Stat Papers (2012) 53:133–149
DOI 10.1007/s00362-010-0321-x

REGULAR ARTICLE

Multivariate normal distribution approaches for dependently truncated data

Takeshi Emura · Yoshihiko Konno

Received: 28 May 2009 / Revised: 17 March 2010 / Published online: 13 April 2010

Abstract Many statistical methods for truncated data rely on the independence assumption regarding the truncation variable. In many application studies, however, the dependence between a variable X of interest and its truncation variable L plays a fundamental role in modeling data structure. For truncated data, typical interest is in estimating the marginal distributions of (L, X) and often in examining the degree of the dependence between X and L. To relax the independence assumption, we present a method of fitting a parametric model on (L, X), which can easily incorporate the dependence structure on the truncation mechanisms. Focusing on a specific example for the bivariate normal distribution, the score equations and Fisher information matrix are provided. A robust procedure based on the bivariate t-distribution is also considered. Simulations are performed to examine finite-sample performances of the proposed method. Extension of the proposed method to doubly truncated data is briefly discussed.

Keywords Correlation coefficient · Truncation · Maximum likelihood · Missing data · Multivariate analysis · Parametric bootstrap

Mathematics Subject Classification (2000) 62F10 · 62H12 · 62N01 · 62N02

T. Emura
Institute of Statistics, National Chiao-Tung University,
Hsin-Chu, Taiwan, ROC

Y. Konno (✉)
Japan Women's University, Tokyo, Japan
e-mail: konno@fc.jwu.ac.jp

1 Introduction

Ever since the pioneering work of Cohen (1959, 1961), statistical analysis of randomly truncated data has been an important topic in both applied and theoretical statistics. Randomly truncated data is commonly seen in studies of education, epidemiology, astronomy and engineering. As an instance, in analysis of test scores in educational research, only observations with the scores above a threshold may appear in the sample. Such sampling scheme occurs when the variables of interest can be observed if their values satisfy a certain inclusion criterion. These samples that do not satisfy the inclusion criterion can never been observed and even their existence is unknown. In this sense, truncated data are fundamentally different from missing data where the indicator variables for missing subjects are typically available.

A parametric approach for truncation data under the normal distribution is well known (Cohen 1959, 1961; Hansen and Zeger 1980). They considered the case where a variable X^O of interest can be included in the sample if it exceeds a deterministic value l. Assuming that the value l is known, they formulated the problem of estimating parameters that determine the distribution of X^O. They presented the maximum likelihood estimator (MLE) based on the conditional density of X^O given $X^O \geq l$ under the normal distribution. They also studied some asymptotic properties of the MLE, including an explicit formula of the Fisher information matrix.

In many application studies, a variable of interest is truncated by another random truncation variable. In left-truncated data, a variable X^O can be included in the sample only if it exceeds another random variable L^O. Such type of data is commonly seen in medical and astronomical research (Klein and Moeschberger 2003). In left-truncated data, the estimation for the distribution function $F_{X^O}(x) = \Pr(X^O \leq x)$ or some parameter $\boldsymbol{\theta} = \boldsymbol{\theta}(F_{X^O})$ is of primary interest, and the distribution for L^O is considered nuisance. Conversely, left-truncated data can be regarded as right-truncated data if L^O is the variable of interest and X^O is truncation variable.

Construction of a nonparametric estimator for $F_{X^O}(x)$ is first proposed by Lynden-Bell (1971) under left-truncation. Asymptotic properties of Lynden-Bell's estimator have been extensively studied by Woodroofe (1985), Wang et al. (1986) and He and Yang (1998) to name but a few. It is well known that the independence between L^O and X^O is a fundamental assumption for the consistency of Lynden-Bell's estimator. Tsai (1990), Martin and Betensky (2005), Chen et al. (1996) and Emura and Wang (2010) present methods for testing the independence assumption. For positive random variables L^O and X^O, Lakhal-Chaieb et al. (2006) proposed to model the dependency between L^O and X^O using copula models.

Compared with the nonparametric and semiparametric inference, there is not much in literature on the analysis of truncated data based on parametric modeling when L^O is considered random. Although Lynden-Bell's nonparametric approach has validity under any form of the underlying distribution, it has some unfortunate disadvantage; it relies on the assumption that L^O is independent of X^O. This assumption may be doubtful in many practical examples (Tsai 1990; Martin and Betensky 2005; Emura and Wang 2010). A semipararmetric approach proposed by Lakhal-Chaieb et al. (2006) is an alternative in such examples if L^O and X^O are positive random variables and the association between L^O and X^O can be modeled via an Archimedean copula model.

However, a class of Archimedian copula models does not include many important examples such as bivariate normal distribution and bivariate t-distribution.

Truncated data in many practical applications may be suitably modeled by the bivariate normal distribution. For example, the National Center Test for University Admissions is a type of standardized test used by public and some private universities in Japan. Many universities have their own cut-off scores, for example, by 120 points in the sum of Japanese (X^O) and English (Y^O) for acceptance of students with majors in the school of humanities. Here, one can define a left-truncation variable $L^O = 120 - Y^O$ so that the acceptance criteria for the university can be written as $L^O \leq X^O$. In this example, it is natural to assume that L^O and X^O may be negatively correlated through unobserved factors related to students' intellectual ability. Fitting a bivariate normal distribution for L^O and X^O would be a sensible way since it characterizes important quantities, such as the mean scores or the correlation coefficient. Other examples of dependent truncation include the two-stage course placement system discussed in Schiel and Harmston (2000) and Emura and Konno (2009).

In this article, we consider parametric approaches to truncated data, which take into account dependence in the truncation variable. In particular, Sect. 2 presents a framework of likelihood inference under truncation. Section 3 considers a method of fitting the bivariate normal distribution, where the explicit formula of the Fisher information matrix is provided. A goodness-of-fit procedure is also considered. Section 4 presents a robust procedure using the bivariate t-distribution. In Sect. 5, the performance of the present approaches is studied via simulations. Section 6 discusses an extension of the proposed method to doubly truncated data. Section 7 concludes this article.

