

Test for homogeneity with unordered paired observations*

Jiahua Chen

*Department of Statistics
University of British Columbia
Vancouver, BC Canada V6T 1Z4
e-mail: jhchen@stat.ubc.ca*

Pengfei Li

*Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
e-mail: pengfei.li@uwaterloo.ca*

Jing Qin

*National Institute of Allergy and Infectious Diseases
National Institutes of Health, Maryland 20892, U.S.A.
e-mail: jingqin@niaid.nih.gov*

and

Tao Yu

*Department of Statistics and Applied Probability
National University of Singapore, 117546 Singapore
e-mail: stayt@nus.edu.sg*

Abstract: In some applications, an experimental unit is composed of two distinct but related subunits. The response from such a unit is (X_1, X_2) but we observe only $Y_1 = \min\{X_1, X_2\}$ and $Y_2 = \max\{X_1, X_2\}$, i.e., the subunit identities are not observed. We call (Y_1, Y_2) unordered paired observations. Based on unordered paired observations $\{(Y_{1i}, Y_{2i})\}_{i=1}^n$, we are interested in whether the marginal distributions for X_1 and X_2 are identical. Testing methods are available in the literature under the assumptions that $\text{var}(X_1) = \text{var}(X_2)$ and $\text{cov}(X_1, X_2) = 0$. However, by extensive simulation studies, we observe that when one or both assumptions are violated, these methods have inflated type I errors or much lower powers. In this paper, we study the likelihood ratio test statistics for various scenarios and explore their limiting distributions without these restrictive assumptions. Furthermore, we develop Bartlett correction formulae for these statistics to enhance their precision when the sample size is not large. Simulation studies and real-data examples are used to illustrate the efficacy of the proposed methods.

*Jiahua Chen has been supported in part by the National Natural Science Foundation of China Grant 11871419 and the Natural Sciences and Engineering Research Council of Canada Grant RGPIN-2019-04204. Pengfei Li has been supported in part by the Natural Sciences and Engineering Research Council of Canada Grant RGPIN-2020-04964. Tao Yu has been supported in part by Singapore Ministry Education Academic Research Tier 1 Funds (R-155-000-134-112; R-155-123-112; R-155-000-202-114) and the Ministry of Education of Singapore Grant MOE2014-T2-1-072.

MSC2020 subject classifications: Primary 62F03; Secondary 62F05.

Keywords and phrases: Adjusted limiting distribution, Bartlett correction, correlated unordered pairs, likelihood ratio test, two-sample test.

Received January 2020.

Contents

1	Introduction	1662
2	Main results	1664
2.1	Unordered uncorrelated paired data	1664
2.2	Unordered correlated pair data	1666
3	Sketch of the proofs of Theorems 1 and 2	1667
3.1	Reparameterization	1667
3.2	Outline of the proofs	1668
4	Adjusted limiting distributions	1669
5	Simulation studies	1671
5.1	Data generation	1671
5.2	Results	1672
6	Real-data examples	1675
6.1	Data from karyotype analysis	1675
6.2	C-band area of human chromosome data	1676
7	Discussion	1677
8	Technical details	1677
8.1	Some preparation	1677
8.2	Some useful lemmas	1678
8.3	Proof of Theorem 1	1686
8.4	Proof of Theorem 2	1689
	Acknowledgments	1692
	References	1692

1. Introduction

In some applications, an experimental unit is made of two distinct but related subunits. The response from such a unit is (X_1, X_2) but we observe only $Y_1 = \min\{X_1, X_2\}$ and $Y_2 = \max\{X_1, X_2\}$; that is, the subunit identities are not observed or unobservable. We call (Y_1, Y_2) unordered paired observations. We assume that $(X_{1i}, X_{2i})^\tau$, for $i = 1, \dots, n$, are independent and identically distributed (i.i.d.) normal random vectors:

$$\begin{pmatrix} X_{1i} \\ X_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right). \quad (1.1)$$

We say that $\{(Y_{1i}, Y_{2i})\}_{i=1}^n$ are *uncorrelated* when $\rho = 0$ and *correlated* when $\rho \neq 0$. This paper studies the homogeneity testing of the marginal distributions of X_{1i} and X_{2i} :

$$H_0 : (\mu_1, \sigma_1^2) = (\mu_2, \sigma_2^2) \quad \text{versus} \quad H_a : (\mu_1, \sigma_1^2) \neq (\mu_2, \sigma_2^2). \quad (1.2)$$

Unordered paired data occur in many applications, and there is a long research history. For instance, Hinkley (1973) analyzed such a data set from human genetics. The genetic blueprint of an individual is contained in 23 pairs of chromosomes. Each member of the pair is inherited from the corresponding chromosome pair of a parent. If we do not know the chromosome correspondences between the offspring and the parents, we lose the parental identities and end up with unordered paired observations. Olkin and Viana (1995) provide more examples. In visual acuity studies, we may record only a subject's extreme acuities (the "best" and "worst" acuities) without recording the corresponding eyes. In twin experiments, we obtain unordered paired observations without a label for each member of a twin pair; see Ernst et al. (1996) and Shekar et al. (2006) and the references therein. Furthermore, unordered data of a higher dimension are collected in various scientific disciplines. For example, Davies and Phillips (1988) provided an example of unordered data of dimension k . In the interim analysis of a double-blinded clinical trial of k treatments, we get the k order statistics without knowledge of the corresponding treatments; see also van der Meulen (2005) and Miller et al. (2009). In diffusion tensor (DT) brain imaging, the eigenvalues of the DT estimates for each brain voxel are viewed as unordered triples; see Yu et al. (2013) and the references therein.

With unordered paired observations, a fundamental question is whether or not X_{1i} and X_{2i} have the same distribution. Under Model (1.1), this is equivalent to testing the hypothesis specified in (1.2). Hinkley (1973) proposed a likelihood ratio test (LRT) procedure under the assumption that $\rho = 0$ and $\sigma_1^2 = \sigma_2^2$. Li and Qin (2011) investigated this problem in a semiparametric setup. Other approaches can be found in Moore II (1973), Lauder (1977), Moore II et al. (1979), Carothers (1981), Efron et al. (1971), and Qin and Zhang (2005), among others. All these works assume that X_{1i} and X_{2i} are independent with equal variance. These assumptions may not hold in applications, and they can be severely violated, as evidenced by the examples in Section 6. Ignoring the dependence structure and/or imposing an incorrect equal-variance assumption can lead to unreliable inference conclusions: the type I error may be severely inflated or the power markedly decreased.

This paper focuses on tests for (1.2). In particular, we study the LRT in four scenarios: (1) $\rho = 0$ and $\sigma_1^2 = \sigma_2^2$; (2) $\rho = 0$; (3) $\sigma_1^2 = \sigma_2^2$; and (4) no assumption on ρ , σ_1^2 , and σ_2^2 .

Investigating the asymptotic behavior of these LRT statistics is technically challenging. The well-developed theory (Wilks, 1938; Chernoff, 1954; Self and Liang, 1987; Drton, 2009) is not applicable because of the undesirable mathematical properties (see (2.3) in Section 2) of the log-likelihood function. In addition, an important byproduct of the theory for the corresponding LRT statistics is the asymptotic behavior of the maximum likelihood estimators (MLEs) for $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. Interestingly, we have shown that the asymptotic behavior depends on whether $\rho = 0$ is known or ρ is unknown. The convergence rates of these parameter estimates depend on the scenario.

We observe that the limiting distributions of the LRT statistics under H_0 are not sufficiently accurate approximations to their finite-sample distributions

when n is not large. To enhance the approximation precision of the limiting distributions, we adjust the statistics based on the Bartlett correction (Bartlett, 1937; Lawley, 1956). Simulation results confirm the efficacy of the adjustment.

We organize the rest of the paper as follows. Section 2 introduces the LRT statistics for (1.2) and studies their asymptotic behavior under H_0 . Section 3 provides the sketch of the proofs. Section 4 presents the adjusted limiting distributions of our statistics for data of limited sample size. Section 5 contains simulation studies, and Section 6 gives real-data examples. Some discussion is provided in Section 7, and the technical details are relegated to Section 8.

2. Main results

The LRT is an essential tool in statistical inference, especially under the parametric model assumption; see Wilks (1938); Chernoff (1954); Self and Liang (1987); Drton (2009), and the references therein. In this section, we present LRT statistics and study their properties for testing (1.2) under model assumptions on ρ and whether or not $\sigma_1^2 = \sigma_2^2$.

We first derive the log-likelihood function with unordered paired observations. For any $y_1 < y_2$, we have

$$\begin{aligned} P(Y_1 \leq y_1, Y_2 \leq y_2) &= P(\{X_1 \leq y_1, X_2 \leq y_2\} \cup \{X_1 \leq y_2, X_2 \leq y_1\}) \\ &= P(X_1 \leq y_1, X_2 \leq y_2) + P(X_1 \leq y_2, X_2 \leq y_1) \\ &\quad - P(\{X_1 \leq y_1, X_2 \leq y_2\} \cap \{X_1 \leq y_2, X_2 \leq y_1\}) \\ &= P(X_1 \leq y_1, X_2 \leq y_2) + P(X_1 \leq y_2, X_2 \leq y_1) \\ &\quad - P(X_1 \leq y_1, X_2 \leq y_1). \end{aligned}$$

Therefore, the joint density function of (Y_1, Y_2) is given by

$$\phi(y_1, y_2; \boldsymbol{\theta}) + \phi(y_2, y_1; \boldsymbol{\theta}),$$

where $\phi(x_1, x_2; \boldsymbol{\theta})$ denotes the bivariate normal density function with parameters $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)^\tau$ specified in (1.1). The log-likelihood function based on $\{(Y_{1i}, Y_{2i})\}_{i=1}^n$ and Model (1.1) is:

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^n \log\{\phi(Y_{1i}, Y_{2i}; \boldsymbol{\theta}) + \phi(Y_{2i}, Y_{1i}; \boldsymbol{\theta})\}. \quad (2.1)$$

This likelihood function is the basis for our subsequent development.

2.1. Unordered uncorrelated paired data

In this section, we assume that $\rho = 0$ is known; problem (1.2) is reduced to $H_0 : \mu_1 = \mu_2, \sigma_1 = \sigma_2$. We define

$$\hat{\boldsymbol{\theta}} = \arg \sup_{\boldsymbol{\theta}} \{\ell_n(\boldsymbol{\theta}) : \rho = 0\},$$

$$\begin{aligned} \tilde{\boldsymbol{\theta}} &= \arg \sup_{\boldsymbol{\theta}} \{\ell_n(\boldsymbol{\theta}) : \sigma_1 = \sigma_2 = \sigma, \rho = 0\}, \\ \check{\boldsymbol{\theta}} &= \arg \sup_{\boldsymbol{\theta}} \{\ell_n(\boldsymbol{\theta}) : (\mu_1, \sigma_1) = (\mu_2, \sigma_2), \rho = 0\}, \end{aligned}$$

and we use the notational convention that the entries of $\hat{\boldsymbol{\theta}}$ are $\hat{\mu}_1, \hat{\mu}_2$, and so on. Note that $\hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}$, and $\check{\boldsymbol{\theta}}$ are MLEs of $\boldsymbol{\theta}$ under various constraints. The LRT statistics for testing the null hypothesis (1.2) against two alternatives, specified by $\sigma_1 = \sigma_2$ and $\sigma_1 \neq \sigma_2$ respectively, are given by

$$R_{n,1} = 2\{\ell_n(\tilde{\boldsymbol{\theta}}) - \ell_n(\hat{\boldsymbol{\theta}})\}, \quad R_{n,2} = 2\{\ell_n(\check{\boldsymbol{\theta}}) - \ell_n(\hat{\boldsymbol{\theta}})\}. \tag{2.2}$$

Theorem 1 below establishes the asymptotic distributions of $R_{n,1}$ and $R_{n,2}$ as well as the convergence rates of $\tilde{\boldsymbol{\theta}}$ and $\check{\boldsymbol{\theta}}$ under H_0 . For presentational continuity, we relegate its proof to Section 8. Let $\xrightarrow{\mathcal{D}}$ denote ‘‘convergence in distribution.’’ We use $0.5\chi_0^2 + 0.5\chi_1^2$ for an equal mixture of χ_0^2 and χ_1^2 , with χ_0^2 being the distribution with a point mass at zero.

Theorem 1. *Assume Model (1.1) and $\rho = 0$. Under H_0 that $\mu_1 = \mu_2 = \mu_0$ and $\sigma_1^2 = \sigma_2^2 = \sigma_0^2$, as $n \rightarrow \infty$, we have*

(a) $(\tilde{\mu}_1 - \mu_0)^2, (\tilde{\mu}_2 - \mu_0)^2$, and $\tilde{\sigma} - \sigma_0$ are all of order $O_p(n^{-1/2})$, and

$$R_{n,1} \xrightarrow{\mathcal{D}} 0.5\chi_0^2 + 0.5\chi_1^2;$$

(b) $(\hat{\mu}_j - \mu_0)^2, (\hat{\sigma}_j - \sigma_0)^2$ for $j = 1, 2$ are all of order $O_p(n^{-1/2})$, and

$$R_{n,2} \xrightarrow{\mathcal{D}} R \equiv \sup_{x_1, x_2} \{2\mathbf{x}^\tau \mathbf{w} - \mathbf{x}^\tau \mathbf{x}\},$$

where $\mathbf{x}^\tau = (x_1^2, x_2^2, 2x_1x_2)$ and $\mathbf{w}^\tau = (w_1, w_2, w_3)$ with w_1, w_2, w_3 being three i.i.d. $N(0, 1)$ random variables.

