Multivariate Parametric Approaches for Dependently Left-truncated Data

TAKESHI EMURA[1]

National Chiao-Tung University, Hsin-Chu, Taiwan, R.O.C.

and

YOSHIHIKO KONNO

Japan Women's University, Tokyo, Japan

Abstract

Many statistical methods for left-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, a 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 models 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 estimating equations and Fisher information matrix are provided. It is demonstrated that the current approach is adaptive to many important sampling designs, including a two-stage course placement system. Simulations are performed to examine finite-sample performances of the proposed method under two practical sampling designs.

Key words and phrases: Correlation coefficient; Truncation; Maximum likelihood; Missing data; Multivariate analysis

---

1 Correspondence concerning this article should be addressed to Takeshi Emura, Institute of Statistics, National Chiao Tung University, 1001 University Road, Hsinchu, Taiwan 300, ROC. E-mail: temura@stat.nctu.edu.tw; takeshi0909326@hotmail.com

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 & 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 pre-truncated 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 X^O can be included in the sample only if it exceeds another random variable L^O. Such type of data is called left-truncated data and is commonly seen in medical and astronomical research (Klein & Moeschberger 2003). Construction of nonparametric estimator for $F_{X^O}(x) = \Pr(X^O \leq x)$ is first proposed by Lynden-Bell (1971) for analyzing quasars data. Asymptotic properties of Lynden-Bell's estimator have been extensively studied; Woodroofe (1985), Wang et al. (1986). It is well known that independence between L^O and X^O is a fundamental assumption for the consistency of Lynden-Bell's estimator. Tsai (1991) and Chen, Tsai and Chao (1996) present methods for testing the independence assumption.

Compared with the nonparametric 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 from X^O, which many practical examples do not satisfy (Tsai 1990; Martin & Betensky 2005).

Many practical applications in behavioral and educational research can be formulated by left-truncated models where the independence assumption may not hold. For example, the National Center Test for University Admissions is a type of standardized test used by public and some private universities in Japan. The test is intended to measure the academic knowledge that is important for their future undergraduate study. Some university has a cut-off score, 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. Then one can define a left-truncation variable $L^O = 120 - Y^O$ so that the acceptance criteria for the university, $X^O + Y^O \geq 120$, can be written as $L^O \leq X^O$. In this example, it is natural to assume that X^O and L^O may be negatively correlated through unobserved factors related to students' intellectual ability.

In a retrospective sampling of the scores (X^O, Y^O) only from the admitted students, the samples do not include those pair (X^O, Y^O) satisfying $X^O + Y^O < 120$. Estimation of the population (pre-truncated) distribution only from the truncated sample may be useful to determine the subsequent year's cut-off value and the prediction of future applicants. We will return to the model for the National Center Test for University Admissions in Section 5.1 for further investigation. Another important application of random truncation model is the two-stage placement system discussed in Schiel & Harmston (2000), which will be discussed in Section 4 and Section 5.2.

In this article, we consider parametric approaches to left-truncated data, which do not require the independence assumption. In particular, Section 2 presents a general idea of likelihood inference on parametric modeling while Section 3 contains analytical results on the normal distribution case. Section 4 investigates a sampling design known as a two-stage placement system. In Section 5, the performance of the present approaches is studied via simulations. Section 6 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. Therefore, 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 $c(\boldsymbol{\theta}) = \Pr(L^O \leq X^O)$ is the inclusion probability and $\mathbf{1}[l \leq x]$ is the indicator function of the set in square 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

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

Let $\hat{\boldsymbol{\theta}}$ represent the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ based on (1). In principle, any parametric modeling on $f_{\boldsymbol{\theta}}(l, x)$ is feasible as long as the MLE is obtained uniquely. However, it should be noted that, due to the truncation, the likelihood function tends to be complicated and there is no explicit solution for $\hat{\boldsymbol{\theta}}$ in most common distributions. In general, a simple functional form for $c(\boldsymbol{\theta})$ results in the likelihood function that is mathematically tractable. As an instance, Marshall and Olkin's (1967) bivariate exponential model on a pair (L^O, X^O) is specified by

$$\Pr(L^O > l, X^O > x) = \exp\{-\lambda_L l - \lambda_X x - \lambda_{LX} \max(l, x)\}\, \mathbf{1}[\, l \ge 0, x \ge 0\,],$$

where $(\lambda_L, \lambda_X, \lambda_{LX})$ are positive parameters. In this model, a simple form $c(\boldsymbol{\theta}) = (\lambda_X + \lambda_{LX})^{-1}(\lambda_L + \lambda_X + \lambda_{LX})^{-1}$ can be obtained. In Section 3 and subsequent discussions we focus our discussion on the bivariate normal distribution that has a relatively simple form in $c(\boldsymbol{\theta})$ and that can be applied to common sampling designs in behavioral and educational research.

