult to obtain due to truncation. Define the observed Fisher information matrix $i_n(\hat{\theta})$. Replacing θ and $I(\theta)$ with $\hat{\theta}$ and $i_n(\hat{\theta})/n$, respectively, Eq. (6) is approximated by

$$\frac{1}{\sqrt{n}} \sum_j Z_j \hat{V}_j(l, x; \hat{\theta}),$$

where $\hat{V}_j(l, x; \theta) = \mathbf{I}(L_j \leq l, X_j \leq x) - F_{\theta}(l, x)/c(\theta) - \mathbf{g}'_{\theta}(l, x)\{i_n(\theta)/n\}^{-1}\dot{l}_j(\theta)$. The following lemma shows that the substitution does not alter the asymptotic behavior; the proof is given in [Appendix D](#).

Lemma 2. Suppose that (R1) through (R8) listed in Appendix B hold. Under H_0 ,

$$\frac{1}{\sqrt{n}} \sum_j Z_j \hat{V}_j(l, x; \hat{\theta}) = \frac{1}{\sqrt{n}} \sum_j Z_j V_j(l, x; \theta) + o_p(1).$$

Therefore, the distribution of K in (3) can be approximated by

$$\frac{1}{n} \sup_{l \leq x} \left| \sum_j Z_j \hat{V}_j(l, x; \hat{\theta}) \right|.$$

Also, the distribution of C in (4) can be approximated by

$$\frac{1}{n^2} \sum_i \left[\sum_j Z_j \hat{V}_j(L_i, X_i; \hat{\theta}) \right]^2 = \left\| \frac{\mathbf{Z}' \hat{\mathbf{V}}(\hat{\theta})}{n} \right\|^2,$$

where $\hat{\mathbf{V}}(\hat{\theta})$ is a matrix whose (j, i) element is $\hat{V}_j(L_i, X_i; \hat{\theta})$, $\mathbf{Z} = (Z_1, \dots, Z_n)'$ and $\|\mathbf{a}\|^2 = \mathbf{a}'\mathbf{a}$ for a vector \mathbf{a} . Once the matrix $\hat{\mathbf{V}}(\hat{\theta})$ is calculated from observed data, one can easily generate an approximate version of C by multiplying $\hat{\mathbf{V}}(\hat{\theta})/n$ by \mathbf{Z} and then taking the norm. For finite sample accuracy, the standard normal multiplier $Z_j \sim N(0, 1)$ typically does a better job than other types, such as a two-point distribution with $\Pr(Z_j = 1) = \Pr(Z_j = -1) = 0.5$ (see Section 1 of the supplementary material).

The testing algorithm is as follows. Modifications of the algorithm for the statistic K in (3) are straightforward.

Algorithm based on the multiplier method:

Step 0 : Calculate the statistic $C = \sum_i \{\hat{F}(L_i, X_i) - F_{\hat{\theta}}(L_i, X_i)/c(\hat{\theta})\}^2$ and matrix $\hat{\mathbf{V}}(\hat{\theta})$.

Step 1 : Generate $Z_j^{(b)} \sim N(0, 1)$; $b = 1, 2, \dots, B, j = 1, 2, \dots, n$.

Step 2 : Calculate $\{C^{(b)}; b = 1, 2, \dots, B\}$,

where $C^{(b)} = \|(\mathbf{Z}^{(b)})' \hat{\mathbf{V}}(\hat{\theta})/n\|^2$ and $\mathbf{Z}^{(b)} = (Z_1^{(b)}, \dots, Z_n^{(b)})'$.

Step 3 : Reject H_0 at the $100\alpha\%$ significance level if $\sum_{b=1}^B \mathbf{I}(C^{(b)} \geq C)/B < \alpha$.

The algorithm based on the parametric Bootstrap method (Efron and Tibshirani, 1993) is provided for comparison with the multiplier method.

Algorithm based on the parametric Bootstrap method:

Step 0 : Calculate the statistic $C = \sum_i \{\hat{F}(L_i, X_i) - F_{\hat{\theta}}(L_i, X_i)/c(\hat{\theta})\}^2$.

Step 1 : Generate $(L_j^{(b)}, X_j^{(b)})$ which follows the density $f_{\hat{\theta}}(l, x)$, subject to $L_j^{(b)} \leq X_j^{(b)}$, for $b = 1, 2, \dots, B, j = 1, 2, \dots, n$.

Step 2 : Calculate $\{C^{*(b)}; b = 1, 2, \dots, B\}$,

where $C^{*(b)} = \sum_i \{\hat{F}^{(b)}(L_i^{(b)}, X_i^{(b)}) - F_{\hat{\theta}^{(b)}}(L_i^{(b)}, X_i^{(b)})/c(\hat{\theta}^{(b)})\}^2$ and where $\hat{F}^{(b)}(l, x)$ and $\hat{\theta}^{(b)}$ are the empirical CDF and MLE based on $\{(L_j^{(b)}, X_j^{(b)}); j = 1, 2, \dots, n\}$.

Step 3 : Reject H_0 at the $100\alpha\%$ significance level if $\sum_{b=1}^B \mathbf{I}(C^{*(b)} \geq C)/B < \alpha$.

In Step 1 of the multiplier method, it is fairly easy to generate $Z_j^{(b)} \sim N(0, 1)$. On the other hand, Step 1 of the parametric Bootstrap approach is more difficult since data are generated from a given bivariate density function subject to truncation. Usually, the following accept–reject algorithm is used: (i) data (L, X) from the density $f_{\hat{\theta}}(l, x)$ is generated; (ii) if $L \leq X$, we accept the sample and set $(L_j^{(b)}, X_j^{(b)}) = (L, X)$; otherwise we reject (L, X) and return to (i). This algorithm can be time-consuming especially when the acceptance rate $c(\hat{\theta})$ is small.

In Step 2, the multiplier method only needs to multiply the standard normal vector to $\hat{\mathbf{V}}(\hat{\theta})$. On the other hand, the parametric Bootstrap method needs to calculate $\hat{\theta}^{(b)}$ for $b = 1, 2, \dots, B$. This is very time consuming since each maximization process requires repeated evaluations of the likelihood function.

Beside the computational time spent, the parametric Bootstrap method can produce erroneous results when some of $\hat{\theta}^{(b)}$, $b = 1, 2, \dots, B$, are local maxima. If $\hat{\theta}^{(b)}$ is a local maxima, it is usually not close to $\hat{\theta}$, and the Bootstrap version of the Cramér–von Mises statistic $C^{*(b)}$ tends to be very large. Although this error is subject to the ability of computer programs for numerical maximizations, it is not always easy to locate global maxima for a large number of resampling steps (usually $B \geq 1000$ times). The multiplier method is free from this error and more reliably implemented than the parametric Bootstrap.

Finally, we give the validity of the method along the lines of [Kojadinovic et al. \(2011\)](#). Let $G_n(l, x) = \sqrt{n}\{\hat{F}(l, x) - F_{\theta}(l, x)/c(\theta)\}$ and $G_n^{(b)}(l, x) = n^{-1/2} \sum_j Z_j^{(b)} \hat{V}_j(l, x; \theta)$ for $b = 1, 2, \dots, B$. The proof of the following theorem is given in [Appendix E](#).

Theorem 2. Suppose that (R1) through (R8) listed in [Appendix B](#) hold. Under H_0 , as $n \rightarrow \infty$,

$$(G_n, G_n^{(1)}, \dots, G_n^{(B)}) \rightarrow (G_{\theta}, G_{\theta}^{(1)}, \dots, G_{\theta}^{(B)})$$

in $D\{(-\infty, \infty)^2\}^{\otimes(B+1)}$, where G_{θ} is the mean zero Gaussian process whose covariance for $(l, x), (l^*, x^*) \in (-\infty, \infty)^2$ is given as

$$\begin{aligned} \text{Cov}\{G_{\theta}(l, x), G_{\theta}(l^*, x^*)\} &= E\{V_j(l, x; \theta)V_j(l^*, x^*; \theta)\} \\ &= F_{\theta}(l \wedge l^*, x \wedge x^*)/c(\theta) - F_{\theta}(l, x)F_{\theta}(l^*, x^*)/c(\theta)^2 - \mathbf{g}'_{\theta}(l, x)I^{-1}(\theta)\mathbf{g}_{\theta}(l^*, x^*), \end{aligned}$$

where $a \wedge b \equiv \min(a, b)$, and $G_{\theta}^{(1)}, \dots, G_{\theta}^{(B)}$ are independent copies of G_{θ} .