Deriving the asymptotic null distributions of $R_{n,1}$ and $R_{n,2}$ is technically challenging. We make the following comments. Let $\mu = (\mu_1 + \mu_2)/2$ and $\Delta = (\mu_1 - \mu_2)/2$ so that $\mu_1 = \mu + \Delta$ and $\mu_2 = \mu - \Delta$; we have

$$\frac{\partial \ell_n(\mu + \Delta, \mu - \Delta, \sigma_1, \sigma_2, \rho)}{\partial \Delta} \Big|_{\Delta=0, \sigma_1=\sigma_2} = 0. \tag{2.3}$$

This fact implies that the Fisher information matrix of $\boldsymbol{\theta}$ under the null hypothesis degenerates and undermines the basis for the elegant classical results (Wilks, 1938; Chernoff, 1954; Self and Liang, 1987; Drton, 2009). The crucial step in obtaining the asymptotic null distribution of the LRT is a quadratic approximation in $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$ to the log-likelihood ratio function. Following this path, we need to consider a fourth-order Taylor expansion to obtain a quadratic approximation in $(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^2$ and so on. Fortunately, we find that the sandwich technique of Chen and Chen (2001) and Chen et al. (2001) overcomes the technical obstacles caused by (2.3).

2.2. Unordered correlated pair data

In this section, we study the LRTs for (1.2) with ρ being an unknown parameter. Define

$$\begin{aligned}\hat{\boldsymbol{\theta}}^* &= \arg \sup_{\boldsymbol{\theta}} \{\ell_n(\boldsymbol{\theta})\}, \\ \tilde{\boldsymbol{\theta}}^* &= \arg \sup_{\boldsymbol{\theta}} \{\ell_n(\boldsymbol{\theta}) : \sigma_1 = \sigma_2 = \sigma\}, \\ \check{\boldsymbol{\theta}}^* &= \arg \sup_{\boldsymbol{\theta}} \{\ell_n(\boldsymbol{\theta}) : (\mu_1, \sigma_1) = (\mu_2, \sigma_2)\}.\end{aligned}$$

Similarly to the strategy for (2.2), we define the LRT statistics for (1.2) with ρ being an unknown parameter:

$$R_{n,1}^* = 2\{\ell_n(\tilde{\boldsymbol{\theta}}^*) - \ell_n(\check{\boldsymbol{\theta}}^*)\}, \quad R_{n,2}^* = 2\{\ell_n(\hat{\boldsymbol{\theta}}^*) - \ell_n(\check{\boldsymbol{\theta}}^*)\}. \quad (2.4)$$

Theorem 2 below establishes the asymptotic distributions of $R_{n,1}^*$ and $R_{n,2}^*$ as well as the convergence rates of $\tilde{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\theta}}^*$ under their respective H_0 . The proof is given in Section 8.

Theorem 2. *Assume Model (1.1) but do not assume $\rho = 0$. Under H_0 that $\mu_1 = \mu_2 = \mu_0$, $\sigma_1^2 = \sigma_2^2 = \sigma_0^2$, and $\rho = \rho_0$, as $n \rightarrow \infty$, we have*

(a) $(\tilde{\mu}_1^* - \mu_0)^2$, $(\tilde{\mu}_2^* - \mu_0)^2$, $(\tilde{\sigma}^* - \sigma_0)$, and $(\tilde{\rho}^* - \rho_0)$ are all of order $O_p(n^{-1/4})$, and

$$R_{n,1}^* \xrightarrow{\mathcal{D}} 0.5\chi_0^2 + 0.5\chi_1^2;$$

(b) $(\hat{\mu}_1^* - \mu_0)^2$, $(\hat{\mu}_2^* - \mu_0)^2$, $\hat{\sigma}_1^* - \sigma_0$, $\hat{\sigma}_2^* - \sigma_0$, and $\hat{\rho}^* - \rho_0$ are all of order $O_p(n^{-1/4})$, and

$$R_{n,2}^* \xrightarrow{\mathcal{D}} R^* \equiv \max\{w_1^2 + (w_2^+)^2, w_1^2 + (w_3^+)^2\},$$

where w_1 , w_2 , and w_3 are three i.i.d. $N(0, 1)$ random variables.

The limiting cumulative distribution function (c.d.f.) of $R_{n,2}^*$ is given by:

$$\begin{aligned}P(R^* \leq x) &= P\left(\max\{w_1^2 + (w_2^+)^2, w_1^2 + (w_3^+)^2\} \leq x\right) \\ &= \int_0^x \Phi^2(\sqrt{x-y})(2\pi y)^{-1/2} \exp(-y/2) dy\end{aligned}$$

for $x \geq 0$ with $\Phi(\cdot)$ being the c.d.f. of the standard normal distribution. We use this expression to evaluate the asymptotic quantile and the p-value for the corresponding test.

3. Sketch of the proofs of Theorems 1 and 2

3.1. Reparameterization

We first propose a reparameterization scheme that plays a fundamental role in our proof. Recall that the log-likelihood function based on $\{(Y_{1i}, Y_{2i})\}_{i=1}^n$ is given by

$$\begin{aligned} \ell_n(\boldsymbol{\theta}) &= \sum_{i=1}^n \log\{\phi(Y_{1i}, Y_{2i}; \boldsymbol{\theta}) + \phi(Y_{2i}, Y_{1i}; \boldsymbol{\theta})\} \\ &= \sum_{i=1}^n \log\{\phi(X_{1i}, X_{2i}; \boldsymbol{\theta}) + \phi(X_{2i}, X_{1i}; \boldsymbol{\theta})\}. \end{aligned}$$

Let $Z_{1i} = (X_{1i} + X_{2i})/2$ and $Z_{2i} = (X_{1i} - X_{2i})/2$. We define

$$\begin{aligned} \mu &= \mathbb{E}(Z_{1i}) = (\mu_1 + \mu_2)/2, \\ \Delta &= \mathbb{E}(Z_{2i}) = (\mu_1 - \mu_2)/2, \\ \sigma_+^2 &= \text{VAR}(Z_{1i}) = (1/4)(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2), \\ \sigma_-^2 &= \text{VAR}(Z_{2i}) = (1/4)(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2), \\ \xi\sigma_+\sigma_- &= \text{COV}(Z_{1i}, Z_{2i}) = (1/4)(\sigma_1^2 - \sigma_2^2), \end{aligned}$$

and let $\beta_0 = \Delta - \mu(\sigma_-/\sigma_+)\xi$, $\beta_1 = (\sigma_-/\sigma_+)\xi$, $\eta^2 = (1 - \xi^2)\sigma_-^2$. With this reparameterization, testing (1.2) is equivalent to testing $H_0 : \beta_0 = \beta_1 = 0$, and we can decompose the log-likelihood function as

$$\ell_n(\boldsymbol{\theta}) = \ell_{n,1}^*(\mu, \sigma_+) + \ell_{n,2}^*(\beta_0, \beta_1, \eta),$$

where

$$\ell_{n,1}^*(\mu, \sigma_+) = \sum_{i=1}^n \log\{\phi(Z_{1i}; \mu, \sigma_+)\}, \tag{3.1}$$

$$\begin{aligned} \ell_{n,2}^*(\beta_0, \beta_1, \eta) &= \sum_{i=1}^n \log \left\{ 0.5\phi(Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta) \right. \\ &\quad \left. + 0.5\phi(Z_{2i}; -\beta_0 - \beta_1 Z_{1i}, \eta) \right\}, \end{aligned} \tag{3.2}$$

and $\phi(x; \mu, \sigma)$ denotes the density function of $N(\mu, \sigma^2)$.

Based on this transformation and existing results from finite mixture models, we give intuitive explanations of the convergence rates given in Theorems 1 and 2. For simplicity, we consider the case $\sigma_1^2 = \sigma_2^2 = \sigma^2$, which implies $\xi = 0$ and hence $\beta_1 = 0$. Therefore, (3.1) can be viewed as the log-likelihood based on n observations from $N(\mu, \sigma_+^2)$, and (3.2) can be viewed as the log-likelihood based on n observations from the mixture model $0.5N(\beta_0, \eta^2) + 0.5N(-\beta_0, \eta^2)$, where

$$\sigma_+^2 = 0.5\sigma^2 + 0.5\rho\sigma^2, \quad \eta^2 = 0.5\sigma^2 - 0.5\rho\sigma^2. \tag{3.3}$$

As a consequence, testing (1.2) is equivalent to testing $H_0 : \beta_0 = 0$ based on observations from $0.5N(\beta_0, \eta^2) + 0.5N(-\beta_0, \eta^2)$, i.e., testing homogeneity in a two-component normal mixture model with equal variance.

- When $\rho = 0$ is known, we immediately have $\sigma_+^2 = \eta^2$. The MLEs of (μ, σ_+^2) obtained from (3.1) have an $n^{-1/2}$ convergence rate; so does that of η^2 . On the other hand, $\beta_0^2 + \eta^2$ is the second moment of $0.5N(\beta_0, \eta^2) + 0.5N(-\beta_0, \eta^2)$; it can be estimated by maximizing (3.2) and can achieve the $n^{-1/2}$ convergence rate. Hence, the convergence rate for the MLE of β_0^2 is $n^{-1/2}$, or equivalently, that for the MLE of β_0 is $n^{-1/4}$. These comply with the convergence rates given in Part (a) of Theorem 1.
- When ρ is unknown, $\sigma_+^2/\eta^2 = (1 + \rho)/(1 - \rho)$ is unknown; the estimation of σ_+^2 is unable to determine the estimation of η^2 . The MLEs of (μ, σ_+^2) based on (3.1) continue to have the $n^{-1/2}$ convergence rates. However, β_0 and η^2 need to be estimated from (3.2), which is the likelihood of a normal mixture model. When η^2 is unknown, this model is not strongly identifiable (Chen and Chen, 2003) or not identifiable in the second order (Ho and Nguyen, 2016). The MLEs of β_0 and η^2 can achieve convergence rates of only $n^{-1/8}$ and $n^{-1/4}$ respectively; see Theorem 1 of Chen and Li (2009) and Proposition 3.16 of Ho and Nguyen (2019). These are in line with the convergence rates stated in Part (a) of Theorem 2.

3.2. Outline of the proofs

We now give an outline of the proofs of Theorems 1 and 2; the details are provided in Section 8. Without loss of generality, we assume that under H_0 , the true value of $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ is $(0, 0, 2, 2, 0)$, or equivalently that of $(\mu, \sigma_+^2, \beta_0, \beta_1, \eta^2)$ is $(0, 1, 0, 0, 1)$; the rationale for this assumption can be found in Section 8.

We first show that the MLEs for $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)^\tau$ are consistent. This is achieved by showing that for any estimator $\bar{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, if it leads to a sufficiently large value of the likelihood function, namely $\ell_n(\bar{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0) \geq C$ for some constant $C > -\infty$, then under the null model, $\bar{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$. We show that this conclusion is generally valid whether or not the assumption(s) $\sigma_1 = \sigma_2$ and/or $\rho = 0$ are imposed. Therefore, this ensures the consistency of the MLEs of $\boldsymbol{\theta}$ derived from different assumptions.

We then derive the asymptotic distributions of the LRTs through their asymptotic quadratic forms based on the consistency of the MLEs of $\boldsymbol{\theta}$. We need the following notation:

$$B_i = \left(\frac{(Z_{1i}^2 - 1)(Z_{2i}^2 - 1)}{2}, Z_{1i}(Z_{2i}^2 - 1), -\frac{(Z_{2i}^4 - 6Z_{2i}^2 + 3)}{12} \right)^\tau, \quad (3.4)$$

$$D_i = \left(Z_{1i}, \frac{Z_{1i}^2 + Z_{2i}^2 - 2}{2}, \frac{Z_{2i}^2 - Z_{1i}^2}{4}, \frac{(Z_{1i}^2 - 1)(Z_{2i}^2 - 1)}{2}, Z_{1i}(Z_{2i}^2 - 1) \right)^\tau. \quad (3.5)$$

Applying the central limit theorem, we have

$$n^{-1/2} \sum_{i=1}^n B_i \xrightarrow{D} N(0, \Sigma_B) \quad \text{and} \quad n^{-1/2} \sum_{i=1}^n D_i \xrightarrow{D} N(0, \Sigma_D), \quad (3.6)$$

where

$$\Sigma_B = \text{VAR}(B_i) = \text{diag}(1, 2, 1/6) \quad \text{and} \quad \Sigma_D = \text{VAR}(D_i) = \text{diag}(1, 1, 1/4, 1, 2).$$

With some algebra, we are able to obtain the asymptotic quadratic forms of the LRTs; these together with (3.6) lead to the asymptotic distribution of the corresponding LRTs claimed in Theorems 1 and 2.

- If $\rho = 0$ is known and $\sigma_1^2 = \sigma_2^2$, we have

$$R_{n,1} = 4n^{-1} \left\{ \left(\sum_{i=1}^n D_i[3] \right)^+ \right\}^2 + o_p(1),$$

where $D_i[k]$ denotes the k th element of the vector D_i , and this convention is applicable to all vectors hereafter.

- If $\rho = 0$ is known without $\sigma_1^2 = \sigma_2^2$, we have

$$R_{n,2} = \sup_{\mathbf{t}_2} \left\{ 2\mathbf{t}_2^\tau \sum_{i=1}^n D_{i2} - n\mathbf{t}_2^\tau \Sigma_{D2} \mathbf{t}_2 \right\} + o_p(1), \quad (3.7)$$

where $\mathbf{t}_2 = (\beta_0^2, \beta_1^2, \beta_0\beta_1)^\tau$, $D_{i2} = (D_i[3], D_i[4], D_i[5])^\tau$, and $\Sigma_{D2} = \text{VAR}(D_{i2})$.

- If ρ is unknown, but $\sigma_1^2 = \sigma_2^2$, we have

$$R_{n,1}^* = 6n^{-1} \left\{ \left(\sum_{i=1}^n B_i[3] \right)^+ \right\}^2 + o_p(1).$$

- If ρ is unknown, we have

$$\begin{aligned} R_{n,2}^* &= (1/2)n^{-1} \left\{ \sum_{i=1}^n B_i[2] \right\}^2 \\ &+ \max \left[n^{-1} \left\{ \left(\sum_{i=1}^n B_i[1] \right)^+ \right\}^2, 6n^{-1} \left\{ \left(\sum_{i=1}^n B_i[3] \right)^+ \right\}^2 \right] \\ &+ o_p(1). \end{aligned}$$

4. Adjusted limiting distributions

One drawback of general asymptotic results is that they may offer poor approximations to the corresponding finite-sample distributions. The convergence rates

of the parameter estimators given in Theorems 1 and 2 are much lower than those of the MLEs from the regular parametric models. This adversely affects the approximation accuracy of the asymptotic distributions to the finite-sample distributions of the LRT statistics. To improve the approximation precision when n is not very large, we use the Bartlett correction.