2.2 Large Sample Analysis

Let $l_1(\boldsymbol{\theta}) = \log\{c(\boldsymbol{\theta})^{-1} f_{\boldsymbol{\theta}}(L_1, X_1)\}$ be a log-likelihood function based on a truncated observation and let $\dot{l}_1(\boldsymbol{\theta})$ be the first derivative with respect to the parameter $\boldsymbol{\theta}$. Throughout the article, we denote a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Psi}$ by $N(\boldsymbol{\mu}, \boldsymbol{\Psi})$. If $l_1(\boldsymbol{\theta})$ is twice differentiable and certain boundedness conditions on the third derivatives are satisfied, it can be shown that, for sufficiently large n,

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

where $I(\boldsymbol{\theta}) = E\{\dot{l}_1(\boldsymbol{\theta})\dot{l}_1(\boldsymbol{\theta})' \mid L^O \le X^O\}$ is the Fisher information matrix based on a single observation and the notation "$\rightarrow$" signifies the convergence in distribution (Van der Vaart 1998, Theorem 5.42). Except for simple models, such as the one discussed in Section 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 asymptotic approximation for $I(\boldsymbol{\theta})$.

Let $\hat{c} = c(\hat{\boldsymbol{\theta}})$ be an estimate of the inclusion probability $c(\boldsymbol{\theta})$. By the invariance property of the MLE, $\hat{c}$ is also the MLE for the inclusion probability. By the delta method, we obtain the asymptotic distribution as

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

Thus, the standard error of $\hat{c}$ can be obtained from the asymptotic variance by replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$. We do not address issues of regularity conditions for the asymptotic theory, but we explore the implications of the asymptotic results by giving an explicit formula of $I(\boldsymbol{\theta})$ in the subsequent discussions.

3. LIKELIHOOD INFERENCE UNDER NORMAL POPULATION

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

3.1 Left-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 $\boldsymbol{\theta}' = (\mu_L, \mu_X, \sigma_L^2, \sigma_X^2, \sigma_{LX})$ be unknown parameters, where $\sigma_L^2 > 0$, $\sigma_X^2 > 0$ and $-1 < \rho_{LX} < 1$. It is easy to show that the inclusion probability satisfies

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

where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. By combining the likelihood of n truncated samples, we obtain the log-likelihood function:

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

where

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

Central to the analysis of the likelihood arguments is the likelihood equations. 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 likelihood equations can be written as

$$\dot{l}(\boldsymbol{\theta}) = -n\frac{\dot{c}(\boldsymbol{\theta})}{c(\boldsymbol{\theta})} + \sum_j \mathbf{U}_j(\boldsymbol{\theta}) = \mathbf{0}, \tag{5}$$

where, $\dot{l}(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\dot{c}(\boldsymbol{\theta}) = \partial c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. The maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ under the bivariate normal model can be obtained by solving equation (5). We recommend solving the equations using the Newton-Raphson algorithm with the starting values chosen to be the sample mean and sample covariance.

If some components of $\boldsymbol{\theta}' = (\mu_L, \mu_X, \sigma_L^2, \sigma_X^2, \sigma_{LX})$ are known *a priori*, the log-likelihood function (4) is maximized under the known components. For example, if the covariance structure in (2) is known, we redefine the mean parameters

$\boldsymbol{\theta}' = (\mu_L, \mu_X)$ and maximize (4) for $\boldsymbol{\theta}$ with $(\sigma_L^2, \sigma_X^2, \sigma_{LX})$ being fixed. This example will serve as a theoretical tool to investigate the effect of truncation, which are discussed in detail in Section 3.2. Another useful example is the likelihood estimator $\boldsymbol{\theta}' = (\mu_L, \mu_X, \sigma_L^2, \sigma_X^2)$ under $\sigma_{LX} = 0$, the independence assumption. Specifically, one can solve the likelihood equation

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

where

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

The solution obtained by solving equation (6) is consistent for $\boldsymbol{\theta}$ only under the independence assumption, but it may still be workable when the model deviation from the independence is subtle. Roughly speaking, when the test of $H_0 : \sigma_{LX} = 0$ is not statistically significant based on the likelihood test for $\sigma_{LX} = 0$, this method may be used. This point will be further investigated in Section 5.1 based on simulation studies.

3.2 The Effect of Truncation

As we have seen in Section 2.2, the asymptotic theory of MLE guarantees the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}$ and $\hat{c} = c(\hat{\boldsymbol{\theta}})$. For truncated data, it is also 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 describe the impact of truncation on the estimators under the bivariate normal model (2). To simplify the discussion, we consider the case where the covariance structure for (L^o, X^o) is known. 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

$$\tilde{I}\{c(\boldsymbol{\theta})\} = \begin{bmatrix} \sigma_L^2 & \sigma_{LX} \\ \sigma_{LX} & \sigma_X^2 \end{bmatrix}^{-1} - \frac{w\{c(\boldsymbol{\theta})\}}{\sigma_L^2 + \sigma_X^2 - 2\sigma_{LX}} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{7}$$

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 $\phi(x) = \dot{\Phi}(x)$. It can be shown that $w(\cdot)$ is a decreasing function of the inclusion probability $c = c(\boldsymbol{\theta})$. It has the properties that $\lim_{c \downarrow 0} w(c) = 1$ and $\lim_{c \uparrow 1} w(c) = 0$ and the slope is at most $\dot{w}(1/2) = \sqrt{2/\pi}(1 - 4/\pi) < 0$ (Figure 1). A closer look reveals that the second term in the Fisher information (7) reflects the loss of information due to truncation. For instance, $w(0.50) = 0.64$ signifies that there is 64% loss of information when 50% of data is truncated (Figure 1). Of course, in absence of truncation, $w(1) = 0$ and the information loss vanishes. Since $w(\cdot)$ is a decreasing function, the loss of information decreases as the inclusion probability increases. Mathematically speaking, $\tilde{I}\{c\} - \tilde{I}\{c'\}$ is positive semi-definite for $c > c'$. In data analysis, one may use $w(\hat{c})$ as a measure to examine the truncation effect.