2 Likelihood inference

2.1 Likelihood construction

Let (L^O, X^O) be a pair of random variables having a density function

$$f_{\boldsymbol{\theta}}(l, x) = \frac{\partial^2}{\partial l \partial x} \Pr(L^O \leq l, X^O \leq x)$$

where $\boldsymbol{\theta}$ is a vector of parameters. In a truncated sample, a pair (L^O, X^O) is not directly observable but it is observed only if $L^O \leq X^O$ holds. That is, X^O is left-truncated by L^O, and L^O is right-truncated by X^O. Assume that $c(\boldsymbol{\theta}) = \Pr(L^O \leq X^O) > 0$ holds for every $\boldsymbol{\theta}$. The joint density of the observed pair (L, X) can be written as $c(\boldsymbol{\theta})^{-1} f_{\boldsymbol{\theta}}(l, x)\mathbf{1}(l \leq x)$ where $\mathbf{1}(l \leq x)$ is the indicator function of the set in brackets. For observed data $\{(L_j, X_j); j = 1, 2, \ldots, n\}$ subject to $L_j \leq X_j$, the likelihood function has the form

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

Due to truncation, the form of $L(\boldsymbol{\theta})$ tends to be complicated. If $\hat{\boldsymbol{\theta}}$ denotes the maximum likelihood estimator (MLE) that maximizes (1), almost all well-known bivariate distributions on (L^O, X^O) do not provide the explicit solution for $\hat{\boldsymbol{\theta}}$. Furthermore, the explicit formula for $c(\boldsymbol{\theta})$ is usually impossible to obtain since it requires some integration on $\{(l, x) : l \leq x\}$. An interesting exception is the Marshall and Olkin (1967) bivariate exponential model specified by

$$\Pr(L^O > l, X^O > x) = \exp\{-\lambda_L l - \lambda_X x - \lambda_{LX} \max(l, x)\} \quad (l \geq 0, x \geq 0),$$

where parameters $\boldsymbol{\theta} = (\lambda_L, \lambda_X, \lambda_{LX})'$ satisfy $\lambda_L > 0, \lambda_X > 0$ and $\lambda_{LX} \geq 0$. In this model, a simple form $c(\boldsymbol{\theta}) = (\lambda_L + \lambda_{LX})(\lambda_L + \lambda_X + \lambda_{LX})^{-1}$ can be obtained. In what follows, we focus our discussions on the bivariate normal distribution on $\left(L^O, X^O\right)$ that has a tractable form in $c(\boldsymbol{\theta})$ and that can be applied to common sampling designs in practical applications. In Sect. 4, we introduce a robust method by the bivariate t-distribution that also has a simple form in $c(\boldsymbol{\theta})$.

2.2 Large sample analysis

The large sample theory for the likelihood procedure can be applied when both $f_{\boldsymbol{\theta}}(l, x)$ and $c(\boldsymbol{\theta})$ are sufficiently smooth in $\boldsymbol{\theta}$. Let $l_1(\boldsymbol{\theta}) = \log\{c(\boldsymbol{\theta})^{-1} f_{\boldsymbol{\theta}}(L_1, X_1)\}$ and $\dot{l}_1(\boldsymbol{\theta}) = \partial l_1(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ be log-likelihood and score functions based on a single observation respectively. Throughout the article, we denote by $N(\boldsymbol{\mu}, \boldsymbol{\Psi})$ a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Psi}$. If the third derivative of $l_1(\boldsymbol{\theta})$ is available and certain boundedness conditions on the third derivative are satisfied, it can be shown that, when n tends to infinity,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to_d N(\mathbf{0}, I(\boldsymbol{\theta})^{-1}),$$

where $I(\boldsymbol{\theta}) = E\{\dot{l}_1(\boldsymbol{\theta})\dot{l}_1(\boldsymbol{\theta})' | L^O \leq X^O\}$ is the Fisher information matrix based on a single observation and the notation "$\to_d$" signifies the convergence in distribution (Knight 2000, p. 115). Except for simple models, such as the one that will be discussed in Sect. 3.2, an explicit formula for $I(\boldsymbol{\theta})$ is generally impossible to obtain due to the distorted likelihood function by truncation. We recommend using the observed Fisher information $\sum_j \dot{l}_j(\hat{\boldsymbol{\theta}})\dot{l}_j(\hat{\boldsymbol{\theta}})'/n$ in constructing the asymptotic approximation for $I(\boldsymbol{\theta})$, which is useful for calculating the standard error of $\hat{\boldsymbol{\theta}}$.

Let $\hat{c} = c(\hat{\boldsymbol{\theta}})$ be an estimate of $c(\boldsymbol{\theta})$. By the invariance property of the MLE, $\hat{c}$ is also the MLE of $c(\boldsymbol{\theta})$. By the delta method (Knight 2000, Theorem 3.4), we obtain the convergence result:

$$\sqrt{n}(\hat{c} - c) \to_d N(0, \dot{c}(\boldsymbol{\theta})' I(\boldsymbol{\theta})^{-1} \dot{c}(\boldsymbol{\theta})),$$

where $\dot{c}(\boldsymbol{\theta}) = \partial c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Thus, the standard error of $\hat{c}$ can be obtained from the asymptotic variance by replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$.

We will explore the implications of the asymptotic results by giving an explicit formula of $I(\boldsymbol{\theta})$ in the subsequent discussions. For a purpose of applying the above

asymptotic results, checking the existence of the third derivatives of $f_{\boldsymbol{\theta}}(l, x)$ and $c(\boldsymbol{\theta})$ will usually suffice. For more rigorous conditions, one can refer conditions (B1)–(B6) in page 257 of Knight (2000).

3 Likelihood inference under normal populations

The multivariate normal distribution is perhaps the most common in modeling multivariate responses in applied statistical analysis. To implement the framework of the likelihood analysis developed in Sect. 2, we present an example with detailed analyses for a bivariate normal distribution on $f_{\boldsymbol{\theta}}(l, x)$.