The Bartlett correction was originally proposed by Bartlett (1937) to improve the approximation of the limiting distribution of an LRT under the regular model, and it was generalized by Lawley (1956). Let T_n be the LRT statistic for a hypothesis testing problem, which asymptotically follows a χ_d^2 distribution under H_0 . Define b_1 according to the expansion:

$$d^{-1}E(T_n) = 1 + n^{-1}b_1 + O(n^{-2}).$$

Let \hat{b}_1 be a \sqrt{n} -consistent estimator of b_1 under H_0 ; specifically $\hat{b}_1 = b_1$ if b_1 is known. The Bartlett-corrected LRT is then defined to be $T_n^* = (1 + \hat{b}_1/n)^{-1}T_n$. Barndorff-Nielsen and Hall (1988) showed that under appropriate regularity conditions, the Bartlett correction reduces the error rate of T_n from n^{-1} to n^{-2} . Hence, χ_d^2 better approximates the distribution of T_n^* than that of T_n ; in other words, $(1 + \hat{b}_1/n)\chi_d^2$ provides a better approximation to the distribution of T_n than χ_d^2 does. In short, the idea of the Bartlett correction is to adjust the limiting distribution such that its first moment matches that of the LRT up to order $O(n^{-1})$; this helps to improve the approximation accuracy of the limiting distribution to the LRT.

In this spirit, we search for accurate approximate distributions for $R_{n,1}$, $R_{n,2}$, $R_{n,1}^*$, and $R_{n,2}^*$ as follows. Recall that R and R^* are the limiting distributions of $R_{n,2}$ and $R_{n,2}^*$. Let

$$\begin{aligned} F_{n1} &= (1 - p_n)\chi_0^2 + p_n\chi_1^2, & F_{n2} &= r_n R, \\ F_{n1}^* &= (1 - p_n^*)\chi_0^2 + p_n^*\chi_1^2, & F_{n2}^* &= r_n^* R^*. \end{aligned}$$

We aim to find p_n , r_n , p_n^* , and r_n^* so that the above distributions have first moments very close to the first moments of their corresponding test statistics for a wide range of n values. High-order asymptotic techniques can be used, but they may involve complicated analytical tools with little assurance of the quality of the end products. The computer experiment approach of Chen and Li (2011) is more effective and practical, and it matches the spirit of the data science.

The experiment works as follows. We consider a sufficiently wide range of values for n . For each n , we simulate a large number of data sets, with each data set composed of n i.i.d. unordered paired observations. Due to the invariance property of the LRT statistics, each data set is generated from the standard bivariate normal distribution. Based on these data sets, we obtain the simulated first moments of $R_{n,1}$, $R_{n,2}$, $R_{n,1}^*$, and $R_{n,2}^*$. We choose p_n so that the simulated first moment of $R_{n,1}$ matches the first moment of F_{n1} . We then look for a regression model for p_n versus n . Similar procedures are applied to obtain regression models for r_n , p_n^* , and r_n^* .

Specifically, let us take $R_{n,1}$ for ease of illustration:

Step 1. For every n in $\{10, 20, \dots, 100\}$, generate $N = 50,000$ data sets of size n .

Step 2. Obtain N values of $R_{n,1}$ and therefore its simulated first moment, denoted \hat{p}_n . Match \hat{p}_n with the first moment of F_{n1} to find $p_n = \hat{p}_n$.

Step 3. Fit a regression model to (n, p_n) with p_n being the response and n being the covariate.

We postulate the following nonlinear but parametric regression models:

$$p_n = 0.5 + an^{-b} + \epsilon_n \tag{4.1}$$

$$r_n = 1 + an^{-b} + \epsilon_n \tag{4.2}$$

$$p_n^* = 0.5 + an^{-b} + \epsilon_n \tag{4.3}$$

$$r_n^* = 1 + an^{-b} + \epsilon_n, \tag{4.4}$$

with a and b being regression parameters, and ϵ_n accounting for imperfect fit. Applying Steps 1–2 outlined above leads to the p_n, r_n, p_n^* , and r_n^* values in Table 1. Fitting the nonlinear regression models (4.1)–(4.4) to the data in Table 1 gives us the fitted values of a and b . With these values, we calculate the approximate p-values with the following adjusted limiting distributions:

$$\begin{aligned} (0.5 - 1.440n^{-0.676})\chi_0^2 + (0.5 + 1.440n^{-0.676})\chi_1^2 & \text{ for } R_{n,1}, \\ (1 + 4.589n^{-1.163})R & \text{ for } R_{n,2}, \\ (0.5 - 1.332n^{-0.492})\chi_0^2 + (0.5 + 1.332n^{-0.492})\chi_1^2 & \text{ for } R_{n,1}^*, \\ (1 + 6.325n^{-1.176})R^* & \text{ for } R_{n,2}^*. \end{aligned}$$

For each nonlinear regression model in (4.1)–(4.4), we calculate the squared correlations between the predicted values and the responses. These are 98.9%, 99.4%, 99.7%, and 99.5%, indicating that models (4.1)–(4.4) fit the data in Table 1 very well. We have implemented the four LRT statistics with the proposed adjusting limiting distributions in an R package; it is available upon request.

TABLE 1
Values of p_n, r_n, p_n^* , and r_n^* via computer experiments

n	10	20	30	40	50	60	70	80	90	100
p_n	0.809	0.681	0.634	0.627	0.596	0.587	0.585	0.587	0.568	0.568
r_n	1.312	1.150	1.092	1.070	1.046	1.028	1.030	1.032	1.016	1.012
p_n^*	0.932	0.801	0.749	0.721	0.687	0.674	0.669	0.651	0.649	0.645
r_n^*	1.417	1.194	1.129	1.090	1.062	1.040	1.038	1.028	1.022	1.018

5. Simulation studies

5.1. Data generation

Because of the invariance property, we need only study the LRT tests based on data generated from distributions with standardized parameter values.

To examine the sizes of the tests, we simulate at $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$ in (1.1). We study five cases corresponding to $\rho = -0.5, -0.25, 0, 0.25, \text{ and } 0.5$. To compare the powers of the tests, we set $\mu_1 = 0, \sigma_1 = 1$, and form 20 cases as combinations of $\mu_2 = 1.0, 1.5, \sigma_2 = 1.0, 0.5$ and $\rho = -0.5, -0.25, 0, 0.25, 0.5$.

In each case, we generate (X_1, X_2) from model (1.1) with one of the above parameter settings. Then, we obtain $Y_1 = \min\{X_1, X_2\}$ and $Y_2 = \max\{X_1, X_2\}$. We repeat the process to obtain n unordered pairs (Y_1, Y_2) .

Based on each set of n unordered pairs, we compute the values of $R_{n,1}, R_{n,2}, R_{n,1}^*$, and $R_{n,2}^*$ and carry out the tests for H_0 without checking that the model for generating the data satisfies the conditions for the tests. We record the rejection rates based on 50,000 repetitions; the results are presented in the next section.

5.2. Results

We calculate the rejection rate of each test at the significance levels $\alpha = 10\%, 5\%, \text{ and } 1\%$. The rejection percentages under the null models are summarized in Table 2.

When $\rho = 0$, X_1 and X_2 are independent. The assumptions for all the LRTs, $R_{n,1}, R_{n,2}, R_{n,1}^*$, and $R_{n,2}^*$, are satisfied. However, as shown in the first section of Table 2, if their limiting distributions are applied without adjustment, the resulting tests are inaccurate: their type I errors markedly exceed the nominal significance levels. The adjustment proposed in Section 4 is very helpful. After the adjustment, the type I errors of all the tests are close to the nominal levels. The precision is impressive since the adjustment works well even when n is as small as 25.

When $\rho = \pm 0.25$ or ± 0.5 , the model assumptions for $R_{n,1}$ and $R_{n,2}$ are violated. When we apply the tests, the type I errors are either near zero when $\rho = 0.25$ or 0.5 or seriously inflated when $\rho = -0.25$ or -0.5 . In contrast, because of their invariance property, $R_{n,1}^*$ and $R_{n,2}^*$ continue to perform well: with their limiting distributions adjusted, they have satisfactory precision in the type I errors.

To further illustrate the effects of the adjustment on the limiting distributions, Figure 1 presents the type I errors (%) of our LRTs at the 5% significance level when $100 \leq n \leq 2500$ and $\rho = 0$. The trends for the 10% and 1% significance levels are similar and are omitted. The plots show that the type I errors of $R_{n,1}, R_{n,2}$ after the adjustment are within a 0.2% band of the nominal level for large n and a 0.4% band otherwise; similar results are observed for $R_{n,1}^*$. For $R_{n,1}^*$, the approximation accuracy shows no clear improvement as n increases, but the type I errors are between 5% and 5.4%, which is sufficiently accurate for typical applications. We have also run the simulation for $n = 3000, 3500, \dots, 5000$. The trends are similar to the results for $n = 2500$ and are omitted.

TABLE 2
 Simulated Type I errors (%) of LRTs based on limiting distributions/adjusted limiting distributions

Levels	n = 25			n = 75		
	10%	5%	1%	10%	5%	1%
	$\rho = 0$					
$R_{n,1}$	13.7/10.7	7.3/5.7	1.8/1.4	11.3/ 9.9	5.9/5.1	1.3/1.2
$R_{n,2}$	12.9/10.6	6.9/5.2	1.6/1.0	10.8/10.2	5.6/5.2	1.2/1.0
$R_{n,1}^*$	15.9/10.5	8.1/5.5	1.8/1.1	13.4/10.4	7.0/5.5	1.5/1.1
$R_{n,2}^*$	13.5/10.1	7.4/5.0	1.8/1.1	11.1/10.1	5.9/5.2	1.2/1.0
	$\rho = 0.25$					
$R_{n,1}$	1.2/0.8	0.5/0.3	0.1/0.1	0.1/0.0	0.0/0.0	0.0/0.0
$R_{n,2}$	3.8/3.0	1.9/1.4	0.4/0.3	1.8/1.7	0.7/0.7	0.1/0.1
$R_{n,1}^*$	15.9/10.5	8.1/5.5	1.8/1.1	13.4/10.4	7.0/5.5	1.5/1.1
$R_{n,2}^*$	13.5/10.1	7.4/5.0	1.8/1.1	11.1/10.1	5.9/5.2	1.2/1.0
	$\rho = 0.5$					
$R_{n,1}$	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0
$R_{n,2}$	0.7/0.5	0.3/0.2	0.0/0.0	0.1/0.1	0.0/0.0	0.0/0.0
$R_{n,1}^*$	15.9/10.5	8.1/5.5	1.8/1.1	13.4/10.4	7.0/5.5	1.5/1.1
$R_{n,2}^*$	13.5/10.1	7.4/5.0	1.8/1.1	11.1/10.1	5.9/5.2	1.2/1.0
	$\rho = -0.25$					
$R_{n,1}$	53.7/47.2	38.6/33.0	15.2/12.7	83.1/80.9	71.6/69.1	43.6/41.3
$R_{n,2}$	39.0/34.0	25.5/21.2	8.6/6.2	67.6/66.3	53.6/52.0	27.3/25.6
$R_{n,1}^*$	15.9/10.5	8.1/5.5	1.8/1.1	13.4/10.4	7.0/5.5	1.5/1.1
$R_{n,2}^*$	13.5/10.1	7.4/5.0	1.8/1.1	11.1/10.1	5.9/5.2	1.2/1.0
	$\rho = -0.5$					
$R_{n,1}$	92.6/89.9	84.5/80.5	57.5/52.4	100.0/99.9	99.9/99.8	98.5/98.3
$R_{n,2}$	80.1/76.2	67.1/61.3	37.3/30.2	99.7/99.6	99.0/98.9	94.5/93.9
$R_{n,1}^*$	15.9/10.5	8.1/5.5	1.8/1.1	13.4/10.4	7.0/5.5	1.5/1.1
$R_{n,2}^*$	13.5/10.1	7.4/5.0	1.8/1.1	11.1/10.1	5.9/5.2	1.2/1.0

Next, we compare the powers of $R_{n,1}$, $R_{n,2}$, $R_{n,1}^*$, and $R_{n,2}^*$ under the alternatives. All combinations of n , ρ , μ , and σ are incorporated, as described in Section 5.1. Their powers, summarized in Table 3, are computed at the 5% significance level based on the adjusted limiting distributions. We observe that when $\rho = 0$, $R_{n,1}$ and $R_{n,2}$ have higher powers than $R_{n,1}^*$ and $R_{n,2}^*$; when $\rho = 0.25$, $R_{n,1}$ and $R_{n,2}$ have higher powers in most cases; when ρ is increased to 0.5, $R_{n,1}^*$ and $R_{n,2}^*$ are much more powerful; when $\rho = -0.25$ and -0.5 , $R_{n,1}$ and $R_{n,2}$ are more powerful, but at the cost of the inflated type I errors reported in Table 2; a test with a markedly inflated type I error is generally not recommended.

From Tables 2 and 3, we also observe that both the type I errors and the powers of $R_{n,1}$ and $R_{n,2}$ decrease as ρ increases, whereas the powers of $R_{n,1}^*$ and $R_{n,2}^*$ increase as ρ increases. We use the reparameterization in Section 3.1 to provide some insight into these results. For simplicity, we consider the case where $\sigma_1^2 = \sigma_2^2 = \sigma^2$, under which testing (1.2) is equivalent to testing $H_0 : \beta_0 = 0$ based on observations from the mixture model $0.5N(\beta_0, \eta^2) + 0.5N(-\beta_0, \eta^2)$.

The tests based on $R_{n,1}$ and $R_{n,2}$ are established assuming $\rho = 0$ is known. In view of (3.3), this is equivalent to assuming $\sigma_+^2 = 0.5\sigma^2 + 0.5\rho\sigma^2$ and $\eta^2 = 0.5\sigma^2 - 0.5\rho\sigma^2$ are equal. Thus, when the true value of ρ is positive, the approach behind these two tests overestimates η^2 . The estimator based on the likelihood (3.2), on the other hand, is asymptotically unbiased for the variance of the mixture distribution $0.5N(\beta_0, \eta^2) + 0.5N(\beta_0^2, \eta^2)$, which equals $\beta_0^2 + \eta^2$.