Under H_1 , $(G_n^{(1)}, \dots, G_n^{(B)})$ are no longer approximate independent copies of G_n . In particular, the goodness-of-fit process $G_n(l, x)$ usually converges in probability to ∞ at some (l, x) while the multiplier realization $G_n^{(b)}$ converges weakly to a tight Gaussian process. This produces consistency of the goodness-of-fit test.

3.3. Computational aspects

To compute the proposed goodness-of-fit statistic based on the multiplier method, one needs to calculate $F_{\theta}(l, x)$, $\dot{c}(\theta)$, $\dot{F}_{\theta}(l, x)$, and $\dot{l}_j(\theta)$. Although these can be calculated for each specific parametric model, the formulas are not always easy. In what follows, we describe how to calculate these quantities for the bivariate normal, bivariate Poisson, and zero-modified Poisson models discussed in [Examples 1–3 of Appendix A](#).

For the bivariate normal, bivariate Poisson, and zero-modified Poisson models, their formulas are written respectively as:

$$\begin{aligned} F_{\theta}(l, x) &= \int_{-\infty}^{(l-\mu_L)/\sigma_L} \left[\Phi \left\{ \frac{x - \mu_X - \sigma_{LX}S/\sigma_L}{\sqrt{\sigma_X^2 - \sigma_{LX}^2/\sigma_L^2}} \right\} - \Phi \left\{ \frac{\mu_L + \sigma_{LS} - \mu_X - \sigma_{LX}S/\sigma_L}{\sqrt{\sigma_X^2 - \sigma_{LX}^2/\sigma_L^2}} \right\} \right] \phi(s) ds, \\ F_{\theta}(l, x) &= e^{-\lambda_L - \lambda_X - \alpha} \sum_{u=0}^l \sum_{v=u}^x \sum_{w=0}^u \frac{\lambda_L^{u-w} \lambda_X^{v-w} \alpha^w}{(u-w)!(v-w)!w!}, \\ F_{\theta}(l, x) &= \begin{cases} (1-p)e^{-\lambda_X} \sum_{u=0}^x (\lambda_X^u/u!) & \text{if } l = 0; \\ e^{-\lambda_X} \left(\sum_{u=0}^x \lambda_X^u/u! - p \right) & \text{if } l = 1, \end{cases} \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution for $N(0, 1)$, and $\phi(x) = d\Phi(x)/dx$. Since it is not very easy to obtain analytical expressions for $\dot{c}(\theta)$ and $\dot{F}_{\theta}(l, x)$, we propose to use $\{c(\theta + h\mathbf{1}) - c(\theta)\}/h$ and $\{F_{\theta+h\mathbf{1}}(l, x) - F_{\theta}(l, x)\}/h$, where h is a small value and $\mathbf{1}' = (1, \dots, 1)$. Computer programs, such as the “numDeriv” package ([Gilbert, 2010](#)) in R, are also useful. We conducted a simulation to examine the correctness of the proposed numerical derivative (simulation results are given in the supplementary material). The results show that the proposed numerical derivative with $h = 10^{-8}$ is virtually identical to both the analytical derivative and the numerical derivative from the numDeriv package. The proposed numerical derivative has less programming effort than the analytical one.

The formula of $\dot{l}_j(\theta)$ under the bivariate normal model is given in [Emura and Konno \(2012\)](#). Under the bivariate Poisson model, one has $\dot{l}_j(\theta) = (\dot{l}_j^1(\theta), \dot{l}_j^2(\theta), \dot{l}_j^3(\theta))'$ where

$$\begin{aligned} \dot{l}_j^1(\theta) &= -\frac{\partial c(\theta)/\partial \lambda_L}{c(\theta)} - 1 + \frac{\mathbf{I}(L_j \geq 1)}{B_j(\theta)} \sum_{w=0}^{L_j-1} \frac{\lambda_L^{L_j-1-w} \lambda_X^{X_j-w} \alpha^w}{(L_j-1-w)!(X_j-w)!w!}, \\ \dot{l}_j^2(\theta) &= -\frac{\partial c(\theta)/\partial \lambda_X}{c(\theta)} - 1 + \frac{\mathbf{I}(X_j \geq 1)}{B_j(\theta)} \sum_{w=0}^{\min\{L_j, X_j-1\}} \frac{\lambda_L^{L_j-w} \lambda_X^{X_j-1-w} \alpha^w}{(L_j-w)!(X_j-1-w)!w!}, \\ \dot{l}_j^3(\theta) &= -1 + \frac{\mathbf{I}(L_j \geq 1)}{B_j(\theta)} \sum_{w=1}^{L_j} \frac{\lambda_L^{L_j-w} \lambda_X^{X_j-w} \alpha^{w-1}}{(L_j-w)!(X_j-w)!(w-1)!}, \end{aligned}$$

where $B_j(\boldsymbol{\theta}) = \sum_{w=0}^{L_j} \frac{\lambda_L^{L_j-w} \lambda_X^{X_j-w} \alpha^w}{(L_j-w)!(X_j-w)!w!}$. Under the zero-modified Poisson model, one has $\dot{l}_j(\boldsymbol{\theta}) = (\dot{l}_j^1(\boldsymbol{\theta}), \dot{l}_j^2(\boldsymbol{\theta}))'$, where

$$\begin{aligned} \dot{l}_j^1(\boldsymbol{\theta}) &= \frac{L_j}{p(1-p)} - \frac{1}{1-p} + \frac{e^{-\lambda_X}}{1-pe^{-\lambda_X}}, \\ \dot{l}_j^2(\boldsymbol{\theta}) &= -\frac{pe^{-\lambda_X}}{1-pe^{-\lambda_X}} + \frac{X_j}{\lambda_X} - 1. \end{aligned}$$

3.4. Generalization to other estimators for $\boldsymbol{\theta}$

Although the proposed goodness-of-fit procedure is developed in the case where $\boldsymbol{\theta}$ is estimated by the MLE, it is not difficult to modify it for more general estimators. The fundamental requirement is that the estimator is asymptotically linear as in (2). A particularly interesting example in the dependently truncated data is the shrinkage “testimator” (Waiker et al., 1984). Suppose that the parameter can be written as $\boldsymbol{\theta}' = (\boldsymbol{\theta}^*, \rho)$, where ρ is the correlation between L^0 and X^0 , and $\hat{\boldsymbol{\theta}}^*$ is the MLE under the assumption of $\rho = 0$. Let $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$, and

$$\hat{\boldsymbol{\theta}}^{TEST} = \begin{pmatrix} \hat{\boldsymbol{\theta}}^* \\ 0 \end{pmatrix} + \left\{ \hat{\boldsymbol{\theta}} - \begin{pmatrix} \hat{\boldsymbol{\theta}}^* \\ 0 \end{pmatrix} \right\} \mathbf{I}\{2[l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}^*, 0)] > q_n\},$$

where $q_1 \geq q_2 \geq \dots$ is a sequence of positive constants. If q_n is the cutoff value for the chi-square distribution with one degree of freedom, $\hat{\boldsymbol{\theta}}^{TEST} = \hat{\boldsymbol{\theta}}$ if $H_0 : \rho = 0$ is rejected or $\hat{\boldsymbol{\theta}}^{TEST} = (\hat{\boldsymbol{\theta}}^*, 0)$ if H_0 is accepted by the likelihood ratio test. By definition, $\hat{\boldsymbol{\theta}}^{TEST}$ is shrunk to $(\hat{\boldsymbol{\theta}}^*, 0)$, borrowing strength from the smaller variance of $(\hat{\boldsymbol{\theta}}^*, 0)$ than that of $\hat{\boldsymbol{\theta}}$. Emura and Konno (2012) show by simulations that under the bivariate normal model, there is substantial efficiency gain near $\rho = 0$. This is closely related to a superefficient phenomenon due to Hodges (see Lehmann and Casella, 1993). The proof of the following Lemma is given in Appendix F.

