Received: 26 February 2018 | Revised: 22 February 2019 | Accepted: 5 November 2019

DOI: 10.1111/sjos.12442

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Hypothesis testing for quantitative trait locus effects in both location and scale in genetic backcross studies

Guanfu Liu[1] | **Pengfei Li**[2] | **Yukun Liu**[3] | **Xiaolong Pu**[3]

[1]School of Statistics and Information, Shanghai University of International Business and Economics

[2]Department of Statistics and Actuarial Science, University of Waterloo

[3]Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, School of Statistics, East China Normal University

Correspondence
Yukun Liu, Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, School of Statistics, East China Normal University, 3663 Zhongshanbei Street, Shanghai 200062, China.
Email: ykliu@sfs.ecnu.edu.cn

Funding information
National Natural Science Foundation of China, 11771144, 11771145, 11801359, 11971300; Natural Science Foundation of Shanghai, 19ZR1420900; Natural Sciences and Engineering Research Council of Canada, RGPIN-2015-06592; the 111 Project, B14019

Abstract

Testing the existence of a quantitative trait locus (QTL) effect is an important task in QTL mapping studies. Most studies concentrate on the case where the phenotype distributions of different QTL groups follow normal distributions with the same unknown variance. In this paper we make a more general assumption that the phenotype distributions come from a location-scale distribution family. We derive the limiting distribution of the likelihood ratio test (LRT) for the existence of the QTL effect in both location and scale in genetic backcross studies. We further identify an explicit representation for this limiting distribution. As a complement, we study the limiting distribution of the LRT and its explicit representation for the existence of the QTL effect in the location only. The asymptotic properties of the LRTs under a local alternative are also investigated. Simulation studies are used to evaluate the asymptotic results, and a real-data example is included for illustration.

KEYWORDS

backcross study, explicit representation, likelihood ratio test, limiting distribution, quantitative trait locus

1 | INTRODUCTION

Quantitative trait locus (QTL) mapping is an important tool for analyzing the genetic factors contributing to the variations of quantitative traits in humans, plants, and animals. A starting point of QTL mapping studies is to test the existence of a QTL effect. If the QTL effect exists, we proceed to identify the locations and estimate the genetic effect of the QTLs.

A popular method for testing the existence of a QTL effect is the interval mapping method developed by Lander and Botstein (1989). Consider a putative QTL, say **Q**, located between two flanking markers in a backcross design, with **M** and **N** being the left and right flanking markers, respectively. For individuals in the backcross population, the possible genotypes are *MM* and *Mm* at **M**, *NN* and *Nn* at **N**, and *QQ* and *Qq* at **Q**. Hence, the individuals in the backcross population have four marker genotypes: MM/NN, MM/Nn, Mm/NN, or Mm/Nn. For each individual, we can observe the genotypes of the two flanking markers and the phenotype, but we cannot observe the QTL genotype, which must be inferred from the marker information. The phenotype data can be divided into four groups according to their marker genotypes. A testing method based on these data for detecting a QTL in the interval **M**–**N** is referred to as an interval mapping method.

In this paper, we consider the backcross study described above and assume that no interference or double recombination occurs between two marker-QTL intervals. Let f_1 and f_2 be the phenotype probability density functions corresponding to the two genotypes *QQ* and *Qq*, respectively. We use $y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2}, y_{31}, \dots, y_{3n_3}$, and $y_{41}, \dots, y_{4n_4}$ to denote the phenotype data corresponding to the marker genotypes MM/NN, MM/Nn, Mm/NN, and Mm/Nn, respectively. Denote by r and r_1 the recombination frequencies between **M** and **N**, and between **M** and **Q**, respectively. According to Doerge, Zeng, and Weir (1997) and Wu, Ma, and Casella (2007),

$$\begin{aligned}
&y_{1j} \overset{iid}{\sim} f_1(y), \quad j = 1, \dots, n_1, \\
&y_{2j} \overset{iid}{\sim} \theta f_1(y) + (1-\theta) f_2(y), \quad j = 1, \dots, n_2, \\
&y_{3j} \overset{iid}{\sim} (1-\theta) f_1(y) + \theta f_2(y), \quad j = 1, \dots, n_3, \\
&y_{4j} \overset{iid}{\sim} f_2(y), \quad j = 1, \dots, n_4,
\end{aligned} \tag{1}$$

where $\theta = r_1/r$. Since the two flanking markers are prespecified, the recombination frequency r is generally known. However, the location of **Q** is unknown, so neither the recombination frequency r_1 nor the parameter θ is known. The problem of testing the existence of a QTL effect is equivalent to testing $H_0: f_1 = f_2$ versus $H_1: f_1 \neq f_2$.

The likelihood ratio test (LRT) has been commonly applied to test the existence of a QTL effect when f_1 and f_2 in Equation (1) are from the same parametric distribution family (Chen & Chen, 2005; Wu et al., 2007). Due to the mixture model structure in Equation (1) and the fact that θ appears under only the alternative model, determining the critical values of the LRT has been a long-standing problem in its application. Under the assumption that f_1 and f_2 are normal distributions with the same unknown variance, Feingold, Brown, and Siegmund (1993) and Rebai, Goffinet, and Mangin (1994) proposed approximation methods, which do not have rigorous theoretical support, to determine the critical values of the LRT. Under the same setup, Chen and Chen (2005) rigorously showed that the LRT statistic converges in distribution to the supremum of a chi-square process under the null hypothesis. They further suggested using numerical approximation to determine the critical values of the LRT from its limiting distribution. Simulation shows that their method provides more accurate critical values than those of Feingold et al. (1993) and Rebai et al. (1994). Similar limiting distributions for the LRT were obtained by Wu, Chen, and Liu (2008) and Kim, Cui, and Zhao (2013) for the case where f_1 and f_2 are from a one-parameter (mean parameter) exponential family. Under the same assumption on f_1 and f_2 as in Chen and Chen (2005), Chang, Wu, Wu, and Casella (2009) developed a score test and showed that the maximum of the squared score statistic is asymptotically equivalent to the LRT statistic.

In summary, the aforementioned works concentrate on the QTL effect in the mean parameter only. As Weller and Wyler (1992) and Korol, Ronin, Tadmor, Bar-Zur, and Kirzhner (1996) have pointed out, a QTL effect may be economically more critical in the variance than in the mean (e.g., for earliness, flowering time, ripening time under machine harvesting, time for chickens to hatch, and seed dormancy). To the best of our knowledge, Korol et al. (1996) were the first to take the variance effect into account in interval mapping studies: they investigated the power of the LRT through simulations. Their study was based on the known-QTL-location and normality assumptions; no theoretical results were provided on how to determine critical values of the LRT.

In this paper, we fill the gap in genetic backcross studies by studying the LRT procedure for the existence of a QTL effect in both location and scale with an unknown QTL location. The normality assumption on f_1 and f_2 is quite natural, but it can be restrictive. We instead assume that f_1 and f_2 come from a general location-scale distribution family, and they may have different locations and/or scales. Testing $H_0 : f_1 = f_2$ under the above setup in Equation (1) is essentially testing homogeneity in four samples. There has been much research into the asymptotic properties of the LRT for homogeneity in the mixture model in the one-sample case; see Dacunha-Castelle and Gassiat (1999), Liu and Shao (2003), Garel (2005), Gu, Koenker, and Volgushev (2018), and the references therein. However, the mixture model with component densities from a general location-scale distribution family has some undesirable properties. The likelihood function of the unknown parameters is unbounded (Hathaway, 1985), and the Fisher information on the mixing proportion can be infinity (Chen & Li, 2009). Because of these two nonregularities, the existing elegant asymptotic results of the LRT for homogeneity in the one-sample mixture model are not directly applicable to the mixture model with component densities from a general location-scale distribution family. Taking advantage of the specific four-sample structure in Equation (1), we successfully derive the asymptotic properties of the maximum likelihood estimators (MLEs) of the unknown parameters and the LRT statistic under the null hypothesis that no QTL exists. To the best of our knowledge, we are the first to consider the asymptotic results of the LRT for homogeneity in the mixture model with component densities from a general location-scale distribution family.

