PSYCHOMETRIKA
https://doi.org/10.1007/s11336-018-9603-3

GENERATING MULTIVARIATE ORDINAL DATA VIA ENTROPY PRINCIPLES

Yen Lee and David Kaplan

UNIVERSITY OF WISCONSIN - MADISON

When conducting robustness research where the focus of attention is on the impact of non-normality, the marginal skewness and kurtosis are often used to set the degree of non-normality. Monte Carlo methods are commonly applied to conduct this type of research by simulating data from distributions with skewness and kurtosis constrained to pre-specified values. Although several procedures have been proposed to simulate data from distributions with these constraints, no corresponding procedures have been applied for discrete distributions. In this paper, we present two procedures based on the principles of maximum entropy and minimum cross-entropy to estimate the multivariate observed ordinal distributions with constraints on skewness and kurtosis. For these procedures, the correlation matrix of the observed variables is not specified but depends on the relationships between the latent response variables. With the estimated distributions, researchers can study robustness not only focusing on the levels of non-normality but also on the variations in the distribution shapes. A simulation study demonstrates that these procedures yield excellent agreement between specified parameters and those of estimated distributions. A robustness study concerning the effect of distribution shape in the context of confirmatory factor analysis shows that shape can affect the robust χ^2 and robust fit indices, especially when the sample size is small, the data are severely non-normal, and the fitted model is complex.

Key words: Non-normal data generation, Entropy, Discrete data.

1. Introduction

The normality assumption frequently underlies statistical inferences, but real-world data sets are commonly far from normally distributed (Blair, 1981; Bradley, 1982; Micceri, 1989; Pearson & Please, 1975). The effects of violating normality on statistical inferences depend on many different factors, such as analysis method, the degree of non-normality, model complexity, sample size. Because of the complexity of the problem, robustness research conducted with Monte Carlo methods is commonly implemented to understand the effects of non-normality under different conditions.

Many different approaches have been proposed to generate multivariate non-normal data that follow a distribution with specific (marginal) skewness, $\beta_1 = \{\beta_{11}, \beta_{12}, \ldots, \beta_{1m}\}$, (marginal) kurtosis $\beta_2 = \{\beta_{21}, \beta_{22}, \ldots, \beta_{2m}\}$, and correlation matrix $\mathbf{\Sigma}$, where $\beta_1, \beta_2 \in \mathbb{R}^m$, and $\mathbf{\Sigma} \in \mathbb{R}^{m\text{x}m}$ when there are m variables (e.g. Fleishman, 1978; Headrick & Sawilosky, 1999; Headrick, 2010, Mair, Satorra, & Bentler, 2012; Mattson, 1997; Ruscio & Kaczetow 2008). Most approaches consist of three steps. First, estimating the univariate marginal distributions with specific β_1 and β_2. Second, linking the marginal distributions to a joint multivariate distribution with a specified $\mathbf{\Sigma}$. Third, generating data following the joint multivariate distribution.

Although the procedures described above provide great flexibility, most of them are designed to generate data only from continuous distributions. Among them, the method proposed by Ruscio and Kaczetow (2008) is the only one which is able to generate discrete ordinal data which are

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11336-018-9603-3) contains supplementary material, which is available to authorized users.

Correspondence should be made to Yen Lee, Department of Educational Psychology, University of Wisconsin - Madison, 859 Education Sciences, 1025 W. Johnson Street, Madison, WI 53706-1796, USA. Email: ylee373@wisc.edu

Published online: 22 January 2018

frequently seen in social and behavior research. However, these procedures have another important problem other than the lack of ability to generate discrete ordinal data. These procedures all specify correlations between the observed variables instead of between the underlying latent responses. In most situations, researchers are interested in the relationships/correlations among underlying latent responses (represented by $\boldsymbol{\Sigma}^*$) instead of the ones among the observed items (represented by $\boldsymbol{\Sigma}$). Since the data generated by all the procedures described above specify $\boldsymbol{\Sigma}$ but not $\boldsymbol{\Sigma}^*$, they are not suitable for generating discrete data typically observed in these fields and thus other data generating procedures are necessary.

1.1. General Discrete Data Generation

In the social and behavioral sciences, observed data are assumed to be generated by using coarse measurement instruments, in which a set of ordinal items with relatively few categories are used to measure continuous constructs and thus the observed data are discretely distributed (Bollen, 1989; Hipp & Bollen, 2003; Micceri, 1989; Weng & Cheng, 2004). In order to reflect the discreteness generated through this process, a three-step data generation approach is typically applied.

To generate data following an m-dimensional discrete $k = \{k_1, k_2, \ldots, k_m\}$ categories multivariate distribution, p, the first step is to select an m-dimensional continuous multivariate distribution, p^*, which represents the joint distribution of the underlying variables, y^* with specific parameters, such as a specific correlation matrix, $\boldsymbol{\Sigma}^*$ which obeys a structure of the relationships among y^*. The p^* and p are referred to as the latent response and observed ordinal responses distributions, respectively, following Muthén (1983) and Muthén (1984). The marginal latent response distributions of the marginal latent variables, y_j^*, are referred to as p_j^*, where $j = 1, 2, \ldots, m$.

The second step is to select a threshold set $\tau = \{\tau_1, \tau_2, \ldots, \tau_m\}$ of p^*. The $\tau_j = \left\{\tau_{0j}, \ldots, \tau_{k_j j}\right\}$ represents the threshold set of the p_j^*. The τ determines the relationship between the y^* and the observed variables, y. For the marginal latent response variable y_j^*'s and observed variables y_j's, the relationship is formalized as

$$y_j = i_j, \text{ if } \tau_{(i_j-1)j} < y_j^* \leq \tau_{i_j j}, \tag{1}$$

where $i_j = 1, 2, \ldots, k_j$ is the scores of the y_j. $\tau_{0j} \equiv -\infty$ and $\tau_{k_j j} \equiv \infty$. The probability of $y_j = i_j$ is referred to as $p_{i_j j} \equiv F_j(\tau_{i_j j}) - F_j(\tau_{(i_j-1)j})$, where F_j is the marginal cumulative density function (CDF) of y_j^*, and the $F_j(\tau_{i_j j})$ is the probability that $y_j^* \leq \tau_{i_j j}$.

Through the relationship in Eq. 1, τ_j determines each marginal probability mass function (PMF), p_j. Furthermore, τ determines the joint PMF, p, of y when p^* with all parameters specified is assumed. Therefore, the last step, to generate data following p, is commonly achieved by generating y^* following p^*, then transforming y^* to y according to Eq. (1).

The second step which decides p when p^* is specified is the main focus of this paper. When p^* is specified, p is fully determined by the selection of the τ and vice versa. Therefore, estimating p is equivalent to estimating τ in this condition.

There are many different possible candidates p^*'s and τ's in the literature. The most common p^* is the standard multivariate normal distribution (MVN). Many researchers have chosen τ's to yield specific p's to investigate the effect of non-normality according to specific response patterns (e.g. Ethington, 1987; Muthén & Kaplan, 1985; 1992; Yang-Wallentin, Joreskog, & Luo, 2010) or to yield p's having specific β_1's (e.g. Babakus, Ferguson, & Joreskog, 1987; Olsson, 1979). In recent years, researchers have started to select other distributions for p^*'s in robustness research. For example, Hipp and Bollen (2003) apply distributions estimated via the power method, and Flora and Curran (2004) apply the generalized lambda distributions combined with specific τ's to generate specific p's.

From the previous paragraph, we see that researchers frequently select the p by selecting the marginal p_j's, according to substantive considerations in robustness research. For example, Muthén and Kaplan (1985) selected the p_j's which are commonly encountered in the social and behavioral sciences when investigating the impact of ordinal data in exploratory factor analysis, a commonly applied method in the social and behavioral sciences. However, most conclusions from robustness research are presented in terms of β_1's and β_2's instead of p. We argue that this is a mismatch between the research conclusions and the manipulated factors since β_1 and β_2 are only part of the characteristics of p. There is no one-to-one relationship between the combination of β_{1j} and β_{2j} and p_j and thus p. Therefore, the conclusions obtained regarding β_1 and β_2 might be a special case but not a general rule.

1.2. Relationship Between the Specific β_{1j}'s and β_{2j}'s and p's

To eliminate the mismatch, we focus on estimating p with specific β_1 and β_2 which could be considered as the information we have about p. In order to do so, we need to discuss the properties of β_1, β_2 and their relationship with p. When p is specified, the β_1 and β_2 are specified. However, the inverse relationship is not true. The β_{1j} and β_{2j} are nonlinear expressions constructed from the marginal first four moments and could be the same if p's have the same first four moments but different higher moments.

When the number of categories, $k_j = 5$, the marginal first four moments of the p_j and p_j have a one-to-one relationship. When $k_j < 5$, the marginal first four moments of the p_j are determined when fewer moments are determined. For instance, the first four moments of Bernoulli distributions ($k_j = 2$) are determined if any of the odd moments is specified. Therefore, β_{1j} is sufficient to determine the p_j of Bernoulli distributions (Wilkins, 1944). Specifying both β_{1j} and β_{2j} is sufficient to determine the p_j when $k_j = 3$ in most cases (We will discuss this in "Appendix A"). Therefore, β_{1j} and β_{2j} are not sufficient to determine the p_j's when $k_j > 3$ in most conditions. Thus, there could be an infinite number of p_j's which have the same β_{1j} and β_{2j} when $k_j > 3$. This implies that there are an infinite number of p's with the same β_1 and β_2 even when the p^*'s are the same.

A consequence of an infinite number of p's with the same β_1 and β_2 is that it would impede generalizing the conclusions obtained from one distribution to all others. This problem could be partly solved if multiple p's having the same β_1 and β_2 could be analyzed when conducting Monte Carlo research. Then, the p's could be selected according to either (a) probability theory or (b) researchers' knowledge of the ideal shape of p. However, a review of the extant literature indicates there is no method that can generate p's systematically with a specific β_1 and β_2 combination according to either (a) or (b). Therefore, we propose two procedures that can estimate p's with the specific β_1 and β_2 based on two principles wherein researchers can obtain greater control and flexibility, and the results of Monte Carlo research could be more informative. The procedures we propose rest on the concepts of maximum entropy (MaxEnt) and minimum cross-entropy (MinxEnt) principle. The details of our procedures and the notation used in this paper are discussed in the following sections.

2. Notation

The general notation and the focused problem in this paper are defined in this section. Following the notation in the previous sections, p^* is used to indicate the m-dimensional continuous latent response distribution of y^*. Its probability density function (PDF) and CDF are denoted as $f(y_1^*, \ldots y_m^*)$ and $F(y_1^*, \ldots y_m^*)$, respectively. The p_j^* is the distribution of y_j^*. Its PDF is denoted as $f_j(y_j^*)$ and CDF is denoted as $F_j(y_j^*)$, where j is from 1 to m.

Similarly, p is the m-dimensional observed ordinal distribution of y. The p_j is the marginal distribution of y_j. The $p_j = \left\{p_{1j}, \ldots, p_{k_j j}\right\}$ indicates the marginal probability mass function (PMF). The $p_{i_j j}$ is the probability of $y_j = i_j$, where i_j is from 1 to k_j, the number of categories of y_j.

The CDF, $F_j(\tau_{i_j j})$ denotes the probability that $y_j^* \leq \tau_{i_j j}$. Thus, $p_{i_j j}$ also equals $F_j(\tau_{i_j j}) - F_j(\tau_{(i_j-1)j})$, which is the same as the probability of $\tau_{(i_j-1)j} < y_j^* \leq \tau_{i_j j}$. The $\tau_j = \left\{\tau_{0j}, \ldots, \tau_{k_j j}\right\}$ is the threshold set of p_j and $\tau_{0j} \leq \tau_{1j} \leq \tau_{2j} \ldots \leq \tau_{(k_j-1)j} \leq \tau_{k_j j}$.