Lemma 3. Suppose that $\boldsymbol{\theta} = (\boldsymbol{\theta}^*, 0)$ is inside the parameter space and that (R1) through (R7) in Appendix B hold. Also, suppose that either one of the following conditions holds: (i) $q_n \rightarrow 0$; (ii) $q_n = q > 0$ for all n and $\rho \neq 0$. Then,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^{TEST} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_j I^{-1}(\boldsymbol{\theta}) \dot{l}_j(\boldsymbol{\theta}) + o_p(1).$$

Remark 1. According to condition (R1) in Appendix B, the parameter space for $\boldsymbol{\theta} = (\boldsymbol{\theta}^*, \rho)$ should be an open set. This condition ensures that “the true parameter” is located inside the parameter space. When the correlation on (L^0, X^0) is introduced by the random effect, as mentioned in Example 2 of Appendix A, we have $\rho \in (0, 1)$ and the case of $\rho = 0$ is defined as the limit $\rho \downarrow 0$ which is outside the parameter space. Under such circumstances, the likelihood ratio test for $H_0 : \rho = 0$ becomes one-sided and the statistic does not have the asymptotic chi-squared distribution with one degree of freedom. As a result, the statement in Lemma 3 does not hold.

Theorem 3. Suppose that (R1) through (R8) listed in Appendix B hold. Further suppose that either one of the following conditions holds: (i) $q_n \rightarrow 0$; (ii) $q_n = q > 0$ for all n and $\rho \neq 0$. Then, Eq. (5) holds when $\hat{\boldsymbol{\theta}}$ is replaced by $\hat{\boldsymbol{\theta}}^{TEST}$.

The proof of Theorem 3 is similar to that of Theorem 1 based on the result of Lemma 3. Therefore, the algorithm developed in Section 3.2 is applicable by replacing $\hat{\boldsymbol{\theta}}$ with $\hat{\boldsymbol{\theta}}^{TEST}$.

4. Simulation results

To investigate the performance of the proposed goodness-of-fit test, we conducted extensive simulations using R. All results reported in this section are based on the standard normal multiplier $Z_j \sim N(0, 1)$ and the results based on the two-point multiplier are given in Section 2 of the supplementary material.

4.1. Simulations under the null distribution

In the first part, we have chosen the same design as Emura and Konno (2012) in which (L^0, X^0) follows the bivariate normal distribution with the mean vector and covariance matrix given respectively by

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_L \\ \mu_X \end{bmatrix} = \begin{bmatrix} 120 - 62.63 \\ 60.82 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_L^2 & \sigma_{LX} \\ \sigma_{LX} & \sigma_X^2 \end{bmatrix} = \begin{bmatrix} 19.64^2 & 19.64 \times 16.81\rho_{LX} \\ 19.64 \times 16.81\rho_{LX} & 16.81^2 \end{bmatrix},$$

Table 1
Performances of the proposed goodness-of-fit test at the 100α% level based on 1000 replications under the bivariate normal model with $n = 400$.

		$\rho_{X^0Y^0}$				
		0.70	0.35	0.00	−0.35	−0.70
$\hat{\theta}$:MLE	Rejection rate for $\alpha = 0.10$	0.115	0.096	0.086	0.095	0.103
	Rejection rate for $\alpha = 0.05$	0.055	0.048	0.043	0.051	0.056
	Rejection rate for $\alpha = 0.01$	0.011	0.006	0.011	0.011	0.010
	Mean of C	0.0606	0.0592	0.0568	0.0582	0.0578
	Mean of $E_{data}(C^{(b)})$	0.0590	0.0590	0.0585	0.0592	0.0646
	SD of C	0.0290	0.0244	0.0219	0.0217	0.0217
	Mean of $SD_{data}(C^{(b)})$	0.0282	0.0251	0.0228	0.0229	0.0314
$\hat{\theta}^{TEST}$:Shrinkage	Rejection rate for $\alpha = 0.10$	0.108	0.099	0.111	0.100	0.086
	Rejection rate for $\alpha = 0.05$	0.052	0.060	0.052	0.057	0.046
	Rejection rate for $\alpha = 0.01$	0.010	0.012	0.009	0.012	0.010
	Mean of C	0.0602	0.0607	0.0615	0.0587	0.0565
	Mean of $E_{data}(C^{(b)})$	0.0592	0.0595	0.0617	0.0592	0.0612
	SD of C	0.0286	0.0260	0.0238	0.0224	0.0210
	Mean of $SD_{data}(C^{(b)})$	0.0284	0.0253	0.0247	0.0228	0.0268

Note: The mean and SD of the Cramér–von Mises test statistics C are calculated. $E_{data}(C^{(b)})$ and $SD_{data}(C^{(b)})$ are the mean and SD of $\{C^{(b)}; b = 1, 2, \dots, 1000\}$, respectively, based on 1000 multiplier realizations.

where $\rho_{LX} = 0.70, 0.35, 0, -0.35, \text{ or } -0.70$. In this design, the population parameters of Japanese test scores (mean = 60.82, SD = 16.81) and English test scores (mean = 62.63, SD = 19.64) are determined by the records of the National Center Test for University for 2008 in Japan. We set the null hypothesis $H_0 : f \in \mathfrak{S}$, where $\mathfrak{S} = \{f_{\theta} \mid \theta \in \Theta\}$ is a family of bivariate normal distributions with $\theta' = (\mu_L, \mu_X, \sigma_L^2, \sigma_X^2, \sigma_{LX})$.

The R mvnrm routine in MASS package (Ripley et al., 2011) is used to generate random samples from the bivariate normal distribution. Truncated data $\{(L_j, X_j); j = 1, 2, \dots, n\}$ subject to $L_j \leq X_j$ represent those sample whose sum of Japanese and English scores is above 120. We set $n = 400$ as the common number of prospective students in any one department in Japanese universities (Emura and Konno, 2012). For the truncated data, we calculate the MLE $\hat{\theta}$ (and $\hat{\theta}^{TEST}$) and the Cramér–von Mises type statistic C given by (4). Here, $\hat{\theta}$ is obtained by minimizing $-\log L(\theta)$ using the non-linear minimization routine nlm in R. For each data realization, we compute $\{C^{(b)}; b = 1, 2, \dots, 1000\}$ under H_0 . Then we record the rejection/acceptance status of the goodness-of-fit test at the 100α% level. We also record the sample mean and standard deviation (SD) of $\{C^{(b)}; b = 1, 2, \dots, 1000\}$ to be compared with the sampling distribution of C . The results for 1000 repetitions are shown in Table 1. In all configurations, the rejection rates are in good agreement with the selected nominal sizes ($\alpha = 0.01, 0.05, \text{ and } 0.10$). Also, the sample mean and SD of C are close to their resampling versions based on $\{C^{(b)}; b = 1, 2, \dots, 1000\}$. The results provide empirical evidence that the multiplier methods for approximating the null distribution works well under the bivariate normal model.