We focus on the data structure in Equation (1) with f_1 and f_2 from a general location-scale distribution family. Our contributions can be summarized as follows. First, we establish the convergence rates of the MLEs of the location and scale parameters in f_1 and f_2, and we show that the LRT statistic converges in distribution to the supremum of a χ^2_2 process, under the null hypothesis that no QTL exists. We further identify an explicit representation of the limiting distribution, which can be used to rapidly calculate the critical values of the LRT. Second, as a complement, we study the limiting distribution of the LRT and its explicit representation for the existence of a QTL effect in the location only (i.e., f_1 and f_2 have the same unknown scale parameter). Third, we derive the local powers of the above two LRTs under a series of local alternatives. The local power results indicate that the two LRTs are consistent under the given local alternatives. We emphasize that the existing asymptotic results under the normality assumption with a common unknown variance rely on the assumption that the mean parameter space is bounded (Chang et al., 2009; Chen & Chen, 2005). Our asymptotic results do not depend on this assumption whether or not a QTL effect in the scale exists. Fourth, based on the limiting distributions of two LRTs, the approximation of Davies (1987) can be applied to approximate the critical values of the LRTs. We conduct simulations to show that the empirical type I errors of two LRTs based on the explicit representations are quite close to the nominal levels, and they are closer than those based on Davies's method. Further, the LRT in both the location and scale is uniformly more powerful than existing nonparametric tests such as the multiple-sample Kolmogorov–Smirnov test (Kiefer, 1959) and the multiple-sample Anderson–Darling test (Scholz & Stephens, 1987).

The paper is organized as follows. Section 2 presents the large-sample properties of the MLEs of the unknown parameters and the LRTs where (1) f_1 and f_2 may have different locations and/or scales and (2) f_1 and f_2 have the same unknown scale. We give explicit representations of the limiting distributions of the LRTs in these two cases, and we study their local powers under a series of local alternatives. Section 3 investigates the finite-sample performance of the LRTs via simulation studies, and Section 4 analyzes a real-data set. Section 5 concludes with a discussion. For clarity, the proofs are postponed to the Appendix A or the Appendix S1.

2 | MAIN RESULTS

Suppose we have the observations $\{y_{ij}, i = 1, \dots, 4, j = 1, \dots, n_i\}$ from (1) with $f_1(y)$ and $f_2(y)$ from the same location-scale distribution family. That is, $f_1(y) = f(y; \mu_1, \sigma_1)$ and $f_2(y) = f(y; \mu_2, \sigma_2)$ with $f(y; \mu, \sigma) = \sigma^{-1} f((y-\mu)/\sigma; 0, 1)$, where $f(\cdot; 0, 1)$ is a known probability density function, and μ and σ are the location and scale parameters, respectively. Under this setup, testing the existence of the QTL effect in both location and scale is equivalent to testing

$$H_0 : (\mu_1, \sigma_1) = (\mu_2, \sigma_2). \tag{2}$$

2.1 | Asymptotic properties under the null

The LRT is one of the most important tools in statistical inference, especially for parametric models (Chernoff, 1954; Self & Liang, 1987; Wilks, 1938). In this subsection, we establish the LRT statistics for testing (2). Based on the observed data in Equation (1), the log-likelihood function of $(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2)$ is

$$\begin{aligned} l_n(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2) = & \sum_{j=1}^{n_1} \log f_1(y_{1j}) + \sum_{j=1}^{n_2} \log\{\theta f_1(y_{2j}) + (1-\theta) f_2(y_{2j})\} \\ & + \sum_{j=1}^{n_3} \log\{(1-\theta) f_1(y_{3j}) + \theta f_2(y_{3j})\} + \sum_{j=1}^{n_4} \log f_2(y_{4j}). \end{aligned}$$

The MLEs of the unknown parameters under the null and full models are respectively

$$(\hat{\mu}_0, \hat{\sigma}_0) = \arg\max_{\mu, \sigma} l_n(0.5, \mu, \mu, \sigma, \sigma), \tag{3}$$

and

$$(\hat{\theta}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2) = \arg \max_{\theta,\mu_1,\mu_2,\sigma_1,\sigma_2} l_n(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2).$$

Then the LRT statistic for testing (2) is defined to be

$$R_n = 2\{l_n(\hat{\theta}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2) - l_n(0.5, \hat{\mu}_0, \hat{\mu}_0, \hat{\sigma}_0, \hat{\sigma}_0)\}.$$

We reject H_0 when R_n exceeds some critical value to be determined by its limiting distribution. In the definition of the MLEs in Equation (3) under the null model, we arbitrarily set the parameter θ to 0.5; other choices of θ do not change the resulting likelihood or LRT, since θ does not appear in the null model.

We define some notation to ease the presentation of the asymptotic properties of $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2)$ and R_n. Let the total sample size be $n = \sum_{i=1}^4 n_i$. We assume that the n_i's go to ∞ at the same rate. That is, n_i/n goes to a constant p_i with $p_i > 0$, $i = 1, \ldots, 4$. In the genetic backcross studies described in Section 1, the p_i's are related to r, the recombination frequency between two markers **M** and **N**, in the following way (see Wu et al., 2007):

$$(p_1, p_2, p_3, p_4) = \left(\frac{1-r}{2}, \frac{r}{2}, \frac{r}{2}, \frac{1-r}{2}\right).$$

Let z_{hk} $(h, k = 1, 2)$ be independent and identically distributed standard normal random variables, and define for $h = 1, 2$,

$$Z_h(\theta) = \frac{\sqrt{1-r}}{\sqrt{1+4r\theta(\theta-1)}} z_{h1} + \frac{\sqrt{r}(2\theta-1)}{\sqrt{1+4r\theta(\theta-1)}} z_{h2}, \tag{4}$$

where $0 \le \theta \le 1$. It is clear that the $\{Z_h(\theta) : \theta \in [0, 1]\}$ $(h = 1, 2)$ are independent and both are Gaussian processes with zero mean, unit variance, and covariance function

$$\mathrm{Cov}\,(Z_h(\theta_1), Z_h(\theta_2)) = \frac{1 + r\{4\theta_1\theta_2 - 2(\theta_1 + \theta_2)\}}{\sqrt{\{4r\theta_1(\theta_1 - 1) + 1\}\{4r\theta_2(\theta_2 - 1) + 1\}}}.$$

Let $\gamma = \arccos \sqrt{1-r}$. Define three sets of angles,

$$A_1 = [-\gamma, \gamma] \cup [\pi - \gamma, \pi] \cup [-\pi, -\pi + \gamma],$$

$$A_2 = [\gamma, \pi/2] \cup [-\pi + \gamma, -\pi/2],$$

$$A_3 = [\pi/2, \pi - \gamma] \cup [-\pi/2, -\gamma], \tag{5}$$

which form a partition of $[-\pi, \pi]$ and are depicted in Figure 1.

In Theorem 1 we establish the root-n consistency of the MLE of $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ and the limiting distribution of R_n. For presentational continuity, we have put the long proof in the Appendices A and S1.

Theorem 1. *Suppose that $f(y; \mu, \sigma)$ satisfies Conditions A1–A7 in the Appendix A and that n_i/n goes to $p_i \in (0, 1)$ as $n \to \infty$, $i = 1, \ldots, 4$. Under the null distribution $f(y; \mu_0, \sigma_0)$, we have that*

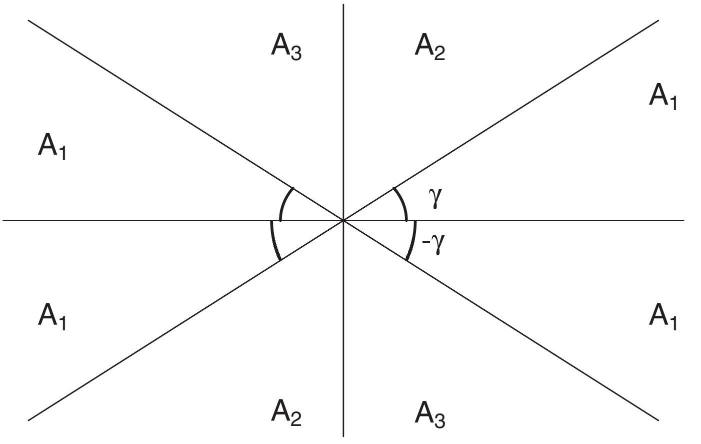

FIGURE 1 Graphical presentation of the sets A_1, A_2, and A_3 defined in Equation (5).

