STAPRO: 8314 Model 3G pp. 1–8 (col. fig: NIL)

ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

UNCORRECTED PROOF

Estimation of minimum and maximum correlation coefficients

Jim X. Xiang

Janssen Research and Development, 1000 Route 202, Raritan, NJ, 08869-0602, United States

ARTICLE INFO

Article history:
Received 23 January 2017
Accepted 19 August 2018
Available online xxxx

MSC:
primary 60F05
secondary 62G30

Keywords:
Minimum and maximum correlation coefficients
Strong law of large number
Asymptotic normality
L-statistics

ABSTRACT

To generate correlated data for given marginal distributions, it is essential that the desired Pearson correlation coefficient is between the minimum and maximum correlation coefficients. In this paper, we consider estimation of the minimum and maximum correlation coefficients of continuous random variables X and$Y$. A strong law of large numbers and asymptotic normality are established for the estimators studied in this paper.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

In the clinical research, we often deal with correlated data. For example, in a clinical trial to treat healthy post-menopausal women with vasomotor symptoms, from the US Food and Drug Administration (FDA), the changes of the vasomotor symptoms from the baseline to week 4 and week 12 are recommended as the co-primary endpoints. Another example is about the informative censoring in the survival analysis. In this type of studies, the survival and censoring variables are not independent. Moreover, they may not have the same distributions, for instance, the survival distribution is Weibull and the censoring distribution is exponential.

Over the last decade, the Monte Carlo simulation has played an important role in the drug development (Chang, 2011). To carry out Monte Carlo simulation in the examples just mentioned above, the correlated data is generated and the statistical procedure is evaluated accordingly. For example, to determine the sample size for the treatment of healthy post-menopausal women with vasomotor symptoms, we need to take into account the correlation structures of the two co-primary endpoints (Sozu and others, 2006). Dukic and Marić (2013) and Xiang (2015) proposed algorithms to generate correlated data for given marginal distributions and a specified Pearson correlation coefficient (simply correlation coefficient). It is essential that the specified correlation coefficient should be between the minimum and maximum correlation coefficients. Estimation of the minimum and maximum correlation coefficients becomes important for generating correlated data. For some distributions, there are closed-form expressions for the minimum correlation coefficients (see Dukic and Marić, 2013; De Veaux, 1976). However, in most cases the closed-form expressions do not exist.

Let $\{X_i\}, i = 1, \ldots, n$, be iid with $X_i \sim F(x)$ and $\{Y_i\}, i = 1, \ldots, n$, be iid with $Y_i \sim G(x)$. A sorting algorithm was studied by DeGroot and Goel (1980) and was linked to the minimum and maximum correlation coefficients by Whitt (1976) and Demirtas and Hedeker (2011). The algorithm is based on the rearrangement theorem which can be found in

E-mail address: jxiang@its.jnj.com.

https://doi.org/10.1016/j.spl.2018.08.010
0167-7152/© 2018 Elsevier B.V. All rights reserved.

Please cite this article in press as: Xiang J.X., Estimation of minimum and maximum correlation coefficients. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.08.010.

Hardy et al. (1952, Theorem 368). To estimate the minimum correlation coefficient, we sort $\{X_i\}$ in ascending order and $\{Y_i\}$ in descending order and pair the sorted data to a bivariate sequence (X_i', Y_i'), $i = 1, 2, \ldots, n$. Thus the sample correlation coefficient of (X_i', Y_i') is an estimate of the minimum correlation coefficient. To estimate the maximum correlation coefficient, we sort both $\{X_i\}$ and $\{Y_i\}$ in ascending order instead. Demirtas and Hedeker (2011) showed numerically the estimators from the sorting algorithm is convergence to the minimum and maximum correlation coefficients. In this paper, we study the large sample property for a class of estimators including the estimators derived from the sorting algorithm. A strong law of large numbers and asymptotic normality are established.

The rest of the paper is organized as follows. In Section 2, we establish a strong law of large numbers. In Section 3 we will show the asymptotic normality. Some calculations of the minimum and maximum correlation coefficients are illustrated in Section 4. Some discussions are included in Section 5.