In the second part, we carried out simulations under the bivariate Poisson model and zero-modified Poisson model. Hence, the null hypothesis is $H_0 : f \in \mathfrak{S}$, where $\mathfrak{S} = \{f_{\theta} \mid \theta \in \Theta\}$ is either a family of bivariate Poisson model or zero-modified Poisson model where the form of f_{θ} is given in Examples 2 and 3 of Appendix A. For the bivariate Poisson model, we set $(\lambda_L, \lambda_X) = (1, 1)$ or $(1, 2)$, and set α so that the correlation $\rho_{X^0Y^0} = \{\alpha/(\lambda_L\lambda_X)\}^{-1/2}$ becomes 0.35 or 0.7. Unlike the bivariate normal model, the bivariate Poisson model only permits positive association on (L^0, X^0) . For the zero-modified Poisson model, we set $\lambda_X = 1$ or $\lambda_X = 2$ and $p = 0.3$ or $p = 0.7$. Simulation results for $n = 400$ are summarized in Table 2. The rejection rates essentially have the correct values ($\alpha = 0.01, 0.05, \text{ and } 0.10$) and the sample mean and SD of the statistic C are correctly approximated by the multiplier method under both the bivariate Poisson and zero-modified Poisson models. Unlike the bivariate normal model, one cannot apply $\hat{\theta}^{TEST}$ under the bivariate Poisson model since the independence is defined as a limit $\alpha \downarrow 0$ (see Remark 1).

In the final part, we compare the computational time between the proposed and the parametric Bootstrap methods under the bivariate normal and bivariate Poisson models. For a fixed dataset, we calculate the required computational time (in seconds) of the two competing methods using the routine proc.time() in R. As shown in Table 3, the required computational time for the multiplier method is much smaller than that for the parametric Bootstrap method for all entries. In particular, the use of the multiplier method under the bivariate Poisson distribution reduces the computational time by 1000 times.

4.2. Power property

To investigate the power of the proposed test, we generated data from the bivariate t -distribution (Lang et al., 1989) while we performed the goodness-of-fit test under the null hypothesis of the bivariate normal distribution. The mean and scale matrix of the bivariate t -distribution are chosen to be the same as the mean and covariance matrix of the bivariate normal model in Section 4.1. As the degree of freedom parameter ν is related to the discrepancy from the null hypothesis, we draw empirical power curves based on 1000 repetitions for selected values of ν . Fig. 1 shows the empirical power curves under $\rho_{X^0Y^0} = -0.35$ with $n = 400$ and 800. The empirical power increases as $1/\nu$ gets larger and it gets close to the

Table 2

Performances of the proposed goodness-of-fit test at the 100α% level based on 1000 replications under the bivariate Poisson model and zero-modified Poisson model with $n = 400$.

Bivariate Poisson	$\lambda_L = 1, \lambda_X = 1$		$\lambda_L = 1, \lambda_X = 2$	
	$\rho_{X^0Y^0} = 0.70$	$\rho_{X^0Y^0} = 0.35$	$\rho_{X^0Y^0} = 0.70$	$\rho_{X^0Y^0} = 0.35$
Rejection rate for $\alpha = 0.10$	0.105	0.086	0.092	0.103
Rejection rate for $\alpha = 0.05$	0.046	0.049	0.046	0.059
Rejection rate for $\alpha = 0.01$	0.007	0.012	0.008	0.012
Mean of C	0.0606	0.0546	0.0718	0.0684
Mean of $E_{data}(C^{(b)})$	0.0613	0.0636	0.0712	0.0695
SD of C	0.0422	0.0392	0.0411	0.0450
Mean of $SD_{data}(C^{(b)})$	0.0421	0.0517	0.0404	0.0431
Zero-modified Poisson	$\lambda_X = 1$		$\lambda_X = 2$	
	$p = 0.3$	$p = 0.7$	$p = 0.3$	$p = 0.7$
Rejection rate for $\alpha = 0.10$	0.103	0.104	0.090	0.069
Rejection rate for $\alpha = 0.05$	0.053	0.056	0.046	0.034
Rejection rate for $\alpha = 0.01$	0.010	0.010	0.006	0.008
Mean of C	0.0620	0.0524	0.0686	0.0616
Mean of $E_{data}(C^{(b)})$	0.0614	0.0522	0.0691	0.0704
SD of C	0.0563	0.0463	0.0496	0.0484
Mean of $SD_{data}(C^{(b)})$	0.0551	0.0452	0.0503	0.0501

Note: The mean and SD of the Cramér–von Mises test statistics C are calculated. $E_{data}(C^{(b)})$ and $SD_{data}(C^{(b)})$ are the mean and SD of $\{C^{(b)}; b = 1, 2, \dots, 1000\}$, respectively, based on 1000 multiplier realizations.

Table 3

A comparison of the multiplier method and parametric Bootstrap with $n = 400$.

	Elapsed time (in seconds)		Resampling mean and SD (in parenthesis)	
	Multiplier	Parametric Bootstrap	Multiplier	Parametric Bootstrap
Bivariate normal				
$\rho_{X^0Y^0} = 0.70$	23.17	1402.31	0.0621 (0.0314)	0.0603 (0.0304)
$\rho_{X^0Y^0} = 0.35$	22.84	1463.01	0.0606 (0.0269)	0.0581 (0.0252)
$\rho_{X^0Y^0} = 0.00$	23.04	1539.29	0.0597 (0.0219)	0.0586 (0.0225)
$\rho_{X^0Y^0} = -0.35$	23.37	1380.09	0.0598 (0.0234)	0.0577 (0.0209)
$\rho_{X^0Y^0} = -0.70$	26.77	1296.32	0.0591 (0.0225)	0.0569 (0.0217)
Bivariate Poisson				
$\lambda_L = 1, \lambda_X = 1$				
$\rho_{X^0Y^0} = 0.70$	12.48	10563.95	0.0608 (0.0413)	0.0614 (0.0426)
$\rho_{X^0Y^0} = 0.35$	12.40	12265.27	0.0505 (0.0353)	0.0543 (0.0404)
$\lambda_L = 1, \lambda_X = 2$				
$\rho_{X^0Y^0} = 0.70$	12.41	10632.34	0.0703 (0.0381)	0.0703 (0.0380)
$\rho_{X^0Y^0} = 0.35$	12.62	13138.48	0.0676 (0.0395)	0.0692 (0.0413)

Note: Elapsed time is calculated by `proc.time()` routine in R. The number of resamplings is 1000 for both multiplier and parametric Bootstrap methods.

nominal sizes ($\alpha = 0.10$ or 0.05) as $1/\nu \rightarrow 0$. The curves for $n = 800$ dominate those for $n = 400$. The empirical power curves for $\rho_{X^0Y^0} = 0.35$ and 0.00 , not shown here, are very similar to those in Fig. 1. In general, the value of $\rho_{X^0Y^0}$ does not have much influence on the power properties.

5. Data analysis

The proposed method is illustrated using the law school data available in Efron and Tibshirani (1993). The data consist of the average score of the LSAT (the national law test) and average GPA (grade point average) for $N = 82$ American law schools. We denote a pair of LSAT and GPA by $(LSAT_j, GPA_j)$ for $j = 1, 2, \dots, N$. The average scores of the LSAT and GPA for the 82 schools are 597.55 and 3.13, respectively. The correlation coefficient between the LSAT and GPA is 0.76.