Insert Figure 1 here.

Unfortunately, when the covariance structure is unknown, the Fisher information becomes untractable due to the complicated form of the second moments on a truncated pair (L, X). We would like to conjecture that the effect of truncation on Fisher information is similar in this general case. To complete our understanding of the truncation effect, we will further study the effect of truncation through the simulation studies in Section 5.2.

4. APPLICATION TO A TWO-STAGE COURSE PLACEMENT SYSTEM

In a two-stage course placement system discussed in Schiel & Harmston (2000), all incoming students are tested with an initial screening test. For example, ACT mathematics scores can be used in the post-secondary schools to examine whether to place students into either a standard or remedial mathematics course. Those scoring below a pre-specified cutoff point K must take the placement test for further grouping by skill levels. An example of the placement test is COMPASS algebra test which is a computer adaptive testing system.

Under this design, researchers may obtain a pair of the screening test S^O and the placement test P^O only when $S^O \le K$ holds. For example, consider a group of students who join a placement test by satisfying $S^O \le K$. After the placement test, investigators can collect students' scores for the placement test. For these students, investigators may also retrospectively ascertain the test scores for the screening test by means of questionnaires. However, for those students who satisfy $S^O > K$ and do not join the placement test, researchers may not have a direct assess to the screening test score. Therefore researchers can obtain a pair of (S^O, P^O) subject to $S^O \le K$.

We consider a problem of estimating the joint distribution of (S^O, P^O) only using data $(S_j, P_j);\ j = 1, 2, \dots, n$ subject to $S_j \le K$. Information on those pairs (S^O, P^O) with $S^O > K$ (those who do not join the placement test) is unavailable under this design. To specify the problem, we suppose that (S^O, P^O) follows the bivariate normal distribution:

$$\begin{pmatrix} \text{Screening test } (S^O) \\ \text{Placement test } (P^O) \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_{\mathrm{S}} \\ \mu_P \end{bmatrix}, \begin{bmatrix} \sigma_{\mathrm{S}}^2 & \sigma_{\mathrm{SP}} \\ \sigma_{\mathrm{SP}} & \sigma_{\mathrm{P}}^2 \end{bmatrix} \right). \tag{8}$$

Major practical interest in this setting is to estimate the mean parameters (μ_S, μ_P). Due to the truncation, the observed data $(S_j, P_j);\ j = 1, 2, \ldots, n$ with $S_j \le K$ no longer follows the model (8). To adapt to the left-truncation model discussed in Section 2.1, we let $X^O = P^O - S^O$ and $L^O = P^O - K$. With these transformations, one can observe a pair (L^O, X^O) when $L^O \le X^O$ where

$$\begin{pmatrix} L^O \\ X^O \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_{\mathrm{P}} - K \\ \mu_P - \mu_S \end{bmatrix}, \begin{bmatrix} \sigma_{\mathrm{P}}^2 & \sigma_{\mathrm{P}}^2 - \sigma_{\mathrm{SP}} \\ \sigma_{\mathrm{P}}^2 - \sigma_{\mathrm{SP}} & \sigma_{\mathrm{P}}^2 + \sigma_{\mathrm{S}}^2 - 2\sigma_{\mathrm{SP}} \end{bmatrix} \right).$$

The maximum likelihood estimates $\hat{\boldsymbol{\theta}}' = (\hat{\mu}_X, \hat{\mu}_L, \hat{\sigma}_X^2, \hat{\sigma}_L^2, \hat{\sigma}_{LX})$ based on $\{(L_j, X_j);\ j = 1, 2, \ldots, n\}$ can be easily converted to obtain $(\hat{\mu}_S, \hat{\mu}_P, \hat{\sigma}_S^2, \hat{\sigma}_P^2, \hat{\sigma}_{SP})$ via the transformations:

$$\begin{pmatrix} \mu_S \\ \mu_P \\ \sigma_S^2 \\ \sigma_P^2 \\ \sigma_{SP} \end{pmatrix} = \begin{pmatrix} K + \mu_L - \mu_X \\ K + \mu_L \\ \sigma_X^2 + \sigma_L^2 - 2\sigma_{LX} \\ \sigma_L^2 \\ \sigma_L^2 - \sigma_{LX} \end{pmatrix}. \tag{9}$$

Also, the inclusion probability $\hat{c} = c(\hat{\boldsymbol{\theta}})$ estimates the probability of the screening cut-off $\Pr(S^O \le K)$.