The joint probability distribution $p = \left\{p_{i_1 i_2,\ldots,i_m}\right\}$; the i_j is from 1 to k_j and j is from 1 to m. The $p_{i_1 i_2,\ldots,i_m}$ is the probability of $y_1 = i_1, y_2 = i_2, \ldots, y_m = i_m$. It also equals the probability of $\tau_{(i_1-1)1} < y_1^* \leq \tau_{i_1 1}, \tau_{(i_2-1)2} < y_2^* \leq \tau_{i_2 2}, \ldots, \tau_{(i_m-1)m} < y_m^* \leq \tau_{i_m m}$. The $p_{i_1 i_2,\ldots,i_m}$ has to satisfy the normalizing constraint:

$$1 = \sum_{i_1=1}^{k_1} \ldots \sum_{i_j=1}^{k_j} \ldots \sum_{i_m=1}^{k_m} (p_{i_1 i_2 \ldots i_{j-1} i_j i_{j+1} \ldots i_m}) \tag{2}$$

The $p_{i_j j}$ of y_j is calculated by the following formula.

$$p_{i_j j} = \sum_{i_1=1}^{k_1} \ldots \sum_{i_{j-1}=1}^{k_{j-1}} \sum_{i_{j+1}=1}^{k_{j+1}} \ldots \sum_{i_m=1}^{k_m} (p_{i_1 i_2 \ldots i_{j-1} i_j i_{j+1} \ldots i_m}) \tag{3}$$

The target constraints of the β_1 and β_2 of p are defined marginally in p as follows (Ramachandran & Tsokos 2009):

$$\beta_{1j} = \frac{E(y_j - \mu_j)^3}{\sigma_j^3} \tag{4}$$

$$\beta_{2j} = \frac{E(y_j - \mu_j)^4}{\sigma_j^4} - 3 \tag{5}$$

where μ_j and σ_j are the mean and standard deviation of p_j. Eqs. (4) and (5) can be transformed into the following expressions for discrete variables (Lee, n.d.):

$$\beta_{1j} = \frac{\left(\sum_{i_j=1}^{k_j} i_j^3 p_{i_j j} - 3\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} \sum_{i_j=1}^{k_j} i_j p_{i_j j} + 2\left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^3\right)}{\left(\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} - \left(\sum_{i_j-1}^{k_j} i_j p_{i_j j}\right)^2\right)^{3/2}} \tag{6}$$

$$\beta_{2j} = \frac{\left(\sum_{i_j=1}^{k_j} i_j^4 p_{i_j j} - 4\sum_{i_j=1}^{k_j} i_j^3 p_{i_j j} \sum_{i_j=1}^{k_j} i_j p_{i_j j} + 6\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j}\left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2 - 3\left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^4\right)}{\left(\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} - \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2\right)^2} - 3 \tag{7}$$

As mentioned earlier, there are an infinite number of p's which can satisfy Eqs. (6) and (7). The details of the MaxEnt and MinxEnt procedures which we propose to select one of them are introduced in the following sections.

3. The Maximum Entropy (MaxEnt) Procedure

Shannon (1948) proposed the idea of *information entropy* (hereafter, entropy), $H(p)$, to quantify the uncertainty of probability distributions. For a one-dimensional or a marginal distribution p_j, Shannon's entropy is defined as

$$H_j(p_j) \equiv -\sum_{i_j=1}^{k_j} p_{i_j j} \ln(p_{i_j j}) \tag{8}$$

Note that when $p_{i_j j} = 0$, then $p_{i_j j} \ln(p_{i_j j})$ is defined as zero in Eq. (8). The term $\ln(p_{i_j j})$ assures that $p_{i_j j} \geq 0$ and therefore satisfies the inequality $\tau_{0j} \leq \tau_{1j} \leq \tau_{2j} \ldots \leq \tau_{(k_j-1)j} \leq \tau_{k_j j}$ simultaneously.

Equation (8) can be extended to measure the multivariate distribution p's by defining

$$H(p) \equiv -\sum_{i_1=1}^{k_1} \ldots \sum_{i_m=1}^{k_m} p_{i_1 \ldots i_m} \ln(p_{i_1 \ldots i_m}) \tag{9}$$

The entropy $H(p)$ reflects the amount of uncertainty in p and decreases when more information about p is known. The MaxEnt principle suggests choosing the MaxEnt distribution; the p that has maximum $H(p)$ among all the distributions subject to the same information/constraints. By choosing the MaxEnt distribution, we choose the most prudent distribution since it is the one estimated with the least information and having the greatest uncertainty (Golan, Judge, & Miller, 1997; Jaynes, 1957; Kapur & Kesavan, 1992). In addition, the MaxEnt distribution is the most frequently appearing and thus the most typical one in the probability distribution space of the distributions satisfying the constraints (Golan et al., 1997; Jaynes, 1982; Wu, 1997; Zellner & Highfield, 1988); moreover, the distribution tends to be flat (Jaynes 1957; 1982). Zero probabilities are avoided, which is consistent with the expectation of p's when we only have limited information. These characteristics make the MaxEnt principle a preferred guide to choose p when only β_1 and β_2 are specified.

In this paper, two estimation methods following the MaxEnt principle are proposed. The first method is the marginal approach which estimates the p_1 to p_m individually and the second method is a global approach which estimates the τ when p^* with all parameters are specified is assumed.

3.1. The Marginal Approach

For the marginal approach, we estimate p_j when β_{1j} and β_{2j} are specified, and thus the problem we are solving is defined as follows:

$$\underset{p_j}{\operatorname{argmax}}\; H_j(p_j), \tag{10}$$

subjects to the constraints

$$\beta_{1j} = c_{1j}, \beta_{2j} = c_{2j}, \sum_{i_j=1}^{k_j} p_{i_j j} = 1 \tag{11}$$

for $j = 1, 2, \ldots, m$. The c_{1j} and c_{2j} refer to the pre-specified β_{1j} and β_{2j} values. We solve the problem l times if there are l different combinations of $\left\{\beta_{1j}, \beta_{2j}, k_j\right\}$. For example, if we want to estimate a 5 dimensional p whose $\beta_1 = \{0, 0, 0, -1, -1\}$, $\beta_2 = \{0, 0, 0, 3, 3\}$ and $k = \{5, 5, 5, 6, 6\}$, we only need to solve the problem 2 times since there are only two different $\left\{\beta_{1j}, \beta_{2j}, k_j\right\}$ combinations: $\{0, 0, 5\}$ and $\{-1, 3, 6\}$.

The partial derivatives of $H_j(p_j)$, β_{1j}, and β_{2j} would be implemented in the optimization and are derived as follows:

$$\frac{\partial H_j(p_j)}{\partial p_{i_j j}} = -(1 + \ln p_{i_j j}) \tag{12}$$

$$\begin{aligned}\frac{\partial \beta_{1j}(p_j)}{\partial p_{i_j j}} = &\left\{\left[i_j^3 - 3i_j^2 \sum_{i_j=1}^{k_j} i_j p_{i_j j} - 3i_j \sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} + 6i_j \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2\right]\right.\\
&\left.\times \left[\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} - \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2\right]^{-\frac{3}{2}}\right\}\\
&- \left\{\frac{3}{2}\left[\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} - \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2\right]^{-\frac{5}{2}} \left[i_j^2 - 2i_j \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)\right]\right.\\
&\left.\left[\sum_{i_j=1}^{k_j} i_j^3 p_{i_j j} - 3\left(\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} \sum_{i_j=1}^{k_j} i_j p_{i_j j}\right) + 2\left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^3\right]\right\}\end{aligned} \tag{13}$$

$$\begin{aligned}\frac{\partial \beta_{2j}(p_j)}{\partial p_{i_j j}} = &\left\{\left[i_j^4 - 4i_j^3 \sum_{i_j=1}^{k_j} i_j p_{i_j j} - 4i_j \sum_{i_j=1}^{k_j} i_j^3 p_{i_j j} + 6i_j^2 \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2\right.\right.\\
&\left. + 12i_j \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)\left(\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j}\right) - 12i_j \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^3\right]\\
&\left.\times \left[\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} - \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2\right]^{-2}\right\}\\
&- \left\{2\left[\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} - \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2\right]^{-3} \left[i_j^2 - 2i_j \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)\right]\right.\\
&\times \left(\sum_{i_j=1}^{k_j} i_j^4 p_{i_j j} - 4\sum_{i_j=1}^{k_j} i_j^3 p_{i_j j} \sum_{i_j=1}^{k_j} i_j p_{i_j j} + 6\sum_{i_j=1}^{k_j} i_j^2 p_{i_j j} \left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^2\right.\\
&\left.\left. - 3\left(\sum_{i_j=1}^{k_j} i_j p_{i_j j}\right)^4\right)\right\}\end{aligned} \tag{14}$$

After p_j is estimated, τ_j is obtained by applying the inverse CDF function of the p_j^*. When m τ_j's are obtained, the τ is known and we can further categorize the p^* to yield p. When p is determined, its properties, such as $\boldsymbol{\Sigma}$, are decided.

3.2. The Global Approach

In the global approach, we focus on estimate τ instead of $p_{i_1 \dots i_m}$'s when β_1 and β_2 are specified to simplify the problem. Estimating τ reduces the number of unknowns we need to estimate for optimization. If we want to optimize Eq. (9) directly, we have to estimate $\prod_{j=1}^{m} k_j$ unknown $p_{i_1 \dots i_m}$'s in p. The number is very large when m and k are large, and thus makes the problem hard to be solved. In contrast, only $\sum_{j=1}^{m} (k_j - 1)$ unknown elements in τ need to be estimated. The difference between the number of unknowns shows that constraints which make $p_{i_1 \dots i_m}$ equals the probability of $\tau_{(i_1-1)1} < y_1^* \le \tau_{i_1 1}, \tau_{(i_2-1)2} < y_2^* \le \tau_{i_2 2}, \dots, \tau_{(i_m-1)m} < y_m^* \le \tau_{i_m m}$ are necessary to be imposed when solving $p_{i_1 \dots i_m}$'s. Without imposing the constraints, we will not be able to obtain a p whose relationship with p^* is defined by τ. However, adding constraints usually makes optimization problem harder to solve. Therefore, we estimate τ when p^* is assumed in the global MaxEnt approach. In this approach, the number of unknowns to be solved is relatively small and the relationships between p and p^* defined by τ can be guaranteed without any constraints other than Eqs. (6) and (7).

The problem we are solving here is defined as follows:

$$\underset{\tau}{\operatorname{argmax}}\ H(\tau) \tag{15}$$

subject to the constraints

$$\beta_{1j} = c_{1j}, \beta_{2j} = c_{2j}, \sum_{i_1=1}^{k_1} \dots \sum_{i_m=1}^{k_m} p_{i_1 \dots i_m} = 1, \tag{16}$$

$$\tau_{0j} \le \tau_{1j} \le \dots \le \tau_{k_j j}, \text{ where } j \text{ is from 1 to } m \tag{17}$$

The $H(\tau)$ is converted from $H(p)$ and is defined as follows:

$$\begin{aligned} H(\tau) = &- \sum_{i_1=1}^{k_1} \dots \sum_{i_m=1}^{k_m} \int_{\tau_{(i_1-1)1}}^{\tau_{i_1 1}} \dots \int_{\tau_{(i_m-1)m}}^{\tau_{i_m m}} f(y_1^*, \dots y_m^*) dy_m^* \dots dy_1^* \\ &\times \ln \left(\int_{\tau_{(i_1-1)1}}^{\tau_{i_1 1}} \dots \int_{\tau_{(i_m-1)m}}^{\tau_{i_m m}} f(y_1^*, \dots y_m^*) dy_m^* \dots dy_1^* \right) \end{aligned} \tag{18}$$

where $f(y_1^*, \dots y_m^*)$ is the joint PDF of y^* and $p_{i_1 \dots i_m} = \int_{\tau_{(i_1-1)1}}^{\tau_{i_1 1}} \dots \int_{\tau_{(i_m-1)m}}^{\tau_{i_m m}} f(y_1^*, \dots y_m^*) dy_m^* \dots dy_1^*$