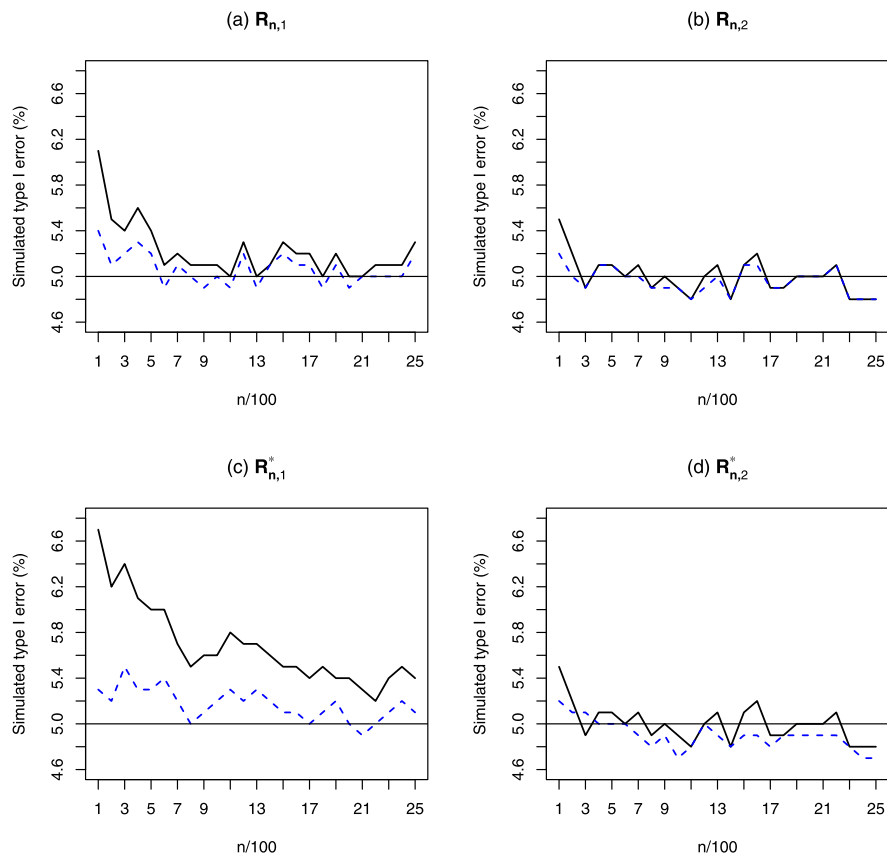


FIG 1. Simulated type I errors (%) at the 5% significance level when $100 \leq n \leq 1500$ and $\rho = 0$. The solid and dashed lines are the rates before and after the adjustments, respectively.

With η^2 underestimated and $\beta_0^2 + \eta^2$ unbiasedly estimated, we obtain an underestimated β_0^2 , asymptotically. As ρ gets larger, the overestimation of η^2 and underestimation of β_0^2 become more severe. This explains the decrease in the type I errors and powers of $R_{n,1}$ and $R_{n,2}$. Similar arguments are applicable to the case where $\rho < 0$.

Tests based on $R_{n,1}^*$ and $R_{n,2}^*$ are established without a known- ρ assumption. The test problems become testing $H_0 : \beta_0 = 0$ based on observations from the mixture model:

$$0.5N(\beta_0, \eta^2) + 0.5N(-\beta_0, \eta^2) = \eta \{0.5N(\beta_0/\eta, 1) + 0.5N(-\beta_0/\eta, 1)\}.$$

The powers of these tests are determined by the value of $|\beta_0|/\eta = |\mu_1 - \mu_2|/\sqrt{2(1 - \rho)\sigma^2}$. Therefore, if the true value of ρ increases, the powers of $R_{n,1}^*$ and $R_{n,2}^*$ increase. This is in line with the trend observed in Table 3.

TABLE 3
Powers (%) of $R_{n,1}$, $R_{n,2}$, $R_{n,1}^*$, and $R_{n,2}^*$ at the 5% significance level

σ	μ	$n = 25$				$n = 75$			
		$R_{n,1}$	$R_{n,2}$	$R_{n,1}^*$	$R_{n,2}^*$	$R_{n,1}$	$R_{n,2}$	$R_{n,1}^*$	$R_{n,2}^*$
$\rho = 0$									
1.0	1.0	28.1	18.3	8.3	6.3	57.6	41.8	11.2	8.0
1.0	1.5	67.0	49.7	19.2	11.3	97.5	93.0	40.2	24.8
0.5	1.0	46.9	85.2	12.3	70.5	88.2	99.9	21.7	99.6
0.5	1.5	92.2	99.2	39.2	90.6	100.0	100.0	79.7	100.0
$\rho = 0.25$									
1.0	1.0	7.2	6.2	10.4	7.2	6.7	6.0	16.7	10.5
1.0	1.5	38.8	27.0	29.6	17.5	70.9	56.9	63.9	44.8
0.5	1.0	22.4	77.3	16.4	78.2	43.2	99.8	32.5	99.9
0.5	1.5	80.9	98.5	54.0	95.5	99.7	100.0	93.5	100.0
$\rho = 0.5$									
1.0	1.0	1.0	1.9	15.8	9.8	0.1	1.0	32.8	20.0
1.0	1.5	17.7	13.1	54.7	34.6	22.4	16.6	93.7	83.2
0.5	1.0	8.4	71.8	24.3	91.3	7.6	99.6	53.6	100.0
0.5	1.5	66.0	98.1	76.4	99.5	95.7	100.0	99.5	100.0
$\rho = -0.25$									
1.0	1.0	65.1	45.6	7.3	5.9	97.7	93.1	9.0	6.8
1.0	1.5	90.0	76.1	14.2	9.0	100.0	99.9	27.1	16.5
0.5	1.0	75.7	92.1	10.2	68.3	99.7	100.0	16.6	99.5
0.5	1.5	97.9	99.7	29.5	87.8	100.0	100.0	64.5	100.0
$\rho = -0.5$									
1.0	1.0	93.8	81.0	6.7	5.7	100.0	100.0	8.1	6.4
1.0	1.5	99.0	94.3	11.3	7.9	100.0	100.0	19.8	12.2
0.5	1.0	94.9	97.8	9.0	73.9	100.0	100.0	13.3	99.8
0.5	1.5	99.7	100.0	23.3	90.6	100.0	100.0	50.3	100.0

6. Real-data examples

6.1. Data from karyotype analysis

This example considers 40 unordered pairs of the lengths of the longer and shorter arms of chromosome II of *Larix decidua* from 40 specimens; so $n = 40$. The data are available in Table 1 of Matérn and Simak (1968). The test results from $R_{n,1}$, $R_{n,2}$, $R_{n,1}^*$, and $R_{n,2}^*$ for (1.2) are as follows:

- $R_{n,1} = 14.91$ and $R_{n,2} = 17.71$. Calibrated by the adjusted limiting distributions, the asymptotic p -values of $R_{n,1}$ and $R_{n,2}$ are 7×10^{-5} and 2×10^{-4} .
- $R_{n,1}^* = 1.08$ and $R_{n,2}^* = 16.69$. Calibrated by the adjusted limiting distributions, the asymptotic p -values of $R_{n,1}^*$ and $R_{n,2}^*$ are 0.21 and 4×10^{-4} .

The maximum likelihood estimate of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ is found to be

$$(\hat{\mu}_1^*, \hat{\mu}_2^*, \hat{\sigma}_1^*, \hat{\sigma}_2^*, \hat{\rho}^*) = (62.05, 65.55, 3.50, 8.20, -0.73).$$

Note that $\hat{\rho}^* = -0.73$ suggests strong negative correlation between X_{1i} and X_{2i} . As revealed in the simulation studies reported in the bottom section of Table 2, $R_{n,1}$ and $R_{n,2}$ are therefore not reliable because they are designed for $\rho = 0$. Moreover, the fitted values $\hat{\mu}_1^*$ and $\hat{\mu}_2^*$ are very close, but $\hat{\sigma}_1^*$ and $\hat{\sigma}_2^*$ are significantly different. Hence, $R_{n,1}^*$ is unsuitable because it is designed for the case where $\sigma_1 = \sigma_2$. We recommend $R_{n,2}^*$, which is designed to detect departures from either equal-mean or equal-variance hypotheses.

Matérn and Simak (1968) proposed a method for testing $H_0 : \mu_1 = \mu_2$, and the outcome is not significant at the 5% level. We observe that their test is different from ours, since their H_0 does not require $\sigma_1^2 = \sigma_2^2$. Our test based on $R_{n,2}^*$ rejects $(\mu_1, \sigma_1^2) = (\mu_2, \sigma_2^2)$ with high significance. In fact, we can use the classical LRT for their hypothesis. In particular, let

$$\tilde{R}_n = 2 \left\{ \ell_n(\hat{\boldsymbol{\theta}}^*) - \ell_n(\check{\boldsymbol{\theta}}^0) \right\},$$

where $\ell_n(\boldsymbol{\theta})$ is as in (2.1), $\hat{\boldsymbol{\theta}}^*$ is given in Section 2.2, and

$$\check{\boldsymbol{\theta}}^0 = \arg \sup_{\boldsymbol{\theta}} \{ \ell_n(\boldsymbol{\theta}) : \mu_1 = \mu_2 \}.$$

We observe (proof omitted) that the model under $H_0 : \mu_1 = \mu_2$ is regular when $(\mu_1, \sigma_1^2) \neq (\mu_2, \sigma_2^2)$. Therefore, \tilde{R}_n has a χ_1^2 limiting distribution. The test results are $\tilde{R}_n = 3.21$ with p -value = 0.07, which is in line with the conclusion of Matérn and Simak (1968). In summary, both \tilde{R}_n and Matérn and Simak (1968) indicate that μ_1 and μ_2 are not empirically distinguishable based on the given data; however, $R_{n,2}^*$ suggests that $(\mu_1, \sigma_1^2) = (\mu_2, \sigma_2^2)$ should be rejected with high significance. The difference in the two distributions comes from the difference in their variances.

6.2. C-band area of human chromosome data

This example consists of normalized measurements of the C-band area on the No. 9 chromosome pair (Mason et al., 1975). The measurements are based on three groups: the father, mother, and offspring. These groups respectively have 40, 18, and 31 unordered pairs of normalized measurements of the C-band area. The data are available in Table 1 of Lauder (1977). We analyze the group of fathers as an example; the analysis of the other groups is similar. We constructed $R_{n,1}$, $R_{n,2}$, $R_{n,1}^*$, and $R_{n,2}^*$ and the corresponding p -values from the adjusted limiting distributions. The results are as follows:

- $R_{n,1} = 6.51$ and $R_{n,2} = 9.47$ with $n = 40$. Calibrated by the adjusted limiting distributions, the asymptotic p -values of $R_{n,1}$ and $R_{n,2}$ are 6.6×10^{-3} and 8.9×10^{-3} .
- $R_{n,1}^* = 10.74$ and $R_{n,2}^* = 13.48$ with $n = 40$. Calibrated by the adjusted limiting distributions, the asymptotic p -values of $R_{n,1}^*$ and $R_{n,2}^*$ are 7.5×10^{-4} and 1.9×10^{-3} .

The maximum likelihood estimate of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ is found to be

$$(\hat{\mu}_1^*, \hat{\mu}_2^*, \hat{\sigma}_1^*, \hat{\sigma}_2^*, \hat{\rho}^*) = (86.75, 68.58, 10.55, 8.29, 0.46).$$

Note that $\hat{\rho}^* = 0.46$ suggests strong positive correlation between X_{1i} and X_{2i} . Moreover, $\hat{\mu}_1^*$ and $\hat{\mu}_2^*$ are quite different whereas $\hat{\sigma}_1^* \approx \hat{\sigma}_2^*$. These suggest that

$R_{n,1}^*$ is the most suitable test while $R_{n,2}^*$ is also a possibility. Note that $R_{n,1}^*$ is sharper than $R_{n,2}^*$ with a smaller p-value.

Lauder (1977) used the aforementioned data to study the transmission of the heritage properties on the chromosome from parents to children. The validity of the analysis hinges on $(\mu_1, \sigma_1) \neq (\mu_2, \sigma_2)$ in the model for the unordered pair data of the father or the mother, but the paper did not verify this assumption. Our work fills this gap and provides numerical evidence to support the analysis.

7. Discussion

We have proposed methods for testing the homogeneity of unordered pair data. With the LRT approach based on the likelihood of the bivariate normal random vectors, we considered the testing problem (1.2) for four scenarios: (1) $\rho = 0$ and $\sigma_1^2 = \sigma_2^2$; (2) $\rho = 0$; (3) $\sigma_1^2 = \sigma_2^2$; and (4) no assumption for these parameters. We derived both the limiting distributions of the LRTs and the convergence rates of the MLEs of the unknown parameters. Furthermore, in the spirit of the Bartlett correction, we proposed the adjusted limiting distribution for each LRT. By simulation, we demonstrated that the adjusted limiting distributions provide accurate approximations to the finite-sample distributions of the corresponding LRTs even when the sample size is as small as $n = 25$.

Throughout this paper, we have focused on unordered *paired* observations from *normal* random vectors. This leads to many interesting but open-ended research topics. For example, the problem becomes more challenging when the data are n unordered k -tuples. Our methods and theory serve as a useful starting point for the general problem. One obstacle is that the joint density function of an unordered k -tuple involves $k!$ terms. This makes the asymptotic expansion of the log-likelihood extremely complicated and therefore introduces tremendous difficulty in the technical development. Another obstacle is that the reparameterization in Section 3.1 plays an important role in the theoretical development. For unordered k -tuples with $k > 2$, we have yet to find a comparable reparameterization. Because of these difficulties, we leave this theoretical development to future research. It would also be interesting to consider unordered k -tuples that follow the general exponential family distributions. In addition to the aforementioned technical difficulties, we may need to model the correlation structure among the k -tuples. We expect that copula models (Nelsen, 2006) will be useful for this problem.