If investigators are only interested in the marginal distribution of the screening test S^O, the analysis can be conducted using only marginal distribution for $S_j;\ j = 1, 2, \ldots, n$ subject to $S_j \le K$. The marginal log-likelihood for $S_j;\ j = 1, 2, \ldots, n$ is given by

$$l(\mu_S, \sigma_S^2) = -\frac{n}{2}\log(2\pi) - n\log\Phi\left(\frac{K - \mu_S}{\sigma_S}\right) - \frac{n}{2}\log\sigma_S^2 + \frac{1}{2\sigma_S^2}\sum_j (S_j - \mu_S). \tag{10}$$

Using the method discussed in Cohen (1959, 1961), we obtain the marginal MLE $(\tilde{\mu}_S, \tilde{\sigma}_S^2)$ that maximizes (10). Also, the MLE of the inclusion probability $\Pr(S^O \le K)$

is estimated by $\tilde{c} = \Phi(\tilde{\xi})$ where $\tilde{\xi} = (K - \tilde{\mu}_S)/\tilde{\sigma}_S$ (Hansen & Zeger, 1980). Note that this approach leaves the distribution for P^O completely unspecified. Now we establish the relationship between the proposed likelihood estimators and the marginal MLE.

Proposition 1 *Let* $\hat{\boldsymbol{\theta}}' = (\hat{\mu}_S, \hat{\mu}_P, \hat{\sigma}_S^2, \hat{\sigma}_P^2, \hat{\sigma}_{SP})$ *be the MLE , which solves equations (5) and (9). Then* $\hat{\mu}_S = \tilde{\mu}_S$, $\hat{\sigma}_S^2 = \tilde{\sigma}_S^2$ *and* $\hat{c} = \tilde{c}$ *.*

In Appendix B, we give a proof of Proposition 1 by showing that the joint and marginal likelihood approaches lead to the same estimating equation for (μ_S, σ_S^2) . One important implication of Proposition 1 is the invariance of the conclusion: estimators for the marginal distribution for S^O do not depend on the methods used irrespective of the degree of correlation.

5. NUMERICAL ANALYSIS

The simulation studies are conducted to investigate the empirical performances of the proposed approaches under two different sampling designs: Section 5.1 is concerned with the comparison between the proposed parametric approaches and the existing nonparametric approaches under the models for a university entrance examination: Section 5.2 investigates the effect of the inclusion probability on the likelihood estimators under a two-stage placement system.

5.1 Comparison with Nonparametric Estimators

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/center_exam/20exam/20heikin.html), Japanese score (X^O) and English score (Y^O) follows a bivariate normal distribution

$$\begin{pmatrix} X^O \\ Y^O \end{pmatrix} \sim N\left(\begin{bmatrix} 60.82 \\ 62.63 \end{bmatrix}, \begin{bmatrix} 19.64^2 & (19.64)(16.81)\rho_{XY} \\ (19.64)(16.81)\rho_{XY} & 16.81^2 \end{bmatrix} \right). \tag{11}$$

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. Since the correlation between X^O and Y^O is not available publicly, we consider four levels of correlations ρ_{XY}=0 (not correlated), ρ_{XY}=0.25 (weakly correlated), ρ_{XY}=0.50 (moderately correlated) and ρ_{XY}=0.75 (strongly correlated). These four models are used to generate data in the subsequent studies.

Now we consider a hypothetical department in a university having a policy to accept students whose sum of Japanese and English score reach 120. We can define $L^O = 120 - Y^O$ so that the acceptance criteria can be written as $L^O \le X^O$. Accordingly, the probability of sample inclusion is $c = \Pr(X^O + Y^O \ge 120) = \Pr(L^O \le X^O)$, which depends on the correlation ρ_{XY}. Thus, the observations used in this simulation consist of $\{(L_j, X_j);\ j = 1, 2, \ldots, n\}$ subject to $L_j \le X_j$, which are the scores for the admitted students. We chose $n = 100$ and $n = 200$ since these numbers are popular in many departments of Japanese universities.

To compare the performances of the proposed parametric estimators with the existing nonparametric estimators, we generated 1,000 dataset of n truncated samples under (11). For each run, we computed the MLE of $\hat{\mu}_X$ based on equation (5) to estimate the mean Japanese score 60.82. For comparison, we also computed the nonparametric MLE defined by

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

where $\hat{F}_X(x)$ is the Lynden-Bell's nonparametric estimator. This method is valid only under the assumption of $H_0 : L^O \perp X^O$, and, of course, this assumption is valid under $\rho_{XY} = 0$. We are also interested in the estimator $\hat{\mu}_X^0$ under the independent normal distribution of $\rho_{XY} = 0$, which is a solution of (6). Note that $\hat{\mu}_X^0$ requires the strongest assumption among the three estimators. The means and mean squared errors (MSEs) of $\hat{\mu}_X^0$, $\hat{\mu}_X$, $\hat{\mu}_X^{NP}$ were then computed over the 1,000 runs. To investigate the performance of estimating the inclusion probability, we computed $\hat{c}^0$, $\hat{c}$ and $\hat{c}^{NP}$ in a similar fashion and compared their performance, where $\hat{c}^{NP}$ is the nonparametric estimator proposed by He & Yang (1998). Numerical results on the two estimands, $\mu_X = 60.82$ and $c = \Pr(X + Y \geq 120)$, are presented in this article.