The partial derivative of $H(\tau)$ is derived as follows:

$$\begin{aligned} \frac{\partial H(\tau)}{\partial \tau_{i_j j}} = &- \sum_{i_1=1}^{k_1} \dots \sum_{i_{(j-1)}}^{k_{(j-1)}} \sum_{i_{(j+1)}}^{k_{(j+1)}} \dots \sum_{i_m=1}^{k_m} \int_{\tau_{(i_1-1)1}}^{\tau_{i_1 1}} \dots \int_{\tau_{(i_{(j-1)}-1)(j-1)}}^{\tau_{i_{(j-1)}(j-1)}} \int_{\tau_{(i_{(j+1)}-1)(j+1)}}^{\tau_{i_{(j+1)}(j+1)}} \dots \int_{\tau_{(i_m-1)m}}^{\tau_{i_m m}} \\ &\times f(y_1^*, \dots, y_{j-1}^*, \tau_{ij}, y_{j+1}^*, \dots, y_m^*) dy_m^* \dots dy_{j+1}^* dy_{j-1}^* \dots dy_1^* \\ &\times \ln \left(\frac{\int_{\tau_{(i_1-1)1}}^{\tau_{i_1 1}} \dots \int_{\tau_{(i_j-1)j}}^{\tau_{i_j j}} \dots \int_{\tau_{(i_m-1)m}}^{\tau_{i_m m}} f(y_1^*, \dots, y_j^*, \dots, y_m^*) dy_m^* \dots dy_j^* \dots dy_1^*}{\int_{\tau_{(i_1-1)1}}^{\tau_{i_1 1}} \dots \int_{\tau_{i_j j}}^{\tau_{(i_j+1)j}} \dots \int_{\tau_{(i_m-1)m}}^{\tau_{i_m m}} f(y_1^*, \dots, y_j^*, \dots, y_m^*) dy_m^* \dots dy_j^* \dots dy_1^*} \right) \end{aligned} \tag{19}$$

The β_1 and β_2 constraints of Eqs. (6) and (7) could be shown as functions of τ by substituting $p_{i_j j}$ by $\int_{\tau_{(i_j-1)j}}^{\tau_{i_j j}} y_j^* dy_j$. The corresponding partial derivatives are as follows:

$$\begin{aligned}
\frac{\partial \beta_{1j}(\tau)}{\partial \tau_{i_j j}} = f_j(\tau_{i_j}) \Bigg\{ \Bigg[& (i_j^3 - i_{j+1}^3) - 3(i_j^2 - i_{j+1}^2) \sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \\
& - 3(i_j - i_{j+1}) \sum_{i_j=1}^{k_j} i_j^2 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \\
& + 6(i_j - i_{j+1}) \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^2 \Bigg] \\
& \times \left[\sum_{i_j=1}^{k_j} i_j^2 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j - \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^2 \right]^{-\frac{3}{2}} \\
& - \frac{3}{2} \left[\sum_{i_j=1}^{k_j} i_j^2 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j - \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^2 \right]^{-\frac{5}{2}} \\
& \times \left[(i_j^2 - i_{j+1}^2) - 2(i_j - i_{j+1}) \sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right] \\
& \times \Bigg[\sum_{i_j=1}^{k_j} i_j^3 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j - 3 \left(\sum_{i_j=1}^{k_j} i_j^2 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right) \\
& + 2 \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^3 \Bigg] \Bigg\}
\end{aligned} \tag{20}$$

$$\begin{aligned}
\frac{\partial \beta_{2j}(\tau)}{\partial \tau_{i_j j}} = f_j(\tau_{i_j}) \Bigg\{ \Bigg[& (i_j^4 - i_{j+1}^4) - 4(i_j^3 - i_{j+1}^3) \sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \\
& - 4(i_j - i_{j+1}) \sum_{i_j=1}^{k_j} i_j^3 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j + 6(i_j^2 - i_{j+1}^2) \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^2 \\
& + 12(i_j - i_{j+1}) \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right) \left(\sum_{i_j=1}^{k_j} i_j^2 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right) \\
& - 12(i_j - i_{j+1}) \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^3 \Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\times \left[\sum_{i_j=1}^{k_j} i_j^2 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j - \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^2 \right]^{-2} \\
&- 2 \left[\sum_{i_j=1}^{k_j} i_j^2 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j - \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^2 \right]^{-3} \\
&\times \left[(i_j^2 - i_{j+1}) - 2(i_j - i_{j+1}) \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right) \right] \\
&\times \left[\sum_{i_j=1}^{k_j} i_j^4 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j - 4 \sum_{i_j=1}^{k_j} i_j^3 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right. \\
&+ 6 \sum_{i_j=1}^{k_j} i_j^2 \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^2 \\
&\left. \left. - 3 \left(\sum_{i_j=1}^{k_j} i_j \int_{\tau_{(i_j-1)j}}^{\tau_{i_j}} y_j^* dy_j \right)^4 \right] \right\}
\end{aligned}
\tag{21}
$$

When optimizing, Eq. (18) is set as $-\infty$ when Eq. (17) is violated. After τ is estimated, p is determined by applying the CDF function of p^*. Once p is determined, its properties, such as $\mathbf{\Sigma}$, are known. The data are generated by simulating y^* then transforming them according to Eq. (1).

It is important to note that the entropy of multivariate distributions, $H(p)$, is a joint distributional measurement of uncertainty. Therefore, we expect that our global approach would be able to estimate the MaxEnt distribution, p, precisely. In contrast, our marginal approach which estimates p_j instead of p, in turn, would accelerate computational speed but will sacrifice the property of maximum entropy. The difference between $H(\tau)$ and $H(p_j)$, the entropy of p estimated by the marginal MaxEnt approach, is important when choosing the approach. To understand the difference, a numerical evaluation is conducted and discussed later.

4. The Minimum Cross-Entropy (MinxEnt) Procedure

When conducting robustness research situated in the social and behavioral sciences, most researchers have, in fact, the ideal distributional shapes which they wish to investigate, see e.g. Muthén and Kaplan (1985, in the context of factor analysis). It would be valuable if this information could be used to estimate p's with specific β_1 and β_2. When the ideal shapes are able to formulated as PMF's of ideal distributions, which we denote as, q, the Kullback–Leibler (1951) cross-entropy principle (MinxEnt) could be implemented to estimate the target p. The cross-entropy, $D(p:q)$, is a measure of the divergence between p and q.

The Kullback–Leibler's cross-entropy in dimension j is defined as:

$$
D(p_j; q_j) \equiv \sum_{i_j=1}^{k_j} p_{i_j j} \ln \left(\frac{p_{i_j j}}{q_{i_j j}} \right) \tag{22}
$$

where the p_j and q_j are the target and the ideal marginal PMF of the y_j, respectively. The $q_j = \{q_{1j}, \ldots, q_{k_j j}\}$. The term $\ln(\frac{p_{i_j j}}{q_{i_j j}})$ assures that $p_{i_j j} \geq 0$ and satisfies the inequality $\tau_{0j} \leq \tau_{1j} \leq \tau_{2j} \ldots \leq \tau_{k_j j}$. As with the MaxEnt procedure, $(p_{i_j j}) \ln(p_{i_j j}/q_{i_j j}) = 0$ when $p_{i_j j} = 0$. In this formula, every $q_{i_j j}$ is assumed to be greater than zero; otherwise, it would be the same as setting the q_j with a fewer number of categories.

Note that $D(p_j : q_j)$ decreases when the p_j is similar to the q_j and it decreases to zero if and only if $p_j = q_j$. When setting the ideal distribution as q_j, the p_j which is closest to the q_j and also satisfies the constraints of the β_{1j} and β_{2j} are estimated by following the MinxEnt principle. The MinxEnt principle suggests choosing the p that has the minimum $D(p_j : q_j)$ from the distribution space subject to the constraints. The MinxEnt procedure estimates the p_j instead of p since it is more common that the researchers' ideal distributions are marginal rather than joint.

When a q_j is specified, we estimate p_j by solving the problem defined as follows:

$$\underset{p_j}{\text{argmin}}\ D(p_j : q_j) \tag{23}$$

subjects to Eq. (11).

The partial derivative of $D(p_j : q_j)$ is derived as:

$$\frac{\partial D(p_j : q_j)}{\partial p_{i_j j}} = \ln p_{i_j j} - \ln q_{i_j j} + 1. \tag{24}$$

The partial derivatives of the constraints are the same as Eqs. (13) and (14). After $p_{i_j j}$ is estimated, $\tau_{i_j j}$ can be obtained by applying the inverse CDF of the y_j^*. The τ can further categorize p^* to get p if the PDF of y^* is known. Once p is determined, its properties, such as $\mathbf{\Sigma}$, are decided.

When we compare Eqs. (8) and (10) with Eqs. (22) and (23), we could see that the MaxEnt procedure is a special case of the MinxEnt procedure when we set uniform distribution as q_j. Since the setting of q_j represents our preference of the shape of p_j, this setting implies no preference of any category of the target PMF, and thus it also reflects the lack of information about p_j.

5. Computational Software: An R package

An R package (R Core Team, 2014) is provided to estimate p's and to generate data according to the MaxEnt and MinxEnt procedures. As there is no analytical solution to estimate p's or τ's for both procedures, a numerical optimization algorithm is necessary. Specifically, we solve the optimization problems via an augmented Lagrangian minimization algorithm (ALM) using the package "alabama" (Varadhan, 2015). The objective function of ALM is a combination of the Lagrangian function and the quadratic penalty function of the constraints. The quadratic penalty function largely preserves smoothness and thus facilitates optimization.

The ALM can be separated into inner iterative loops and outer iterative loops. The inner iterative loop optimizes the combination of the Lagrangian function and the quadratic penalty function with specified weight. The outer iterative loop increases the weight in each iteration. The details of this method can be found in Madsen, Nielsen, and Tingleff (2004) and in Nocedal and Wright (2006). The optimization process of inner loop stops when the relative tolerance (the difference between the objective function values of two consecutive iterations) is smaller than 10^{-8}, and the outer loop stops when the relative tolerance is smaller than 10^{-7}. The default starting values we applied are the $p_{i_j j}$ (marginal MaxEnt procedure and MinxEnt procedure) or τ_{ijj} (global MaxEnt procedure) of the uniform distribution, which is the MaxEnt distribution subject only to the normalizing constraint. It is worth noting that the feasible regions of the two optimization problems are disjoint due to the nonlinear constraints. The disjoint nature makes finding the global optimal solution a difficult task and heavily depends on the starting values. Therefore, multiple starting values are preferred when conducting the purposed procedures. In our package, multiple random starting points sets can be used to find the optimal solution.

In our global MaxEnt approach, the p^* is assumed to be the MVN. Its PDF and CDF are evaluated using the package "mvtnorm" (Genz et al., 2015) to calculate the Eq. (18) to Eq. (21). The "mvtnorm" package is developed especially for evaluating the CDF and PDF of the MVN. It implements the algorithm which evaluates several univariate CDF of the normal distribution instead of evaluating the high dimensional CDF. This transformation not only preserves the accuracy but also increases the efficiency (Genz & Bretz, 2002). The details of this algorithm could be seen in Genz (1992). Our R program is attached as an appendix, and the accuracy of the package is described in the next section.

6. Evaluation of the Procedures

In this section, we evaluate the accuracy of our marginal, global MaxEnt and MinxEnt procedures by the differences between the β_1 and β_2 of the estimated p and their specified values. The difference of $H(p)$ optimized via the marginal MaxEnt approach, $H(p_j)$, and the one via global MaxEnt approach, $H(\tau)$ was also compared.

As we mentioned, we expect that the $H(p_j)$ will be smaller than $H(\tau)$. We further expect that the deviance increases when the dependence increases, as in Kapur and Kesavan (1992) who showed that $H(p) \leq \sum_{j=1}^{m} H_j(p_j)$. The equality holds only when p_j's are independent and the deviance increases when the dependence increases. Therefore, the deviance amounts in different dependence conditions are evaluated to understand the applicability of the marginal MaxEnt procedure.

When the β_1 and β_2 are set the same across marginals, six bivariate five categories $p's$ with different β_1 and β_2 combinations $(\beta_1, \beta_2) = \{(0, 0), (0, 3), (-2, 3), (0, 10), (3, 10), (0, -1)\}$ are estimated by our marginal, global MaxEnt and MinxEnt procedures. We consider $\beta_2 = 10$, $\beta_2 = 3$ and $\beta_2 = 0$ or -1 as severe, moderate, and trivial non-normality. The non-normality level increases when absolute value of β_1 increases within the same β_2 value. The latent response distribution, p^*, is set as MVN with $\mu = 0$, $\sigma = 1$ when correlation varies from 0 to 0.9 in increments of 0.1. The varying degrees of correlation represent dependence in MVN which affects the dependence in p.

For the MinxEnt procedure, we specify seven different q_j's to evaluate the performance of the MinxEnt procedure. The specified q_j's are bimodal (Prior 2), truncated (Prior 3), unimodal symmetric with different leptokurtotic levels (Prior 4 and Prior 5) and unimodal left skewed with different skewed and leptokurtotic levels (Prior 6 to Prior 8). The specified q_j's are shown in the ideal distribution part in Table 1.