8. Technical details

8.1. Some preparation

Throughout the proofs, we use a generic $\boldsymbol{\theta}$ for the parameters, which may be interpreted as $\boldsymbol{\theta} = (\mu, \sigma_+, \beta_0, \beta_1, \eta)^\top$ when necessary. Recall that $(\mu, \sigma_+, \beta_0, \beta_1, \eta)$ is defined in Section 3.1, the log-likelihood $\ell(\boldsymbol{\theta})$ can be written as $\ell(\boldsymbol{\theta}) =$

$\ell_{n,1}^*(\mu, \sigma_+) + \ell_{n,2}^*(\beta_0, \beta_1, \eta)$, where

$$\ell_{n,1}^*(\mu, \sigma_+) = \sum_{i=1}^n \log\{\phi(Z_{1i}; \mu, \sigma_+)\},$$

and $\ell_{n,2}^*(\beta_0, \beta_1, \eta)$ can be equivalently written as

$$\ell_{n,2}^*(\beta_0, \beta_1, \eta) = \sum_{i=1}^n \log\{0.5\phi(Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta) + 0.5\phi(-Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta)\}.$$

Under H_0 in Theorem 1, which includes the assumption that $\rho = 0$, suppose that the true parameter values of the data-generating distribution are $\mu_1 = \mu_2 = \mu_*$, $\sigma_1^2 = \sigma_2^2 = \sigma_*^2$. We may then, in our proofs, work with the transformed data

$$X_1^* = \sqrt{2}(X_1 - \mu_*)/\sigma_*, \quad X_2^* = \sqrt{2}(X_2 - \mu_*)/\sigma_*.$$

After the transformation, the algebraic form of the likelihood does not change, but the true parameter values of the data-generating distribution become $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 2$. Without loss of generality, based on the above invariance property, we may assume that the true parameters $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 2$ under H_0 .

Under H_0 in Theorem 2, without loss of generality, the same assumption is applicable to μ and σ . We now reveal that by the same invariance principle we may also assume $\rho = 0$ as long as the true value $\rho \neq \pm 1$. When $\rho \neq \pm 1$, we simply let

$$(X_1^{**}, X_2^{**}) = \left\{ X_1^*, (X_2^* - \rho X_1^*)/\sqrt{1 - \rho^2} \right\}.$$

The distribution-generated data $\{X_1^{**}, X_2^{**}\}$ now has the true parameter values $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 2$, and $\rho = 0$ under H_0 .

With the above standardization operation, for both Theorems 1 and 2, we study the asymptotic null properties under the assumption that Z_{1i} and Z_{2i} are independent normal random variables with the standard parameter values:

$$(\mu, \sigma_+, \beta_0, \beta_1, \eta) = (0, 1, 0, 0, 1).$$

8.2. Some useful lemmas

In this section, we present three lemmas. Lemma 1 is a technical preparation that provides an upper bound on the number of observations in a set. Lemma 2 indicates that under the null model any estimator of θ with a large likelihood value is consistent for θ . Lemma 3 strengthens Lemma 2 by providing specific convergence rates.

Lemma 1. *As $n \rightarrow \infty$, we have, almost surely,*

$$\sup_{\beta_0, \beta_1} \sum_{i=1}^n \mathbb{I}(|Z_{2i} - \beta_0 - \beta_1 Z_{1i}| \leq 1/4) \leq (1/4)n,$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Proof. Note that

$$n^{-1} \sum_{i=1}^n \mathbb{I}(|Z_{2i} - \beta_0 - \beta_1 Z_{1i}| \leq 1/4)$$

is the empirical measure of the two-dimensional stripe formed by the inequality

$$|Z_2 - \beta_0 - \beta_1 Z_1| \leq 1/4.$$

This class of stripes can divide n points in two-dimensional space into at most a polynomial number of different subsets. By Pollard (1990), this property implies the uniform strong law of large numbers:

$$\sup_{\beta_0, \beta_1} \left| n^{-1} \sum_{i=1}^n \mathbb{I}(|Z_{2i} - \beta_0 - \beta_1 Z_{1i}| \leq 1/4) - P(|Z_2 - \beta_0 - \beta_1 Z_1| \leq 1/4) \right| \rightarrow 0 \tag{8.1}$$

almost surely.

The distribution of $Z_2 - \beta_0 - \beta_1 Z_1$ is normal with variance at least 1. Based on this, we have $P(|Z_2 - \beta_0 - \beta_1 Z_1| \leq 1/4) \leq 0.2$ for any β_0, β_1 . Hence, almost surely,

$$\sum_{i=1}^n \mathbb{I}(|Z_{2i} - \beta_0 - \beta_1 Z_{1i}| \leq 1/4) \leq 0.2n + o(n) \leq 0.25n.$$

This completes the proof. □

Lemma 2. *Suppose an estimator $\bar{\theta}$ satisfies*

$$\begin{aligned} \ell_n(\bar{\theta}) - \ell_n(\theta_0) &= \{\ell_{n,1}^*(\bar{\mu}, \bar{\sigma}_+) + \ell_{n,2}^*(\bar{\beta}_0, \bar{\beta}_1, \bar{\eta})\} - \{\ell_{n,1}^*(0, 1) + \ell_{n,2}^*(0, 0, 1)\} \\ &= \{\ell_{n,1}^*(\bar{\mu}, \bar{\sigma}_+) - \ell_{n,1}^*(0, 1)\} + \{\ell_{n,2}^*(\bar{\beta}_0, \bar{\beta}_1, \bar{\eta}) - \ell_{n,2}^*(0, 0, 1)\} \\ &\geq C > -\infty \end{aligned} \tag{8.2}$$

for some constant C . Then under the null model, $\bar{\theta} = \theta_0 + o_p(1) = (0, 1, 0, 0, 1)^\tau + o_p(1)$.

Proof. The classical consistency proof of the MLE such as that in Wald (1949) is essentially done for models with a compact parameter space. If necessary, one may first compactify the parameter space or show that, with probability one, the MLE is inside a compact parameter subspace. We adopt the second approach by showing that as $n \rightarrow \infty$, $\epsilon_0 \leq \bar{\eta} \leq M_0$ and $\epsilon_0 \leq \bar{\sigma}_+ \leq M_0$ almost surely for some positive ϵ_0 and M_0 . For convenience, we specifically choose $M_0 = \exp(4)$ and ϵ_0 such that when $\eta < \epsilon_0$, $\log(\eta) + (1/64)/\eta^2 \geq 2$. The existence of such an ϵ_0 is obvious.

We start with $\bar{\eta}$. Note that we have decomposed $\ell_n(\bar{\theta}) - \ell_n(\theta_0)$ into a sum of two terms. For the first term, according to the classical result for the LRT under regular models (Serfling, 2000), it is clear that

$$\sup_{\mu, \sigma_+} \{\ell_{n,1}^*(\mu, \sigma_+) - \ell_{n,1}^*(0, 1)\} = O_p(1). \tag{8.3}$$

When in the second term the variance parameter $\eta > M_0 = \exp(4)$, we have

$$\sum_{i=1}^n \log\{0.5\phi(Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta) + 0.5\phi(-Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta)\} \leq -n \log M_0 = -4n.$$

By the law of large numbers, we have

$$n^{-1} \ell_{n,2}^*(0, 0, 1) \geq -(1/2) \log(2\pi) - \mathbb{E}(Z_2^2) \geq -2$$

almost surely. This implies that **when** $\eta > M_0$,

$$\ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \leq -2n \quad (8.4)$$

and subsequently, uniformly for η in this range,

$$\ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \rightarrow -\infty.$$

Together with (8.3), we have, whenever $\eta > M_0$,

$$\ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}_0) \rightarrow -\infty$$

in probability. Since the lemma condition clearly states that $\bar{\eta}$ does not have the above property, it cannot be in this range. That is, we conclude that $\bar{\eta} \leq M_0$.

Suppose $\eta < \epsilon_0$. In this case, for all i , we have

$$\log\{0.5\phi(Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta) + 0.5\phi(-Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta)\} \leq -\log(\eta).$$

For i such that

$$\min\{|Z_{2i} + \beta_0 + \beta_1 Z_{1i}|, |Z_{2i} - \beta_0 - \beta_1 Z_{1i}|\} > 1/4, \quad (8.5)$$

we have

$$\log\{0.5\phi(Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta) + 0.5\phi(-Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta)\} \leq -\log(\eta) - (1/32)/\eta^2.$$

By Lemma 1, uniformly in β_0 and β_1 and almost surely, at most $(1/2)n$ of the i 's satisfy (8.5). Therefore, when $\eta < \epsilon_0$,

$$\ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \leq -\{\log(\eta) + (1/64)/\eta^2\}n \leq -2n \quad (8.6)$$

uniformly in β_0 and β_1 and almost surely. Hence, for all $\eta < \epsilon_0$, we also have

$$\ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \rightarrow -\infty.$$

In conclusion, the $\bar{\eta}$ value satisfying the lemma condition must almost surely fall within the interval $[\epsilon_0, M_0]$. That is, this result shows that we may reduce the parameter space of (β_0, β_1, η) to $\mathbb{R}^2 \times [\epsilon_0, M_0]$ for asymptotic considerations.

Next, we consider $\bar{\sigma}_+$. For a sufficiently small $\epsilon > 0$, let

$$B_\epsilon = \{(\beta_0, \beta_1, \eta) : \beta_0^2 + \beta_1^2 + (\eta^2 - 1)^2 \leq \epsilon^2\}$$

be a ball centered at the true value $(0, 0, 1)$. On the reduced parameter space $\mathbb{R}^2 \times [\epsilon_0, M_0]$, the density function

$$0.5\phi(Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta) + 0.5\phi(-Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta)$$

satisfies the conditions specified in Wald (1949) for the consistency of the MLE. For instance, it is a continuous density function with its limit being 0 whenever β_0 or β_1 goes to infinity. The side conclusion stated in Wald (1949) is

$$\sup_{(\beta_0, \beta_1, \eta) \notin B_\epsilon} \ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \leq -\delta n \rightarrow -\infty \tag{8.7}$$

for some $\delta > 0$.

Define

$$g(z_1, z_2; \epsilon) = \sup_{(\beta_0, \beta_1, \eta) \in B_\epsilon} \{ \phi(z_2; \beta_0 + \beta_1 z_1, \eta) + \phi(-z_2; \beta_0 + \beta_1 z_1, \eta) \}.$$

Then $\lim_{\epsilon \rightarrow 0^+} g(z_1, z_2; \epsilon) = \phi(z_2; 0, 1)$. This easily leads to

$$\lim_{\epsilon \rightarrow 0^+} E\{\log g(Z_1, Z_2; \epsilon)\} - E\{\log \phi(Z_2; 0, 1)\} = 0.$$

Hence, we can find a sufficiently small $\epsilon > 0$ such that

$$E\{\log g(Z_1, Z_2; \epsilon)\} \leq E\{\log \phi(Z_2; 0, 1)\} + 1/2.$$

By the law of large numbers, almost surely, we have

$$\begin{aligned} & \sup_{(\beta_0, \beta_1, \eta) \in B_\epsilon} \ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \\ & \leq \sum_{i=1}^n \{ \log g(Z_{1i}, Z_{2i}; \epsilon) - \log g(Z_{2i}; 0, 1) \} \leq n. \end{aligned} \tag{8.8}$$

We remark that the upper bounds in the above two inequalities are much larger than necessary, but they are small enough for our proof.

By (8.7) and (8.8), we have shown that

$$\begin{aligned} & \sup_{(\beta_0, \beta_1, \eta)} \ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \\ & \leq \sum_{i=1}^n \{ \log g(Z_{1i}, Z_{2i}; \epsilon) - \log g(Z_{2i}; 0, 1) \} \leq n. \end{aligned} \tag{8.9}$$

Similarly to the proof of (8.4) and (8.6), we can show that, when $\sigma_+ \notin [\epsilon_0, M_0]$,

$$\ell_{n,1}^*(\mu, \sigma_+) - \ell_{n,1}^*(0, 1) \leq -2n \tag{8.10}$$

uniformly in μ , almost surely.

Using (8.9) and (8.10), we obtain, when $\sigma_+ \notin [\epsilon_0, M_0]$,

$$\ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}_0) = \{ \ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \} + \{ \ell_{n,1}^*(\mu, \sigma_+) - \ell_{n,1}^*(0, 1) \}$$

$$\leq n + (-2n) = -n \rightarrow -\infty.$$

Therefore, the $\bar{\sigma}_+$ value satisfying the lemma condition must almost surely fall within the interval $[\epsilon_0, M_0]$. When we combine this with the results for $\bar{\eta}$, the effective parameter space for $(\mu, \beta_0, \beta_1, \sigma_+, \eta)$ is reduced to $\mathbb{R}^3 \times [\epsilon_0, M_0] \times [\epsilon_0, M_0]$ from the asymptotic point of view. Within this range, the joint density function

$$\phi(Z_{1i}; \mu, \sigma_+) \{0.5\phi(Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta) + 0.5\phi(-Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta)\}$$

satisfies the conditions specified in Wald (1949) for the consistency of the MLE. Therefore, the lemma condition ensures the consistency of $\bar{\theta}$. This completes the proof. \square

Next, we strengthen the results of Lemma 2. We first define some notation. Let

$$A_i = (Z_{1i}, (Z_{1i}^2 - 1)/2)^\tau, \\ B_i = ((Z_{2i}^2 - 1)/2, (Z_{1i}^2 - 1)(Z_{2i}^2 - 1)/2, Z_{1i}(Z_{2i}^2 - 1)/2, -(Z_{2i}^4 - 6Z_{2i}^2 + 3)/12)^\tau.$$

It can be seen that $\mathbb{E}(A_i) = 0$, $\mathbb{E}(B_i) = 0$, A_i and B_i are uncorrelated, and

$$\Sigma_A = \text{VAR}(A_i) = \text{diag}(1, 1/2); \quad \Sigma_B = \text{VAR}(B_i) = \text{diag}(1/2, 1, 2, 1/6).$$

Further, we introduce two parameter vectors of lengths 2 and 4:

$$\mathbf{s}_1 = (\mu, \sigma_+^2 - 1)^\tau; \quad \mathbf{s}_2 = (\beta_0^2 + \beta_1^2 + (\eta^2 - 1), \beta_1^2, \beta_0\beta_1, \beta_0^4)^\tau. \quad (8.11)$$

In the following, we use $|\mathbf{x}|$ and $\|\mathbf{x}\|$ to denote the L_1 and L_2 norms of the vector \mathbf{x} , respectively.