Table 1 summarizes the results of simulation studies on estimating $\mu_X = 60.82$. The estimator $\hat{\mu}_X$ that fits the bivariate normal distribution is approximately unbiased in all cases. On the other hand, the biases of $\hat{\mu}_X^{NP}$ and $\hat{\mu}_X^0$ are negligible only under the case of ρ_{XY}=0.00 in the first row. As expected, the biases of $\hat{\mu}_X^{NP}$ and $\hat{\mu}_X^0$ increase as the level of correlation departs from ρ_{XY}=0.00. In all the three estimators, the MSEs decrease as the sample gets large under the independent model. However, the MSEs of $\hat{\mu}_X^{NP}$ and $\hat{\mu}_X^0$ remain large due to the wrong model assumption for moderately and strongly correlated cases. Somewhat surprisingly, if the correlation is weak and sample size is small in the second row, the MSEs of the inconsistent estimator $\hat{\mu}_X^{NP}$ and $\hat{\mu}_X^0$ remain still smaller than the consistent estimator $\hat{\mu}_X$. The large variance of $\hat{\mu}_X$ causes the inflation in the MSEs.

Insert Table 1 here.

Table 2 presents the simulation results on estimating the inclusion probability $c = \Pr(X + Y \geq 120)$. The performance is similar to those in Table 1. In presence of correlation, estimator $\hat{c}$ that adjusts to the dependent truncation dominates the other two estimators that assume independence. On the other hand, the estimator $\hat{c}^0$ attains the smallest MSEs under independence since it imposes the strongest model assumption.

Insert Table 2 here.

Now we focus our discussion on the independence case in the first column of Table 1. The MSEs for the MLE $\hat{\mu}_X$ are much larger than those for $\hat{\mu}_X^0$. This is, of course, because $\hat{\mu}_X$ requires additional estimation of the correlation parameters while $\hat{\mu}_X^0$ is calculated under the true correlation ρ_{XY}=0.00. Furthermore, the MSEs for $\hat{\mu}_X$ are still much larger than the nonparametric estimator $\hat{\mu}_X^{NP}$. This shows the undesirable feature of modeling dependence between L and X when they are actually independent. On the other hand, it is reasonable that $\hat{\mu}_X^0$ is the best estimator in terms of the MSE since it requires strongest assumption among the three. These results show that the independence assumption is quite informative in parameter estimation.

In general, we found that the estimator $\hat{\mu}_X$ adjusts for the dependent truncation mechanism in all cases and it gives nearly unbiased estimates. Due to modeling of the dependency on truncation, $\hat{\mu}_X$ tends to have large variability, especially in small sample sizes. If the independence is known *a priori*, either the nonparametric estimator $\hat{\mu}_X^{NP}$ or parametric estimator $\hat{\mu}_X^0$ is recommended to reduce the variability.

5.2 Numerical Studies on the Effect of Truncation

In the subsequent simulation studies we adopt a two-stage placement system in Section

4 with the cutoff value being fixed at $K = 50$. Random pairs of (S, P) were generated from the normal distribution

$$\begin{pmatrix} \text{Screening test } (S^O) \\ \text{Placement test } (P^O) \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_S \\ 50 \end{bmatrix}, \begin{bmatrix} 100 & 10\times10\times0.5 \\ 10\times10\times0.5 & 100 \end{bmatrix} \right), \tag{12}$$

where the inclusion probability is $c = \Phi\{(50 - \mu_S)/10\}$. Five levels of c, namely 0.84, 0.69, 0.50, 0.31, 0.16 are chosen by setting μ_S=40, 45, 50, 55 and 60 respectively. For $n = 500$ and 1000 truncated samples in each run, we calculated the MLE $(\hat{\mu}_S, \hat{\mu}_P, \hat{\sigma}_S^2, \hat{\sigma}_P^2, \hat{\sigma}_{SP})$. Over 1,000 simulation runs we computed the MSEs for the mean parameters $(\hat{\mu}_S, \hat{\mu}_P)$.

Table 3 shows the results of simulation studies. If we compare the MSEs for $n = 500$, the MSE clearly increases as the inclusion probability decreases. It is worth mentioning that each MSE is calculated under the same sample sizes, but different inclusion probability. A possible explanation of the poor performance in the case of $c = 0.16$ is that available data only represent a very small fraction (approximately 16%) of the whole population. Similar relationships between the MSEs and inclusion probability are found for $n = 1000$. As expected, the MSEs for $n = 1000$ are approximately half of those for $n = 500$, suggesting that the estimator has regular convergence rates.

Insert Table 3 here.

In general, the simulation studies reveal that the accuracy of the likelihood estimator based on truncated samples critically depends on the inclusion probability. The results are consistent with the implication from the analytical formula of the Fisher information in Section 3.2.