(i) $\hat{\mu}_h = \mu_0 + O_p(n^{-1/2})$ *and* $\hat{\sigma}_h = \sigma_0 + O_p(n^{-1/2})$, $h = 1, 2$;

(ii) as $n \to \infty$, *the LRT statistic* $R_n \to^d R = \sup_{0 \le \theta \le 1}\{Z_1^2(\theta) + Z_2^2(\theta)\}$, *where* $\overset{d}{\to}$ *stands for convergence in distribution and* $Z_1(\theta)$ *and* $Z_2(\theta)$ *are defined in Equation (4).*

Further, let ρ_h^2 *with* $\rho_h > 0$ *and* $h = 1, 2$ *be independent random variables from* χ_2^2. *Define* $\eta = (U_1 + U_2)/2 - (\pi/4)$, *where* U_1 *and* U_2, *independent of* ρ_1^2 *and* ρ_2^2, *are independent random variables from the uniform distribution on* $[-3\pi/4, 5\pi/4]$. *Then,*

$$R \overset{d}{=} \frac{\rho_1^2 + \rho_2^2}{2} + \rho_1\rho_2\{I_{A_1}(\eta) + I_{A_2}(\eta)\cos(2\eta - 2\gamma) + I_{A_3}(\eta)\cos(2\eta + 2\gamma)\}, \tag{6}$$

where $X \overset{d}{=} Y$ *indicates that the two random variables X and Y have the same distribution.*

Developing the asymptotic results of R_n is technically challenging for two reasons. First, θ is a nuisance parameter that appears only in the alternative model and hence is not identifiable under the null model. This invalidates many elegant asymptotic results for the classical LRT method (Chernoff, 1954; Self & Liang, 1987; Wilks, 1938). Second, the presence of scale parameters in the model also complicates the derivation; see the comments in Chen and Chen (2003) and Chen and Li (2009). The log-likelihood functions for the second and third groups of observations are unbounded. Using the first and fourth groups of observations, we are able to show in Lemma 1 in the Appendix A(see the Appendix S1 for a detailed proof) that any estimator of $(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2)$ with a large likelihood value is consistent for μ_h and σ_h, $h = 1, 2$. This implies the consistency of $\hat{\mu}_h$ and $\hat{\sigma}_h$ without the condition that the parameter space for μ_h and/or σ_h is bounded.

The limiting distribution of R_n involves the supremum of the χ_2^2 process. It may be difficult to calculate the critical values in general (Adler, 1990). The explicit representation of the limiting distribution in Equation (6) considerably eases this burden. For a large number N, we can generate N groups of $(\rho_1^2, \rho_2^2, U_1, U_2)$, from which we obtain N realizations of R: $R^{(1)}, \ldots, R^{(N)}$. The quantiles of R can be well approximated by those of $R^{(1)}, \ldots, R^{(N)}$. This method provides a fast way to obtain the critical values of R: it takes less than 1 min when $N = 100,000$. The approximation method of Davies (1987) can be used to find the approximate quantiles of the supremum of the χ_2^2 process. The simulation studies in Section 3.2 show that our explicit representation results in an LRT with a more accurate type I error rate than that from Davies's method.

It is worth pointing out that the regularity conditions on $f(y;\mu,\sigma)$ are not restrictive. The location-scale distributions generated by the commonly used normal, logistic, extreme-value, and t distributions all satisfy Conditions A1–A7; see Appendix S1. Hence, the results in Theorem 1 are applicable to situations where $f(y;\mu,\sigma)$ comes from any of these distributions.

As a comparison, we further consider the asymptotic properties of the LRT test under the assumption that $\sigma_1 = \sigma_2$. The LRT test statistic in this case is defined as

$$R_n^* = 2\{l_n(\hat{\theta}^*, \hat{\mu}_1^*, \hat{\mu}_2^*, \hat{\sigma}^*, \hat{\sigma}^*) - l_n(0.5, \hat{\mu}_0, \hat{\mu}_0, \hat{\sigma}_0, \hat{\sigma}_0)\},$$

where

$$(\hat{\theta}^*, \hat{\mu}_1^*, \hat{\mu}_2^*, \hat{\sigma}^*) = \arg \max_{\theta,\mu_1,\mu_2,\sigma} l_n(\theta, \mu_1, \mu_2, \sigma, \sigma).$$

We present the asymptotic properties of $(\hat{\theta}^*, \hat{\mu}_1^*, \hat{\mu}_2^*, \hat{\sigma}^*)$ and R_n^* in Theorem 2. Its proof is similar to that of Theorem 1 and is omitted to save space.

Theorem 2. *Assume the conditions of Theorem 1. Under the null distribution $f(y;\mu_0,\sigma_0)$, we have*

(i) $\hat{\mu}_h^* = \mu_0 + O_p(n^{-1/2})$ *$(h = 1, 2)$ and* $\hat{\sigma}^* = \sigma_0 + O_p(n^{-1/2})$;
(ii) as $n \to \infty$, the LRT statistic $R_n^* \to^d R^* = \sup_{0\le\theta\le1}\{Z_1^2(\theta)\}$, where $Z_1(\theta)$ is defined in Equation (4).

Further, suppose ρ^2 and η^* are two independent random variables that follow χ_2^2 and the uniform distribution on $[-\pi,\pi]$, respectively. Then

$$R^* \overset{d}{=} \rho^2\{I_{A_1}(\eta^*) + I_{A_2}(\eta^*)\cos^2(\eta^* - \gamma) + I_{A_3}(\eta^*)\cos^2(\eta^* + \gamma)\}, \tag{7}$$

where the A_i's are defined in Equation (5).

Compared with the results in Chen and Chen (2005), Theorem 2 makes two additional contributions. First, the results are applicable to the more general location-scale distribution family, whereas the results of Chen and Chen (2005) are restricted to the normal family. Second, from Lemma 1 in the Appendix A, $(\hat{\mu}_1^*, \hat{\mu}_2^*, \hat{\sigma}^*)$ is consistent without the assumption that the parameter space for (μ_1, μ_2) is bounded. Hence, the asymptotic result for R_n^* does not depend on this restrictive assumption. The explicit representation in Equation (7) can be used as in Equation (6) to calculate the critical values of R^*. Zhang, Chen, and Li (2008) also identified a representation for the χ_1^2 process in Part (ii) of Theorem 2. Our representation in Equation (7), obtained by polar transformations, is a refined version of theirs.

2.2 | Local power analysis

Our previous analysis guarantees that in theory the LRT with the proposed critical value determining strategy can control its type I error asymptotically when the null hypothesis is true. We may wonder how it performs when the null hypothesis is violated. Asymptotic local power analysis has become an important and increasingly used tool for this purpose.

To investigate the asymptotic local power of R_n and R_n^*, we consider the following local alternative

$$H_A^n : \theta = \theta_0, \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_0 - n^{-1/2}\delta_\mu \\ \mu_0 + n^{-1/2}\delta_\mu \end{pmatrix}, \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix} = \begin{pmatrix} \sigma_0 - n^{-1/2}\delta_\sigma \\ \sigma_0 + n^{-1/2}\delta_\sigma \end{pmatrix}, \quad (8)$$

where δ_μ and δ_σ are both positive constants. Define $\tau(\theta) = 1 + 4r\theta(\theta - 1)$,

$$T = \frac{\partial f(y_{11}; 0, 1)/\partial \mu}{f(y_{11}, 0, 1)}, \quad \text{and} \quad U = \frac{\partial f(y_{11}; 0, 1)/\partial \sigma}{f(y_{11}; 0, 1)}.$$

Let $\sigma_T^2 = E(T^2)$ and let $\boldsymbol{A}$ be the covariance matrix of (T, U), where the expectation and covariance are taken with respect to $f(y; 0, 1)$. Further, let $\chi_m^2(c)$ denote the non-central chi-square distribution with noncentrality parameter c and m degrees of freedom.