3.1 Truncation under normal distribution

We assume that pre-truncated variables follow a bivariate normal distribution

$$\begin{pmatrix} L^O \\ X^O \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_L \\ \mu_X \end{bmatrix}, \begin{bmatrix} \sigma_L^2 & \sigma_{LX} = \rho_{LX}\sigma_L\sigma_X \\ \sigma_{LX} = \rho_{LX}\sigma_L\sigma_X & \sigma_X^2 \end{bmatrix} \right). \tag{2}$$

Let $\boldsymbol{\theta}' = \left(\mu_L, \mu_X, \sigma_L^2, \sigma_X^2, \sigma_{LX}\right)$ be unknown parameters, where $\sigma_L^2 > 0, \sigma_X^2 > 0$ and $-1 < \rho_{LX} < 1$. Then, the inclusion probability is given by

$$c(\boldsymbol{\theta}) = \Pr\left(L^O \le X^O\right) = \Phi\left(\frac{\mu_X - \mu_L}{\sqrt{\sigma_X^2 + \sigma_L^2 - 2\sigma_{LX}}}\right), \tag{3}$$

where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. Since $\sigma_X^2 + \sigma_L^2 - 2\sigma_{LX} > 0$ always holds, it follows that $0 < c(\boldsymbol{\theta}) < 1$ for any $\boldsymbol{\theta}$. By combining the likelihood of n truncated samples, we obtain the log-likelihood function:

$$l(\boldsymbol{\theta}) = -n\log(c(\boldsymbol{\theta})) - n\log(2\pi) - \frac{n}{2}\log(\sigma_L^2\sigma_X^2 - \sigma_{LX}^2) - \frac{1}{2}\sum_j D_j^2(\boldsymbol{\theta}), \tag{4}$$

where

$$D_j^2(\boldsymbol{\theta}) = \frac{\sigma_X^2(L_j - \mu_L)^2 - 2\sigma_{LX}(L_j - \mu_L)(X_j - \mu_X) + \sigma_L^2(X_j - \mu_X)^2}{\left(\sigma_L^2\sigma_X^2 - \sigma_{LX}^2\right)}.$$

Let

$$\mathbf{U}_j(\boldsymbol{\theta}) = \frac{1}{\sigma_L^2\sigma_X^2 - \sigma_{LX}^2} \begin{bmatrix} \sigma_X^2(L_j - \mu_L) - \sigma_{LX}(X_j - \mu_X) \\ -\sigma_{LX}(L_j - \mu_L) + \sigma_L^2(X_j - \mu_X) \\ -\sigma_X^2/2 + \sigma_X^2 D_j^2(\boldsymbol{\theta})/2 - (X_j - \mu_X)^2/2 \\ -\sigma_L^2/2 + \sigma_L^2 D_j^2(\boldsymbol{\theta})/2 - (L_j - \mu_L)^2/2 \\ \sigma_{LX} - \sigma_{LX} D_j^2(\boldsymbol{\theta}) + (L_j - \mu_L)(X_j - \mu_X) \end{bmatrix}.$$

Then the score functions can be written as

$$-n\frac{\dot{c}(\boldsymbol{\theta})}{c(\boldsymbol{\theta})} + \sum_i \mathbf{U}_i(\boldsymbol{\theta}) = \mathbf{0}. \tag{5}$$

We recommend finding $\hat{\boldsymbol{\theta}}$ by a Newton-type iteration algorithm in which both (4) and (5) are used. The algorithm can be easily done using "nlm" command in R with the starting value chosen to the solution to (5) under $c(\boldsymbol{\theta}) = 1$. This is very easy to obtain from the sample means and sample covariance matrix.

If one can assume the independence between L^O and X^O, the maximum likelihood estimator for $\boldsymbol{\theta}' = (\mu_L, \mu_X, \sigma_L^2, \sigma_X^2)$ under $\sigma_{LX} = 0$ should be used. In this case, Eq. 5 reduces to

$$-n\frac{\dot{c}^*(\boldsymbol{\theta})}{c^*(\boldsymbol{\theta})} + \sum_i \mathbf{U}_i^*(\boldsymbol{\theta}) = \mathbf{0}, \tag{6}$$

where

$$c^*(\boldsymbol{\theta}) = \Phi\left(\frac{\mu_X - \mu_L}{\sqrt{\sigma_X^2 + \sigma_L^2}}\right), \mathbf{U}_i^*(\boldsymbol{\theta}) = \begin{bmatrix} (L_i - \mu_L)/\sigma_L^2 \\ (X_i - \mu_X)/\sigma_X^2 \\ -\sigma_L^2/2 + (L_i - \mu_L)^2/\left(2\sigma_L^4\right) \\ -\sigma_X^2/2 + (X_i - \mu_X)^2/\left(2\sigma_X^4\right) \end{bmatrix}.$$

Solving (6) still requires some iteration algorithms. The solution obtained by solving (6), denoted by $\hat{\boldsymbol{\theta}}^*$, is consistent for $\boldsymbol{\theta}$ only under the independence assumption. Instead of this unfortunate advantage, $\hat{\boldsymbol{\theta}}^*$ has higher efficiency than that of $\hat{\boldsymbol{\theta}}$ under the independence assumption. In fact, by simulations provided in Sect. 5, the gain in efficiency turns out to be substantial.

To take advantage of the high efficiency of $\hat{\boldsymbol{\theta}}^*$ under the independence as well as to take into account the dependency, it is natural to consider the preliminary test for $H_0 : \sigma_{LX} = 0$ using the likelihood ratio $2[l(\hat{\boldsymbol{\theta}}) - l\{(\hat{\boldsymbol{\theta}}^*, 0)\}]$. If the test rejects $H_0 : \sigma_{LX} = 0$, then we adopt $\hat{\boldsymbol{\theta}}$. Otherwise we accept $H_0 : \sigma_{LX} = 0$ and adopt $\hat{\boldsymbol{\theta}}^*$. This motivates a statistic, similar to the two stage shrinkage "testimator" (Waiker et al. 1984) of the form:

$$\hat{\boldsymbol{\theta}}^{TEST} = \hat{\boldsymbol{\theta}}^{*} + (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{*})\mathbf{I}(2[l(\hat{\boldsymbol{\theta}}) - l\{(\hat{\boldsymbol{\theta}}^{*}, 0)\}] > q) \tag{7}$$

where a constant q may be chosen to be the cutoff value for the chi-square distribution with one degree of freedom. By definition, $\hat{\boldsymbol{\theta}}^{TEST}$ is shrunk to $\hat{\boldsymbol{\theta}}^{*}$, whereby it borrows strength from the small variability of $\hat{\boldsymbol{\theta}}^{*}$. The testimator for the inclusion probability is defined similarly as $\hat{c}^{TEST} = c(\hat{\boldsymbol{\theta}}^{TEST})$.