We consider a situation that one can only observe a vector $(LSAT_j, GPA_j)$ whose sum of LSAT and $100 \times$ GPA are above a threshold $600 + 100 \times 3.0 = 900$. The number of such samples is $n = 49$, and the inclusion rate is therefore $n/N = 0.598$. In this sampling design, we observe $\{(L_j, X_j); j = 1, 2, \dots, 49\}$, subject to $L_j \leq X_j$, where $L_j = 900 - 100 \times GPA_j$ and $X_j = LSAT_j$. Under the (working) independence assumption between LSAT and GPA, the Lynden-Bell's nonparametric estimator (Lynden-Bell, 1971) for the mean LSAT score is defined by

$$\hat{\mu}_X^{NP} = \int_{-\infty}^{\infty} x d\hat{F}_X(x),$$

where

$$\hat{F}_X(x) = 1 - \prod_{u \leq x} \left\{ 1 - \frac{\sum_j \mathbf{I}(L_j \leq u, X_j = u)}{\sum_j \mathbf{I}(L_j \leq u \leq X_j)} \right\}.$$

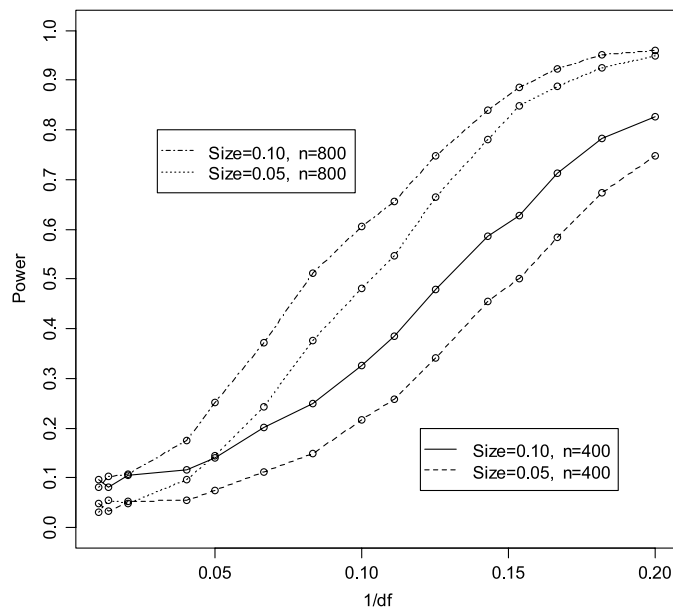


Fig. 1. Empirical power curves for the goodness-of-fit test under the null hypothesis of bivariate normal distribution when data are generated from the bivariate t -distribution with degree of freedom ν (denoted as df). Two sizes ($\alpha = 0.10$ and 0.05) are chosen.

The resulting estimate is $\hat{\mu}_X^{NP} = 618.63$, which seems somewhat larger than the average LSAT score 597.55. This bias may result from the wrong independence assumption between LSAT and GPA.

We fit the bivariate normal distribution models for the truncated data. The p -values of the goodness-of-fit test for the bivariate normality are 0.884 for the multiplier method (required computational time = 1.25 seconds) and 0.645 for the parametric Bootstrap method (required computational time = 222.46 seconds). Both methods reach the same conclusion that there is no evidence to reject the bivariate normality assumption.

We proceed to the parametric analysis under bivariate normality. The estimated population mean of LSAT is $\hat{\mu}_X = 591.32$. This is much closer to the average LSAT score 597.55 than the estimate based on Lynden-Bell's method. The estimated inclusion probability $\hat{c}(\theta) = 0.545$ is also close to the inclusion rate $n/N = 0.598$. Our study shows that the parametric analysis of the dependently truncated data produces reliable results when the goodness-of-fit test favors the fitted model. Note that all the analysis results are easily reproduced by the R codes in the supplementary material.

A reviewer has pointed out the difference of the p -values between the two methods. This is explained by the right skewness of the resampling distribution for the multiplier method compared to that for the parametric Bootstrap (Fig. 2). The difference of the two resampling distributions may be attributed to a slight departure of the data generating system for the LSAT and GPA values from the bivariate normal model. In particular, a few ties in the LSAT and GPA values indicate that the data do not exactly follow the bivariate normal model. This implies that the two resampling procedures can yield different powers in real applications.

6. Conclusion and discussion

The main objective of the present paper is to develop a new goodness-of-fit procedure for parametric models based on dependently truncated data. The method utilizes the multiplier central limit theorem and has the advantage of being less computationally demanding than the parametric Bootstrap procedure by avoiding the complicated resampling scheme of the parametric Bootstrap under dependent truncation. Note that the method is easily implemented by the R codes available in the supplementary material.

Although many studies have already applied the multiplier method in many different contexts (Spiekerman and Lin, 1998; Jin et al., 2003; Bücher and Dette, 2010; Kojadinovic and Yan, 2011), the computational advantage for dependent truncation models is remarkable. The idea can be applied to a variety of problems in which the Bootstrap resampling involves truncation. In particular, generalizations of the proposed method to even more complicated truncation schemes, such as double truncation and multivariate truncation, deserves further investigation.

Acknowledgments

The authors would like to thank the associate editor and the two reviewers for helpful comments. The research was financially supported by the grant NSC100-2118-M008-006 (the first author) and by the Japan Society for the Promotion of Science through Grants-in-Aid for Scientific Research (C)(No. 21500283) (the second author).

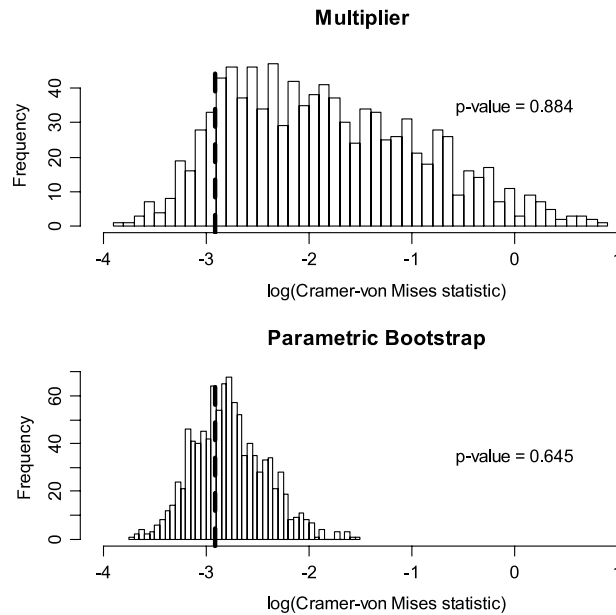


Fig. 2. Two histograms for the resampling distributions based on the multiplier and parametric Bootstrap methods based on 1000 resamplings. The dotted line signifies the value of the Cramér–von Mises type statistic in “log” scale.

Appendix A. Examples of parametric models

Example 1. Bivariate t - and normal distributions.

Let ν be a positive number and

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_L \\ \mu_X \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_L^2 & \sigma_{LX} \\ \sigma_{LX} & \sigma_X^2 \end{bmatrix},$$

where the positive definite matrix $\boldsymbol{\Sigma}$ is called scale matrix. The density of the bivariate t -distribution is given by

$$f_{\boldsymbol{\theta}}(l, x) = \frac{(\sigma_L^2 \sigma_X^2 - \sigma_{LX}^2)^{-1/2} \Gamma\{(v+2)/2\}}{\pi \Gamma(v/2) \nu} \left\{ 1 + \frac{Q^2(l, x | \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\nu} \right\}^{-(v+2)/2},$$

where Γ is the gamma function,

$$Q^2(l, x | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\sigma_L^2 \sigma_X^2}{\sigma_L^2 \sigma_X^2 - \sigma_{LX}^2} \left\{ \left(\frac{l - \mu_L}{\sigma_L} \right)^2 - 2\sigma_{LX} \frac{(l - \mu_L)(x - \mu_X)}{\sigma_L^2 \sigma_X^2} + \left(\frac{x - \mu_X}{\sigma_X} \right)^2 \right\},$$

and $\boldsymbol{\theta}' = (\boldsymbol{\mu}', \sigma_L^2, \sigma_X^2, \sigma_{LX}, \nu)$ are unknown parameters. Using the fact that $L^0 - X^0$ is also a t -distribution (Fang et al., 1990), one can derive

$$c(\boldsymbol{\theta}) = \Pr(L^0 \leq X^0) = \Psi \left(\frac{\mu_X - \mu_L}{\sqrt{\sigma_X^2 + \sigma_L^2 - 2\sigma_{LX}}}; \nu \right),$$

where $\Psi(\cdot; \nu)$ is the cumulative distribution function for the standard t -distribution with ν degree of freedom. In the special case of $\nu \rightarrow \infty$, $L^0 - X^0$ is a normal distribution, and the inclusion probability is $c(\boldsymbol{\theta}) = \Phi \left(\frac{\mu_X - \mu_L}{\sqrt{\sigma_X^2 + \sigma_L^2 - 2\sigma_{LX}}} \right)$. Tables for $\Psi(\cdot; \nu)$ and $\Phi(\cdot)$ are available in well-known software packages, such as `pt` and `pnorm` routines in R. \square