Theorem 3. *Assume the conditions of Theorem 1. Under the local alternative H_A^n in Equation (8), we have*

(i) $R_n \xrightarrow{d} \sup_{\theta \in [0,1]} \{\chi_2^2 \left(\boldsymbol{\rho}_{\theta_0}^\tau(\theta)\boldsymbol{\rho}_{\theta_0}(\theta)\right)\}$, *where*

$$\boldsymbol{\rho}_{\theta_0}(\theta) = -\{1 + 2r(2\theta_0\theta - \theta_0 - \theta)\}\{\tau(\theta)\}^{-\frac{1}{2}}\sigma_0^{-1}\boldsymbol{A}^{\frac{1}{2}} \begin{pmatrix} \delta_\mu \\ \delta_\sigma \end{pmatrix};$$

(ii) $R_n^* \xrightarrow{d} \sup_{\theta \in [0,1]} \{\chi_1^2 \left(\rho_{\theta_0}^{*2}(\theta)\right)\}$, *where*

$$\rho_{\theta_0}^*(\theta) = -\{1 + 2r(2\theta_0\theta - \theta_0 - \theta)\}\{\tau(\theta)\}^{-\frac{1}{2}}\sigma_0^{-1}\sigma_T\delta_\mu.$$

For convenience of presentation, the proof of Theorem 3 is deferred to the Appendix A. Theorem 3 indicates that the two LRTs R_n and R_n^* are both consistent under the local alternative H_A^n. Note that δ_σ appears in the limiting distribution of R_n but not in that of R_n^*. Hence, we expect that R_n is more powerful than R_n^* when σ_1 and σ_2 are significantly different, i.e., δ_σ is significantly different from 0. This is confirmed in the simulation study.

3 | SIMULATION STUDY

3.1 | Setup

In this section, we conduct Monte Carlo simulations to provide insight into the following questions:

(a) Do the limiting distributions of R_n and R_n^* provide accurate approximations to their finite-sample distributions?
(b) How do R_n and R_n^* perform in terms of statistical power for detecting the existence of QTL effects?

In our simulation studies, we concentrate on the normal and logistic kernels. We consider four total sample sizes, $n = 50$, 100, 200, and 300, and three values 5, 10, and 20 for d, the intermarker distance of the single interval. We use the Haldane map function $r = 0.5(1 - e^{-2d/100})$ to determine the recombination frequency r. The sample-size vector (n_1, n_2, n_3, n_4) is generated from a multinomial distribution $Multinom(n; (1-r)/2, r/2, r/2, (1-r)/2)$. In the simulations, we set the significance level to $\alpha = 5\%$ and 1%. To save space, the simulation results for $d = 10$ are omitted. Since detecting the existence of the QTL effects is essentially testing the homogeneity of the distributions in four samples, multiple-sample nonparametric tests can be applied. We choose the multiple-sample analog of the Kolmogorov–Smirnov test (Kiefer, 1959; denoted *KS*) and the multiple-sample Anderson–Darling test (Scholz & Stephens, 1987; denoted *AD*) as competitors. We study the finite-sample performance of the two LRTs, R_n and R_n^*, by comparing them with the two nonparametric tests.

3.2 | Comparison of type I errors

We first check the performance of the limiting distributions. There are two methods to obtain the quantiles of R_n and R_n^* from their limiting distributions: the first (denoted "Ours") is based on the explicit representations in Equations (6) and (7) and the second (denoted "Davies") is the approximation method of Davies (1987) for the supremum of the χ^2 process. When we apply the first method, we generate $N = 100,000$ realizations from the explicit representations in Equations (6) and (7), respectively, to determine the critical values of R_n and R_n^*. We summarize the empirical type I errors of R_n and R_n^* from the above two methods and those of *KS* and *AD* calibrated by their limiting distributions.

For the simulation results, the data are generated from $N(0,1)$ and $Logistic(0,1)$, that is, $f_1 = f_2 = N(0,1)$ and $f_1 = f_2 = Logistic(0,1)$, respectively. Here $N(\mu, \sigma^2)$ denotes normal distribution with mean μ and variance σ^2 and Logistic(μ, σ) denotes logistic distribution with location and scale parameters being μ and σ, respectively. All the empirical type I errors in Table 1 are calculated based on 10,000 repetitions. We can see that the empirical type I errors of the two LRTs based on the explicit representations in Equations (6) and (7) are quite close to the nominal levels, and they are closer than those based on Davies's method. The empirical type I errors of R_n and R_n^* based on the explicit representations indicate that the limiting distributions of R_n and R_n^* provide accurate approximations to their finite-sample distributions. The simulated type I errors of *AD* are also quite close to the nominal level, while *KS* seems to be quite conservative. As the sample size increases, all the empirical type I errors become closer to the nominal levels.

3.3 | Comparison of powers

In this subsection, we compare the powers of our LRT methods with those of *KS* and *AD*. We consider two values for θ, 0.5 and 0.7, and six combinations of f_1 and f_2:

Case I: $f_1 = N(0,1)$ and $f_2 = N(0.5, 1)$;
Case II: $f_1 = N(0,1)$ and $f_2 = N(0.5, 1.25^2)$;
Case III: $f_1 = N(0.5, 0.75)$ and $f_2 = N(0.5, 1.25)$;
Case IV: $f_1 = \text{Logistic}(0,1)$ and $f_2 = \text{Logistic}(1,1)$;
Case V: $f_1 = \text{Logistic}(0,1)$ and $f_2 = \text{Logistic}(0.8, 1.35)$;
Case VI: $f_1 = \text{Logistic}(0.5,1)$ and $f_2 = \text{Logistic}(0.5, 1.5)$.

TABLE 1 Empirical type I errors (%) of R_n, R_n^*, Kolmogorov–Smirnov test (*KS*), and Anderson–Darling test (*AD*). Here "Ours" means our method based on the explicit representations and "Davies" means Davies's approximation method. The random samples are generated from model (1) with $f_1 = f_2 = N(0, 1)$ and $f_1 = f_2 = \text{Logistic}(0, 1)$, respectively.

$f_1 = f_2 = N(0,1)$												
$\alpha = 5\%$	$d = 5$						$d = 20$					
n	R_n		R_n^*				R_n		R_n^*			
	Ours	Davies	Ours	Davies	*KS*	*AD*	Ours	Davies	Ours	Davies	*KS*	*AD*
50	5.77	6.42	5.18	5.71	3.12	5.31	5.77	6.74	5.88	6.84	4.30	5.10
100	5.52	6.04	5.32	5.87	3.48	5.17	5.70	6.68	5.46	6.46	5.19	5.41
200	5.08	5.68	5.19	5.59	3.89	4.81	5.36	6.15	4.93	5.80	4.64	4.59
300	4.62	5.13	4.95	5.48	4.18	4.83	4.51	5.26	4.92	5.78	4.93	5.00
$\alpha = 1\%$	$d = 5$						$d = 20$					
n	R_n		R_n^*				R_n		R_n^*			
	Ours	Davies	Ours	Davies	*KS*	*AD*	Ours	Davies	Ours	Davies	*KS*	*AD*
50	1.30	1.50	1.30	1.50	0.47	0.77	1.20	1.40	1.20	1.38	0.57	0.91
100	1.05	1.17	1.13	1.30	0.40	0.90	1.34	1.54	1.33	1.59	0.94	1.22
200	0.92	1.05	1.13	1.34	0.56	0.87	1.10	1.21	0.99	1.24	0.92	0.98
300	0.91	0.97	0.90	1.04	0.69	0.84	1.00	1.14	0.99	1.16	0.93	1.05
$f_1 = f_2 = \text{Logistic}(0,1)$												
$\alpha = 5\%$	$d = 5$						$d = 20$					
n	R_n		R_n^*				R_n		R_n^*			
	Ours	Davies	Ours	Davies	*KS*	*AD*	Ours	Davies	Ours	Davies	*KS*	*AD*
50	5.80	6.22	5.25	5.72	3.09	5.34	5.89	6.86	5.67	6.77	3.96	4.76
100	5.09	5.63	5.10	5.68	3.59	5.22	5.04	5.95	4.91	5.81	4.90	5.21
200	4.99	5.49	4.92	5.58	4.06	5.11	5.11	6.04	5.19	6.14	5.11	5.09
300	5.06	5.53	4.97	5.43	4.59	4.97	4.99	5.87	4.95	6.03	5.07	5.23
$\alpha = 1\%$	$d = 5$						$d = 20$					
n	R_n		R_n^*				R_n		R_n^*			
	Ours	Davies	Ours	Davies	*KS*	*AD*	Ours	Davies	Ours	Davies	*KS*	*AD*
50	1.33	1.52	1.12	1.29	0.37	0.61	1.13	1.37	1.16	1.48	0.51	0.69
100	1.00	1.19	0.97	1.14	0.52	0.90	0.98	1.15	0.91	1.06	0.76	0.96
200	1.21	1.28	0.95	1.07	0.66	0.97	1.24	1.43	1.04	1.17	0.78	0.92
300	1.08	1.25	0.94	1.08	0.75	0.91	1.12	1.20	1.06	1.21	0.96	0.94