Lemma 3. *Under the conditions of Lemma 2 and the null hypothesis, we have*

- (a) $\ell_{n,1}^*(\bar{\mu}, \bar{\sigma}_+) - \ell_{n,1}^*(0, 1) = \bar{\mathbf{s}}_1^\tau \sum_{i=1}^n A_i - \frac{n}{2} \{\bar{\mathbf{s}}_1^\tau \Sigma_A \bar{\mathbf{s}}_1\} \{1 + o_p(1)\} + o_p(1)$;
- (b) $\ell_{n,2}^*(\bar{\beta}_0, \bar{\beta}_1, \bar{\eta}) - \ell_{n,2}^*(0, 0, 1) \leq \bar{\mathbf{s}}_2^\tau \sum_{i=1}^n B_i - \frac{n}{2} \{\bar{\mathbf{s}}_2^\tau \Sigma_B \bar{\mathbf{s}}_2\} \{1 + o_p(1)\} + o_p(1)$;
- (c) $\bar{\mu}$, $\bar{\sigma}_+^2 - 1$, $\bar{\beta}_0^4$, $\bar{\beta}_1^2$, and $(\bar{\eta}^2 - 1)^2$ are $O_p(n^{-1/2})$.

Proof. We first prove (a). By Lemma 2, we have $(\bar{\mu}, \bar{\sigma}_+) = (0, 1) + o_p(1)$. We obtain (a) by expanding $\ell_{n,1}^*(\bar{\mu}, \bar{\sigma}_+)$ at $(\bar{\mu}, \bar{\sigma}_+) = (0, 1)$ to the second order and then assessing the asymptotic orders via the weak law of large numbers.

To prove (b), we first denote

$$\delta_i(\beta_0, \beta_1, \eta) = \{\phi(Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta) + \phi(-Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta)\} / \{2\phi(Z_{2i}; 0, 1)\} - 1$$

and then write

$$\ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) = \sum_{i=1}^n \log\{1 + \delta_i(\beta_0, \beta_1, \eta)\}.$$

Applying the inequality $\log(1+x) \leq x - x^2/2 + x^3/3$, we have

$$\begin{aligned} & \ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \\ & \leq \sum_{i=1}^n \delta_i(\beta_0, \beta_1, \eta) - (1/2) \sum_{i=1}^n \delta_i^2(\beta_0, \beta_1, \eta) + (1/3) \sum_{i=1}^n \delta_i^3(\beta_0, \beta_1, \eta). \end{aligned} \tag{8.12}$$

Next, we delineate $\delta_i(\beta_0, \beta_1, \eta)$ when (β_0, β_1, η) are in an $o_p(1)$ neighborhood of $(0, 0, 1)$, given $(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\eta}) = (0, 0, 1) + o_p(1)$ as proved in Lemma 2. We perform two main steps. In the *first step*, we obtain the fourth-order Taylor expansion of $\delta_i(\beta_0, \beta_1, \eta)$; in the *second step*, we assess the asymptotic orders of the terms in the expansion and put them into appropriate order expressions.

We start with the *first step*. Let the partial derivatives be

$$\delta_i^{(s,t,k)}(\beta_0, \beta_1, \eta) = \frac{\partial^{s+t+k} \delta_i(\beta_0, \beta_1, \eta)}{\partial \beta_0^s \partial \beta_1^t \partial (\eta^2)^k}.$$

Expanding both $\phi(\pm Z_{2i}; \beta_0 + \beta_1 Z_{1i}, \eta)$ to the fourth order at $(\beta_0, \beta_1, \eta) = (0, 0, 1)$, we get

$$\delta_i(\beta_0, \beta_1, \eta) = \sum_{s+t+k=1}^4 \frac{\beta_0^s \beta_1^t (\eta^2 - 1)^k}{s!t!k!} \delta_i^{(s,t,k)}(0, 0, 1) + \epsilon_{in}^{(1)}, \tag{8.13}$$

where the summation is over all non-negative integer combinations of s, t, k summing to 4 and $\epsilon_{in}^{(1)}$ is the remainder term in the Taylor expansion. Let $\epsilon_n^{(1)} = \sum_{i=1}^n \epsilon_{in}^{(1)}$; then

$$\epsilon_n^{(1)} = O_p(n^{1/2}) \sum_{s+t+k=5} \beta_0^s \beta_1^t (\eta^2 - 1)^k.$$

Next, we argue that every term in the above summation part is of order $o_p(n^{1/2})|\mathbf{s}_2|$. For instance, when $s = t = 1$ and $k = 3$, we have

$$|\beta_0 \beta_1 (\eta^2 - 1)^3| = o_p(1) |\beta_0 \beta_1| = o_p(|\mathbf{s}_2|),$$

helped by the fact that we are investigating the region of $\eta^2 - 1 = o_p(1)$. The orders of the other terms can be similarly assessed. Hence,

$$\epsilon_n^{(1)} = o_p(n^{1/2})|\mathbf{s}_2|.$$

In the *second step*, we first show that every term in the summation part of (8.13) satisfying $s+2t+2k \geq 5$ is also of order $o_p(n^{1/2})|\mathbf{s}_2|$. We use $s = t = k = 1$ as an example. Note that

$$|\beta_0 \beta_1 (\eta^2 - 1)| = o_p(1) |\beta_0 \beta_1| = o_p(|\mathbf{s}_2|).$$

For notational simplicity, let $\delta_i^{(s,t,k)} = \delta_i^{(s,t,k)}(0, 0, 1)$. It is easy to check that $\delta_i^{(s,t,k)}$ has zero mean and finite variance, so

$$\sum_{i=1}^n \delta_i^{(s,t,k)} = O_p(n^{1/2}).$$

Therefore, we have

$$\sum_{i=1}^n \frac{\beta_0^s \beta_1^t (\eta^2 - 1)^k}{s!t!k!} \delta_i^{(s,t,k)} = o_p(n^{1/2}|\mathbf{s}_2|).$$

The proofs for the other $s + 2t + 2k \geq 5$ terms are similar. Hence, we may write

$$\delta_i(\beta_0, \beta_1, \eta) = \sum_{s+2t+2k=1}^4 \frac{\beta_0^s \beta_1^t (\eta^2 - 1)^k}{s!t!k!} \delta_i^{(s,t,k)} + \epsilon_{in}^{(2)} \tag{8.14}$$

and still have

$$\sum_{i=1}^n \epsilon_{in}^{(2)} = o_p(n^{1/2}|\mathbf{s}_2|). \tag{8.15}$$

By straightforward algebra, we find

$$\sum_{s+2t+2k=1}^4 \frac{\beta_0^s \beta_1^t (\eta^2 - 1)^k}{s!t!k!} \delta_i^{(s,t,k)} = \mathbf{s}_2^\tau B_i - 1.5\{\beta_0^2 + (\eta^2 - 1)\}^2 B_i[4] \tag{8.16}$$

where the unwanted term $B_i[4]$ is the fourth element of vector B_i . Its coefficient is easily verified to be $\{\beta_0^2 + (\eta^2 - 1)\}^2 = o_p(|\mathbf{s}_2|)$. This allows us to obtain a neater expression by absorbing it into the higher-order term, concluding that

$$\delta_i(\beta_0, \beta_1, \eta) = \mathbf{s}_2^\tau B_i + \epsilon_{in}^{(3)} \tag{8.17}$$

such that

$$\sum_{i=1}^n \epsilon_{in}^{(3)} = o_p(n^{1/2}|\mathbf{s}_2|) = o_p(1) + o_p(n\|\mathbf{s}_2\|^2). \tag{8.18}$$

In short, we have shown that

$$\sum_{i=1}^n \delta_i(\beta_0, \beta_1, \eta) = \mathbf{s}_2^\tau \sum_{i=1}^n B_i + o_p(1) + o_p(n\|\mathbf{s}_2\|^2). \tag{8.19}$$

The above algebraic manipulations are typical of the techniques employed in Chen and Chen (2001) and Chen et al. (2001). The same techniques, which are tedious but not sophisticated, give

$$\sum_{i=1}^n \delta_i^2(\beta_0, \beta_1, \eta) = \mathbf{s}_2^\tau \left\{ \sum_{i=1}^n B_i B_i^\tau \right\} \mathbf{s}_2 + o_p(1) + o_p(n\|\mathbf{s}_2\|^2), \tag{8.20}$$

$$\sum_{i=1}^n \delta_i^3(\beta_0, \beta_1, \eta) = \sum_{i=1}^n |\mathbf{s}_2^\tau B_i|^3 + o_p(1) + o_p(n\|\mathbf{s}_2\|^2). \tag{8.21}$$

We provide some details below for (8.20); the result in (8.21) can be similarly obtained.

Note that

$$\sum_{i=1}^n \delta_i^2(\beta_0, \beta_1, \eta) = \sum_{i=1}^n (\mathbf{s}_2^\tau B_i)^2 + 2 \sum_{i=1}^n (\mathbf{s}_2^\tau B_i) \epsilon_{in}^{(3)} + \sum_{i=1}^n \{\epsilon_{in}^{(3)}\}^2. \tag{8.22}$$

Repeating the procedures for assessing the order of $\sum_{i=1}^n \epsilon_{in}^{(3)}$ in (8.18), we similarly have

$$\sum_{i=1}^n \{\epsilon_{in}^{(3)}\}^2 = o_p(n)(\|\mathbf{s}_2\|^2) = o_p(1)(n^{1/2}\|\mathbf{s}_2\|)^2 = o_p(1) + o_p(n\|\mathbf{s}_2\|^2) \tag{8.23}$$

and

$$\sum_{i=1}^n (\mathbf{s}_2^\tau B_i) \epsilon_{in}^{(3)} = o_p(n)(\|\mathbf{s}_2\|^2) = o_p(1) + o_p(n\|\mathbf{s}_2\|^2). \tag{8.24}$$

Combining (8.22)–(8.24) gives (8.20).

Together with the weak law of large numbers (8.20) and (8.21) lead to

$$\sum_{i=1}^n \delta_i^2(\beta_0, \beta_1, \eta) = n\mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2 + o_p(1) + o_p(n\|\mathbf{s}_2\|^2), \tag{8.25}$$

$$\sum_{i=1}^n \delta_i^3(\beta_0, \beta_1, \eta) = o_p(1) + o_p(n\|\mathbf{s}_2\|^2). \tag{8.26}$$

Combining (8.19)–(8.26) with (8.12), we have

$$\ell_{n,2}^*(\beta_0, \beta_1, \eta) - \ell_{n,2}^*(0, 0, 1) \leq \mathbf{s}_2^\tau \sum_{i=1}^n B_i - (n/2)\mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2 \{1 + o_p(1)\} + o_p(1).$$

Recall that $(\bar{\beta}_0, \bar{\beta}_1, \bar{\eta}) = (0, 0, 1) + o_p(1)$, so the above upper bound is applicable to $\ell_{n,2}^*(\bar{\beta}_0, \bar{\beta}_1, \bar{\eta}) - \ell_{n,2}^*(0, 0, 1)$. This completes the proof of (b).

Finally, we come to (c). Combining (a) and (b) and the conditions in Lemma 2, we have

$$\begin{aligned} C &\leq \{\ell_{n,1}^*(\bar{\mu}, \bar{\sigma}_+) - \ell_{n,1}^*(0, 1)\} + \{\ell_{n,2}^*(\bar{\beta}_0, \bar{\beta}_1, \bar{\eta}) - \ell_{n,2}^*(0, 0, 1)\} \\ &\leq \sum_{i=1}^n \{\bar{\mathbf{s}}_1^\tau A_i + \bar{\mathbf{s}}_2^\tau B_i\} - (n/2)\{\bar{\mathbf{s}}_1^\tau \Sigma_A \bar{\mathbf{s}}_1 + \bar{\mathbf{s}}_2^\tau \Sigma_B \bar{\mathbf{s}}_2\} \{1 + o_p(1)\} + o_p(1), \end{aligned} \tag{8.27}$$

which is possible only if both $\bar{\mathbf{s}}_1 = O_p(n^{-1/2})$ and $\bar{\mathbf{s}}_2 = O_p(n^{-1/2})$. This leads to the order assessments in (c) and completes the proof of the entire lemma. \square

Remark: The MLE of θ always satisfies the condition (8.2) with $C = 0$. Hence, Lemma 2 implies the consistency of the MLE of θ , and Lemma 3 establishes the asymptotic order of the MLE.

8.3. Proof of Theorem 1

The difference between Theorems 1 and 2 is that in the former we consider $\rho_0 = 0$ to be known when formulating the test statistic. This makes it helpful to reorganize the entries of A_i and B_i and the corresponding entries of \mathbf{s}_1 and \mathbf{s}_2 . After that, we will refine the results in Lemma 3 and apply the refined results to prove (a) and (b).