6. CONCLUSIONS

This article illustrates the ability of parametric modeling to deal with left-truncated data. This approach can handle the dependent truncation mechanism by a simple likelihood construction. Therefore, the present approach is of greatest value when the independence assumption does not hold, where the Lynden-Bell's nonparametric approach cannot apply. We have demonstrated the impact of inclusion probability on the estimator using the analytic expression of Fisher information and the simulation studies. Both the analytic and simulation analyses show that the inclusion probability plays a key role in the efficiency of the proposed estimator. This nature suggests that, in analyzing truncated data, researchers need to pay attention on the inclusion probability as well as the sample size in the usual statistical practice.

An important issue which is not discussed in this article is the robustness properties of the likelihood methods. In general, results from parametric methods are sensitive to the extreme values and they lead to bias in estimation. Some modification based on robust statistics (Huber, 1974) is desired so that the method can work under contaminated data. A relatively simple way of adjusting for the extreme values may be obtained by fitting the multivariate contaminated normal or multivariate t-distribution distributions (Little and Rubin, 2002) on (L^O, X^O).

Recent research has been focusing on more complicated independent truncation mechanisms. Efron and Petrosian (1999) proposed a nonparametric estimator for $F_X(x)$ for double-truncation where both left (L^O) and right (R^O) truncation play a role in the inclusion criteria, i.e., X^O is in the sample only if it belongs to the interval $[L^O, R^O]$. Their method developed for nonparametric analysis also relies on the independence assumption between X^O and $[L^O, R^O]$. Truncation models discussed in

Marchetti and Stanghllini (2007) is the multivariate generalization of the double-truncation. They treat the interval $[L^O, R^O]$ as fixed in each truncated components and essentially do not model the dependent truncation effect.

An extension of the present approach to double-truncation setting may be possible. Consider the case where both L^O and R^O are random. We assume the multivariate normal distribution

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

Under this model, the inclusion 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:

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

Under this structure, the inclusion probability is simplified as

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

where $\boldsymbol{\theta}' = (\mu_L, \mu_X, \mu_R, \sigma_L^2, \sigma_X^2, \sigma_R^2, \sigma_{LX}, \sigma_{XR})$. 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.

ACKNOWLEDGMENTS

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.17500185)

REFERENCES

Chen, C. -H., Tsai, W. -Y. and Chao, W. -H. (1996), "The Product-moment Correlation Coefficient and Linear Regression for Truncated Data.", *Journal of the American Statistical Association*, 91, 1181-1186.

Cohen, A. C. (1959), "Simplified Estimators for the Normal Distribution when Samples are Singly Censored or Truncated", *Technometrics*, 1, 217-237.

Cohen, A. C. (1961), "Tables for Maximum Likelihood Estimators: Singly Truncated and Singly Censored Samples", *Technometrics*, 3, 535-541.

Efron , B. and Petrosian, V. (1999), "Nonparametric Methods for Doubly Truncated Data." *Journal of the American Statistical Association*, 94, 824-834.

Marchetti, G. M. and Stanghellini (2007), "A Note on Distortions Induced by Truncation with Applications to Linear Regression Systems," *Statistics & Probability Letters*, 78, 824-829.

Hansen, J. N. and Zeger S. (1980), "The Asymptotic Variance of the Estimated Proportion Truncated from a Normal Population" , *Technometrics*, 22, 271-274.

He, S. and Yang, G. L. (1998), "Estimation of the Truncation Probability in the Random Truncation Model," *Annals of Statistics*. 26, 1011-1027.

Huber, P. (1974), "*Robust Statistics*." New York: Springer.

Klein, J. P. and Moeschberger, M. L. (2003). *"Survival Analysis: Techniques for Censored and Truncated Data."* New York: Springer.

Little J. A. and Rubin, D. B. (2002). "*Statistical Analysis with Missing Data*", New Jersey: John Wiley & Sons, Inc.

Lynden-Bell, D. (1971), "A Method of Allowing for Known Observational Selection in Small Samples Applied to 3RC Quasars," *Monthly Notices. Roy. Astronom. Soc.*,155, 95-118.

Marshall, A. W. and Olkin, I. (1967). "A Multivariate Exponential Distribution," *Journal of the American Statistical Association*, **62**, 30-44.

Martin, E. C. and 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.

Schiel, J. L. and Harmston, M. (2000), "Validating Two-Stage Course Placement Systems When Data are Truncated", *ACT Research Report Series 2000-3*, Iowa City, IA: ACT.

Tsai, W. -Y. (1990), "Testing the Association of Independence of Truncation Time and Failure Time," *Biometrika*, 77, 169-177.

Van der Vaart. A. W. (1998), "*Asymptotic Statistics,*" Cambridge Series in Statistics and Probabilistic Mathematics, Cambridge: Cambridge University Press.

Wang, M. C., Jewell, N. P. and Tsai, W. Y. (1986), "Asymptotic Properties of the Product-limit Estimate and Right Censored Data," *Annals of Statistic*, 13, 1597-1605.

Woodroofe, M. (1985), "Estimating a Distribution Function With Truncated Data," *The Annals of Statistics,* 13, 163-177.