3.2 The effect of truncation

For truncated data, it is important to investigate the effect of truncation since data are more or less subject to information loss. In this subsection, we derive some useful formulas which systematically describe the impact of truncation on the estimators under the bivariate normal model (2). We consider the case where the covariance matrix for (L^{O}, X^{O}) is known, where the calculations become tractable. It is shown in Appendix A that the Fisher information matrix for estimating the mean parameters $\boldsymbol{\theta}' = (\mu_L, \mu_X)$ is explicitly written as

$$\tilde{I}\{c(\boldsymbol{\theta})\} = \begin{bmatrix} \sigma_L^2 & \sigma_{LX} \\ \sigma_{LX} & \sigma_X^2 \end{bmatrix}^{-1} - \frac{w\{c(\boldsymbol{\theta})\}}{\sigma_L^2 + \sigma_X^2 - 2\sigma_{LX}} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{8}$$

where $w(\cdot) : (0, 1) \to [0, 1]$ is defined to be

$$w(c) = \frac{\Phi^{-1}(c)\phi\{\Phi^{-1}(c)\}}{c} + \frac{\phi\{\Phi^{-1}(c)\}^2}{c^2},$$

where $\phi(x) = \dot{\Phi}(x)$. As shown in Fig. 1, $w(\cdot)$ is a strictly decreasing function of c with the maximum slope of $\dot{w}(1/2) = \sqrt{2/\pi}\,(1 - 4/\pi) < 0$. Also, it follows that $\lim_{c\downarrow 0} w(c) = 1$ and $\lim_{c\uparrow 1} w(c) = 0$. Notice that, given the covariance matrix, Eq. 8 depends on $\boldsymbol{\theta}' = (\mu_L, \mu_X)$ as a function of $c(\boldsymbol{\theta})$. Considering that the first term in (8) is the Fisher information matrix for the complete data, the second term in (8) reflects the loss of information due to truncation. For instance, $w(0.50) = 0.64$ signifies that there is 64% loss of information when the probability of truncation is 50% (Fig. 1). Taking the limit $c(\boldsymbol{\theta}) \uparrow 1$ in (8) corresponds to the complete data and $\lim_{c\uparrow 1} w(c) = 0$ implies that the information loss vanishes. The monotonicity of the information loss can be expressed as the fact that $\tilde{I}\{c\} - \tilde{I}\{c'\}$ is positive semi-definite for $c > c'$. In data analysis, one may use $w(\hat{c})$ as a measure to examine the truncation effect.

Establishing the monotone relationship between c and the Fisher information matrix under more general models is not only a mathematically interesting problem, but also is useful in practical applications. It can be conjectured that the level of c is related to the Fisher information matrix in a similar fashion as (8). Unfortunately, when the covariance matrix is unknown under the bivariate normal distribution, the Fisher information matrix becomes untractable due to the complicated form of the second moments on a truncated pair (L, X). Instead, we will investigate the relationship between c and the

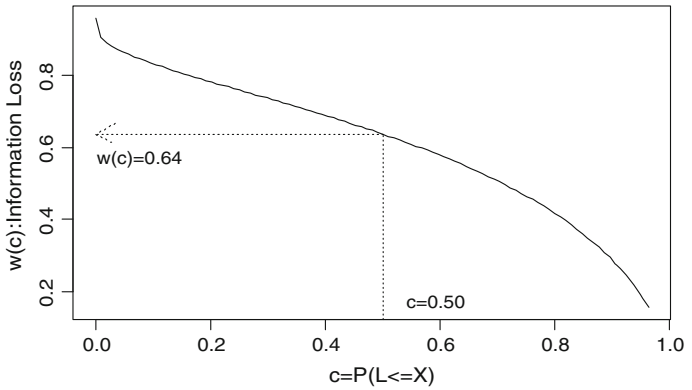

Fig. 1 The graph of the function $w(c) = \frac{\Phi^{-1}(c)\phi\{\Phi^{-1}(c)\}}{c} + \frac{\phi\{\Phi^{-1}(c)\}^2}{c^2}$ that describes the information loss under the inclusion probability c. About 64% information loss is expected when 50% of the population is truncated

accuracy of $\hat{\boldsymbol{\theta}}$ by simulations in Sect. 5.3. We will also verify the correctness of the Fisher information matrix in (8) by the simulations.

3.3 Goodness-of-fit procedure

The proposed method relies on the assumption of bivariate normality. One of the popular classes of goodness-of-fit tests consists of comparing the distance between $\hat{F}(l,x) = \sum_j I(L_j \le l, X_j \le x)/n$ and $F_{\hat{\boldsymbol{\theta}}}(l,x)/c(\hat{\boldsymbol{\theta}})$, where

$$F_{\boldsymbol{\theta}}(l,x) = \iint\limits_{u\le x, u\le v\le x} f_{\boldsymbol{\theta}}(u,v)dudv$$
$$= \int\limits_{-\infty}^{(l-\mu_L)/\sigma_L} \left[\Phi\left\{ \frac{x-\mu_X-\sigma_{LX}s/\sigma_L}{\sqrt{\sigma_X^2-\sigma_{LX}^2/\sigma_L^2}} \right\} - \Phi\left\{ \frac{\mu_L+\sigma_L s-\mu_X-\sigma_{LX}s/\sigma_L}{\sqrt{\sigma_X^2-\sigma_{LX}^2/\sigma_L^2}} \right\} \right] \phi(s)ds$$

is derived for bivariate normality. The Kolmogorov-Smirnov type test is based on

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

The calculation of K requires of the numerical integrations for $F_{\boldsymbol{\theta}}(l,x)$ at $n \sim n^2$ different points in $\{(l,x); l \le x\}$. A computationally simpler alternative is to use the Cramér-von-Mises type statistics

$$C = \iint\limits_{l\le x} \left\{ \hat{F}(l,x) - F_{\hat{\boldsymbol{\theta}}}(l,x)/c(\hat{\boldsymbol{\theta}}) \right\}^2 d\hat{F}(l,x) = \sum_j \left\{ \hat{F}(L_j,X_j) - F_{\hat{\boldsymbol{\theta}}}(L_j,X_j)/c(\hat{\boldsymbol{\theta}}) \right\}^2.$$

This requires exactly n evaluations of the numerical integrations. In either case, the null distributions have not been derived and depend on the true value of $\boldsymbol{\theta}$. To obtain the approximate p-values of the two tests, one can perform the following parametric bootstraps:

Step 1: For some large integer B, generate independently and identically distributed pairs of $\left(L_j^{(b)}, X_j^{(b)}\right), j = 1, 2, \ldots, n, b = 1, 2, \ldots, B$, which follows the truncated distribution $F_{\hat{\boldsymbol{\theta}}}(l, x)/c(\hat{\boldsymbol{\theta}})$.
Step 2: For each $b = 1, 2, \ldots, B$, compute statistics K (or C) using data $\left\{\left(L_j^{(b)}, X_j^{(b)}\right); j = 1, 2, \ldots, n\right\}$, which is denoted by $K^{(b)}$ (or $C^{(b)}$).
Step 3: Approximate p-values for the test is given by

$$B^{-1}\sum_{b=1}^{B}\mathbf{I}\left(K^{(b)} \geq K\right) \quad \left(\text{or } B^{-1}\sum_{b=1}^{B}\mathbf{I}\left(C^{(b)} \geq C\right)\right).$$

In Step 1, using B = 1,000 would usually suffice as suggested by Efron and Tibshirani (1993), which is also checked by subsequent simulations in Sect. 5.4. A pair $\left(L_j^{(b)}, X_j^{(b)}\right)$ in Step 1 may be easily generated as a truncated sample from the pre-truncated distribution of $F_{\hat{\boldsymbol{\theta}}}(l, x)$.