When $\rho_0 = 0$ is known, we have $\sigma_+ = \sigma_-$. Let

$$\mathbf{t} = (\mu, \beta_0^2/2 + \sigma_+^2 - 1, \beta_0^2, \beta_1^2, \beta_0\beta_1)^\tau. \quad (8.28)$$

Every entry of \mathbf{s}_1 and \mathbf{s}_2 is a linear combination of the entries of \mathbf{t} , possibly with an $O_p(\|\mathbf{t}\|^2)$ difference when these parameter values approach their default null values. We enumerate these entries as follows. The first entry of \mathbf{s}_1 is $\mathbf{s}_1[1] = \mathbf{t}[1]$, and the second is $\mathbf{s}_1[2] = \mathbf{t}[2] - \mathbf{t}[3]/2$. For the entries of \mathbf{s}_2 , we have

$$\mathbf{s}_2[1] = \beta_0^2 + \beta_1^2 + (\eta^2 - 1) = \mathbf{t}[2] + \mathbf{t}[3]/2 - \beta_1^2(\sigma_+^2 - 1) = \mathbf{t}[2] + \mathbf{t}[3]/2 + O_p(\|\mathbf{t}\|^2).$$

For the others, $\mathbf{s}_2[2] = \mathbf{t}[4]$, $\mathbf{s}_2[3] = \mathbf{t}[5]$, and $\mathbf{s}_2[4] = (\mathbf{t}[3])^2 = O_p(\|\mathbf{t}\|^2)$. In summary,

$$\mathbf{s}_1 = (\mathbf{t}[1], \mathbf{t}[2] - \mathbf{t}[3]/2)^\tau, \quad \mathbf{s}_2 = (\mathbf{t}[2] + \mathbf{t}[3]/2, \mathbf{t}[4], \mathbf{t}[5], 0)^\tau + O_p(\|\mathbf{t}\|^2). \quad (8.29)$$

Because every entry of \mathbf{s}_1 and \mathbf{s}_2 is virtually a linear combination of the entries of \mathbf{t} , we can reorganize the entries of A_i and B_i into D_i such that

$$\mathbf{s}_1^\tau A_i + \mathbf{s}_2^\tau B_i = \{\mathbf{t}^\tau + O_p(\|\mathbf{t}\|^2)\} D_i. \quad (8.30)$$

Recall that D_i is defined in (3.5):

$$D_i = (Z_{1i}, (Z_{1i}^2 + Z_{2i}^2 - 2)/2, (Z_{2i}^2 - Z_{1i}^2)/4, (Z_{1i}^2 - 1)(Z_{2i}^2 - 1)/2, Z_{1i}(Z_{2i}^2 - 1))^\tau$$

with $\mathbb{E}(D_i) = 0$ and $\text{VAR}(D_i) = \Sigma_D = \text{diag}(1, 1, 1/4, 1, 2)$. Using the central limit theorem, we have

$$n^{-1/2} \sum_{i=1}^n D_i \xrightarrow{\mathcal{D}} N(0, \Sigma_D). \quad (8.31)$$

With the above preparation, we refine the results in Lemma 3 to the following lemma.

Lemma 4. *Assume the conditions of Lemma 3 and let $\bar{\rho} = 0$. If, under the null model,*

$$\ell_n(\bar{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0) \geq C > -\infty,$$

we then have

- (a) $\ell_n(\bar{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0) \leq \bar{\mathbf{t}}^\tau \sum_{i=1}^n D_i - (n/2) \bar{\mathbf{t}}^\tau \Sigma_D \bar{\mathbf{t}} \{1 + o_p(1)\} + o_p(1)$;
- (b) $\bar{\mu}$, $\bar{\sigma}_+^2 - 1$, $\bar{\beta}_0^2$, and $\bar{\beta}_1^2$ are $O_p(n^{-1/2})$.

Proof. Combining the results of (a) and (b) in Lemma 3, we have

$$\ell_n(\bar{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0) \leq \sum_{i=1}^n \{\bar{\mathbf{s}}_1^T A_i + \bar{\mathbf{s}}_2^T B_i\} - (n/2) \{\bar{\mathbf{s}}_1^T \Sigma_A \bar{\mathbf{s}}_1 + \bar{\mathbf{s}}_2^T \Sigma_B \bar{\mathbf{s}}_2\} \{1 + o_p(1)\} + o_p(1). \tag{8.32}$$

With (8.29) and (8.30), after some algebra, (8.32) can be simplified to

$$\ell_n(\bar{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0) \leq \bar{\mathbf{t}}^T \sum_{i=1}^n D_i - (n/2) \bar{\mathbf{t}}^T \Sigma_D \bar{\mathbf{t}} \{1 + o_p(1)\} + o_p(1), \tag{8.33}$$

as claimed in part (a) of the lemma.

For (b), with the lemma assumption $\ell_n(\bar{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0) \geq C > -\infty$, (8.33) implies that $\bar{\mathbf{t}}$ has the order $O_p(n^{-1/2})$. The order assessment in (b) then follows the definition of \mathbf{t} in (8.28). \square

We are now ready for Theorem 1. Note that the MLEs in both Theorem 1(a) and 1(b) satisfy the conditions of Lemma 4 with $C = 0$. Hence, the order conclusions of the MLEs in both Theorem 1(a) and 1(b) have been established in Lemma 4. We next derive the limiting distributions.

We rewrite $R_{n,1}$ defined in (2.2) as

$$R_{n,1} = 2\{\ell_n(\tilde{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\} - 2\{\ell_n(\check{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\}$$

with $\tilde{\boldsymbol{\theta}}$ being the maximum point of the reduced model where $(\mu_1, \sigma_1) = (\mu_2, \sigma_2)$. Since the reduced model is regular, by standard techniques such as those in Serfling (2000):

$$2\{\ell_n(\check{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\} = n^{-1} \left\{ \left(\sum_{i=1}^n D_i[1] \right)^2 + \left(\sum_{i=1}^n D_i[2] \right)^2 \right\} + o_p(1). \tag{8.34}$$

Next, note that $\tilde{\boldsymbol{\theta}}$ is the maximum point of the reduced model where $\sigma_1 = \sigma_2 = \sigma$. This makes $\beta_1 = \xi = 0$ and subsequently for \mathbf{t} under the reduced model,

$$\mathbf{t} = (\mu, \beta_0^2/2 + (\sigma_+^2 - 1), \beta_0^2, 0, 0)^T.$$

Nevertheless, Lemma 4 is applicable to the above form of \mathbf{t} as long as it is close to its counterpart in the null model. Applying Lemma 4(a) to $\tilde{\boldsymbol{\theta}}$, we have

$$\begin{aligned} & 2\{\ell_n(\tilde{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\} \\ & \leq 2 \sum_{i=1}^n \tilde{\mathbf{t}}^T D_i - n \tilde{\mathbf{t}}^T \Sigma_D \tilde{\mathbf{t}} + o_p(1) \\ & \leq \sup \left\{ 2 \sum_{i=1}^n \mathbf{t}^T D_i - n \mathbf{t}^T \Sigma_D \mathbf{t} : \mathbf{t}[3] \geq 0, \mathbf{t}[4] = 0, \mathbf{t}[5] = 0 \right\} + o_p(1) \end{aligned}$$

$$\leq n^{-1} \left[\left(\sum_{i=1}^n D_i[1] \right)^2 + \left(\sum_{i=1}^n D_i[2] \right)^2 + 4 \left\{ \left(\sum_{i=1}^n D_i[3] \right)^+ \right\}^2 \right] + o_p(1). \quad (8.35)$$

Note that the range of the supremum conforms to the form of \mathbf{t} in the reduced model and the fact that $\mathbf{t}[3] = \beta_0^2 \geq 0$. The specific coefficient values are due to the value of Σ_D .

To derive the limiting distribution of $R_{n,1}$, we need to show that the upper bound in (8.35) can be attained. Let $\check{\boldsymbol{\theta}}$ be the estimator of $\boldsymbol{\theta}$ such that the corresponding

$$\check{\mathbf{t}} = n^{-1} \left(\sum_{i=1}^n D_i[1], \sum_{i=1}^n D_i[2], 4 \left\{ \sum_{i=1}^n D_i[3] \right\}^+, 0, 0 \right)^\tau.$$

With some straightforward algebra, the corresponding $\check{\boldsymbol{\theta}}$ values of $\check{\mathbf{t}}$ exist and satisfy

$$\check{\mu}_1 = O_p(n^{-1/4}), \quad \check{\mu}_2 = O_p(n^{-1/4}), \quad \check{\sigma}^2 - 1 = O_p(n^{-1/2}). \quad (8.36)$$

If we apply the Taylor expansion, the order assessment in (8.36) leads to the following approximation:

$$\begin{aligned} & 2\{\ell_n(\check{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\} \\ &= n^{-1} \left[\left(\sum_{i=1}^n D_i[1] \right)^2 + \left(\sum_{i=1}^n D_i[2] \right)^2 + 4 \left\{ \left(\sum_{i=1}^n D_i[3] \right)^+ \right\}^2 \right] + o_p(1). \end{aligned} \quad (8.37)$$

Since $\check{\boldsymbol{\theta}}$ is the maximum point of $\ell_n(\boldsymbol{\theta})$, $2\{\ell_n(\check{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\}$ is not smaller than the value in (8.37). The sandwich technique of Chen and Chen (2001) and Chen et al. (2001) or the squeeze theorem can be applied to obtain

$$\begin{aligned} & 2\{\ell_n(\check{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\} \\ &= n^{-1} \left[\left(\sum_{i=1}^n D_i[1] \right)^2 + \left(\sum_{i=1}^n D_i[2] \right)^2 + 4 \left\{ \left(\sum_{i=1}^n D_i[3] \right)^+ \right\}^2 \right] + o_p(1). \end{aligned} \quad (8.38)$$

Combining (8.34) and (8.38) gives

$$R_{n,1} = 4n^{-1} \left\{ \left(\sum_{i=1}^n D_i[3] \right)^+ \right\}^2 + o_p(1).$$

By (8.31), $2n^{-1/2} \sum_{i=1}^n D_i[3] \xrightarrow{D} N(0, 1)$. Hence, $R_{n,1}$ has the limiting distribution $0.5\chi_0^2 + 0.5\chi_1^2$. This completes the proof of part (a).

We now prove conclusion (b). In this case, the range of \mathbf{t} has only an intrinsic restriction as seen in the expression

$$\mathbf{t} = (\mu, \beta_0^2/2 + (\sigma_+^2 - 1), \beta_0^2, \beta_1^2, \beta_0\beta_1)^\tau.$$

Let $\mathbf{t}_1 = (\mu, \beta_0^2/2 + (\sigma_+^2 - 1))^\tau$ and $\mathbf{t}_2 = (\beta_0^2, \beta_1^2, \beta_0\beta_1)^\tau$. It can be seen that \mathbf{t}_2 lies on a two-dimensional manifold. Nonetheless, the upper bound developed in Lemma 4 remains valid for $\hat{\boldsymbol{\theta}}$. We partition D_i into $D_{i1} = (D_i[1], D_i[2])^\tau$ and $D_{i2} = (D_i[3], D_i[4], D_i[5])^\tau$ with covariance matrices Σ_{D1} and Σ_{D2} . With these preparations, we have

$$\begin{aligned} & 2\{\ell_n(\hat{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\} \\ & \leq 2\hat{\mathbf{t}}_1^\tau \sum_{i=1}^n D_{i1} + 2\hat{\mathbf{t}}_2^\tau \sum_{i=1}^n D_{i2} - n \left(\hat{\mathbf{t}}_1^\tau \Sigma_{D1} \hat{\mathbf{t}}_1 + \hat{\mathbf{t}}_2^\tau \Sigma_{D2} \hat{\mathbf{t}}_2 \right) + o_p(1) \\ & \leq n^{-1} \left(\sum_{i=1}^n D_{i1} \right)^\tau \left(\sum_{i=1}^n D_{i1} \right) + \sup_{\mathbf{t}_2} \left\{ 2\mathbf{t}_2^\tau \sum_{i=1}^n D_{i2} - n\mathbf{t}_2^\tau \Sigma_{D2} \mathbf{t}_2 \right\} + o_p(1). \end{aligned} \tag{8.39}$$

The supremum is taken over \mathbf{t}_2 with the intrinsic restriction respected. Similarly to (8.38), the upper bound (8.39) is attained at some feasible parameter value. Hence,

$$\begin{aligned} & 2\{\ell_n(\hat{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\} \\ & = n^{-1} \left(\sum_{i=1}^n D_{i1} \right)^\tau \left(\sum_{i=1}^n D_{i1} \right) + \sup_{\mathbf{t}_2} \left\{ 2\mathbf{t}_2^\tau \sum_{i=1}^n D_{i2} - n\mathbf{t}_2^\tau \Sigma_{D2} \mathbf{t}_2 \right\} + o_p(1). \end{aligned} \tag{8.40}$$

Combining (8.34) and (8.40), we get

$$R_{n,2} = \sup_{\mathbf{t}_2} \left\{ 2\mathbf{t}_2^\tau \sum_{i=1}^n D_{i2} - n\mathbf{t}_2^\tau \Sigma_{D2} \mathbf{t}_2 \right\} + o_p(1).$$

The intrinsic restriction due to the specific form of $\mathbf{t}_2 = (\beta_0^2, \beta_1^2, \beta_0\beta_1)^\tau$ and (8.31) leads to the nonstandard form of the limiting distribution in the theorem.

8.4. Proof of Theorem 2

The test problem in Theorem 2 is different from that of Theorem 1 because we do not assume knowledge of the ρ_0 value. The parameter vector is now $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)^\tau$ including the correlation coefficient ρ .

Since the MLEs in Theorem 2(a) and (b) both satisfy the conditions of Lemma 3 with $C = 0$, Lemma 3(c) can be used to establish the order con-

clusions of the MLEs in Theorem 2. Next, we focus on the limiting distribution results. We start with some preparation.