Example 2. Random effect model

Suppose that $L^0 = U + W$ and $X^0 = V + W$, where U, V and W are independent random variables. In this model, the correlation between L^0 and X^0 can be explained by the common latent variable W , called random effect. The model is particularly convenient in the present context since $c(\boldsymbol{\theta})$ has a simple form

$$c(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} S_V(u) dF_U(u),$$

where $F_U(u) = \Pr(U \leq u)$ and $S_V(v) = \Pr(V \geq v)$. Hence $c(\boldsymbol{\theta})$ is free from the distribution of W . For bivariate discrete distributions, the above integration is replaced by summation. Since the integrant function $S_V(v)$ can be easily calculated by software packages, computation of $c(\boldsymbol{\theta})$ only requires one dimensional numerical integration or summation. For instance, if U, V and W are independent Poisson random variables, whose means are $\boldsymbol{\theta} = (\lambda_L, \lambda_X, \alpha)'$, it leads to a classical bivariate Poisson model (Holgate, 1964). The density function and inclusion probability are given respectively by

$$f_{\boldsymbol{\theta}}(l, x) = e^{-\lambda_L - \lambda_X - \alpha} \sum_{w=0}^l \frac{\lambda_L^{l-w} \lambda_X^{x-w} \alpha^w}{(l-w)!(x-w)!w!},$$

$$c(\boldsymbol{\theta}) = \Pr(L^0 \leq X^0) = e^{-\lambda_L - \lambda_X} \sum_{l=0}^{\infty} \frac{\lambda_L^l}{l!} S_V(l).$$

$S_V(l) = \sum_{x=l}^{\infty} \lambda_X^x / x!$ is calculated from packages, for example, using “ppois” in R. □

Example 3. Zero-modified Poisson model

Suppose that L^0 is a binary variable taking $L^0 = 1$ with probability p and $L^0 = 0$ with probability $1 - p$. Also suppose that X^0 is a Poisson distribution with parameter λ_X , which is independent of L^0 . The joint density function is

$$f_{\boldsymbol{\theta}}(l, x) = p^l (1-p)^{1-l} \frac{(\lambda_X)^x e^{-\lambda_X}}{x!}, \quad l = 0, 1; x = 0, 1, 2, \dots,$$

where $\boldsymbol{\theta} = (p, \lambda_X)'$, and the inclusion probability is

$$c(\boldsymbol{\theta}) = \Pr(L^0 \leq X^0) = 1 - pe^{-\lambda_X}.$$

The observed distribution of X^0 is,

$$\Pr(X^0 = x | L^0 \leq X^0) = \begin{cases} c(\boldsymbol{\theta})^{-1} \lambda_X^x e^{-\lambda_X} / x! & x = 1, 2, \dots; \\ c(\boldsymbol{\theta})^{-1} (1-p)e^{-\lambda_X} & x = 0. \end{cases}$$

Re-parameterizing by $p^* = (1 - pe^{-\lambda_X})^{-1} \geq 1$, the above distribution is the zero-modified Poisson distribution with the zero-modification parameter p^* (Dietz and Böhning, 2000). A special case of $p = 1$ corresponds to the zero-truncated Poisson distribution.

Appendix B. Regularity conditions

By modifying the regularity conditions (e.g., p. 257 of Knight, 2000) with the truncated data, the following conditions are sufficient for verifying the asymptotic results:

- (R1) The parameter space Θ is an open subset of \mathbb{R}^p .
- (R2) The set $A = \{(l, x); f_{\boldsymbol{\theta}}(l, x) > 0, l \leq x\}$ does not depend on $\boldsymbol{\theta}$.
- (R3) For some $\varepsilon > 0$ and for every $\boldsymbol{\theta} \in \Theta$, it holds that $c(\boldsymbol{\theta}) = \Pr(L^0 \leq X^0) \geq \varepsilon$.
- (R4) $\boldsymbol{\theta} \mapsto f_{\boldsymbol{\theta}}(l, x)$ is three times continuously differentiable (all of the third partial derivatives exist and they are continuous) for every $l \leq x$. Also, $c(\boldsymbol{\theta})$ is three times continuously differentiable.
- (R5) $E \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log \frac{f_{\boldsymbol{\theta}}(L^0, X^0)}{c(\boldsymbol{\theta})} \middle| L^0 \leq X^0 \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$.
- (R6) $I(\boldsymbol{\theta}) = -E \left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log \frac{f_{\boldsymbol{\theta}}(L^0, X^0)}{c(\boldsymbol{\theta})} \middle| L^0 \leq X^0 \right]$ exists and is positive definite.
- (R7) For some $M_{ijk, \boldsymbol{\theta}}(l, x)$ satisfying $E[M_{ijk, \boldsymbol{\theta}}(L^0, X^0) | L^0 \leq X^0] < \infty$,

$$|\partial^3 f_{\boldsymbol{\theta}}(l, x) / \partial \theta_i \partial \theta_j \partial \theta_k| \leq M_{ijk, \boldsymbol{\theta}}(l, x).$$
- (R8) A map $F_{\boldsymbol{\theta}} : \Theta \mapsto D\{(-\infty, \infty)^2\}$ is Fréchet differentiable. That is,

$$\|F_{\boldsymbol{\theta}+\mathbf{h}} - F_{\boldsymbol{\theta}} - \mathbf{h}'\dot{F}_{\boldsymbol{\theta}}\|_{\infty} = o(\|\mathbf{h}\|) \quad \text{as } \mathbf{h} \rightarrow \mathbf{0}.$$

Remarks on condition (R8): The Fréchet differentiability of $F_{\boldsymbol{\theta}}$ requires that $\dot{F}_{\boldsymbol{\theta}}$ is uniformly bounded on $(-\infty, \infty)^2$. Under the bivariate normal model, partial derivatives $\dot{F}_{\boldsymbol{\theta}}$ have explicit bounds (given in Section 4 of the supplementary material).

Appendix C. Proof of Theorem 1

Under conditions (R3) and (R4), all the partial derivatives of $1/c(\boldsymbol{\theta})$ exist and they are continuous. This implies that $1/c(\boldsymbol{\theta})$ is differentiable in the sense that for $\mathbf{h} \in \mathbf{R}^p$,

$$\left| \frac{1}{c(\boldsymbol{\theta} + \mathbf{h})} - \frac{1}{c(\boldsymbol{\theta})} + \mathbf{h}' \frac{\dot{c}(\boldsymbol{\theta})}{c(\boldsymbol{\theta})^2} \right| = o(\|\mathbf{h}\|) \quad \text{as } \mathbf{h} \rightarrow \mathbf{0}. \tag{C.1}$$

By straightforward calculations,

$$\begin{aligned} & \|\varphi(\boldsymbol{\theta} + \mathbf{h}) - \varphi(\boldsymbol{\theta}) - \mathbf{h}' \mathbf{g}_\theta\|_\infty \\ &= \left\| \frac{F_{\boldsymbol{\theta}+\mathbf{h}} - F_\theta - \mathbf{h}' \dot{F}_\theta}{c(\boldsymbol{\theta})} + F_\theta \left(\frac{1}{c(\boldsymbol{\theta} + \mathbf{h})} - \frac{1}{c(\boldsymbol{\theta})} + \mathbf{h}' \frac{\dot{c}(\boldsymbol{\theta})}{c(\boldsymbol{\theta})^2} \right) + (F_{\boldsymbol{\theta}+\mathbf{h}} - F_\theta) \left(\frac{1}{c(\boldsymbol{\theta} + \mathbf{h})} - \frac{1}{c(\boldsymbol{\theta})} \right) \right\|_\infty \\ &\leq \frac{1}{\varepsilon} \|F_{\boldsymbol{\theta}+\mathbf{h}} - F_\theta - \mathbf{h}' \dot{F}_\theta\|_\infty + \left| \frac{1}{c(\boldsymbol{\theta} + \mathbf{h})} - \frac{1}{c(\boldsymbol{\theta})} + \mathbf{h}' \frac{\dot{c}(\boldsymbol{\theta})}{c(\boldsymbol{\theta})^2} \right| + \left\| \mathbf{h}' \dot{F}_\theta + o(\|\mathbf{h}\|) \right\| \left(-\mathbf{h}' \frac{\dot{c}(\boldsymbol{\theta})}{c(\boldsymbol{\theta})^2} + o(\|\mathbf{h}\|) \right) \Big|_\infty. \end{aligned}$$