The results show that the β_1 and β_2 of the p estimated by our marginal, global MaxEnt and MinxEnt procedures are consistent with the specified values with absolute differences smaller than 10^{-7}. The deviances between $H(\tau)$ and $H(p_j)$ are consistent with our expectation, which means that the $H(\tau)$ is consistently greater than $H(p_j)$ except when correlation is 0. However, all the differences are under 0.1% of $H(\tau)$ even when the correlation is as high as .90. This implies that the marginal approach is quite robust and accurate under the conditions we specified.

In conclusion, all of our approaches are quite accurate under the condition we specified (p^* is bivariate normal distributed and p is a bivariate distribution with five categories). Regarding choosing between marginal and global MaxEnt procedures, we recommend applying the marginal

TABLE 1.
Five categories p_j's of six combinations of skewness and kurtosis with eight different ideal distributions (q_j).

p_j	p_{1j}	p_{2j}	p_{3j}	p_{4j}	p_{5j}
Ideal distribution (q_j)					
Prior1	20.00	20.00	20.00	20.00	20.00
Prior2	42.50	5.00	5.00	5.00	42.50
Prior3	5.00	42.50	5.00	42.50	5.00
Prior4	5.00	20.00	50.00	20.00	5.00
Prior5	5.00	7.50	75.00	7.50	5.00
Prior6	5.00	10.00	20.00	30.00	35.00
Prior7	2.50	5.00	15.00	30.00	47.50
Prior8	5.00	5.00	5.00	10.00	75.00

p_j	p_{1j}	p_{2j}	p_{3j}	p_{4j}	p_{5j}	p_{1j}	p_{2j}	p_{3j}	p_{4j}	p_{5j}
Estimated distribution (p_j)										
β_1, β_2			0,0					0,3		
Prior1	8.41	16.54	50.09	16.54	8.41	3.64	9.14	74.45	9.14	3.64
Prior2	12.47	9.08	56.89	9.08	12.47	0.23	0.00	0.01	67.81	31.96
Prior3	5.38	20.66	47.92	20.66	5.38	0.06	0.24	6.11	78.73	14.87
Prior4	5.21	20.83	47.92	20.83	5.21	2.60	10.42	73.96	10.42	2.60
Prior5	11.98	10.06	55.91	10.06	11.98	4.76	7.33	75.81	7.33	4.76
Prior6	1.72	4.84	47.47	30.03	15.93	0.20	0.00	2.36	71.95	25.49
Prior7	0.05	0.00	3.64	62.42	33.88	0.22	0.00	0.94	69.61	29.23
Prior8	0.09	0.00	0.03	57.30	42.57	0.23	0.00	0.01	67.81	31.96
β_1, β_2			− 2,3					0,10		
Prior1	7.39	2.11	3.28	21.73	65.49	1.56	4.39	88.11	4.39	1.56
Prior2	6.29	1.58	8.28	12.19	71.65	0.31	0.00	0.01	82.39	17.29
Prior3	4.84	7.95	6.93	78.93	1.36	11.14	86.21	2.43	0.04	0.18
Prior4	5.98	14.27	79.58	0.17	0.00	1.20	4.81	87.98	4.81	1.20
Prior5	9.21	7.31	82.27	1.21	0.00	1.80	4.04	88.32	4.04	1.80
Prior6	6.52	3.02	3.65	19.61	67.20	0.28	0.04	0.48	83.32	15.88
Prior7	5.96	3.45	4.24	17.82	68.54	0.30	0.00	0.31	82.84	16.55
Prior8	3.91	4.66	6.49	12.03	72.91	0.31	0.00	0.07	82.47	17.16
β_1, β_2			− 3,10					0,− 1		
Prior1	3.11	0.31	1.42	22.17	72.99	15.23	20.32	28.90	20.32	15.23
Prior2	1.90	0.33	4.83	13.46	79.48	20.41	10.10	38.97	10.11	20.41
Prior3	2.58	2.80	5.72	86.54	2.36	8.23	30.81	21.91	30.81	8.23
Prior4	2.63	7.40	89.26	0.70	0.00	0.47	27.84	43.38	27.84	0.47
Prior5	3.93	4.83	89.99	1.25	0.00	20.30	10.33	38.74	10.33	20.30
Prior6	2.87	0.49	1.73	20.96	73.95	0.91	15.73	36.12	23.63	23.61
Prior7	2.60	0.63	2.18	19.49	75.10	0.12	1.99	35.60	35.86	26.43
Prior8	1.31	1.09	3.99	13.45	80.16	6.73	19.34	35.68	12.56	25.69

The $p_{i_j} j$ indicate the probability of $y_j = i_j$.
The numbers shown are estimated to the sixth decimal place but rounded to the second decimal place.

approach especially when m and k are large and p^* is assumed to be MVN. In this condition, our marginal approach leads to trivial deviances even when y_j^*'s are highly dependent but is significantly faster than the global approach.

7. Numerical Study

7.1. Design

A robustness study was conducted in order to (a) evaluate the quality of data generated from the proposed procedures, and to (b) illustrate the application of our procedures. Specifically, confirmatory factor analysis (CFA) was chosen as the illustrative model insofar as it has been the subject of most robustness research in the social and behavioral sciences. In this study, the p^*'s were the MVN. The five categories p's were evaluated because they are commonly encountered in social and behavioral science research and are well studied (Flora & Curran, 2004; Muthén & Kaplan, 1985).

The p's were estimated under four experimental conditions: (a) model complexity, (b) level of non-normality, (c) distribution shape, and (d) sample size. Four models were evaluated to investigate the model complexity condition. Model 1 was a 1-factor model measured with five indicators; Model 2 was a 1-factor model measured with ten indicators; Model 3 was a 2-factor model, with each factor measured by five indicators; and Model 4 was a 2-factor model with each factor measured by ten indicators. The factor loadings were 0.7 across all models, no error correlations were specified, and the factor correlation was 0.6 if applicable. According to the setting, r^*'s (the elements of $\boldsymbol{\Sigma^*}$) have only two different values: 0.49, for items load on the same factor, and 0.294, for items load on different factors.

The degree of non-normality and q_j were specified as in the evaluation of the procedures section. The specified (β_1, β_2) combinations were selected to allow examination of their separate effects. The estimated p_j's of different β_{1j}, β_{2j} and q_j are shown in Table 1 in which the distributions estimated with Prior 1 are the MaxEnt distributions and the ones estimated with Prior 2 to Prior 8 are the MinxEnt distributions. The marginal approach described earlier was used to estimate the MaxEnt distributions.

Random samples of five different sizes: 100, 200, 500, 1000 and 2000 were drawn from each estimated p. This study had the following experimental conditions: four models $\times$ six β_1 and β_2 combinations $\times$ eight distribution shapes $\times$ five sample sizes $=$ 960 independent cells. Data were generated and fitted 1000 times in each cell of the design.

Two analyses were conducted. In Analysis I, we focused on evaluating the parameters of the estimated p's. The precision of β_1 and β_2 of p were examined first to confirm the accuracy of our procedures. After accuracy was confirmed, the impact of the distribution shape on $\boldsymbol{\Sigma}$ was evaluated by deviance (D) and relative deviance (RD). The D and RD are defined as $r - r^*$ and $\frac{r-r^*}{|r^*|} \times 100\%$, where r and r^* denote the element in $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma^*}$, respectively.

In addition, we analyzed the characteristics of the sampling distributions of $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{r}$ to understand the data generated under different conditions by computing the mean and standard deviation (SD) of bias (B), $\widehat{\theta} - \theta$, and relative bias (RB), $\frac{\widehat{\theta}-\theta}{|\theta|} \times 100\%$. When θ was 0, the RB was defined as $\widehat{\theta} \times 100\%$. The θ usually denoted the parameters of p, and the $\widehat{\theta}$ denoted the corresponding estimates. However, we compared $\widehat{\boldsymbol{\Sigma}}$ with $\boldsymbol{\Sigma^*}$ instead of $\boldsymbol{\Sigma}$ since the researcher is interested in the relationships between the underlying constructs. Consistent with previous simulation studies conducted by Flora and Curran (2004), the absolute mean RB less or equal to 5% indicated trivial bias; the one between 5 and 10% indicated moderate bias; and the one greater than 10% indicated substantial bias.

In Analysis II, the data were fitted with the correctly specified CFA models to examine the effects of the four design conditions on the fit statistics. All of the models were fitted with the R package lavaan (Rosseel, 2012) with diagonal weighted least squares with mean and variance adjusted test statistic estimation (WLSMV). The estimator corrects the effects of non-normality and discreteness on the χ^2 values and the fit indices (Asparouhov & Muthén, 2010) and is widely recommended when analyzing non-normal ordinal data (DiStefano & Morgan, 2014; Flora & Curran, 2004). The RB and the empirical Type I error rate (rejection rate when nominal α-level was set as .05, RR) of the robust χ^2 statistic were evaluated. The robust CFI, TLI , NFI, and RMSEA (rCFI, rTLI, rNFI and rRMSEA, respectively), which were computed by the corrected robust χ^2 (Jorgensen, 2016), were evaluated by their acceptance rate (AR) with the cutoff values set as .95, .95, .95, and .08, respectively (Hooper, Coughlan, & Mullen, 2008; Browne & Cudeck, 1993). We further presented the values with 5% type I error rate (95% quantile for rRMSEA and 5% quantile for the other fit indices).

7.2. *Analysis I: The Quality of the Estimated p and the Generated Data*

7.2.1. The Parameters of the Estimated p Consistent with previous results, the β_1 and β_2 of the estimated p's were precisely estimated across all experimental conditions. In contrast, the r's were all attenuated compared to r^*'s under all conditions, as expected.

The D and RD of r are summarized by their values. The r's were attenuated more when the non-normal level increased. The RD varied from $-8.04\%(\beta_1 = 0, \beta_2 = -1)$ to $-47.87\%(-3, 10)$when$r = 0.49$, and it varied from $-8.65\%(0, -1)$ to $-55.05\%(-3, 10)$ when$r = 0.294$. The ranges (the maximum value minus the minimum value) of RD with the same non-normal levels varied from -2.09% (0, 10) to -23.77% (0, 0) when $r = 0.49$ and from -1.07% (0, 10) to -25.30% (0, 0) when $r = 0.294$. The large range of RD when $\beta_1 = 0$ and $\beta_2 = 0$ implied that this condition has the largest differentiating effect across different shapes. The trends of the D and RD when $r = 0.294$ or $r = 0.49$ were similar. The D and RD of r are shown in Table 2.

7.2.2. Characteristics of the Sampling Distributions of $\widehat{\beta_1}$, $\widehat{\beta_2}$ and $\widehat{\boldsymbol{\Sigma}}$ To save space, the results of $\widehat{\beta_1}$, $\widehat{\beta_2}$ of Model 1 and $\widehat{\boldsymbol{\Sigma}}$ of Model 3 are presented. The B's and RB's of β_1 and β_2 are summarized by their values. The absolute mean RB's of the $\widehat{\beta_1}$ and $\widehat{\beta_2}$ increased as sample size decreased, and as the level of non-normality increased, but they were not affected by the model complexity, as expected. When the sample size was 100, the absolute mean RB ranged from 0.02% (0, -1) to 79.08% (0, 10) for the $\widehat{\beta_1}$ and from 0.19% (0, -1) to 39.00% (0, 10) for the $\widehat{\beta_2}$. Across all conditions, 10 out of 48 (22.92%) mean RB's of $\widehat{\beta_1}$ were substantial and 6 out of 48 (12.50%) mean RB's of $\widehat{\beta_1}$ were moderate. Similarly, 12 out of 48 (25.00%) mean RB's of $\widehat{\beta_2}$ were substantial and 6 out of 48 (12.50%) mean RB's of $\widehat{\beta_2}$ were moderate. The absolute values of mean RB's decreased sharply when sample size increased. When sample size was 2000, all except one of the absolute mean RB's were trivial; the maximum absolute mean RB of $\widehat{\beta_1}$ was 5.18% (0, 10), and the one of $\widehat{\beta_2}$ was 3.20% (0, 0). The SD of RB also decreased sharply with increasing sample size which is expected.