4 Robust modeling

An important issue for fitting the bivariate normal distribution is the robustness properties of the likelihood methods. In general, results from fitting the normal distribution are sensitive to model misspecifications and extreme outliers, and they lead to bias in estimation. A simple way to deal with outliers is to exclude or downweight them in some way. The robust procedures proposed by Huber (1974) involve the modification of the score function by certain weighting techniques. The term of truncation effect $-n\dot{c}(\boldsymbol{\theta})/c(\boldsymbol{\theta})$ in (5) distorts the likelihood equation and makes it difficult to directly apply the weighting scheme. For truncated data, none of the above approaches can be applied straightforwardly.

A relatively simple robust procedure may be obtained by fitting the multivariate t-distribution distributions (Lang et al. 1989) on (L^O, X^O). We replace the model (2) by the bivariate t-distribution:

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

where

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

and $\boldsymbol{\theta}' = \left(\mu_L, \mu_X, \sigma_L^2, \sigma_X^2, \sigma_{LX}, \nu\right)$.

The t-distribution supplies the degree of freedom parameter ν for the robustness. The low values of ν results in the heavy tails while $\nu \to \infty$ corresponds to the bivariate normal distribution. If data are contaminated by a few outliers in the tails, one can expect that the bivariate t-distribution fits better than the bivariate normal distribution due to the flexibility of ν. Using the fact that $L^O - X^O$ is also a t-distribution (Fang et al. 1990), one can derive a simple formula of the inclusion probability:

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

where $\Psi(\cdot; \nu)$ is the cumulative distribution function for the standard t-distribution with ν degree of freedom. Thus, the likelihood inference discussed in Sect. 2 can be applied quite straightforwardly.

Although the degree of freedom parameter provides a tool for achieving the robustness, there exist some types of model deviations which cannot be handled by the t-distribution. Such example includes asymmetric outliers. We leave this as a future topic for investigation.

5 Numerical analysis

Simulation studies are conducted to investigate the performances of the proposed approaches. In Sect. 5.1, we describe the experimental design used in subsequent subsections. Section 5.2 is concerned with the comparison between the proposed parametric approaches and the existing nonparametric approach. Section 5.3 investigates the effect of the inclusion probability on the likelihood estimators, which supplements the theoretical results of Sect. 3.2. Finally, Sect. 5.4 studies the performance of the goodness-of-fit procedure.

5.1 Design

The sum of Japanese and English test scores plays one of the most important factors in the admission to Japanese universities for students with majors in the school of humanities. Based on the record for National Center Test for University for 2008 (http://www.dnc.ac.jp/modules/center_exam/content0196.html), Japanese score (X^O) and English score (Y^O) follows a bivariate normal distribution

$$\begin{pmatrix} X^O \\ Y^O \end{pmatrix} \sim N\left(\begin{bmatrix} 60.82 \\ 62.63 \end{bmatrix}, \begin{bmatrix} 16.81^2 & (19.64)(16.81)\rho_{X^O Y^O} \\ (19.64)(16.81)\rho_{X^O Y^O} & 19.64^2 \end{bmatrix} \right). \quad (9)$$

Here, the mean and standard deviation of X^O are calculated from 481,315 samples and those of Y^O are calculated from 497,101 samples. The correlation between X^O and Y^O is not available publicly and it is denoted as $\rho_{X^O Y^O}$.

Now we consider a hypothetical university having a policy to accept students whose sum of Japanese and English score reaches a pre-specified cut-off point K. Formally, the inclusion criteria is written as

$$X^O + Y^O \geq K$$

One can define $L^O = K - Y^O$ so that the acceptance criteria can be written as $L^O \leq X^O$. Also, the strength of dependency between L^O and X^O is described by the correlation $\rho_{L^O X^O} = -\rho_{X^O Y^O}$. If one is interested in estimating the population mean of the Japanese score, namely $\mu_X = 60.82$, data X^O is said to be left-truncated by L^O. Two scenarios are considered for the model parameters:

(i) Fix the constant at $K = 120$ and move the correlation in the range of $\rho_{X^O Y^O} \in [-0.7, 0.7]$, or equivalently $\rho_{L^O X^O} \in [-0.7, 0.7]$.
(ii) Fix the correlation at $\rho_{X^O Y^O} = 0.5$ and move the cut-off point in the range of $K \in [76.81, 152.51]$ so that $c \in [0.2, 0.9]$.

Scenario (i) is used in Sects. 5.2 and 5.4 while scenario (ii) is used in Sect. 5.3. We choose $n = 400$ since it is one of the common prescribed numbers of students in a department in Japanese universities (years 2003–2009). For instance, the Japan women's university has the annual admission number of 400 in the Faculty of Integrated Arts and Social Sciences. In what follows, we compare the estimators for μ_X based on truncated data $\{(L_j, X_j); j = 1, 2, \ldots, n\}$, which is the scores for the admitted students.