Under (R8) and by (C.1), the last three terms have the order $o(\|\mathbf{h}\|)$. Therefore,

$$\|\varphi(\boldsymbol{\theta} + \mathbf{h}) - \varphi(\boldsymbol{\theta}) - \mathbf{h}' \mathbf{g}_\theta\|_\infty = o(\|\mathbf{h}\|) \quad \text{as } \mathbf{h} \rightarrow \mathbf{0},$$

which leads to the Fréchet differentiability of φ . The Hadamard differentiability of φ immediately follows since it is a weaker condition than the Fréchet differentiability. In fact, the Fréchet differentiability and Hadamard differentiability are the same condition since the domain of φ (i.e., $\Theta \subset \mathbf{R}^p$) is finite dimensional (see Example 3.9.2 of van der Vaart and Wellner, 1996). Applying the functional delta method (p. 297 of van der Vaart, 1998) to the result of Lemma 1, we have

$$\sqrt{n}\{F_{\hat{\boldsymbol{\theta}}}(l, x)/c(\hat{\boldsymbol{\theta}}) - F_\theta(l, x)/c(\boldsymbol{\theta})\} = \frac{1}{\sqrt{n}} \sum_j \mathbf{g}'_\theta(l, x) I^{-1}(\boldsymbol{\theta}) \dot{l}_j(\boldsymbol{\theta}) + o_p(1),$$

where $o_p(1)$ is uniform in (l, x) . Therefore,

$$\begin{aligned} \sqrt{n}\{\hat{F}(l, x) - F_{\hat{\boldsymbol{\theta}}}(l, x)/c(\hat{\boldsymbol{\theta}})\} &= \sqrt{n}\{\hat{F}(l, x) - F_\theta(l, x)/c(\boldsymbol{\theta})\} - \sqrt{n}\{F_{\hat{\boldsymbol{\theta}}}(l, x)/c(\hat{\boldsymbol{\theta}}) - F_\theta(l, x)/c(\boldsymbol{\theta})\} \\ &= \frac{1}{\sqrt{n}} \sum_j \{\mathbf{I}(L_j \leq l, X_j \leq x) - F_\theta(l, x)/c(\boldsymbol{\theta})\} - \frac{1}{\sqrt{n}} \sum_j \mathbf{g}'_\theta(l, x) I^{-1}(\boldsymbol{\theta}) \dot{l}_j(\boldsymbol{\theta}) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_j V_j(l, x; \boldsymbol{\theta}) + o_p(1). \quad \square \end{aligned}$$

Appendix D. Proof of Lemma 2

Let $\hat{\mathbf{W}}_j(l, x; \boldsymbol{\theta}) = \partial \hat{V}_j(l, x; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$. By a Taylor expansion,

$$\hat{V}_j(l, x; \hat{\boldsymbol{\theta}}) - \hat{V}_j(l, x; \boldsymbol{\theta}) = \hat{\mathbf{W}}_j'(l, x; \boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2).$$

Also, by the weak law of large numbers,

$$\hat{V}_j(l, x; \boldsymbol{\theta}) - V_j(l, x; \boldsymbol{\theta}) = -\mathbf{g}'_\theta(l, x) [\{i_n(\boldsymbol{\theta})/n\}^{-1} - I(\boldsymbol{\theta})] \dot{l}_j(\boldsymbol{\theta}) = o_p(1),$$

where $o_p(1)$ is uniform in (l, x) . The preceding two equations lead to

$$\begin{aligned} \left\| \frac{1}{\sqrt{n}} \sum_j Z_j \{\hat{V}_j(l, x; \hat{\boldsymbol{\theta}}) - V_j(l, x; \boldsymbol{\theta})\} \right\|_\infty &= \left\| \frac{1}{\sqrt{n}} \sum_j Z_j \{\hat{\mathbf{W}}_j'(l, x; \boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|) + o_p(1)\} \right\|_\infty \\ &= \left\| (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \frac{1}{\sqrt{n}} \sum_j Z_j \hat{\mathbf{W}}_j(l, x; \boldsymbol{\theta}) \right\|_\infty + o_p(1) = \left\| (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' O_p(1) \right\|_\infty + o_p(1) = o_p(1). \quad \square \end{aligned}$$

Appendix E. Proof of Theorem 2

By Lemmas 1 and 2, $G_n, G_n^{(1)}, \dots, G_n^{(B)}$ are asymptotically uncorrelated. Therefore, to show the joint convergence $(G_n, G_n^{(1)}, \dots, G_n^{(B)}) \rightarrow (G_\theta, G_\theta^{(1)}, \dots, G_\theta^{(B)})$ in $D\{(-\infty, \infty)^2\}^{\otimes(B+1)}$, we only need to show $G_n \rightarrow G_\theta$ and $G_n^{(b)} \rightarrow G_\theta^{(b)}$ in $D\{(-\infty, \infty)^2\}$ for each $b \in \{1, \dots, B\}$.

To show $G_n \rightarrow G_\theta$, we first check the marginal convergence. By Cramér–Wold’s device, this can be proven by, for $a_1, \dots, a_m \in R$ and $(l_1, x_1), \dots, (l_m, x_m)$,

$$\sum_{k=1}^m a_k G_n(l_k, x_k) \rightarrow \sum_{k=1}^m a_k G_\theta(l_k, x_k) \tag{E.1}$$

in distribution. By Theorem 1 and the central limit theorem,

$$\sum_{k=1}^m a_k G_n(l_k, x_k) = \frac{1}{\sqrt{n}} \sum_j \sum_{k=1}^m a_k V_j(l_k, x_k; \theta) + o_p(1) \rightarrow N(0, \Omega(\theta))$$

in distribution, where

$$\begin{aligned} \Omega(\theta) &= \text{Var} \left\{ \sum_{k=1}^m a_k V_j(l_k, x_k; \theta) \right\} = \sum_{k=1}^m \sum_{k'=1}^m a_k a_{k'} E\{V_1(l_k, x_k; \theta) V_1(l_{k'}, x_{k'}; \theta)\} \\ &= \sum_{k=1}^m \sum_{k'=1}^m a_k a_{k'} \text{Cov}\{G_\theta(l_k, x_k), G_\theta(l_{k'}, x_{k'})\}. \end{aligned}$$

By the definition of G_θ , $\sum_k a_k G_\theta(l_k, x_k)$ is a zero mean normal distribution with variance $\Omega(\theta)$. Hence, (E.1) is proven. Next, the tightness of $G_n = \sqrt{n}\{\hat{F} - F_\theta/c(\hat{\theta})\}$ is verified by showing the tightness of $\sqrt{n}\{\hat{F} - F_\theta/c(\hat{\theta})\}$ and $\sqrt{n}\{F_\theta/c(\hat{\theta}) - F_\theta/c(\theta)\}$, respectively. The tightness of the former is deduced from the empirical process theory for the multi-dimensional distribution function (Example 2.5.4 of van der Vaart and Wellner, 1996). The tightness of the latter follows from the functional delta method applied for the Hadamard differentiable map $\varphi(\theta) = F_\theta/c(\theta)$ (Appendix C). Therefore, $G_n \rightarrow G_\theta$ in $D\{(-\infty, \infty)^2\}$ is proven.