TABLE 2 Power (%) comparison of the two likelihood ratio tests, R_n and R_n^*, Kolmogorov–Smirnov test (KS), and Anderson–Darling test (AD). The random samples are generated from model (1), in which $f_1 = N(0,1)$ and $f_2 = N(0.5,1)$ in Case I; $f_1 = N(0,1)$ and $f_2 = N(0.5,1.25^2)$ in Case II; and $f_1 = N(0.5,0.75)$ and $f_2 = N(0.5,1.25)$ in Case III. The significance level is $\alpha = 5\%$

Case I	$\theta = 0.5$		$n = 100$				$n = 200$			100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	53.9	66.9	50.4	42.4	85.8	90.7	79.6	80.0	2.89
	20	45.1	56.8	39.4	37.4	80.1	88.1	73.5	71.6	2.53
	$\theta = 0.7$		$n = 100$				$n = 200$			100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	58.7	69.6	48.9	47.2	87.2	92.4	79.3	81.3	2.91
	20	46.2	58.1	40.1	42.2	81.6	89.0	74.2	74.4	2.61
Case II	$\theta = 0.5$		$n = 100$				$n = 200$			100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	62.0	55.7	43.9	38.5	92.4	85.0	74.5	73.2	3.45
	20	54.1	47.6	35.1	33.3	87.3	79.3	64.3	67.1	3.05
	$\theta = 0.7$		$n = 100$				$n = 200$			100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	61.6	54.0	44.5	40.1	92.7	85.2	73.7	73.8	3.48
	20	55.2	50.4	34.8	36.8	89.2	80.4	65.7	69.8	3.13
Case III	$\theta = 0.5$		$n = 100$				$n = 200$			100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	30.3	4.40	5.70	8.00	59.9	4.20	9.90	11.4	1.54
	20	23.2	4.60	6.00	6.90	51.8	5.40	7.10	9.00	1.36
	$\theta = 0.7$		$n = 100$				$n = 200$			100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	29.4	5.70	7.00	8.10	59.2	4.20	8.60	12.1	1.55
	20	26.1	4.10	6.10	5.50	49.5	4.60	9.30	9.90	1.40

The QTL affects only the location in Cases I and IV; both the location and scale in Cases II and V; and only the scale in Cases III and VI. We consider the same four sample sizes n and three values for d as before. To save space, we present the results only for $n = 100$ and 200; the trends are similar for the other sample sizes. The simulated powers of R_n, R_n^*, KS, and AD for Cases I–III are shown in Table 2, and those for Cases IV–VI are shown in Table 3. For a fair comparison, the critical values of R_n, R_n^*, KS, and AD are obtained by 10,000 repetitions under the null model. All the power calculations are based on 1,000 repetitions.

Under the normal kernel, the performance trends of the simulation results displayed in Table 2 are as follows. When f_1 and f_2 have different means but the same variance, R_n^* is the

TABLE 3 Power (%) comparison of the two likelihood ratio tests, R_n and R_n^*, Kolmogorov–Smirnov test (KS), and Anderson–Darling test (AD). The random samples are generated from model (1), in which f_1 = Logistic(0, 1) and f_2 = Logistic(1, 1) in Case IV; f_1 = *Logistic*(0, 1) and f_2 = Logistic(0.8, 1.35) in Case V; and f_1 = Logistic(0.5, 1) and f_2 = Logistic(0.5, 1.5) in Case VI. The significance level is $\alpha = 5\%$

Case IV	$\theta = 0.5$		$n = 100$			$n = 200$				100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	69.4	78.2	67.1	60.9	95.8	97.2	92.5	91.8	3.83
	20	61.0	69.5	55.9	53.9	88.4	94.6	84.8	86.7	3.36
	$\theta = 0.7$		$n = 100$			$n = 200$				100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	68.9	77.5	64.0	60.5	95.1	98.0	93.9	91.9	3.86
	20	59.0	68.9	58.2	53.7	89.3	93.8	88.5	87.8	3.46
Case V	$\theta = 0.5$		$n = 100$			$n = 200$				100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	62.1	47.6	36.7	36.8	91.1	77.8	69.1	70.3	3.33
	20	55.4	42.2	29.8	35.5	85.5	66.8	56.5	64.9	2.94
	$\theta = 0.7$		$n = 100$			$n = 200$				100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	62.5	45.6	39.4	37.0	93.3	77.4	69.5	72.3	3.36
	20	54.1	37.8	33.2	30.2	86.5	66.7	60.7	64.2	3.02
Case VI	$\theta = 0.5$		$n = 100$			$n = 200$				100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	52.9	5.20	8.10	10.3	85.6	5.00	16.6	23.6	2.78
	20	48.2	6.10	8.50	11.3	77.8	4.90	11.2	19.9	2.45
	$\theta = 0.7$		$n = 100$			$n = 200$				100KL
	d	R_n	R_n^*	KS	AD	R_n	R_n^*	KS	AD	
	5	52.0	5.50	9.50	10.9	84.1	7.20	16.6	24.1	2.80
	20	44.1	4.70	7.70	11.4	79.4	5.70	10.5	22.3	2.52

most powerful of the four tests, and R_n is always more powerful than KS and AD. In contrast, when both the mean and variance of f_1 and f_2 are different, R_n is more powerful than the other three methods, and R_n^* is better than KS and AD. That is, if the QTL affects the mean but not the variance, R_n^* is more powerful at detecting it than the other three methods, while if the QTL affects both the mean and variance, R_n is more powerful. All the powers increase as the sample size increases. When f_1 and f_2 differ only in the variance, the powers of R_n are far greater than those of the other three methods, whereas R_n^* has almost no power. The power of R_n increases as the sample size increases, while that of R_n^* remains almost unchanged and is close to the nominal type I error. This implies that if the QTL affects the variance but not the mean, R_n is more powerful at detecting it and R_n^* seemingly fails.

Under the logistic kernel, the power performance trend is similar to that of the normal kernel. Hence, we omit the analysis.

Comparing the powers of R_n in Tables 2 and 3, we notice that the power of R_n under the logistic kernel is larger than that under the normal kernel in most cases when other settings such as d and θ are the same. To explain this phenomenon, we provide the Kullback–Leibler (KL) information with respect to the null model for all alternative models in the last column of Tables 2 and 3. It is seen that the KL information under Cases IV is larger than that under Case I. This explains why the power of R_n under Case IV is larger. The same interpretation is applicable to the comparison between Case VI and Case III. We also note that the KL information under Cases V is close to that under Case II, which explains why the powers of R_n under these two cases are comparable.

4 | REAL APPLICATION

In this section, we illustrate our test by analyzing the data on the double haploid (DH) population of rice in example 11.3 of Wu et al. (2007). The dataset is available at http://www.buffalo.edu/~cxma/book/. We first give a brief background. Two inbred lines, semi-dwarf IR64 and tall Azucena, were crossed to generate an heterozygous F_1 progeny population. Doubling the haploid chromosomes for the gametes of the F_1 population led to 123 DH plants (Huang et al., 1997). Such a DH population is equivalent to a backcross population because its marker segregation follows 1:1 (Huang et al., 1997). These 123 DH plants were then genotyped for 135 RFLP and 40 isozyme and RAPD markers, which represent a good coverage of 12 rice chromosomes (Huang et al., 1997).