5.2 Comparison with the nonparametric estimator

Under (i), we compare the performances of the proposed parametric estimators with the existing nonparametric estimator. For each run, we computed the four estimators of μ_X, namely $\hat{\mu}_X$, $\hat{\mu}_X^0$, $\hat{\mu}_X^{TEST}$ and $\hat{\mu}_X^{NP}$, defined as follows. The proposed parametric estimator $\hat{\mu}_X$ is calculated based on (5) while $\hat{\mu}_X^0$ is based on (6) under the working independence assumption of $\rho_{L^O X^O} = 0$. The compromise between $\hat{\mu}_X$ and $\hat{\mu}_X^0$ is the $\hat{\mu}_X^{TEST}$ defined in (7), where the cutoff value q is the upper 5% point of the chi-squared distribution with one degree of freedom. The nonparametric estimator is defined by

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

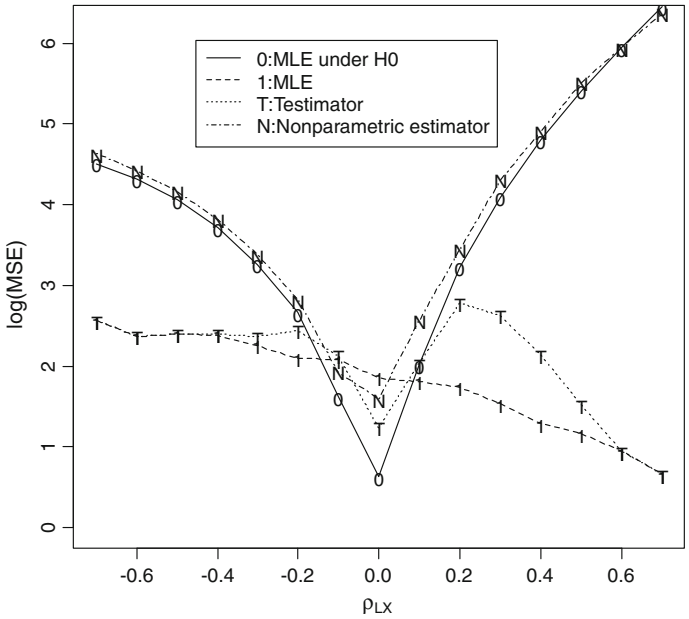

Fig. 2 The mean square errors (MSE) of the four estimators $\hat{\mu}_X$, $\hat{\mu}_X^0$, $\hat{\mu}_X^{TEST}$ and $\hat{\mu}_X^{NP}$ using a log-scale: "0" represents $\hat{\mu}_X^0$ based on (6) under the independent normal distribution of $\rho_{XY} = 0$, "1" represents $\hat{\mu}_X$ calculated based on (5), "T" represents the testimator $\hat{\mu}_X^{TEST}$ defined in (7) and "N" represents the nonparametric estimator $\hat{\mu}_X^0$

where

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

is the Lynden-Bell's nonparametric estimator. Note that $\hat{\mu}_X^{NP}$ is consistent under the independence assumption of $H_0 : L^O \perp X^O$, and that it does not require the normality assumption. Among the three estimators, $\hat{\mu}_X^0$ requires the strongest assumption since it assumes both the normality and independence. The mean squared error (MSE) of $\hat{\mu}_X^0$, $\hat{\mu}_X$, $\hat{\mu}_X^{NP}$ and $\hat{\mu}_X^{TEST}$ was then computed over the 1,000 runs, where the MSE is defined by $\sum_{r=1}^{1000} (\hat{\theta}_r - \mu_X)^2/1000$ for $\hat{\theta}_r$ representing each of the four estimators in rth simulation run.

Figure 2 summarizes the simulation results. In terms of the MSE, the estimator $\hat{\mu}_X$ and $\hat{\mu}_X^{TEST}$ has clear advantage when the dependency between L^O and X^O becomes strong. Especially when $|\rho_{L^O X^O}| \ge 0.2$, the other estimators $\hat{\mu}_X^{NP}$ and $\hat{\mu}_X^0$ that are not adjusted for the dependency leads to very large MSE. On the other hand, $\hat{\mu}_X^{NP}$ and $\hat{\mu}_X^0$ have smaller MSE than that of $\hat{\mu}_X$ and $\hat{\mu}_X^{TEST}$ when L^O and X^O are close to the independence. More specifically, when $\rho_{L^O X^O} = 0$, then the MSE of $\hat{\mu}_X^0$ was 1.867 while that of $\hat{\mu}_X$ was 6.451, leading to substantial difference in efficiency. Notice that

$\hat{\mu}_X^{TEST}$ is a compromise between $\hat{\mu}_X^0$ and $\hat{\mu}_X$. As a result, $\hat{\mu}_X^{TEST}$ takes advantage of the both approaches and provides more stable results than the other three estimators.

In general, we found that the proposed estimators $\hat{\mu}_X$ and $\hat{\mu}_X^{TEST}$ adjust for the dependent truncation in all cases. Instead, due to the additional modeling of the dependency of a truncation variable, $\hat{\mu}_X$ tends to have large variability. If the independence is known a priori, $\hat{\mu}_X^{NP}$, $\hat{\mu}_X^0$, $\hat{\mu}_X^{TEST}$ are recommended to reduce the variability.

5.3 Numerical studies on the effect of truncation

This subsection studies the impact of the inclusion probability on the efficiency of the proposed estimator $\hat{\mu}_X$. If the covariance matrix in (9) is known, then the asymptotic efficiency of $\hat{\mu}_X$ is directly measured by the asymptotic variance:

$$AV_{\mu_X}\{c(\boldsymbol{\theta});\ n\} = \left(\tilde{I}\{c(\boldsymbol{\theta})\}^{-1}/n\right)_{2,2}, \tag{10}$$

where $(\mathbf{B})_{i,j}$ denotes the (i, j) component of a matrix $\mathbf{B}$. Figure 3 depicts the function $AV_{\mu_X}(c; 400)$ calculated under model (9). As expected from (8), $AV_{\mu_X}(c; 400)$ decreases as c increases. To confirm the correctness of $AV_{\mu_X}(c; n)$, we also calculated the MSE of $\hat{\mu}_X$, using 1,000 replications of $\{(L_j, X_j);\ j = 1, 2, \ldots, 400\}$, under the known covariance matrix. As shown in Fig. 3, the MSE of $\hat{\mu}_X$ is very close to the asymptotic variance $AV_{\mu_X}(c; 400)$ for all c examined. This result is natural since the MSE and $AV_{\mu_X}(c; n)$ should be very close for large n from the large sample theory of the MLE.

If the covariance matrix in (9) is unknown, the asymptotic variance for $\hat{\mu}_X$ does not have a closed form expression. We calculated the MSE of $\hat{\mu}_X$ by simulations for selected values of $c(\boldsymbol{\theta})$ as shown in Fig. 3. The MSE decreases as c increases, which follow our conjecture on the impact of the inclusion probability discussed in Sect. 3.2. As expected, the MSE is uniformly larger than that calculated under the known covariance matrix. We also calculated the MSE of $\hat{\mu}_X$, using 1,000 replications of $\{(L_j, X_j);\ j = 1, 2, \ldots, 400\}$ under the bivariate t-distribution with unknown covariance matrix. Again, the MSE gets small when c increases, as in the previous two normal distribution models. The MSE under the bivariate t-distribution is slightly larger than that for the bivariate normal model due to the additional estimation of the degree of freedom parameter.