With the introduction of ρ , it helps to redefine \mathbf{s}_1 , \mathbf{s}_2 , and so on as follows:

$$\mathbf{s}_1 = (\mu, \sigma_+^2 - 1, \beta_0^2 + \beta_1^2 + (\eta^2 - 1))^\tau; \quad \mathbf{s}_2 = (\beta_1^2, \beta_0\beta_1, \beta_0^4)^\tau$$

and the corresponding A_i , B_i as

$$\begin{aligned} A_i &= (Z_{1i}, (Z_{1i}^2 - 1)/2, (Z_{2i}^2 - 1)/2)^\tau, \\ B_i &= ((Z_{1i}^2 - 1)(Z_{2i}^2 - 1)/2, Z_{1i}(Z_{2i}^2 - 1), -(Z_{2i}^4 - 6Z_{2i}^2 + 3)/12)^\tau. \end{aligned}$$

These are almost the quantities with the same names defined/recalled above Lemma 3. The difference is that the first entry of \mathbf{s}_2 is now the third entry of \mathbf{s}_1 . That is, we partition the vector differently here. Note that $\mathbb{E}(A_i) = 0$, $\mathbb{E}(B_i) = 0$, A_i and B_i are still uncorrelated, and

$$\Sigma_A = \text{VAR}(A_i) = \text{diag}(1, 1/2, 1/2); \quad \Sigma_B = \text{VAR}(B_i) = \text{diag}(1, 2, 1/6).$$

Using the central limit theorem, we have

$$n^{-1/2} \sum_{i=1}^n B_i \xrightarrow{\mathcal{D}} N(0, \Sigma_B). \quad (8.41)$$

We are now ready to prove the limiting distribution results. When $(\mu_1, \sigma_1) = (\mu_2, \sigma_2)$ in Theorem 2, the asymptotic expansion of the likelihood ratio is an expansion for regular models:

$$2\{\ell_n(\tilde{\boldsymbol{\theta}}^*) - \ell_n(\boldsymbol{\theta}_0)\} = n^{-1} \left(\sum_{i=1}^n A_i \right)^\tau \Sigma_A^{-1} \left(\sum_{i=1}^n A_i \right) + o_p(1). \quad (8.42)$$

The result of Lemma 3 remains applicable to both $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}^*$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^*$:

$$2\{\ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}_0)\} \leq 2\mathbf{s}_1^\tau \sum_{i=1}^n A_i + 2\mathbf{s}_2^\tau \sum_{i=1}^n B_i - n\{\mathbf{s}_1^\tau \Sigma_A \mathbf{s}_1 + \mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2\} + o_p(1).$$

Since $\sigma_1 = \sigma_2$ in Theorem 2(a), we have

$$\mathbf{s}_1 = (\mu, \sigma_+^2 - 1, \beta_0^2 + (\eta^2 - 1))^\tau; \quad \mathbf{s}_2 = (0, 0, \beta_0^4)^\tau.$$

This leads to

$$\begin{aligned} &2\{\ell_n(\tilde{\boldsymbol{\theta}}^*) - \ell_n(\boldsymbol{\theta}_0)\} \leq \\ &n^{-1} \left(\sum_{i=1}^n A_i \right)^\tau \Sigma_A^{-1} \left(\sum_{i=1}^n A_i \right) + 6n^{-1} \left\{ \left(\sum_{i=1}^n B_i[3] \right)^+ \right\}^2 + o_p(1), \end{aligned} \quad (8.43)$$

where we have $(\sum_{i=1}^n B_i[3])^+$ instead of $(\sum_{i=1}^n B_i[3])$ because of the intrinsic constraint $\mathbf{s}_2[3] = \beta_0^4 \geq 0$. We skip the step of showing that the above upper bound is attainable, since this is now routine.

Combining (8.42) and (8.43) gives

$$R_{n,1}^* = 6n^{-1} \left\{ \left(\sum_{i=1}^n B_i[3] \right)^+ \right\}^2 + o_p(1),$$

which converges to $0.5\chi_0^2 + 0.5\chi_1^2$ in distribution by (8.41). This is conclusion (a).

For $R_{n,2}^*$ in (b), we are not helped by $\sigma_1 = \sigma_2$. Yet

$$2\{\ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}_0)\} \leq 2\mathbf{s}_1^\tau \sum_{i=1}^n A_i + 2\mathbf{s}_2^\tau \sum_{i=1}^n B_i - n\{\mathbf{s}_1^\tau \Sigma_A \mathbf{s}_1 + \mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2\} + o_p(1)$$

remains true for $\boldsymbol{\theta}$ in a small neighborhood of $\boldsymbol{\theta}_0$. Similarly, we still have

$$\begin{aligned} & 2\{\ell_n(\hat{\boldsymbol{\theta}}^*) - \ell_n(\boldsymbol{\theta}_0)\} \\ &= n^{-1} \left(\sum_{i=1}^n A_i \right)^\tau \Sigma_A^{-1} \left(\sum_{i=1}^n A_i \right) + \sup_{\mathbf{s}_2} \left\{ 2\mathbf{s}_2^\tau \sum_{i=1}^n B_i - n\mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2 \right\} + o_p(1). \end{aligned}$$

We skip the proof that this upper bound is attained. Hence,

$$R_{n,2}^* = \sup_{\mathbf{s}_2} \left\{ 2\mathbf{s}_2^\tau \sum_{i=1}^n B_i - n\mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2 \right\} + o_p(1). \tag{8.44}$$

The challenge is to provide an analytical description of the limiting distribution when

$$\mathbf{s}_2 = (\beta_1^2, \beta_0\beta_1, \beta_0^4)^\tau.$$

For this purpose, we highlight the fact that $n^{-1/2} \sum_{i=1}^n B_i$ is asymptotically multivariate normal with mean 0 and covariance matrix $\Sigma_B = \text{diag}(1, 2, 1/6)$. The supremum is hence attained in the range of $\mathbf{s}_2 = O_p(n^{-1/2})$. In the sub-region where $|\beta_0| < n^{-1/7} = o(n^{-1/8})$, we have $\mathbf{s}_2[3] = \beta_0^8 < n^{-8/7} = o(n^{-1})$. Hence,

$$\begin{aligned} & \sup_{\mathbf{s}_2, |\beta_0| < n^{-1/7}} \left\{ 2\mathbf{s}_2^\tau \sum_{i=1}^n B_i - n\mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2 \right\} \\ &= \sup_{\mathbf{s}_2, \beta_0=0} \left\{ 2\mathbf{s}_2^\tau \sum_{i=1}^n B_i - n\mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2 \right\} + o_p(1) \\ &= n^{-1} \left\{ \left(\sum_{i=1}^n B_i[1] \right)^+ \right\}^2 + (1/2)n^{-1} \left\{ \sum_{i=1}^n B_i[2] \right\}^2 + o_p(1). \tag{8.45} \end{aligned}$$

In the other subregion where $|\beta_0| \geq n^{-1/7}$, combined with the restriction $\beta_0\beta_1 = O_p(n^{-1/2})$, we must have $\beta_1 = O_p(n^{-1/3})$. Consequently, in this region, $\mathbf{s}_2[1] = \beta_1^2 = O(n^{-2/3})$. This leads to

$$\mathbf{s}_2[1] \sum_{i=1}^n B_i[1] - n\{\mathbf{s}_2[1]\}^2 = o_p(1).$$

Hence,

$$\begin{aligned} & \sup_{\mathbf{s}_2, |\beta_0| \geq n^{-1/7}} \left\{ 2\mathbf{s}_2^\tau \sum_{i=1}^n B_i - n\mathbf{s}_2^\tau \Sigma_B \mathbf{s}_2 \right\} \\ &= \sup_{\mathbf{s}_2} \left\{ 2\mathbf{s}_2[2] \sum_{i=1}^n B_i[2] - 2n\mathbf{s}_2[2]^2 + 2\mathbf{s}_2[3] \sum_{i=1}^n B_i[3] - (1/6)n\mathbf{s}_2[3]^2 \right\} + o_p(1) \\ &= (1/2)n^{-1} \left\{ \sum_{i=1}^n B_i[2] \right\}^2 + 6n^{-1} \left\{ \left(\sum_{i=1}^n B_i[3] \right)^+ \right\}^2 + o_p(1). \end{aligned} \quad (8.46)$$

Combining (8.44)–(8.46), we find

$$\begin{aligned} R_{n,2}^* &= (1/2)n^{-1} \left\{ \sum_{i=1}^n B_i[2] \right\}^2 \\ &\quad + \max \left[n^{-1} \left\{ \left(\sum_{i=1}^n B_i[1] \right)^+ \right\}^2, 6n^{-1} \left\{ \left(\sum_{i=1}^n B_i[3] \right)^+ \right\}^2 \right] + o_p(1). \end{aligned}$$

Therefore, $R_{n,2}^*$ has the limiting distribution as claimed.

Acknowledgments

The authors thank the editor, the associate editor, and the referees for constructive comments and suggestions that led to significant improvements in the paper.

References

- Barndorff-Nielsen, O. E. and Hall, P. (1988), ‘On the level-error after Bartlett adjustment of the likelihood ratio statistic’, *Biometrika* **75**, 374–378. [MR0946056](#)
- Bartlett, M. S. (1937), ‘Properties of sufficiency and statistical tests’, *Proceedings of The Royal Society A* **160**, 268–282. [MR0024103](#)
- Carothers, A. D. (1981), ‘On determining the parental origins of homologous chromosomes’, *Annals of Human Genetics* **45**, 367–374.

- Chen, H. and Chen, J. (2001), ‘The likelihood ratio test for homogeneity in finite mixture models’, *The Canadian Journal of Statistics* **29**, 201–215. [MR1840705](#)
- Chen, H. and Chen, J. (2003), ‘Tests for homogeneity in normal mixtures in the presence of a structural parameter’, *Statistica Sinica* **13**, 351–365. [MR1977730](#)
- Chen, H., Chen, J. and Kalbfleisch, J. D. (2001), ‘A modified likelihood ratio test for homogeneity in finite mixture models’, *Journal of the Royal Statistical Society: Series B* **63**, 19–29. [MR1811988](#)
- Chen, J. and Li, P. (2009), ‘Hypothesis test for normal mixture models: The EM approach’, *The Annals of Statistics* **37**, 2523–2542. [MR2543701](#)
- Chen, J. and Li, P. (2011), ‘Tuning the EM-test for finite mixture models’, *The Canadian Journal of Statistics* **39**(3), 389–404. [MR2842420](#)
- Chernoff, H. (1954), ‘On the distribution of the likelihood ratio’, *The Annals of Mathematical Statistics* **25**, 573–578. [MR0065087](#)
- Davies, P. and Phillips, A. J. (1988), ‘Nonparametric tests of population differences and estimation of the probability of misidentification with unidentified paired data’, *Biometrika* **75**, 753–760. [MR0995117](#)
- Drton, M. (2009), ‘Likelihood ratio tests and singularities’, *The Annals of Statistics* **37**, 979–1012. [MR2502658](#)
- Efron, B., Miller, R. G. and Brown, B. W. (1971), Spurious appearance of mosaicism in three generations in one family with a 3/B translocation. II. Statistical model of the chromosomal abnormality, Technical Report 27, Department of Statistics, Stanford University.
- Ernst, M. D., Guerra, R. and Schucany, W. R. (1996), ‘Scatterplots for unordered pairs’, *The American Statistician* **50**, 260–265.
- Hinkley, D. V. (1973), ‘Two-sample tests with unordered pairs’, *Journal of the Royal Statistical Society: Series B* **35**, 337–346. [MR0334369](#)
- Ho, N. and Nguyen, X. (2016), ‘Convergence rates of parameter estimation for some weakly identifiable finite mixtures’, *The Annals of Statistics* **44**, 2726–2755. [MR3576559](#)
- Ho, N. and Nguyen, X. (2019), ‘Singularity structures and impacts on parameter estimation in finite mixtures of distributions’, *SIAM Journal on Mathematics of Data Science* **1**, 730–758. [MR4016133](#)
- Lauder, I. J. (1977), ‘Tracing quantitative measurements on human chromosomes in family studies’, *Annals of Human Genetics* **41**, 77–86.
- Lawley, D. N. (1956), ‘A general method for approximating to the distribution of likelihood ratio criteria’, *Biometrika* **43**, 295–303. [MR0082237](#)
- Li, P. and Qin, J. (2011), ‘A new nuisance-parameter elimination method with application to the unordered homologous chromosome pairs problem’, *Journal of the American Statistical Association* **106**, 1476–1484. [MR2896850](#)
- Mason, D., Lauder, I., Rutovitz, D. and Spowart, G. (1975), ‘Measurement of C-bands in human chromosomes’, *Computers in Biology and Medicine* **5**, 179–201.
- Matérn, B. and Simak, M. (1968), ‘Statistical problems in karyotype analysis’, *Hereditas* **59**, 280–288.
- Miller, F., Friede, T. and Kieser, M. (2009), ‘Blinded assessment of treatment

- effects utilizing information about the randomization block length', *Statistics in Medicine* **28**, 1690–1706. [MR2675245](#)
- Moore II, D. (1973), 'Do homologous chromosomes differ? Two statistical tests', *Cytogenetics and Cell Genetics* **12**, 305–314.
- Moore II, D. H., Carrano, A. V. and Mayall, B. H. (1979), 'Do homologous chromosomes differ? A preliminary investigation based on DNA measurements', *Cytogenetics and Cell Genetics* **23**, 108–116.
- Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer, New York. [MR2197664](#)
- Olkin, I. and Viana, M. (1995), 'Correlation analysis of extreme observations from a multivariate normal distribution', *Journal of the American Statistical Association* **90**, 1373–1379. [MR1379480](#)
- Pollard, D. (1990), *Empirical Processes: Theory and Applications. NSF-CBMS Regional Conference Series in Probability and Statistics*, Vol. 2, Institute of Mathematical Statistics. [MR1089429](#)
- Qin, J. and Zhang, J. (2005), 'Marginal likelihood, conditional likelihood and empirical likelihood: Connections and applications', *Biometrika* **92**, 251–270. [MR2201358](#)
- Self, S. G. and Liang, K.-Y. (1987), 'Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions', *Journal of the American Statistical Association* **82**, 605–610. [MR0898365](#)
- Serfling, R. J. (2000), *Approximation Theorems of Mathematical Statistics*, Wiley, New York. [MR0595165](#)
- Shekar, S. N., Banerjee, T. and Biswas, A. (2006), 'Hypotheses on the effect of cadmium on glutathione content of red blood corpuscles', *Twin Research and Human Genetics* **9**, 73–75.
- van der Meulen, E. A. (2005), 'Are we really that blind?', *Journal of Biopharmaceutical Statistics* **15**, 479–489. [MR2190563](#)
- Wald, A. (1949), 'Note on the consistency of the maximum likelihood estimate', *The Annals of Mathematical Statistics* **20**, 595–601. [MR0032169](#)
- Wilks, S. S. (1938), 'The large-sample distribution of the likelihood ratio for testing composite hypotheses', *The Annals of Mathematical Statistics* **9**, 60–62.
- Yu, T., Zhang, C., Alexander, A. L. and Davidson, R. J. (2013), 'Local tests for identifying anisotropic diffusion areas in human brain with DTI', *The Annals of Applied Statistics* **7**, 201–225. [MR3086416](#)