We use chromosome 1 for illustration. The cumulative and pairwise map distances in centiMorgans for 18 markers on chromosome 1 are given in table 11.4 of Wu et al. (2007); this results in 17 intervals. Table 11.4 of Wu et al. (2007) also gives the sample size, sample mean, and sample variance for the observations in each interval. In the analysis of Wu et al. (2007), f_1 and f_2 are assumed to be normal distributions. To check the reasonability of this assumption, we apply the Kolmogorov–Smirnov test for the normality of the first sample and the fourth sample for each of the 17 intervals. The results are summarized in Table 4. As we can see, all the p-values are greater

TABLE 4 Kolmogorov–Smirnov test for the normality of the first sample and the fourth sample for each of the 17 intervals

Interval	**RG472–RG246**	**RG246–K5**	**K5–U10**	**U10–RG532**	**RG532–W1**	**W1–RG173**
First sample	0.6909	0.7656	0.5038	0.4486	0.9314	0.8568
Fourth sample	0.7311	0.7079	0.7211	0.612	0.3305	0.4123
Interval	**RG173–RZ276**	**RZ276–Amy1B**	**Amy1B–RG146**	**RG146–RG345**	**RG345–RG381**	**RG381–RZ19**
First sample	0.7896	0.9132	0.9118	0.9483	0.9598	0.8445
Fourth sample	0.3951	0.5781	0.5247	0.8668	0.9705	0.9821
Interval	**RZ19–RG690**	**RG690–RZ730**	**RZ730–RZ801**	**RZ801–RG810**	**RG810–RG331**	
First sample	0.774	0.6008	0.996	0.9736	0.9632	
Fourth sample	0.9328	0.6159	0.9753	0.6082	0.3977	

TABLE 5 *P*-values of R_n, R_n^*, Kolmogorov–Smirnov test (*KS*), and Anderson–Darling test (*AD*) in the 17 intervals. Here "R_n–Normal" and "R_n^*–Normal" are the two likelihood ratio tests (LRTs) under the normal kernel, and "R_n–Logistic" and "R_n^*–Logistic" are the two LRTs under the logistic kernel

Interval	RG472–RG246	RG246–K5	K5–U10	U10–RG532	RG532–W1	W1–RG173
R_n–Normal	0.683	0.265	0.283	0.209	0.292	0.232
R_n^*–Normal	0.995	0.943	0.641	0.830	0.906	0.757
R_n–Logistic	0.612	0.324	0.311	0.292	0.364	0.370
R_n^*–Logistic	0.965	0.945	0.472	0.703	0.785	0.722
KS	0.285	0.567	0.290	0.141	0.538	0.520
AD	0.369	0.475	0.358	0.083	0.457	0.289
Interval	**RG173–RZ276**	**RZ276–Amy1B**	**Amy1B–RG146**	**RG146–RG345**	**RG345–RG381**	**RG381–RZ19**
R_n–Normal	0.088	0.835	0.711	0.820	0.573	0.161
R_n^*–Normal	0.684	0.876	0.716	0.697	0.304	0.056
R_n–Logistic	0.196	0.823	0.760	0.871	0.650	0.095
R_n^*–Logistic	0.786	0.599	0.651	0.747	0.354	0.032
KS	0.485	0.678	0.521	0.800	0.740	0.089
AD	0.116	0.686	0.524	0.631	0.755	0.117
Interval	**RZ19–RG690**	**RG690–RZ730**	**RZ730–RZ801**	**RZ801–RG810**	**RG810–RG331**	
R_n–Normal	0	0	0	0	0	
R_n^*–Normal	0	0	0	0	0	
R_n–Logistic	0	0	0	0	0	
R_n^*–Logistic	0	0	0	0	0	
KS	0.001	0	0	0	0	
AD	0	0	0	0	0	

than .3 in all 17 intervals. This confirms that it is reasonable to assume that both f_1 and f_2 are normal distributions.

We next calculate R_n and R_n^* for each interval under the normality assumption on f_1 and f_2. From Equation (6)/(7), we generate $N = 100,000$ realizations and use them to calculate the p-values of R_n and R_n^*. Table 5 shows their p-values for the 17 intervals. For comparison purposes, we also include the results from *KS*, *AD*, R_n, and R_n^* under the logistic kernel for f_1 and f_2 for each interval.

We can see that both R_n and R_n^* under the normal kernel suggest overwhelming evidence for the existence of a QTL in the last five intervals. We also observe that the results R_n and R_n^* under the normal kernel are consistent with those from *KS* and *AD*, and also those from R_n and R_n^* under the logistic kernel. These findings may be confirmed by further experiments. It is worth mentioning that in the interval RG173–RZ276, the p-value (0.088) of R_n is much smaller than that (0.684) of R_n^*. At the 10% significance level, R_n declares the existence of a QTL in this interval, while R_n^* fails. Since R_n^* is designed to detect a QTL effect in only the mean, while R_n is able to detect a QTL effect in either the mean, the variance or both, we believe that there exists a QTL effect in only the variance in this interval.

5 | DISCUSSION

In practice, a QTL effect in the variance may be more crucial than a QTL effect in the mean (Korol et al., 1996). In this paper, under the location-scale distribution family, we studied the asymptotic properties of the MLEs of the unknown parameters and the LRT for the existence of a QTL effect under two general setups: (a) f_1 and f_2 may have different locations and/or scales and (b) f_1 and f_2 have the same unknown scale. Our theoretical results do not rely on the assumption that the parameter space for the locations is bounded. Explicit representations for the limiting distributions are presented, which facilitates the determination of the critical values. These results enrich the literature and strengthen existing results on QTL mapping in genetic backcross studies. Our simulation studies show that the explicit representations of the limiting distributions result in LRTs with more accurate type I error rates than those based on the approximation method of Davies (1987). Further, the LRT in both location and scale is uniformly more powerful than the nonparametric tests for the homogeneity of distributions in multiple samples.

The results in this paper are obtained under the assumption that there is no double recombination. When double recombination does occur, the data in each of the four groups are from a mixture distribution in both location and scale. In this situation, the log-likelihood function is unbounded, and hence the MLEs of the unknown parameters are not well defined. We suggest adding a penalty on the scales, leading to a bounded penalized likelihood (Chen, Tan, & Zhang, 2008). The LRT based on the penalized likelihood can be constructed accordingly. We expect that this new LRT will have similar properties to those in Theorem 1; we leave this for future research.

ACKNOWLEDGEMENTS

The authors thank the editor, associate editor, and two referees for constructive comments and suggestions that led to significant improvements in the paper. P.L.'s research is supported in part by NSERC Grant RGPIN-2015-06592. Y.L. and X.P.'s research are supported by grants from the National Natural Science Foundation of China (11801359, 11771144, 11771145, 11971300), the 111 Project (B14019), and Shanghai National Natural Science Foundation (19ZR1420900).

ORCID

Yukun Liu https://orcid.org/0000-0002-9743-9276

REFERENCES

Adler, R. J. (1990). *An introduction to continuity, extrema, and related topics for general gaussian processes*. Lecture Notes—Monograph Series (Vol. 12). Hayward, CA: Institute Mathematical Statistics.

Chang, M. N., Wu, R., Wu, S. S., & Casella, G. (2009). Score statistics for mapping quantitative trait loci. *Statististical Applications in Genetics and Molecular Biology*, *8*(1), 16.

Chen, H., & Chen, J. (2003). Tests for homogeneity in normal mixture in the presence of a structural parameter. *Statistica Sinica*, *13*, 351–365.

Chen, J., & Li, P. (2009). Hypothesis test for normal mixture models: The EM approach. *The Annals of Statistics*, *37*, 2523–2542.

Chen, J., Tan, X., & Zhang, R. (2008). Inference for normal mixtures in mean and variance. *Statistica Sinica*, *18*, 443–465.

Chen, Z., & Chen, H. (2005). On some statistical aspects of the interval mapping for QTL detection. *Statistica Sinica*, *15*, 909–925.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, *25*, 573–578.

Dacunha-Castelle, D., & Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. *The Annals of Statistics*, *27*, 1178–1209.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *74*, 33–43.

Doerge, R. W., Zeng, Z.-B., & Weir, B. S. (1997). Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Statistical Science*, *12*, 195–219.

Feingold, E., Brown, P. O., & Siegmund, D. (1993). Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *American Journal of Human Genetics*, *53*, 234–251.