Similarly, to show $G_n^{(b)} \rightarrow G_\theta$ in $D\{(-\infty, \infty)^2\}$, we first check

$$\sum_{k=1}^m a_k G_n^{(b)}(l_k, x_k) \rightarrow \sum_{k=1}^m a_k G_\theta^{(b)}(l_k, x_k) \tag{E.2}$$

in distribution. By Lemma 2,

$$\sum_{k=1}^m a_k G_n^{(b)}(l_k, x_k) = \frac{1}{\sqrt{n}} \sum_j Z_j^{(b)} \sum_{k=1}^m a_k V_j(l_k, x_k; \theta) + o_p(1) \rightarrow N(0, \Omega(\theta)),$$

where

$$\begin{aligned} \Omega(\theta) &= \text{Var} \left[Z_1^{(b)} \left\{ \sum_{k=1}^m a_k V_j(l_k, x_k; \theta) \right\} \right] = E \left\{ \sum_{k=1}^m a_k V_j(l_k, x_k; \theta) \right\}^2 \\ &= \sum_{k=1}^m \sum_{k'=1}^m a_k a_{k'} E\{V_1(l_k, x_k; \theta) V_1(l_{k'}, x_{k'}; \theta)\} = \sum_{k=1}^m \sum_{k'=1}^m a_k a_{k'} \text{Cov}\{G_\theta(l_k, x_k), G_\theta(l_{k'}, x_{k'})\}. \end{aligned}$$

Hence, (E.2) holds. The tightness of $G_n^{(b)}$ is proven by the same approach as the tightness proof for $G_n \rightarrow G_\theta$. \square

Appendix F. Proof of Lemma 3

It is enough to show that

$$\sqrt{n}(\hat{\theta}^{TEST} - \hat{\theta}) = o_p(1). \tag{F.1}$$

By definition,

$$\sqrt{n}(\hat{\theta}^{TEST} - \hat{\theta}) = \sqrt{n}\{(\hat{\theta}^*, 0) - \hat{\theta}\} \mathbf{I}\{2[l(\hat{\theta}) - l(\hat{\theta}^*, 0)] \leq q_n\}.$$

Under (R1) through (R7), the likelihood ratio test is consistent under $H_1 : \rho \neq 0$. Hence,

$$\lim_{n \rightarrow \infty} \Pr\{2[l(\hat{\theta}) - l(\hat{\theta}^*, 0)] \leq q\} \leq G_{\chi_1}(q), \tag{F.2}$$

where $G_{\chi_1}(\cdot)$ is the percentile of the chi-squared distribution with one degree of freedom. First, we prove (F.1) when $q_n \rightarrow 0$. For any $\varepsilon > 0$, (F.2) and $q_n \rightarrow 0$ together imply

$$\Pr\left(\left\| \sqrt{n}(\hat{\theta}^{TEST} - \hat{\theta}) \right\| > \varepsilon\right) \leq \Pr\{2[l(\hat{\theta}) - l(\hat{\theta}^*, 0)] \leq q_n\} \rightarrow 0 \quad (n \rightarrow \infty),$$

which proves (F.1). Next, we prove (F.1) when $q_n = q > 0$ for all n and $\rho \neq 0$. Under (R1) through (R7), the likelihood ratio test is consistent under $H_1 : \rho \neq 0$ so that

$$\Pr\left(\left\|\sqrt{n}(\hat{\boldsymbol{\theta}}^{TEST} - \hat{\boldsymbol{\theta}})\right\| > \varepsilon\right) \leq \Pr(2[l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}^*, 0)] \leq q) \rightarrow 0 \quad (n \rightarrow \infty).$$

which proves (F.1).

Appendix G. Supplementary data

Supplementary material related to this article can be found online at [doi:10.1016/j.csda.2011.12.022](https://doi.org/10.1016/j.csda.2011.12.022).

References

- Bücher, A., Dette, H., 2010. A note on bootstrap approximations for the empirical copula process. *Statistics & Probability Letters* 80, 1925–1932.
- Chen, C.-H., Tsai, W.-Y., Chao, W.-H., 1996. The product-moment correlation coefficient and linear regression for truncated data. *Journal of American Statistical Association* 91, 1181–1186.
- Cohen, A.C., 1991. *Truncated and Censored Samples: Theory and Applications*. Marcel Dekker, Inc, New York.
- Dietz, E., Böhning, D., 2000. On estimation of the Poisson parameter in zero-modified Poisson models. *Computational Statistics & Data Analysis* 34, 441–459.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, London.
- Emura, T., Konno, Y., 2009. Multivariate parametric approaches for dependently left-truncated data. *Technical Reports of Mathematical Sciences, Chiba University* 25, No. 2.
- Emura, T., Konno, Y., 2012. Multivariate normal distribution approaches for dependently truncated data. *Statistical Papers* 53 (1), 133–149.
- Emura, T., Wang, W., 2010. Testing quasi-independence for truncation data. *Journal of Multivariate Analysis* 101, 223–239.
- Emura, T., Wang, W., Hung, H., 2011. Semi-parametric inference for copula models for truncated data. *Statistica Sinica* 21, 349–367.
- Fang, K.-T., Kotz, S., Ng, K.W., 1990. *Symmetric Multivariate and Related Distributions*. Chapman & Hall, New York.
- Gilberet, P., 2010. *Accurate Numerical Derivatives*. R package version 2010.11-1.
- Holgate, 1964. Estimation of the bivariate Poisson distribution. *Biometrika* 51, 241–245.
- Jin, Z., Lin, D.Y., Wei, L.J., 2003. Rank-based inference for the accelerated failure time model. *Biometrika* 90, 341–353.
- Klein, J.P., Moeschberger, M.L., 2003. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York.
- Knight, K., 2000. *Mathematical Statistics*. Chapman & Hall.
- Kojadinovic, I., Yan, J., 2011. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing* 21, 17–30.
- Kojadinovic, I., Yan, J., Holmes, M., 2011. Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica* 21, 841–871.
- Lakhali-Chaieb, L., Rivest, L.-P., Abdous, B., 2006. Estimating survival under a dependent truncation. *Biometrika* 93, 665–669.
- Lang, K.L., Little, J.A., Taylor, J.M.G., 1989. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* 84, 881–896.
- Lehmann, E.L., Casella, G., 1993. *Theory of Point Estimation*, Second ed. Springer-Verlag, New York.
- Lynden-Bell, D., 1971. 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.
- Martin, E.C., Betensky, R.A., 2005. Testing quasi-independence of failure and truncation via conditional Kendall's tau. *Journal of the American Statistical Association* 100, 484–492.
- Ripley, P., Hornik, K., Gebhardt, A., 2011. *Support Functions and Datasets for Venables and Ripley's MASS*, R package version 7.3-13.
- 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.
- 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.
- Spiekerman, C.F., Lin, D.Y., 1998. Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* 93, 1164–1175.
- Tsai, W.-Y., 1990. Testing the assumption of independence of truncation time and failure time. *Biometrika* 77, 169–177.
- van der Vaart, A.W., 1998. *Asymptotic Statistics*. In: *Cambridge Series in Statistics and Probabilistic Mathematics*, vol. 3. Cambridge University Press, Cambridge.
- van der Vaart, A.W., Wellner, J.A., 1996. *Weak Convergence and Empirical Process*. In: *Springer Series in Statistics*. Springer-Verlag, New York.
- Waiker, V.B., Schuurmann, F.J., Raghunathan, T.E., 1984. On a two-stage shrinkage estimator of the mean of a normal distribution. *Communications in Statistics – Theory and Methods* 13, 1901–1913.
- Wang, M.C., Jewell, N.P., Tsai, W.-Y., 1986. Asymptotic properties of the product-limit estimate and right censored data. *Annals of Statistics* 13, 1597–1605.
- Woodroffe, M., 1985. Estimating a distribution function with truncated data. *Annals of Statistics* 13, 163–177.