5.4 Numerical studies on the goodness-of-fit test

The performance of the parametric bootstrap-based goodness-of-fit tests described in Sect. 3.3 was assessed via simulations. The experimental design is based on (ii). Due to its very high computational cost as each parametric bootstrap iteration requires numerical integrations for $F_{\hat{\boldsymbol{\theta}}}(l, x)/c(\hat{\boldsymbol{\theta}})$, we only present the result from Cramér-von-Mises type statistics. For each run, we test the null hypothesis of the bivariate normality based on B = 1,000 bootstrap replications.

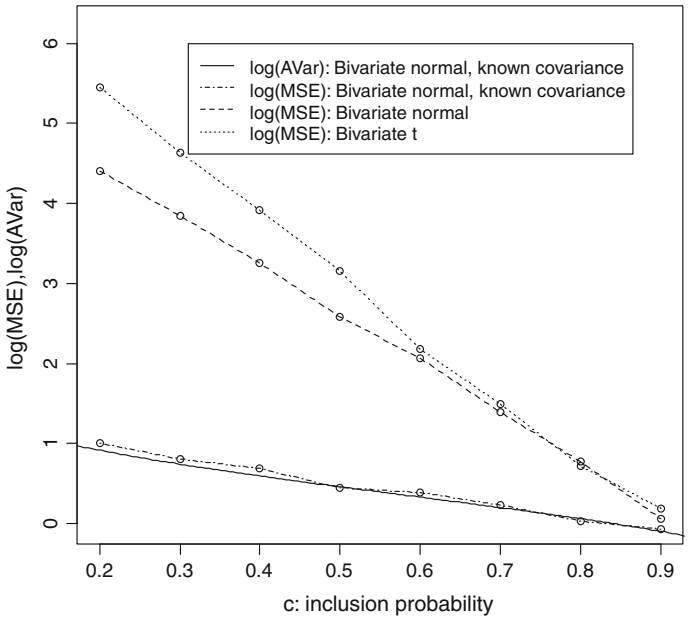

Fig. 3 The log of asymptotic variance, denoted by log(AVar), and the log of mean squared errors, denoted by log(MSE), for estimating μ_X under the three parametric models. The asymptotic variance is calculated by the log of $AV_{\mu_X}(c;n) = \left(\tilde{I}\{c\}^{-1}/n\right)_{2,2}$ in (10). The log of the MSE is defined by $\log\left\{\sum_{r=1}^{1000}\left(\hat{\mu}_X^{(r)} - \mu_X\right)^2/1000\right\}$, where $\hat{\mu}_X^{(r)}$ is the MLE for μ_X based on rth replicated data

Table 1 shows the type I error rates of the test for selected levels of $\alpha = 0.10, 0.05$ and 0.01, which is calculated based on 300 runs. For all the three levels, the type I error rates are in good agreement with their nominal levels. This result is also supported from the empirical suggestion of $B = 1{,}000$ in Efron and Tibshirani (1993). More complete simulations, including numerical studies of the test under alternative hypotheses are beyond the scope of this paper and are left for our future topic for investigation.

Table 1 Type I error rates of the Cramér-von-Mises type goodness-of-fit test at level α based on 300 replications

	ρ_{XOYO}				
	−0.70	−0.35	0.00	0.35	0.70
$\alpha = 0.10$	0.113	0.090	0.107	0.123	0.083
$\alpha = 0.05$	0.056	0.040	0.047	0.060	0.050
$\alpha = 0.01$	0.013	0.007	0.007	0.023	0.013

The cut-off value is obtained by the parametric bootstrap-based procedure based on 1,000 resamplings

6 Extension

Recent research has been focusing on doubly truncated random variables. In doubly truncated data, X^O is in the sample only if $X^O \in [L^O, R^O]$ holds, where (L^O, R^O) is either random or deterministic. Navarro and Ruiz (1996), Sankaran and Sunoj (2004) and Bar-Lev and Boukai (2009) introduce several key quantities that characterize the distribution of X^O when (L^O, R^O) is deterministic. Efron and Petrosian (1999) proposed nonparametric estimator for $F_{X^O}(x)$ when (L^O, R^O) is random and independent of X^O. Truncation models discussed in Marchetti and Stanghellini (2007) is the multivariate generalization of the doubly truncated data. All of the above papers assume that there is no interaction between X^O and the truncation interval.

The present multivariate normal distribution approach is useful for modeling the association between X^O and (L^O, R^O). We assume the model

$$\begin{pmatrix} L^O \\ X^O \\ R^O \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_L \\ \mu_X \\ \mu_R \end{bmatrix}, \begin{bmatrix} \sigma_L^2 & \sigma_{LX} & \sigma_{LR} \\ \sigma_{LX} & \sigma_X^2 & \sigma_{XR} \\ \sigma_{LR} & \sigma_{XR} & \sigma_R^2 \end{bmatrix} \right).$$

Under this model, the probability $c(\boldsymbol{\theta}) = \Pr(L^O \le X^O \le R^O)$ is very complicated, involving an integral on three dimensional space. To facilitate the calculation, one can impose the conditional independence assumption on the density:

$$f_{L^O R^O | X^O}(l, r|x) = f_{L^O}(l|x) f_{R^O}(r|x).$$

Under this structure, the inclusion probability is simplified as

$$c(\boldsymbol{\theta}) = \int \Phi\left(\frac{\mu_X - \mu_L + \sigma_X^t - \rho_{LX}\sigma_L^t}{\sqrt{\sigma_L^2\left(1 - \rho_{LX}^2\right)}} \right) \times \left\{ 1 - \Phi\left(\frac{\mu_X - \mu_R + \sigma_X^t - \rho_{RX}\sigma_R^t}{\sqrt{\sigma_R^2\left(1 - \rho_{RX}^2\right)}} \right) \right\} \phi(t)dt$$

where $\boldsymbol{\theta}' = \left(\mu_L, \mu_X, \mu_R, \sigma_L^2, \sigma_X^2, \sigma_R^2, \sigma_{LX}, \sigma_{RX}\right)$. To implement this procedure, one needs to check the practical utility of the conditional independence assumption and develop numerical procedure to solve the likelihood equations.