Garel, B. (2005). Asymptotic theory of the likelihood ratio test for the identification of a mixture. *Journal of Statistical Planning and Inference.*, *131*, 271–296.

Gu, J., Koenker, R., & Volgushev, S. (2018). Testing for homogeneity in mixture models. *Econometric Theory.*, *34*, 850–895.

Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, *13*, 795–800.

Huang, N., Parco, A., Mew, T., Magpantay, G., McCouch, S., Guiderdoni, E., … Khush, G. S. (1997). RFLP mapping of isozymes, RAPD and QTLs for grain shape, brown planthopper resistance in a doubled haploid rice population. *Molecular Breeding*, *3*, 105–113.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests. *The Annals of Mathematical Statistics*, *30*, 420–447.

Kim, D.-Y., Cui, Y., & Zhao, O. (2013). Asymptotic test of mixture model and its applications to QTL interval mapping. *Journal of Statistical Planning and Inference*, *143*, 1320–1329.

Korol, A. B., Ronin, Y. I., Tadmor, Y., Bar-Zur, A., & Kirzhner, V. M. (1996). Estimating variance effect of QTL: An important prospect to increase the resolution power of interval mapping. *Genetics Research*, *67*, 187–194.

Lander, E. S., & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, *121*, 185–199.

Le Cam, L., & Yang, G. L. (1990). *Asymptotics in statistics: Some basic concepts*. New York, NY: Springer-Verlag.

Liu, X., & Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics*, *31*, 807–832.

Rebai, A., Goffinet, B., & Mangin, B. (1994). Approximated thresholds of interval mapping tests for QTL detection. *Genetics*, *138*, 235–240.

Scholz, F. W., & Stephens, M. A. (1987). K-sample Anderson-darling tests. *Journal of the American Statistical Association*, *82*, 918–924.

Self, S. G., & 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.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York, NY: Wiley.

van der Vaart, A. W. (2000). *Asymptotic statistics*. New York, NY: Cambridge University Press.

Weller, J. I., & Wyler, A. (1992). Power of different sampling strategies to detect quantitative trait loci variance effect. *Theoretical Applied Genetics*, *83*, 582–588.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*, 60–62.

Wu, R., Ma, C.-X., & Casella, G. (2007). *Statistical genetics of quantitative traits: Linkage, maps and QTL*. New York, NY: Springer.

Wu, X., Chen, H., & Liu, Y. (2008). Interval mapping for quantitative trait loci detection in finite mixture models with general kernel functions. *Communications in Statistics Theory and Methods*, *37*, 803–814.

Zhang, H., Chen, H., & Li, Z. (2008). An explicit representation of the limit of the LRT for interval mapping of quantitative trait loci. *Statistics and Probability Letters*, *78*, 207–213.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Liu G, Li P, Liu Y, Pu X. Hypothesis testing for quantitative trait locus effects in both location and scale in genetic backcross studies. *Scand J Statist*. 2020;47:1064–1089. https://doi.org/10.1111/sjos.12442

APPENDIX

Regularity conditions

The asymptotic properties of the LRTs rely on regularity conditions on $f(y;\mu,\sigma)$. We impose the following regularity conditions on $f(y;\mu,\sigma)$ in which the expectations are taken under the null distribution $f(y;\mu_0,\sigma_0)$.

A1. (Integrability) $\int_{\mathbb{R}} |\log f(y;0,1)| f(y;0,1) dy < \infty$.

A2. (Smoothness) The support of $f(y;0,1)$ is $(-\infty,\infty)$, and $f(y;0,1)$ is three times continuously differentiable with respect to y.

A3. (Identifiability) For any two mixing distribution functions ψ_1 and ψ_2 with two supporting points such that

$$\int f(y;\mu,\sigma)\, d\psi_1(\mu,\sigma) = \int f(y;\mu,\sigma)\, d\psi_2(\mu,\sigma),$$

for all y, we must have $\psi_1 = \psi_2$.

A4. (Uniform boundedness) There exists a function g such that

$$\left| \frac{\partial^{(h+l)} f(y;0,1)/\partial\mu^h \partial\sigma^l}{f(y;0,1)} \right|^3 \le g(y), \quad \text{for } h+l \le 2,$$

for all y,where h and l are two nonnegative integers, and

$$\int_{\mathbb{R}} g(y) f(y;0,1) dy < \infty.$$

Moreover, there exists a positive ϵ such that

$$\sup_{\mu^2+|\sigma-1|^2\le\epsilon} \left| \frac{\partial^{(h+l)} f(y;\mu,\sigma)/\partial\mu^h \partial\sigma^l}{f(y;0,1)} \right|^3 \le g(y), \quad \text{for } h+l = 3.$$

A5. (Positive definiteness) The covariance matrix of (T,U) is positive definite, where

$$T = \frac{\partial f(y_{11};0,1)/\partial\mu}{f(y_{11},0,1)} \quad \text{and} \quad U = \frac{\partial f(y_{11};0,1)/\partial\sigma}{f(y_{11};0,1)}.$$

That is, T and U are linearly uncorrelated.

A6. (Tail condition) There exist positive constants v_0, v_1, and β with $\beta > 1$ such that $f(y; 0, 1) \leq \min\{v_0, v_1|y|^{-\beta}\}$.

A7. (Upper bound function) There exist a nonnegative function $s(y; \mu, \sigma)$ and three positive numbers (a_0, b_0, ϵ^*) with $\epsilon^* < 1$, such that (1) $1/\beta < a_0 < 1$ with β in Condition A6, (2) $s(y; 0, 1)$ is continuous in y and satisfies $\int_{\mathbb{R}} |\log s(y; 0, 1)| s(y; 0, 1) dy < \infty$ and $\lim_{y\to\infty} s(y; 0, 1) = 0$, and (3) for $\sigma \in (0, \epsilon^* \sigma_0)$, the function of y $s(y; 0, \sigma)$ is uniformly bounded, $\int s(y; 0, \sigma) dy < 1$, and

$$f(y; 0, \sigma) \leq \begin{cases} \sigma^{-1} s(y; 0, \sigma), & \text{if } |y| \leq \sigma^{1-a_0} \\ \sigma^{b_0} s(y; 0, \sigma), & \text{if } |y| > \sigma^{1-a_0} \end{cases}.$$

Two technical lemmas

Lemma 1 establishes the consistency of the MLEs under the null model; this is the first step in the proof of Theorems 1 and 2. The lemma claims that any estimator of $(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2)$ with a large likelihood value is consistent for μ_h and σ_h, $h = 1, 2$, under the null model. Since both R_n and R_n^* are invariant to location and scale transformations, we assume that $(\mu_0, \sigma_0) = (0, 1)$.

Lemma 1. *Assume the conditions of Theorem 1. Let* $(\overline{\theta}, \overline{\mu}_1, \overline{\mu}_2, \overline{\sigma}_1, \overline{\sigma}_2)$ *be any estimator of* $(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2)$ *such that*

$$l_n(\overline{\theta}, \overline{\mu}_1, \overline{\mu}_2, \overline{\sigma}_1, \overline{\sigma}_2) - l_n(0.5, 0, 0, 1, 1) \geq c > -\infty, \tag{A1}$$

for some constant c for all n. Then under the null model $f(y; 0, 1)$, $\overline{\mu}_1 = o_p(1)$, $\overline{\mu}_2 = o_p(1)$, $\overline{\sigma}_1 - 1 = o_p(1)$, *and* $\overline{\sigma}_2 - 1 = o_p(1)$.

The proof of Lemma 1 is quite long and technically involved. For convenience of presentation, we leave it to the supplementary material.

Remark 1. By the definition of MLE,

$$l_n(\hat{\theta}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2) \geq l_n(0.5, 0, 0, 1, 1).$$

Hence the MLE $(\hat{\theta}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2)$ of $(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2)$ automatically satisfies Condition (A1) with $c = 0$. Therefore the estimators in Theorems 1 to 3 all satisfy this condition automatically.