The shapes of p's also affected the mean RB's. Only trivial absolute mean RB's of $\widehat{\beta_1}$ were observed if data followed the p's which were estimated with Prior 4 and Prior 5 even when sample size was 100. For $\widehat{\beta_2}$, similar situation was observed when data followed p's were estimated with Prior 1. The ranges of mean RB's of $\widehat{\beta_1}$ with the same non-normality level varied from 1.56% (0, -1) to 140.11% (0, 10) and the one of $\widehat{\beta_2}$ varied from 2.12% (0, -1) to 37.49% (0, 10) when sample size was 100. When sample size increased to 2000, the range of mean RB with the same non-normality level varied from 0.09% (0, -1) to 10.11% (0, 10) for $\widehat{\beta_1}$ and ranged from 0.18%

TABLE 2.
The deviation (D) and relative deviation (RD) of different r under different conditions.

r=.49						
β_1, β_2	0,0		0,3		−2,3	
	D	RD	D	RD	D	RD
Prior1	− 0.05	− 11.16	− 0.13	− 27.12	− 0.13	− 26.99
Prior2	− 0.08	− 15.67	− 0.19	− 37.79	− 0.14	− 27.67
Prior3	− 0.05	− 10.66	− 0.17	− 35.60	− 0.16	− 32.63
Prior4	− 0.05	− 10.71	− 0.13	− 26.44	− 0.17	− 33.88
Prior5	− 0.07	− 14.90	− 0.14	− 29.04	− 0.19	− 38.26
Prior6	− 0.07	− 14.19	− 0.17	− 35.02	− 0.13	− 26.77
Prior7	− 0.14	− 29.11	− 0.18	− 36.31	− 0.13	− 26.72
Prior8	− 0.17	− 34.43	− 0.19	− 37.80	− 0.13	− 27.31
β_1, β_2	0,10		− 3, 10		0,−1	
Prior1	− 0.22	− 45.42	− 0.16	− 33.19	− 0.04	− 8.04
Prior2	− 0.23	− 47.08	− 0.16	− 32.96	− 0.06	− 11.37
Prior3	− 0.23	− 46.69	− 0.21	− 41.93	− 0.05	− 9.37
Prior4	− 0.22	− 44.99	− 0.22	− 45.31	− 0.08	− 15.90
Prior5	− 0.23	− 45.99	− 0.23	− 47.87	− 0.06	− 11.26
Prior6	− 0.23	− 46.67	− 0.16	− 32.87	− 0.05	− 11.00
Prior7	− 0.23	− 46.70	− 0.16	− 32.57	− 0.07	− 15.30
Prior8	− 0.23	− 47.00	− 0.16	− 32.90	− 0.05	− 10.59
r=.294						
β_1, β_2	0,0		0,3		−2,3	
	D	RD	D	RD	D	RD
Prior1	− 0.03	− 11.26	− 0.08	− 28.84	− 0.10	− 32.47
Prior2	− 0.05	− 15.81	− 0.12	− 39.65	− 0.10	− 33.75
Prior3	− 0.03	− 10.71	− 0.11	− 37.15	− 0.11	− 38.26
Prior4	− 0.03	− 10.76	− 0.08	− 28.09	− 0.12	− 39.83
Prior5	− 0.04	− 15.04	− 0.09	− 30.84	− 0.13	− 43.50
Prior6	− 0.04	− 14.46	− 0.11	− 36.47	− 0.10	− 32.48
Prior7	− 0.09	− 29.73	− 0.11	− 37.95	− 0.10	− 32.59
Prior8	− 0.11	− 36.01	− 0.12	− 39.66	− 0.10	− 33.64
β_1, β_2	0,10		− 3, 10		0,−1	
Prior1	− 0.15	− 50.63	− 0.12	− 39.89	− 0.03	− 8.65
Prior2	− 0.15	− 50.65	− 0.12	− 40.97	− 0.04	− 12.06
Prior3	− 0.15	− 50.52	− 0.14	− 49.06	− 0.03	− 9.75
Prior4	− 0.15	− 50.15	− 0.16	− 52.97	− 0.05	− 16.36
Prior5	− 0.15	− 51.22	− 0.16	− 55.05	− 0.04	− 11.95
Prior6	− 0.15	− 50.34	− 0.12	− 39.80	− 0.03	− 11.56
Prior7	− 0.15	− 50.32	− 0.12	− 39.78	− 0.05	− 15.93
Prior8	− 0.15	− 50.58	− 0.12	− 41.11	− 0.03	− 11.20

The numbers shown are rounded to the second decimal place.

TABLE 3.
Mean and standard deviation (SD) of bias (B) and relative bias (RB) of $\widehat{\beta_1}$ in each condition when sample size is 100 in 1000 replications.

β_1, β_2	0,0				0,3			
	B	(SD)	RB	(SD)	B	(SD)	RB	(SD)
Prior1	0.00	0.15	0.17	15.10	− 0.00	0.47	− 0.13	47.46
Prior2	− 0.00	0.11	− 0.14	10.58	0.32	0.92	**32.43**	92.37
Prior3	0.00	0.18	0.27	18.20	0.15	0.57	**15.43**	57.35
Prior4	0.00	0.18	0.17	18.43	− 0.01	0.50	− 1.29	50.06
Prior5	− 0.00	0.11	− 0.10	11.21	0.00	0.46	0.09	45.52
Prior6	0.03	0.24	3.21	24.45	0.27	0.82	**26.65**	81.56
Prior7	0.07	0.39	*7.48*	38.68	0.28	0.89	**27.67**	89.29
Prior8	0.14	0.56	**13.66**	55.72	0.30	0.93	**30.43**	92.79
β_1, β_2	− 2,3				0,10			
Prior1	0.01	0.31	0.53	15.25	− 0.03	1.33	− 3.19	133.31
Prior2	0.02	0.31	0.89	15.49	0.79	1.64	**79.08**	164.50
Prior3	0.02	0.35	0.81	17.42	− 0.61	1.43	**− 61.03**	142.57
Prior4	0.02	0.34	0.91	17.05	0.02	1.35	1.75	135.31
Prior5	− 0.02	0.39	− 0.84	19.46	− 0.01	1.37	− 0.69	136.72
Prior6	0.01	0.31	0.55	15.45	0.77	1.60	**77.13**	160.28
Prior7	0.01	0.31	0.60	15.74	0.77	1.61	**77.02**	161.39
Prior8	0.02	0.32	0.80	16.15	0.78	1.63	**78.41**	162.71
β_1, β_2	− 3,10				0,− 1			
Prior1	0.16	0.43	*5.21*	14.45	− 0.00	0.13	− 0.16	12.81
Prior2	0.18	0.51	*6.02*	17.02	0.00	0.11	0.13	11.17
Prior3	0.07	0.60	2.33	19.89	0.00	0.16	0.17	15.56
Prior4	0.04	0.70	1.49	23.44	− 0.00	0.15	− 0.12	15.08
Prior5	0.04	0.73	1.20	24.46	0.00	0.11	0.02	11.09
Prior6	0.16	0.45	*5.25*	14.92	0.01	0.15	0.90	14.72
Prior7	0.16	0.46	*5.38*	15.27	0.01	0.18	1.40	18.28
Prior8	0.18	0.53	*6.11*	17.65	0.01	0.14	0.93	13.64

The numbers shown are rounded to the second decimal place. Substantial RB's are shown in bold and moderate RB's are shown in italic.

(0, − 1) to 3.33% (0, 3) for $\widehat{\beta_2}$. The mean and SD of RB of $\widehat{\beta_1}$ and $\widehat{\beta_2}$ under all conditions when sample size was 100 are shown in Tables 3 and 4.

Because the patterns of B's and RB's when $r^* = .49$ or $r^* = .294$ were very similar, they were analyzed together. All but one of the mean RB's of $\widehat{\boldsymbol{\Sigma}}$ were substantial across all conditions and showed a consistent pattern with the deviation of the $\boldsymbol{\Sigma}$. The decrease depended mainly on the degree of non-normality and not on the model complexity or sample size. The mean RB's decreased more when the level of non-normality increased. The average mean RB's of all the $\widehat{r}$ varied from −8.59% (0, − 1) to − 51.91% (− 3, 10). Within each non-normality level, the range of the mean RB of $\widehat{\boldsymbol{\Sigma}}$ was from 3.78% (0, 10) to 25.38% (0, 0).

7.3. *Analysis II: The Fitted CFA Model*

7.3.1. Improper Solutions For this paper, improper solutions included non-converged solutions and Heywood cases. Improper solutions were excluded in the following analyses. The percentage of improper solutions decreased with an increase in sample size and a decrease in non-normality

TABLE 4.
Mean and standard deviation (SD) of bias (B) and relative bias (RB) of $\widehat{\beta_2}$ in each condition when sample size is 100 in 1000 replications.

β_1, β_2	0,0				0,3			
	B	(SD)	RB	(SD)	B	(SD)	RB	(SD)
Prior1	− 0.04	0.32	− 1.37	10.73	− 0.10	1.08	− 1.64	17.95
Prior2	− 0.02	0.38	− 0.76	12.61	− 2.27	4.20	− **37.90**	69.95
Prior3	− 0.07	0.29	− 2.44	9.70	− 0.65	2.12	− **10.76**	35.39
Prior4	− 0.07	0.29	− 2.19	9.81	− 0.17	1.06	− 2.89	17.71
Prior5	− 0.02	0.38	− 0.55	12.72	− 0.04	1.19	− 0.60	19.87
Prior6	− 0.12	0.42	− 3.96	14.10	− 1.65	3.42	− **27.44**	56.93
Prior7	− 0.38	1.53	− **12.56**	51.00	− 1.90	3.91	− **31.72**	65.15
Prior8	− 1.03	2.74	− **34.41**	91.21	− 2.23	4.23	− **37.10**	70.46
β_1, β_2	− 2,3				0,10			
Prior1	0.17	1.67	2.81	27.81	− 0.49	3.87	− 3.74	29.76
Prior2	0.10	1.59	1.67	26.44	− 5.02	6.65	− **38.58**	51.12
Prior3	0.09	1.73	1.47	28.87	− 3.70	5.81	− **28.48**	44.71
Prior4	0.06	1.62	1.02	26.97	− 0.70	3.77	− *5.36*	28.99
Prior5	0.30	2.02	*5.01*	33.69	− 0.20	4.08	− 1.51	31.39
Prior6	0.16	1.66	2.71	27.66	− 4.72	6.40	− **36.34**	49.26
Prior7	0.14	1.67	2.41	27.85	− 4.96	6.47	− **38.12**	49.80
Prior8	0.08	1.63	1.40	27.17	− 5.07	6.56	− **39.00**	50.46
β_1, β_2	− 3,10				0,− 1			
Prior1	− 0.52	3.48	− 4.02	26.79	− 0.01	0.15	− 0.53	7.54
Prior2	− 1.09	3.95	− *8.42*	30.40	− 0.01	0.18	− 0.27	8.91
Prior3	0.30	4.91	2.34	37.77	− 0.03	0.12	− 1.45	5.84
Prior4	0.14	4.88	1.11	37.56	− 0.03	0.26	− 1.29	13.19
Prior5	0.70	5.73	*5.39*	44.04	− 0.00	0.18	− 0.19	9.02
Prior6	− 0.57	3.53	− 4.40	27.17	− 0.03	0.19	− 1.35	9.70
Prior7	− 0.66	3.62	− *5.11*	27.88	− 0.05	0.33	− 2.31	16.68
Prior8	− 1.20	4.13	− *9.25*	31.80	− 0.02	0.16	− 0.99	8.24

The numbers shown are rounded to the second decimal place. Substantial RB's are shown in bold and moderate RB's are shown in italic.

level or model complexity. Similar findings have been found in Flora and Curran (2004). When $\beta_2 \leq 3$, the percentages of improper solutions were all less than 7%. For Model 1 to Model 4 respectively, the highest improper solution percentages were 18.20% (− 3, 10), 20.20% (0, 10), 56.40% (− 3, 10) and 59.60% (0, 10) when sample size was 100. When sample size increased, the percentage of improper solutions decreased sharply. When sample size was 500, there was only one improper solution across all conditions.

The distribution shape affected the percentage of improper solutions, especially when sample size was small, non-normality level was high and model was complex. When $\beta_2 \leq 3$, the ranges of improper solutions percentages of the same β_1 and β_2 were less than 10%, even for the most complex model. When $\beta_2 = 10$, the ranges of percentages of improper solutions within the same β_1 and β_2 expanded a lot. The largest ranges were 18.10% (− 3, 10), 16.70% (0, 10), 56.00% (− 3, 10) and 56.20% (0, 10) for Model 1 to Model, 4 respectively. The p's estimated with Prior 4 and Prior 5 consistently showed higher improper solution percentages.