Appendix A: 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 (7) follows by letting

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

Appendix B: Proof of Proposition 1

To simplify the notation, we let $|\Omega| = \sigma_S^2\sigma_P^2 - \sigma_{SP}^2$ be the determinant of the covariance matrix for (S_j, P_j) and

$$C(\boldsymbol{\theta}) = \frac{1}{|\Omega|}\sum_j \{\sigma_P^2(S_j - \mu_S)^2 - 2\sigma_{SP}(S_j - \mu_S)(P_j - \mu_P) + \sigma_S^2(P_j - \mu_P)^2\}.$$

The likelihood estimator $(\hat{\mu}_S, \hat{\mu}_P, \hat{\sigma}_S^2, \hat{\sigma}_P^2, \hat{\sigma}_{SP})$ based on the observed data $(S_j, P_j);\ j = 1, 2, \ldots, n$ with $S_j \le K$ solves the likelihood equation:

$$0 = \frac{1}{|\Omega|}\sum_j \{\sigma_P^2(S_j - \mu_S) - \sigma_{SP}(P_j - \mu_P)\} + \frac{nz(\xi)}{\sigma_S}, \quad \text{(A.1)}$$

$$0 = \frac{1}{|\Omega|}\sum_j \{-\sigma_{SP}(S_j - \mu_P) + \sigma_S^2(P_j - \mu_P)\}, \quad \text{(A.2)}$$

$$0 = \frac{n\xi z(\xi)}{2\sigma_S^2} - \frac{n\sigma_P^2}{2|\Omega|} + \frac{\sigma_P^2 C(\boldsymbol{\theta})}{2|\Omega|} - \frac{1}{2|\Omega|}\sum_j (P_j - \mu_P)^2, \quad \text{(A.3)}$$

$$0 = -\frac{n\sigma_S^2}{2|\Omega|} + \frac{\sigma_S^2 C(\boldsymbol{\theta})}{2|\Omega|} - \frac{1}{2|\Omega|}\sum_j (S_j - \mu_S)^2 , \tag{A.4}$$

$$0 = \frac{n\sigma_{SP}}{|\Omega|} - \frac{\sigma_{SP} C(\boldsymbol{\theta})}{|\Omega|} + \frac{1}{|\Omega|}\sum_j (S_j - \mu_S)(P_j - \mu_P) . \tag{A.5}$$

, which are derived from (5) and (9). A simple algebra regarding (A.1) and (A.2) leads to the equation

$$0 = \frac{1}{\sigma_S^2}\sum_j (S_j - \mu_S) + \frac{nz(\xi)}{\sigma_S} . \tag{A.6}$$

Summing up the equations (A.3)-(A.5), we obtain an identity

$$C(\boldsymbol{\theta}) = 2n - n\xi z(\xi) .$$

Using the preceding equation together with equation (A.4), we can derive the equation

$$0 = \frac{n\xi z(\xi)}{2\sigma_S^2} - \frac{n}{2\sigma_S^2} + \frac{1}{2\sigma_S^4}\sum_j (S_j - \mu_S)^2 . \tag{A.7}$$

Now one can easily check that (A.6) and (A.7) are equivalent to the likelihood equation based on marginal likelihood for $S_j;\, j = 1, 2, \ldots, n$ subject to $S_j \le K$. This proves $\hat{\mu}_S = \tilde{\mu}_S$ and $\hat{\sigma}_S^2 = \tilde{\sigma}_S^2$.

To prove $\hat{c} = \tilde{c}$, recall that $\tilde{c} = \Phi\{(K - \tilde{\mu}_S)/\tilde{\sigma}_S\}$ and that $\hat{c} = \Phi\{(K - \hat{\mu}_S)/\hat{\sigma}_S\}$. But since $\hat{\mu}_S = \tilde{\mu}_S$ and $\hat{\sigma}_S^2 = \tilde{\sigma}_S^2$, it follows that $\hat{c} = \tilde{c}$.

Table 1: Means and MSEs of the two proposed estimators $\hat{\mu}_X^0$ and $\hat{\mu}_X$ and the competing estimator $\hat{\mu}_X^{NP}$ (nonparametric estimator) based on 1000 replications when (X, Y) follows the bivariate normal distribution (11) with sample size n=100 and 200.

Model	n		$\hat{\mu}_X^0$	$\hat{\mu}_X$	$\hat{\mu}_X^{NP}$
1) Not correlated μ_X =60.82, ρ_{XY} =0.00, c=0.553	100	Mean	60.761	59.854	60.799
		MSE	7.721	45.495	14.540
	200	Mean	60.806	60.438	60.837
		MSE	3.900	17.173	7.139
2) Weakly correlated μ_X =60.82, ρ_{XY} =0.25, c=0.548	100	Mean	65.250	59.580	65.552
		MSE	24.110	59.102	27.708
	200	Mean	65.228	60.524	65.506
		MSE	21.677	18.886	24.608
3) Moderately correlated μ_X =60.82, ρ_{XY} =0.50, c=0.543	100	Mean	68.436	59.559	68.879
		MSE	60.986	76.347	67.872
	200	Mean	68.386	60.183	68.803
		MSE	58.636	25.723	65.237
4) Strongly correlated μ_X =60.82, ρ_{XY} =0.75, c=0.540	100	Mean	70.755	59.387	71.349
		MSE	100.481	83.046	112.498
	200	Mean	70.773	60.227	71.367
		MSE	99.968	26.880	112.068

Table 2: Means and MSEs of the two proposed estimators $\hat{c}^0$ and $\hat{c}$ and the competing estimator $\hat{c}^{NP}$ (nonparametric estimator) based on 1000 replications when (X, Y) follows the bivariate normal distribution (11) with sample size n=100 and 200.