2. Strong law of large numbers

In this paper we assume that X and Y are continuous random variables. Denote by ρ_{min} and ρ_{max} the minimum and maximum correlation coefficients of X and Y. Let U be a random variable uniformly distributed on [0,1]. From Whitt (1976),

$$\rho_{min} = \frac{1}{\sigma_X \sigma_Y}\left(\int_0^1 F^{-1}(x)G^{-1}(1-x)dx - E(X)E(Y)\right) \tag{2.1}$$

and

$$\rho_{max} = \frac{1}{\sigma_X \sigma_Y}\left(\int_0^1 F^{-1}(x)G^{-1}(x)dx - E(X)E(Y)\right) \tag{2.2}$$

where σ_X^2 and σ_Y^2 are the variances of X and Y. Note that both ρ_{min} and ρ_{max} are determined by the marginal distributions regardless of the joint distribution of X and Y. The empirical distributions of F and G for given $\{X_i\}$, $i = 1, \ldots, m$, $\{Y_i\}$, $i = 1, \ldots, n$, are defined by

$$F_m(x) = \frac{1}{m}\sum_{i=1}^{m} I(X_i \le x) \quad \text{and} \quad G_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le x) \tag{2.3}$$

and the empirical quantile functions are

$$F_m^{-1}(t) = \inf\{x : F_m(x) \ge t\} \quad \text{and} \quad G_n^{-1}(t) = \inf\{x : G_n(x) \ge t\}. \tag{2.4}$$

Replacing the unknown quantities by their estimators in (2.1) and (2.2), we obtain the estimators of the minimum and maximum correlation coefficients

$$\hat\rho_{min} = \frac{1}{s_X s_Y}\left(\int_0^1 F_m^{-1}(x)G_n^{-1}(1-x)dx - \bar X \bar Y\right) \tag{2.5}$$

$$\hat\rho_{max} = \frac{1}{s_X s_Y}\left(\int_0^1 F_m^{-1}(x)G_n^{-1}(x)dx - \bar X \bar Y\right). \tag{2.6}$$

It can be verified that if $m = n$, $\hat\rho_{min}$ and $\hat\rho_{max}$ are equal to the estimators derived from the sorting algorithm described in Section 1. For the sake of simplicity, we assume $m = n$ in the theorems below. A brief discussion is provided for the case of $m \neq n$ in Section 5. The strong law of large numbers of $\hat\rho_{min}$ and $\hat\rho_{max}$ is stated as follows.

Theorem 1. *Let $\{X_i\}$, $i = 1, \ldots, n$, be iid with $X_i \sim F(x)$ and $\{Y_i\}$, $i = 1, \ldots, n$, be iid with $Y_i \sim G(x)$. Suppose that for some $\alpha < 1/2$ and $M > 0$,*

$$|F^{-1}(t)| \le M[t(1-t)]^{-\alpha}, \qquad 0 < t < 1, \tag{2.7}$$

$$|G^{-1}(t)| \le M[t(1-t)]^{-\alpha}, \qquad 0 < t < 1. \tag{2.8}$$

Then, as $n \to \infty$, with probability 1,

$$\hat\rho_{max} \longrightarrow \rho_{max} \text{ and } \hat\rho_{min} \longrightarrow \rho_{min}. \tag{2.9}$$

Remark 2.1. It can be verified that conditions (2.7) and (2.8) imply $E(X_1^2) < \infty$ and $E(Y_1^2) < \infty$. Thus from the classical strong law of large numbers, with probability 1, $\frac{1}{n}\sum_{k=1}^{n} X_k^2 \longrightarrow E(X_1^2)$ and $\frac{1}{n}\sum_{k=1}^{n} Y_k^2 \longrightarrow E(Y_1^2)$.

Proof of Theorem 1. We only provide the proof for $\hat\rho_{max} \longrightarrow \rho_{max}$. The second part of the theorem follows in the same matter. It is easy to see in (2.6), with probability 1, $\bar X \longrightarrow E(X_1)$ and $\bar Y \longrightarrow E(Y_1)$, and from Remark 2.1, both s_X and s_Y are strongly consistent estimators of σ_X and σ_Y. So it remains to show with probability 1,

$$\int_0^1 F_n^{-1}(t)G_n^{-1}(t)dt \longrightarrow \int_0^1 F^{-1}(t)G^{-1}(t)dt. \tag{2.10}$$

UNCORRECTED PROOF

Please cite this article in press as: Xiang J.X., Estimation of minimum and maximum correlation coefficients. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.08.010.

Download English Version:

https://daneshyari.com/en/article/10147249

Download Persian Version:

https://daneshyari.com/article/10147249

Daneshyari.com