7.3.2. RB and RR of Robust χ^2 As mentioned, the mean RB and RR of the robust χ^2 were examined to evaluate the effect of the experimental factors. The RR is considered acceptable if under .10. The degrees-of-freedom of the four models, 5, 35, 34, and 169, were the expected χ^2 values and were used as θ's to calculate RB's.

The mean RB and the RR of the robust χ^2 decreased as model complexity decreased, sample size increased, and level of non-normality decreased. The RB and RR generally behaved better when the β_1 and β_2 values were (0, 0) or (0, −1) than the other combinations in all conditions. Most of substantial mean RB's and unacceptable RR's occurred when $\beta_2 = 10$ across all model complexity with small sample size. When sample size increased to 500, there were only trivial mean RB's and the RR's behaved well in all conditions.

The range of mean RB and RR of the robust χ^2 with the same β_1 and β_2 was smaller when the model complexity decreased, the level of non-normality decreased and sample size increased. When sample size was 100, the largest ranges occurred when the $\beta_1 = 0$ and $\beta_2 = 10$ across all models. In Model 1 to Model 4, the largest range of mean RB (RR) with the same β_1 and β_2 were 91.43% (.23), 100.36% (.50), 153.33% (.72), and 122.31% (.83) when sample size was 100. When sample size was over 500, all the ranges of mean RB and RR were smaller than 10% and 0.10. The details for sample size 100 and 200 are shown in Tables 5 and 6.

7.3.3. AR of Commonly Used Cutoff Criteria and 95% Quantile rRMSEA and 5% Quantile of rCFI, rTLI, and rNFI The percentage of acceptance rates of rRMSEA, rCFI, rTLI and rNFI was essential to model evaluation via applying the recommended cutoff. A fit index was considered acceptable if the AR was greater than .90 and the deviance between the 95%/5% quantile and the recommended cutoff is smaller than .05.

The trends in the rRMSEA, rCFI, rTLI and rNFI were similar to the trend of the robust χ^2. In the other words, all the fit indices performed better when sample size increased, the level of non-normality decreased and model complexity decreased. Across all models, the rRMSEA and rCFI were relatively robust among the fit indices we examined. They tended to have the highest AR and their 95%/5% quantiles have lowest deviance from the recommended cutoff in all conditions among the four fit indices. Followed by the rTLI. The rNFI was the most vulnerable fit index.

The trend in the effect of distribution shape on the fit indices decreased when sample size increased, the level of non-normality decreased and model complexity decreased. The rRMSEA was also the most robust fit index among the four indices in the range of AR and 95%/5% quantile with the same β_1 and β_2. The rRMSEA was not especially robust to the effect of distribution shape when sample size was 100. However, it became the most robust one among the four fit indices with the smallest range of AR and the 95% quantile with the same β_1 and β_2 when the sample size increased. A summary of the results when sample size was 100 and 200 is shown in Table 7.

For the simplest model, Model 1, only 6 (12.50%) 5% quantiles and 9 (18.75%) AR's of rCFI were not in the acceptable range and they all had $\beta_2 = 10$ when sample size was 100. In contrast, none of rRMSEA has AR in the acceptable range across all the non-normality levels with this limited sample size. Although all of AR's were lower than 0.90, only 9 (18.75%) 95% quantile were greater than 0.13 and they all occurred when $\beta_2 = 10$. Except $\beta_1 = 0$ and $\beta_2 = -1$, the rTLI and rNFI performed badly in at least one case in all the non-normality levels.

When sample size increased to 200, the rRMSEA and rCFI performed well across all the non-normality levels. The rTLI only performed badly when $\beta_2 = 10$. In addition to $\beta_2 = 10$, the rNFI also performed badly in one case when $\beta_1 = -2$ and $\beta_2 = 3$. For rTLI and rNFI, respectively, 2 (4.17%) and 1 (2.08%) of 5% quantile, and 7 (14.58%) and 12 (25.00%) of AR's were not in the accepted range. When sample size was 500, the fit indices were all performed well.

TABLE 5.
Mean RB and RR of robust χ^2 statistics of Model 1 and Model 2 with sample size 100 to 200 in 1000 replications.

β_1, β_2	0,0		0,3		−2,3		0, 10		− 3, 10		0,-1	
	RB	RR	RB	RR	RB	RR	RB	RR	RB	RR	RB	RR
Model 1												
$n = 100$												
Prior1	− 0.11	0.05	**11.39**	0.08	7.16	0.06	**78.53**	**0.24**	**18.43**	0.10	2.17	0.05
Prior2	− 1.84	0.04	− 1.13	0.04	**14.30**	0.07	1.94	0.04	**23.82**	**0.10**	1.13	0.05
Prior3	4.55	0.06	− 3.10	0.06	**10.37**	0.06	**12.02**	0.09	**58.90**	**0.19**	1.48	0.06
Prior4	0.41	0.04	**12.91**	0.09	6.27	0.06	**90.53**	**0.27**	**49.83**	**0.20**	1.90	0.04
Prior5	0.11	0.06	3.38	0.06	6.61	0.06	**75.56**	**0.22**	**87.61**	**0.28**	0.07	0.05
Prior6	5.35	0.07	− 2.39	0.06	**11.09**	0.06	1.86	0.05	**18.21**	0.09	2.06	0.05
Prior7	− 1.13	0.05	2.06	0.06	**12.43**	0.08	− 0.90	0.04	**20.07**	0.10	3.78	0.06
Prior8	0.85	0.05	1.87	0.05	**15.93**	0.08	1.21	0.04	**33.04**	**0.12**	1.69	0.05
$n = 200$												
Prior1	0.26	0.05	3.00	0.06	2.37	0.05	**16.55**	0.08	7.43	0.07	− 1.09	0.05
Prior2	− 0.92	0.05	2.87	0.05	5.63	0.05	− 0.37	0.05	9.86	0.06	− 1.03	0.05
Prior3	− 0.20	0.04	− 3.05	0.04	4.71	0.06	0.29	0.06	7.25	0.07	1.89	0.05
Prior4	0.91	0.04	5.74	0.06	1.85	0.06	**22.00**	0.09	**12.09**	0.08	0.66	0.05
Prior5	1.52	0.05	2.79	0.06	0.25	0.04	**12.75**	0.08	**14.32**	0.07	− 0.48	0.04
Prior6	4.99	0.06	− 1.39	0.05	4.71	0.06	− 1.55	0.04	9.03	0.08	1.64	0.06
Prior7	− 5.33	0.04	− 1.68	0.04	5.75	0.06	0.24	0.05	**10.29**	0.06	− 0.26	0.06
Prior8	− 1.29	0.05	2.01	0.05	3.58	0.05	− 2.75	0.03	**19.16**	**0.10**	− 3.11	0.04
Model 2												
$n = 100$												
Prior1	2.84	0.05	9.25	0.08	**11.01**	0.08	**103.59**	**0.53**	**17.25**	**0.14**	1.85	0.03
Prior2	1.92	0.05	2.63	0.04	**13.35**	0.09	5.74	0.06	**22.98**	**0.20**	2.16	0.05
Prior3	6.50	0.08	3.43	0.07	**13.41**	**0.10**	**32.40**	**0.28**	**69.64**	**0.41**	3.62	0.05
Prior4	6.41	0.07	**11.98**	**0.12**	8.99	0.06	**95.49**	**0.55**	**72.74**	**0.46**	3.28	0.04
Prior5	2.19	0.04	9.70	**0.10**	**10.85**	0.08	**105.05**	**0.54**	**101.81**	**0.58**	2.64	0.04
Prior6	6.56	0.06	5.10	0.07	**11.71**	0.09	7.05	0.08	**20.39**	**0.18**	5.03	0.06
Prior7	3.33	0.06	5.36	0.06	**13.02**	0.10	6.00	0.07	**19.70**	**0.16**	6.48	0.05
Prior8	3.44	0.03	2.52	0.03	**15.70**	**0.12**	4.69	0.05	**26.64**	**0.24**	3.49	0.05
$n = 200$												
Prior1	0.29	0.04	3.91	0.07	5.14	0.06	**19.97**	**0.17**	9.62	0.09	0.22	0.04
Prior2	1.89	0.06	3.31	0.05	6.47	0.05	3.58	0.04	**13.79**	**0.11**	0.42	0.04
Prior3	1.60	0.05	1.46	0.06	5.99	0.06	3.27	0.06	**11.05**	0.09	2.20	0.05
Prior4	3.01	0.06	5.22	0.07	4.33	0.06	**20.70**	**0.17**	**16.56**	**0.12**	2.78	0.04
Prior5	1.29	0.05	3.17	0.06	4.15	0.04	**22.56**	**0.17**	**23.45**	**0.13**	1.19	0.05
Prior6	4.44	0.07	2.57	0.06	6.13	0.06	5.25	0.04	**10.72**	**0.10**	3.87	0.06
Prior7	0.45	0.05	3.39	0.06	7.64	0.07	4.84	0.06	**12.10**	**0.11**	4.40	0.06
Prior8	0.78	0.05	2.79	0.05	7.10	0.07	2.84	0.04	**16.89**	**0.15**	2.22	0.05

The numbers shown are rounded to the second decimal place. The RB's and RR's which are not in the defined acceptable range (RB>10%; RR>.10) are shown in bold.

TABLE 6.
Mean RB and RR of robust χ^2 statistics of Model 3 and Model 4 with sample size 100 to 200 in 1000 replications.

β_1, β_2	0,0		0,3		−2,3		0, 10		− 3, 10		0,-1	
	RB	RR	RB	RR	RB	RR	RB	RR	RB	RR	RB	RR
Model 3												
$n = 100$												
Prior1	6.06	0.08	**12.92**	**0.13**	**13.67**	**0.12**	**150.72**	**0.79**	**20.05**	**0.18**	5.26	0.07
Prior2	4.79	0.06	5.58	0.06	**16.44**	**0.13**	9.75	**0.11**	**25.29**	**0.24**	4.88	0.06
Prior3	6.97	0.08	6.62	0.10	**20.00**	**0.16**	**44.79**	**0.52**	**88.04**	**0.65**	7.52	0.08
Prior4	9.16	0.09	**14.72**	**0.16**	**12.49**	0.10	**149.25**	**0.78**	**77.01**	**0.67**	7.61	0.09
Prior5	5.00	0.07	**10.00**	**0.11**	**17.19**	**0.18**	**161.38**	**0.81**	**102.11**	**0.82**	4.84	0.05
Prior6	**11.34**	**0.11**	6.26	0.07	**15.33**	**0.13**	**13.41**	**0.17**	**24.13**	**0.22**	6.87	0.07
Prior7	5.85	0.07	6.03	0.07	**13.96**	**0.11**	**11.75**	**0.14**	**24.52**	**0.22**	7.76	0.08
Prior8	4.52	0.05	5.74	0.06	**19.42**	**0.16**	8.05	0.09	**31.31**	**0.31**	5.62	0.06
$n = 200$												
Prior1	2.33	0.06	3.97	0.07	5.40	0.07	**55.66**	**0.33**	**11.51**	**0.10**	1.72	0.05
Prior2	2.53	0.06	2.91	0.05	6.55	0.07	2.85	0.05	**15.11**	**0.14**	1.98	0.06
Prior3	4.19	0.06	2.16	0.06	8.64	0.10	**10.60**	**0.12**	**20.16**	**0.17**	3.80	0.06
Prior4	4.01	0.06	5.30	0.07	8.22	0.09	**52.98**	**0.36**	**37.03**	**0.27**	5.45	0.08
Prior5	1.91	0.05	4.00	0.06	5.47	0.08	**49.50**	**0.32**	**44.55**	**0.39**	2.18	0.05
Prior6	6.85	0.08	4.08	0.06	7.13	0.07	4.64	0.07	**14.73**	**0.13**	5.28	0.07
Prior7	2.83	0.06	5.34	0.07	7.96	0.08	5.19	0.06	**15.09**	**0.14**	4.75	0.06
Prior8	1.99	0.06	3.22	0.05	9.17	0.08	4.20	0.06	**18.99**	**0.17**	2.90	0.05
Model 4												
$n = 100$												
Prior1	4.88	0.06	**10.72**	**0.19**	9.72	**0.14**	**120.21**	**0.95**	**14.26**	**0.29**	3.84	0.05
Prior2	4.28	0.06	3.83	0.03	**10.11**	**0.13**	7.86	**0.13**	**17.24**	**0.40**	3.18	0.05
Prior3	6.62	0.10	9.32	**0.21**	**12.78**	**0.22**	**44.19**	**0.78**	**75.76**	**0.90**	5.12	0.08
Prior4	6.74	**0.11**	**11.53**	**0.20**	8.22	0.08	**120.81**	**0.96**	**70.93**	**0.92**	5.80	0.07
Prior5	4.41	0.06	9.97	**0.18**	**17.59**	**0.33**	**130.17**	**0.95**	**92.90**	**0.96**	3.64	0.04
Prior6	7.71	**0.11**	6.50	0.10	9.92	**0.15**	**13.89**	**0.30**	**15.66**	**0.34**	5.38	0.07
Prior7	5.01	0.07	5.76	0.07	**10.81**	**0.16**	**11.00**	**0.22**	**17.01**	**0.38**	6.13	0.08
Prior8	3.88	0.05	3.99	0.04	**12.97**	**0.24**	8.91	**0.16**	**21.31**	**0.60**	4.59	0.05
$n = 200$												
Prior1	2.39	0.04	4.50	0.08	5.68	0.08	**55.80**	**0.64**	8.77	**0.14**	1.99	0.04
Prior2	2.60	0.05	3.04	0.05	6.11	0.10	3.59	0.03	**11.47**	**0.21**	2.22	0.04
Prior3	3.55	0.07	2.79	0.05	6.41	0.09	**13.85**	**0.25**	**19.83**	**0.31**	2.75	0.05
Prior4	3.62	0.06	5.71	0.07	4.51	0.05	**56.34**	**0.64**	**33.10**	**0.49**	5.26	0.08
Prior5	2.65	0.05	3.44	0.07	5.43	0.08	**58.73**	**0.66**	**51.85**	**0.71**	1.95	0.03
Prior6	4.91	0.08	4.06	0.06	5.19	0.07	5.20	0.08	**10.07**	**0.18**	3.68	0.06
Prior7	2.89	0.05	5.60	0.09	5.44	0.07	4.23	0.04	**10.48**	**0.18**	4.23	0.08
Prior8	2.45	0.03	2.67	0.04	6.31	0.07	3.67	0.04	**14.03**	**0.29**	2.36	0.06