Model	n		$\hat{c}^0$	$\hat{c}$	$\hat{c}^{NP}$
1) Not correlated μ_X =60.82, ρ_{XY} =0.00, c=0.553	100	Mean	0.549	0.551	0.558
		MSE	0.007	0.028	0.010
	200	Mean	0.552	0.554	0.558
		MSE	0.003	0.014	0.005
2) Weakly correlated μ_X =60.82, ρ_{XY} =0.25, c=0.548	100	Mean	0.709	0.541	0.725
		MSE	0.031	0.028	0.037
	200	Mean	0.714	0.552	0.730
		MSE	0.030	0.014	0.036
3) Moderately correlated μ_X =60.82, ρ_{XY} =0.50, c=0.543	100	Mean	0.826	0.543	0.853
		MSE	0.082	0.028	0.098
	200	Mean	0.827	0.541	0.854
		MSE	0.081	0.014	0.097
4) Strongly correlated μ_X =60.82, ρ_{XY} =0.75, c=0.540	100	Mean	0.904	0.530	0.941
		MSE	0.133	0.029	0.161
	200	Mean	0.907	0.541	0.944
		MSE	0.135	0.014	0.163

Table 3: Comparison of mean squared errors (MSE) under different inclusion probabilities $c(\boldsymbol{\theta})$ =0.84, 0.69, 0.50, 0.31 and 0.16. MSEs are calculated based on 1000 simulated estimators under bivairate normal distribution with $n = 500$ and 1000.

n	$c(\boldsymbol{\theta})$	MSE of $\hat{\mu}_S$	MSE of $\hat{\mu}_P$
500	0.84 (μ_S =40)	0.668	0.350
	0.69 (μ_S =45)	1.786	0.671
	0.50 (μ_S =50)	5.165	1.675
	0.31 (μ_S =55)	14.641	4.577
	0.16 (μ_S =60)	44.271	13.226
1000	0.84 (μ_S =40)	0.346	0.182
	0.69 (μ_S =45)	0.890	0.337
	0.50 (μ_S =50)	2.539	0.851
	0.31 (μ_S =55)	6.119	2.047
	0.16 (μ_S =60)	17.639	5.148

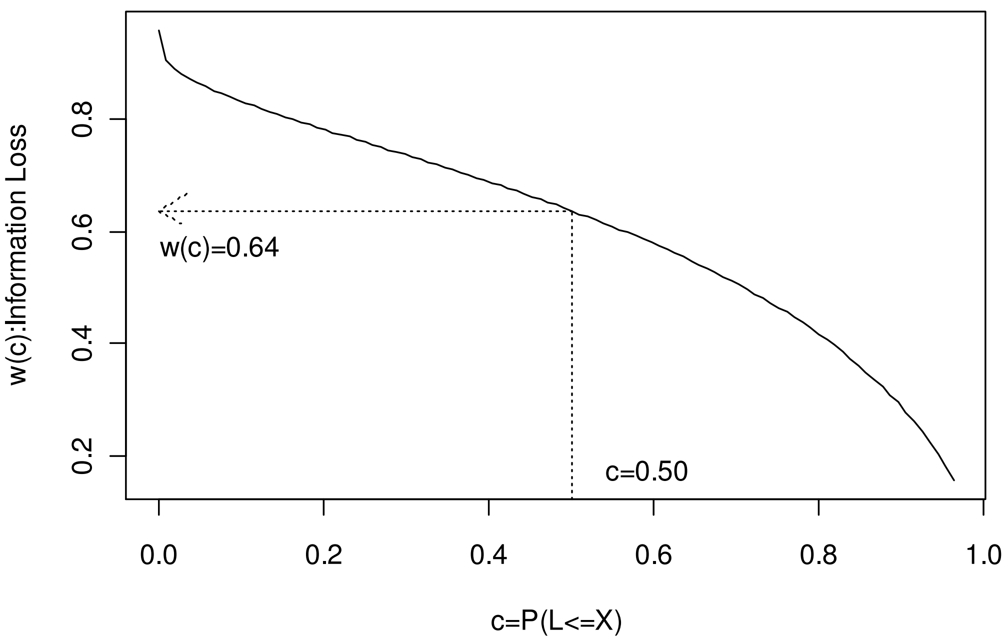

Figure 1. The graph of the function $w(c) = \frac{\Phi^{-1}(c)\phi\{\Phi^{-1}(c)\}}{c} + \frac{\phi\{\Phi^{-1}(c)\}^2}{c^2}$ that describes the information loss under the inclusion probability c. About 64% information loss is expected when 50% of the population is left-truncated.