7 Conclusions

This article illustrates the ability of parametric modeling to deal with truncated data. This approach can handle the dependent truncation by a simple likelihood construction. Therefore, the present approach is of greatest value when the independence between L^O and X^O does not hold. The independence can be empirically checked by using one of quasi-independence tests proposed by Tsai (1990), Chen et al. (1996), Martin and

Betensky (2005) and Emura and Wang (2010). If the test accepts quasi-independence, one can apply Lynden-Bell's nonparametric estimator for the distribution of X^O. If L^O and X^O can be assumed to be normal, then the proposed parametric approach under $\sigma_{LX} = 0$ leads to more efficient estimator for the parameters. If quasi-independence test is rejected, either the proposed method or the copula-based approach of Lakhal-Chaieb et al. (2006) can be applied. Although both approaches can model the dependency of truncation, they can be applicable for very different circumstances. The method of Lakhal-Chaieb et al. (2006) is applied when the dependency is modeled by Archimedean copula (AC) models and L^O and X^O are positive random variables. On the other hand, the proposed bivariate normal distribution is not included in AC models and L^O and X^O are distributed on the real line. The full parametric specification of the proposed approach is a drawback relative to Lakhal-Chaieb et al. (2006) in which the marginal distributions are unspecified. Nevertheless, the full-specification leads to an asymptotically efficient estimator, where the variance is easily estimated by inverting the observed Fisher information. On the other hand, the asymptotic variance of the estimator of Lakhal-Chaieb et al. (2006) has not been derived and jackknife method is recommended. What is more important for practitioners is that the parameters of the multivariate normal distribution are much easier to understand.

Acknowledgments The authors are grateful to the editor and the two referees for many helpful comments on the first version of this paper. The research in the second author was in part supported by the Japan Society for the Promotion of Science through Grants-in-Aid for Scientific Research (C) (No. 21500283).

Appendix A: Derivation of the Fisher information matrix

Let $\boldsymbol{\theta}' = (\mu_L, \mu_X)$. Then, the first derivative of the log-likelihood can be written as

$$\dot{l}_1(\boldsymbol{\theta}) = -\frac{z(\xi)}{\sqrt{\sigma_L^2 + \sigma_X^2 - 2\sigma_{LX}}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \begin{bmatrix} \sigma_L^2 & \sigma_{LX} \\ \sigma_{LX} & \sigma_X^2 \end{bmatrix}^{-1} \begin{pmatrix} L_1 - \mu_L \\ X_1 - \mu_X \end{pmatrix}$$

where $z(x) = \phi(x)/\Phi(x)$ and

$$\xi = \frac{\mu_X - \mu_L}{\sqrt{\sigma_L^2 + \sigma_X^2 - 2\sigma_{LX}}}.$$

By noting that $\dot{z}(x) = -xz(x) - z(x)^2$, we obtain the Fisher information

$$I(\boldsymbol{\theta}) = \frac{\xi z(\xi) + z(\xi)^2}{\sigma_L^2 + \sigma_X^2 - 2\sigma_{LX}} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} \sigma_L^2 & \sigma_{LX} \\ \sigma_{LX} & \sigma_X^2 \end{bmatrix}^{-1}.$$

Equation 8 follows by letting

$$\xi z(\xi) + z(\xi)^2 = \frac{\Phi^{-1}(c)\phi\{\Phi^{-1}(c)\}}{c} + \frac{\phi\{\Phi^{-1}(c)\}^2}{c^2} \equiv w(c).$$

References

Bar-Lev S-K, Boukai B (2009) A characterization of the exponential distribution by means of coincidence of location and truncated densities. Stat Pap 50:403–405

Chen C-H, Tsai W-Y, Chao W-H (1996) The product-moment correlation coefficient and linear regression for truncated data. J Am Stat Assoc 91:1181–1186

Cohen AC (1959) Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics 1:217–237

Cohen AC (1961) Tables for maximum likelihood estimators: singly truncated and singly censored samples. Technometrics 3:535–541

Efron B, Petrosian V (1999) Nonparametric methods for doubly truncated data. J Am Stat Assoc 94: 824–834

Efron B, Tibshirani RJ (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, vol 25, no. 2. Chiba University

Emura T, Wang W (2010) Testing quasi-independence for truncation data. J Multivar Anal 101:223–239

Fang K-T, Kotz S, Ng KW (1990) Symmetric multivariate and related distributions. Chapman & Hall, New York

Hansen JN, Zeger S (1980) The asymptotic variance of the estimated proportion truncated from a normal population. Technometrics 22:271–274

He S, Yang GL (1998) Estimation of the truncation probability in the random truncation model. Ann Statist 26:1011–1027

Huber P (1974) Robust statistics. Springer, New York

Klein JP, Moeschberger ML (2003) Survival analysis: techniques for censored and truncated data. Springer, New York

Knight K (2000) Mathematical statistics. Chapman & Hall, New York

Lakhal-Chaieb L, Rivest L-P, Abdous B (2006) Estimating survival under a dependent truncation. Biometrika 93:665–669

Lang KL, Little JA, Taylor JMG (1989) Robust statistical modeling using the *t* distribution. J Am Stat Assoc 84:881–896

Lynden-Bell D (1971) A method of allowing for known observational selection in small samples applied to 3RC quasars. Mon Not R Astron Soc 155:95–118

Marchetti GM, Stanghellini E (2007) A note on distortions induced by truncation with applications to linear regression systems. Stat Prob Lett 78:824–829

Marshall AW, Olkin I (1967) A multivariate exponential distribution. J Am Stat Assoc 62:30–44

Martin EC, Betensky RA (2005) Testing quasi-independence of failure and truncation via conditional Kendall's Tau. J Am Stat Assoc 100:484–492

Navarro J, Ruiz JM (1996) Failure rate functions for doubly truncated random variables. IEEE Trans Reliab 45:685–690

Sankaran PG, Sunoj SM (2004) Identification of models using failure rate and mean residual life of doubly truncated random variables. Stat Pap 45:97–109

Schiel JL, Harmston M (2000) Validating two-stage course placement systems when data are truncated. ACT research report series 2000-3. ACT, Iowa City

Tsai W-Y (1990) Testing the association of independence of truncation time and failure time. Biometrika 77:169–177

Waiker VB, Schuurmann FJ, Raghunathan TE (1984) On a two-stage shrinkage testimator of the mean of a normal distribution. Commun Stat Theory Method 13:1901–1913

Wang MC, Jewell NP, Tsai WY (1986) Asymptotic properties of the product-limit estimate and right censored data. Ann Stat 13:1597–1605

Woodroofe M (1985) Estimating a distribution function with truncated data. Ann Stat 13:163–177