The numbers shown are rounded to the second decimal place. The RB's and RR's which are not in the defined acceptable range (RB> 10%; RR> .10) are shown in bold.

TABLE 7.
The summary of the 95% (5%) quantile and AR of the fit indices in each model.

$n = 100$	$\bar{Q}$	$\bar{AR}$	Q%	AR%	bQ	β_1, β_2	mAR	β_1, β_2	rQ	β_1, β_2	rAR	β_1, β_2
Model 1												
rRMSEA	0.13	0.81	18.75	100.00	0.23	−3,10	0.64	0,10	0.13	0,10	0.24	0,10
rCFI	0.93	0.94	12.50	18.75	0.66	−3,10	0.69	−3,10	0.29	−3,10	0.26	−3,10
rTLI	0.86	0.86	47.92	70.83	0.32	−3,10	0.61	−3,10	0.58	−3,10	0.24	−3,10
rNFI	0.89	0.78	35.42	70.83	0.65	−3,10	0.43	−3,10	0.27	−3,10	0.36	−3,10
Model 2												
rRMSEA	0.08	0.92	14.58	14.58	0.20	0,10	0.47	−3,10	0.14	0,10	0.49	−3,10
rCFI	0.92	0.92	14.58	14.58	0.53	−3,10	0.44	−3,10	0.43	−3,10	0.54	−3,10
rTLI	0.90	0.90	14.58	29.17	0.39	−3,10	0.41	−3,10	0.56	−3,10	0.53	−3,10
rNFI	0.84	0.32	66.67	100.00	0.47	−3,10	0.02	−3,10	0.41	−3,10	0.66	0,0
Model 3												
rRMSEA	0.09	0.87	12.50	20.83	0.21	0,10	0.23	−3,10	0.13	0,10	0.71	−3,10
rCFI	0.88	0.82	25.00	45.83	0.56	0,10	0.17	−3,10	0.36	−3,10	0.71	−3,10
rTLI	0.85	0.77	41.67	68.75	0.41	0,10	0.16	−3,10	0.48	−3,10	0.64	−3,10
rNFI	0.79	0.09	89.58	100.00	0.49	−3,10	0.00	—	0.33	−3,10	0.36	0,0
Model 4												
rRMSEA	0.06	0.90	8.33	14.58	0.17	0,10	0.17	0,10	0.11	0,10	0.82	−3,10
rCFI	0.88	0.80	27.08	43.75	0.40	−3,10	0.04	−3,10	0.54	−3,10	0.85	−3,10
rTLI	0.87	0.78	29.17	43.75	0.33	−3,10	0.03	−3,10	0.60	−3,10	0.81	−3,10
rNFI	0.72	0.00	100.00	100.00	0.32	−3,10	0.00	—	0.44	−3,10	0.00	0,10
$n = 200$	$\bar{Q}$	$\bar{AR}$	Q%	AR%	bQ	β_1, β_2	mAR	β_1, β_2	rQ	β_1, β_2	rAR	β_1, β_2
Model 1												
rRMSEA	0.08	0.95	0.00	0.00	0.09	0,10	0.91	−3,10	0.02	0,10	0.05	0,10
rCFI	0.97	0.99	0.00	0.00	0.94	−3,10	0.92	−3,10	0.03	−3,10	0.07	−3,10
rTLI	0.95	0.95	4.17	14.58	0.88	−3,10	0.83	−3,10	0.07	−3,10	0.12	−3,10
rNFI	0.95	0.94	2.08	25.00	0.89	−3,10	0.69	−3,10	0.07	−3,10	0.27	−3,10
Model 2												
rRMSEA	0.05	1.00	0.00	0.00	0.09	−3,10	0.93	−3,10	0.04	−3,10	0.07	−3,10
rCFI	0.98	0.99	2.08	2.08	0.81	−3,10	0.90	−3,10	0.17	−3,10	0.10	−3,10
rTLI	0.97	0.99	4.17	2.08	0.75	−3,10	0.88	−3,10	0.22	−3,10	0.12	−3,10
rNFI	0.93	0.72	20.83	60.42	0.73	−3,10	0.06	−3,10	0.20	−3,10	0.75	−3,10
Model 3												
rRMSEA	0.06	0.98	0.00	10.42	0.12	0,10	0.73	−3,10	0.07	0,10	0.27	−3,10
rCFI	0.94	0.96	12.50	12.50	0.69	−3,10	0.61	−3,10	0.27	−3,10	0.38	−3,10
rTLI	0.93	0.94	12.50	14.58	0.59	−3,10	0.58	−3,10	0.36	−3,10	0.39	−3,10
rNFI	0.88	0.45	33.33	72.92	0.61	−3,10	0.00	−3,10	0.29	−3,10	0.58	0,0

The shape had the greatest effect when $\beta_2 = 10$ and sample size was 100. The largest ranges of 95% (5%) quantile (AR) within each β_1 and β_2 combination were .13 (.24), .29 (.26), .58 (.24) and .27 (.36) for the rCFI, rTLI, rNFI and rRMSEA, respectively. The largest ranges shrank when sample size increased.

For Model 2, the rRMSEA and rCFI only performed badly when $\beta_2 = 10$ and sample size was 100. The rRMSEA and rCFI both had 7 (14.58%) of 95%/5% quantile and 7 (14.58%) of AR which were out of the accepted range. In addition to $\beta_2 = 10$, the rTLI also performed badly in one case when $\beta_1 = -2$ and $\beta_2 = 3$. The rNFI performed badly in all non-normality level.

TABLE 7.
continued

$n = 200$	$\bar{Q}$	$\bar{AR}$	Q%	AR%	bQ	β_1, β_2	mAR	β_1, β_2	rQ	β_1, β_2	rAR	β_1, β_2
Model 4												
rRMSEA	0.04	0.99	0.00	6.25	0.09	0,10	0.85	0,10	0.06	0,10	0.15	0,10
rCFI	0.95	0.93	14.58	14.58	0.65	− 3,10	0.30	− 3,10	0.32	−3,10	0.70	− 3,10
rTLI	0.94	0.93	14.58	14.58	0.61	− 3,10	0.29	− 3,10	0.36	− 3,10	0.71	− 3,10
rNFI	0.84	0.05	68.75	100.00	0.52	− 3,10	0.00	—	0.36	− 3,10	0.30	0,0

The numbers shown are rounded to the second decimal place. The $\bar{Q}$ and $\bar{AR}$ columns indicate the mean of the 95% (5%) quantiles and mean of AR of the fit index of the 6 different non-normality levels and 8 distributions. The $Q\%$ and $AR\%$ columns indicate the percentages of quantiles and AR's of fit indices which were out of the accepted range of the 6 different non-normality levels and 8 distributions. The bQ column indicates the maximum 95% quantile of rRMSEA and the minimum 5% quantile of the other fit indices. The mAR indicates the minimum AR of the corresponding fit index. The rQ and rAR columns indicate the maximum range of the 5% or 95% quantile and AR of the fit index of a specified β_1 and β_2. The β_1 and β_2 following bQ, mAR, rQ and rAR indicate the β_1 and β_2 combination wherein the bQ, mAR, rQ and rAR occurred. The — in cell indicates the bQ, mAR, rQ and rAR values occurred in multiple β_1 and β_2 combinations.

When sample size increased to 200, rRMSEA performed well in all non-normality levels and its largest 95% quantile value was only 0.09. The rCFI and rTLI also performed well. Only less than 5% of the 5% quantile and AR of these two fit indices were in the unacceptable range, and they all occurred when $\beta_2 = 10$. In contrast, the rNFI performed badly in at least one case of all non-normality levels. When sample size increased to 500, the rRMSEA, rCFI and rTLI performed well in all conditions. However, the rNFI still performed badly in one case when $\beta_2 = 10$. The rNFI performed well in all conditions when sample size increased to 1000.

The largest ranges of AR on rRMSEA, rCFI, rTLI, and rNFI within each β_1 and β_2 condition were .49 (− 3, 10), .54 (− 3, 10), .53 (− 3, 10), and .75 (− 3, 10), respectively. The largest range of 95%/5% quantile was 0.14 (0, 10), 0.43 (− 3, 10), 0.56 (− 3, 10) and 0.41 (− 3, 10) for rRMSEA, rCFI, rTLI and rNFI, respectively. All the largest range of AR occurred when sample size was 100 except the one of rNFI which occurred when sample size was 200. It was due to the low AR's of rNFI in all conditions when sample size was 100 which was reflected by the low mean AR (.32).

For Model 3, the rRMSEA performed badly when $\beta_2 = 10$ and in one case when $\beta_1 = -2$ and $\beta_2 = 3$ with sample size 100. There were 6 (12.50%) 95% quantile and 10 (20.83%) AR out of the accepted range. There were 12 (25.00%) of 5% quantile and 22 (45.83%) of AR out of the accepted range which all occurred when $\beta_2 \geq 3$. The rTLI performed badly when $\beta_2 \geq 0$ and 20 (41.67%) of 5% quantile and 33 (68.75%) of AR were out of the accepted range. The rNFI performed badly in all non-normality levels. Only 5 (10.42%) of the 5% quantile were in the accepted range. The maximum AR of rNFI was .37, and thus it was unlikely to obtain an accepted rNFI value when sample size was 100 no matter the non-normal level.

When sample size was 200, only 5 (10.42%) AR's of rRMSEA were badly performed and they all occurred when $\beta_2 = 10$. The rCFI and rTLI also performed badly only when the $\beta_2 = 10$; no more than 7 (14.58%) cases had 5% quantile or AR out of the accepted range. In contrast, the rNFI performed badly when the $\beta_2 \geq 0$. There were 16 (33.33%) 5% quantile and 35 (72.93%) AR's out of the accepted range. When sample size increased to 500, the rRMSEA, rCFI and rNFI performed well across all the levels of non-normality. The rNFI still performed badly when $\beta_2 = 10$ with 11 (22.92%) AR's out of the accepted range. The rNFI performed well across all levels of non-normality until sample size increased to 1000.

The largest ranges of AR within each β_1 and β_2 combination of the rRMSEA, rCFI, rTLI, and rNFI were .71 (− 3, 10), .71 (− 3, 10), .64 (− 3, 10), and .58 (− 3, 10). All the largest ranges occurred when the sample size was 100 except the one of rNFI which occurred when sample size was 500. The largest ranges of 95%/5% quantile of the rRMSEA, rCFI, rTLI, and rNFI were .13 (0, 10), .36 (− 3, 10), .48 (− 3, 10), and .33 (− 3, 10), respectively, when sample size was 100.

For Model 4, the trends were almost identical to the trends in Model 3. The largest ranges of AR within each β_1 and β_2 combination of the rRMSEA, rCFI, rTLI, and rNFI were .82 (− 3, 10), .85 (− 3, 10), .81 (− 3, 10), and .90 (− 3, 10), respectively. All the largest ranges occurred when sample size was 100 except the one of the rNFI which occurred when sample size was 500. The largest ranges of 95% / 5% quantile of the rRMSEA, rCFI, rTLI, and rNFI were .11 (0, 10), .54 (− 3, 10), .60 (− 3, 10), .44 (− 3, 10) when sample size was 100. Moreover, all the AR's of rNFI were smaller than .01 and thus it was very unlikely to obtain an accepted rNFI value when sample size was 100.

Considering the effect of the distribution shape, all the fit indices of the p's estimated with Prior 3 and Prior 5 performed worse than the others across all levels of non-normality. When the level of non-normality was severe ($\beta_2 = 10$), the fit indices of the p's estimated with Prior 1 and Prior 4 performed badly as well.

8. Conclusion and Discussion

In this paper, two entropy-based procedures, MaxEnt and MinxEnt, were proposed to generate multivariate ordinal distributions with pre-specified β_1 and β_2 in a systematic way. The p_j is estimated under the marginal MaxEnt approach and MinxEnt procedures and the τ is estimated under the global MaxEnt approach. For marginal MaxEnt approach, only the β_1 and β_2 are need to be pre-specified. In addition to β_1 and β_2, the p^* whose parameters are all specified is also required for global MaxEnt approach. In contrast, the ideal distributions, q_j's, β_1 and β_2 are needed to be pre-specified for MinxEnt procedure.

Two analyses were conducted to examine our procedures and to show its applicability. Analysis I showed that our proposed procedures can estimate p's precisely. Our approaches yielded an excellent agreement between the specified and estimated β_1 and β_2 of p across a wide range of values. In addition, our procedures can also generate data with the required β_1 and β_2 in most situations with satisfactory precision when the sample size is moderate (500). These characteristics make our procedures attractive ways to generate data with pre-specified β_1 and β_2. In detail, our MaxEnt procedure was capable of generating data from prudent and smooth distributions where zero probability is avoided. Our MinxEnt procedure provides an easy way for researchers to generate data from distributions not only satisfying the pre-specified β_1 and β_2 but also close to the distribution shapes which are frequently seen in their research fields. When considering the computing speed, our marginal MaxEnt and MinxEnt procedures are computationally very fast with CPU time less than one minute when five categories p_j's are estimated as programmed in R. The computer authors used to test equips Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz.

Our global MaxEnt approach is capable of being applied when p^* is assumed. Although it is programmed when p^* is set as the MVN in our R package, it can be programmed in R with other p^*'s by interested users. However, it is beyond the scope of this paper to explore other distributions for the p^*'s.

To generate y's through our procedures, the first step is to specify the parameters of p^*, which has the required property, such as specified $\mathbf{\Sigma^*}$. Then, the τ can be estimated by the global MaxEnt approach or be obtained by applying the inverse CDF function of the p^*_j, estimated by marginal MaxEnt approach or MinxEnt procedure. Then, users could generate y^* follows p^*, and apply (1) to transform y^* to y.

It is worth noting that the p^*'s or p_j^*'s are not required to be specified when the marginal MaxEnt approach or MinxEnt procedure is applied to estimate p_j's. The p_j^*'s are only required when estimating τ_j. Therefore, all the continuous non-normal data generation procedures which are able to generate y^* following known p_j^*'s but unknown p^* can be used to generate y through our procedure. The p_j can be estimated by our marginal MaxEnt and MinxEnt procedures, then compute τ_j and transform y^* into y according to Eq. (1). For instance, the copula procedure could be easily applied with these two approaches even when the joint PDF is unknown (see Mair et al. (2012) for a basic introduction of the copula and how to use it to generate data within SEM). This fact provides great flexibility to these two approaches.

It should be noted that the marginal MaxEnt procedure sacrifices the maximum entropy property of the estimated distribution p. As we have shown, the entropy of p decreases when the degree of dependence of y_j^*'s increases when p^* is set as MVN. In our numerical evaluation, the marginal approach is fast and relatively robust. However, it is worth noting that the result depends on our choice of p^*, which is set as the MVN (specifically, bivariate MVN). The result might be different if other p^*'s are chosen. However, we recommend applying the marginal approach to generate data when the MVN is set as p^* and suggests the researcher to evaluate the entropy decrease before choosing between the two procedures.

Our procedures have wide freedom to estimate p's with different shapes but the same β_1 and β_2. In particular, an estimated m-dimensional p can have m different p_j's, whose k_j, β_{1j} and β_{2j} can be specified arbitrarily once the inequality of β_1 and β_2: $\beta_{1j}^2 \leq \beta_{2j} + 2$ is satisfied (Rohatgi & Székely, 1989). The shapes of the distributions with the same β_1 and β_2 could be varied by carefully specifying the q_j when applying the MinxEnt procedure. However, it is worth noting that the estimated distributions are restricted by the setting of q_j, which means that the p is fixed once we specify the same q_j's. In other words, the p's are still restricted by the selection of q_j's.

In Analysis II, our results showed that all the test statistics we examined tended to perform worse when non-normality level increases. However, the distribution setting also impacts their performances, especially when both the β_1 and β_2 are far from 0, the sample size is small, and the model is complex. Under these conditions, the ranges in empirical Type I error rates could be as large as 80% and the ranges of the AR rates could achieve 90%. In other words, the robust χ^2 could perform relatively well in extreme conditions, such as small sample size, severe non-normality, and complex models. For example, the smallest RR of robust χ^2 was 9% in Model 3 and 13% in Model 4 when $\beta_1 = 0$, $\beta_2 = 10$ and sample size was 100. This has not been shown in previous robustness research. Generally speaking, these ranges are reduced when sample size increases, except for the rNFI in the complex models. In complex models, the rNFI is severely underestimated in all distribution shapes when β_2 increases to 3 or higher and sample size is relatively small. Thus, the ranges of the AR rate within those conditions are small.

Three possible reasons can explain the results. First, large β_1 and β_2 imply large probabilities in very few categories and small probabilities in other categories. In this case, the sampling distributions of the generated data are unstable. Second, large values of the β_1 and β_2 can also lead to a substantial number of empty cells in p. The empty cells might result in unreliable estimation (Yang-Wallentin et al., 2010). Third, the different attenuated degree of correlation coefficients caused by the shapes of p's might lead to different baselines of the fit indices (For instance, the chi^2 value of the null model in TLI), and further impact their values. However, all of these explanations are arguable. For instance, we could find the distributions with the worst performing statistics have relatively larger probabilities for the categories have small probabilities when $\beta_1 = 0$ and $\beta_2 = 10$. In addition, we could also observe the conditions within which the correlation coefficients attenuation varies a lot, but the fit indices perform only trivial differences and vice versa. Therefore, these findings deserve additional research.

In conclusion, the results of these studies suggest that our entropy-based procedures are valuable tools for (a) generating data following distributions with specific characteristics or researchers' knowledge subject to pre-specified β_1 and β_2, and (b) investigating the effect of the shape of observed ordinal distributions when the β_1 and β_2 are controlled. With these procedures, new areas of robustness research can be explored and easily tailored to different research areas. Our procedures provide researchers simple methods to conduct the Monte Carlo studies with discrete probability distributions, and open the door to a wide range of possible robustness studies.

Acknowledgments

We acknowledge Professors Stephen Wright and Michael Ferris for their valuable suggestions on solving the optimization problems. Professor Chunming Zhang suggested ways for proving the uniqueness when $k_j = 3$. The discussion between the first author and Professor Chun-Ping Cheng have inspired this paper.

References

Asparouhov, T., Muthén, B. (2010). *Simple second order chi-square correction*. Retrieved from Mplus website: http://www.statjnodel.com/dowrJoad/WLSMV_new_chi21.

Babakus, E., Ferguson, E. J., & Joreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, *24*, 222–228.

Blair, R. C. (1981). A reaction to consequence of failure to meet assumptions underlying fixed effects analysis of variance and covariance. *Review of Educational Research*, *51*, 499–507.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley and Sons Inc.

Bradley, J. V. (1982). The insidious L-shaped distribution. *Bulletin of the Psychonomic Society*, *20*, 85–88.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136–162). Thousand Oaks, CA: Sage Publications.

DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *213*, 425–438.

Ethington, C. A. (1987). The robustness of LISREL estimates in structural equation models with categorical variables. *The Journal of Experimental Education*, *55*, 80–88.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521–532.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics*, *1*(2), 141–149.

Genz, A., & Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, *11*(4), 950–971.

Genz, A., Bretz, F., Miwa, T., Mi, X , Leisch, F., Scheipl, F., Hothorn, T. (2015). *mvtnorm: Multivariate normal and t distributions.* http://CRAN.R-project.org/package=mvtnorm (R package version 1.0-3)

Golan, A., Judge, G., & Miller, D. (1997). *Maximum entropy econometrics: Robust estimation with limited data*. Chichester: Wiley.

Headrick, T. C. (2010). *Statistical simulation: Power method polynomials and other transformations*. Boca Raton, FL: Chapman and Hall.

Headrick, T. C., & Sawilosky, S. S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika*, *64*, 25–35.

Hipp, J. R., & Bollen, K. A. (2003). Model fit in structural equation models with censored, ordinal, and dichotomous variables: Testing vanishing tetrads. *Sociological Methodology*, *33*, 267–305.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*, 53–60.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physics Review*, *106*, 620–630.

Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *IEEE*, *70*, 939–952.

Jorgensen, T. (2016). *lavaanTabular Lavaan output .scaled —.robust*. https://groups.google.com/forum/#!topic/lavaan/rGitXu9h9zY (Online; accessed February 19, 2017).

Kapur, J. N., & Kesavan, H. K. (1992). *Entropy optimization principles with applications*. Boston: Academic Press.

Kullback, S., & Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79–86.

Lee, Y. (2010). Generation of non-normal approximated discrete random variables. Master's thesis, National Chengchi University, Taipei, Taiwan.

Madsen, K., Nielsen, H. B., Tingleff, O. (2004). *Optimization with constraints*. LyngbyIMM, Technical University of Denmark. http://orbit.dtu.dk/files/2721110/imm4213.pdf

Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating nonnormal multivariate data using copulas: Applications to sem. *Multivariate Behavioral Research*, *47*, 547–565.

Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, *32*, 355–373.
Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166.
Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 486–5.
Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.
Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171–189.
Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19–30.
Nocedal, J., & Wright, S. (2006). *Numerical optimization*. Berlin: Springer.
Olsson, U. (1979). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research*, *14*, 485–500.
Pearson, E. S., & Please, N. W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika*, *62*, 223–241.
R Core Team. (2014). *R: A Language and Environment for Statistical Computing Vienna, Austria*. http://www.R-project.org/
Ramachandran, K. M., & Tsokos, C. P. (2009). *Mathematical statistics with applications*. Burlington, MA: Elsevier.
Rohatgi, V. K., & Székely, G. J. (1989). Sharp inequalities between skewness and kurtosis. *Statistics & Probability Letters*, *84*, 297–299.
Rosseel, Y. 2012. lavaan: An R package for structural equation modeling. *Journal of Statistical Software4821-36*. http://jstatsoft.org/v48/i02/
Ruscio, J., & Kaczetow, W. (2008). Simulating multivariate nonnormal data using an iterative technique. *Multivariate Behavioral Research*, *48*, 355–381.
Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.
Varadhan, R. (2015). *alabama: Constrained nonlinear optimization*. http://CRAN.R-project.org/package=alabama (R package version 2015.3-1).
Weng, L., & Cheng, C. (2004). Effects of response order on Likert-type scale. *Educational and Psychological Measurement*, *60*, 908–924.
Wilkins, J. E. (1944). A note on skewness and kurtosis. *The Annal of Mathematical Statistic*, *15*, 333–335.
Wu, N. (1997). *The maximum entropy method*. New York: Springer.
Yang-Wallentin, F., Joreskog, K., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, *17*, 392–423.
Zellner, A., & Highfield, R. A. (1988). Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *Journal of Econometrics*, *37*, 195–209.

Manuscript Received: 10 DEC 2015