In the next lemma, we strengthen the conclusion of Lemma 1 by providing an order assessment. For convenience of presentation, we define some notation. Let

$$T_{ij} = \frac{\partial f(y_{ij}; 0, 1)/\partial\mu}{f(y_{ij}; 0, 1)}, \quad U_{ij} = \frac{\partial f(y_{ij}; 0, 1)/\partial\sigma}{f(y_{ij}; 0, 1)}, \quad i = 1, \ldots, 4, \quad j = 1, \ldots, n_i.$$

Further, let

$$\boldsymbol{a}_{ij} = \begin{pmatrix} T_{ij} \\ U_{ij} \end{pmatrix}, \quad \boldsymbol{a}_i = \sum_{j=1}^{n_i} \boldsymbol{a}_{ij}, \quad \boldsymbol{a} = \sum_{i=1}^{4} \boldsymbol{a}_i, \quad \text{and} \quad \boldsymbol{A} = \begin{pmatrix} \sigma_T^2 & \sigma_{TU} \\ \sigma_{TU} & \sigma_U^2 \end{pmatrix},$$

where $\sigma_T^2 = Var(T_{i1})$, $\sigma_U^2 = Var(U_{i1})$, and $\sigma_{TU} = Cov(T_{i1}, U_{i1})$. We define

$$\overline{m}_1(\theta) = \theta\overline{\mu}_1 + (1-\theta)\overline{\mu}_2, \ \overline{m}_2(\theta) = \theta(\overline{\sigma}_1 - 1) + (1-\theta)(\overline{\sigma}_2 - 1), \tag{A2}$$

and $\overline{\boldsymbol{m}}(\theta) = \left(\overline{m}_1(\theta), \overline{m}_2(\theta)\right)^\tau$.

Lemma 2. *Assume the conditions of Lemma 1. Then under the null model* $f(y;0,1)$, $\overline{\mu}_1 = O_p(n^{-1/2})$, $\overline{\mu}_2 = O_p(n^{-1/2})$, $\overline{\sigma}_1 - 1 = O_p(n^{-1/2})$, *and* $\overline{\sigma}_2 - 1 = O_p(n^{-1/2})$.

Proof. Let

$$l_{n1}(\mu_1,\sigma_1) = \sum_{j=1}^{n_1} \log f_1(y_{1j}), \quad l_{n2}(\theta,\mu_1,\mu_2,\sigma_1,\sigma_2) = \sum_{j=1}^{n_2} \log\{\theta f_1(y_{2j}) + (1-\theta) f_2(y_{2j})\},$$

$$l_{n3}(\theta,\mu_1,\mu_2,\sigma_1,\sigma_2) = \sum_{j=1}^{n_3} \log\{(1-\theta) f_1(y_{3j}) + \theta f_2(y_{3j})\}, \quad l_{n4}(\mu_2,\sigma_2) = \sum_{j=1}^{n_4} \log f_2(y_{4j}).$$

Then

$$l_n(\theta,\mu_1,\mu_2,\sigma_1,\sigma_2) = l_{n1}(\mu_1,\sigma_1) + l_{n2}(\theta,\mu_1,\mu_2,\sigma_1,\sigma_2) + l_{n3}(\theta,\mu_1,\mu_2,\sigma_1,\sigma_2) + l_{n4}(\mu_2,\sigma_2).$$

Next we derive an upper bound for $l_n(\overline{\theta},\overline{\mu}_1,\overline{\mu}_2,\overline{\sigma}_1,\overline{\sigma}_2) - l_n(0.5,0,0,1,1)$. Together with the lower bound c, we get the order assessment of $\overline{\mu}_h$ and $\overline{\sigma}_h$, $h = 1,2$.

From Lemma 1, we have the consistency $\overline{\mu}_h = o_p(1)$ and $\overline{\sigma}_h - 1 = o_p(1)$, $h = 1,2$. Applying a second-order Taylor expansion to $l_{n1}(\overline{\mu}_1,\overline{\sigma}_1) - l_{n1}(0,1)$ around $(0,1)$ and using the law of large numbers, we get

$$l_{n1}(\overline{\mu}_1,\overline{\sigma}_1) - l_{n1}(0,1) = \left\{\overline{\boldsymbol{m}}(1)\right\}^{\tau} \boldsymbol{a}_1 - \frac{n_1}{2}\left\{\overline{\boldsymbol{m}}(1)\right\}^{\tau} \boldsymbol{A} \left\{\overline{\boldsymbol{m}}(1)\right\}\{1 + o_p(1)\}. \tag{A3}$$

Similarly,

$$l_{n4}(\overline{\mu}_2,\overline{\sigma}_2) - l_{n4}(0,1) = \left\{\overline{\boldsymbol{m}}(0)\right\}^{\tau} \boldsymbol{a}_4 - \frac{n_4}{2}\left\{\overline{\boldsymbol{m}}(0)\right\}^{\tau} \boldsymbol{A} \left\{\overline{\boldsymbol{m}}(0)\right\}\{1 + o_p(1)\}. \tag{A4}$$

We now find an upper bound for $l_{n2}(\overline{\theta},\overline{\mu}_1,\overline{\mu}_2,\overline{\sigma}_1,\overline{\sigma}_2) - l_{n2}(0.5,0,0,1,1)$. Write

$$l_{n2}(\overline{\theta},\overline{\mu}_1,\overline{\mu}_2,\overline{\sigma}_1,\overline{\sigma}_2) - l_{n2}(0.5,0,0,1,1) = \sum_{j=1}^{n_2} \log(1+\delta_j)$$

with

$$\delta_j = \frac{\overline{\theta}\left\{f(y_{2j};\overline{\mu}_1,\overline{\sigma}_1) - f(y_{2j};0,1)\right\} + (1-\overline{\theta})\left\{f(y_{2j};\overline{\mu}_2,\overline{\sigma}_2) - f(y_{2j};0,1)\right\}}{f(y_{2j};0,1)}.$$

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

$$l_{n2}(\overline{\theta},\overline{\mu}_1,\overline{\mu}_2,\overline{\sigma}_1,\overline{\sigma}_2) - l_{n2}(0.5,0,0,1,1) \le \sum_{j=1}^{n_2} \delta_j - \sum_{j=1}^{n_2} \delta_j^2/2 + \sum_{j=1}^{n_2} \delta_j^3/3. \tag{A5}$$

Applying a first-order Taylor expansion to $f(y_{2j};\overline{\mu}_h,\overline{\sigma}_h)$, $h = 1,2$, we find that

$$\delta_j = \left\{\overline{\boldsymbol{m}}(\overline{\theta})\right\}^{\tau} \boldsymbol{a}_{2j} + \varepsilon_{nj}$$

and the remainder term $\varepsilon_n = \sum_{j=1}^{n_2} \varepsilon_{nj}$ satisfies

$$\varepsilon_n = O_p(n_2^{1/2}) \sum_{h=1}^{2} \left\{ \mu_h^2 + (\overline{\sigma}_h - 1)^2 \right\} = o_p(n) \left\{ ||\overline{\boldsymbol{m}}(0)||^2 + ||\overline{\boldsymbol{m}}(1)||^2 \right\}. \tag{A6}$$

Here $||\overline{\boldsymbol{m}}(0)||$ and $||\overline{\boldsymbol{m}}(1)||$ denote the L_2 norms of $\overline{\boldsymbol{m}}(0)$ and $\overline{\boldsymbol{m}}(1)$, respectively. Therefore, for the linear term in Equation (A5), we have

$$\sum_{j=1}^{n_2} \delta_j = \left\{ \overline{\boldsymbol{m}}(\overline{\theta}) \right\}^{\tau} \boldsymbol{a}_2 + \varepsilon_n, \tag{A7}$$

where the order of ε_n is assessed in Equation (A6). After some straightforward algebra, we find that the quadratic and cubic terms in Equation (A5) satisfy

$$\sum_{j=1}^{n_2} \delta_j^2 = \sum_{j=1}^{n_2} \left[\left\{ \overline{\boldsymbol{m}}(\overline{\theta}) \right\}^{\tau} \boldsymbol{a}_{2j} \right]^2 + O_p(\varepsilon_n),$$

$$\sum_{j=1}^{n_2} \delta_j^3 = \sum_{j=1}^{n_2} \left[\left\{ \overline{\boldsymbol{m}}(\overline{\theta}) \right\}^{\tau} \boldsymbol{a}_{2j} \right]^3 + O_p(\varepsilon_n).$$

By the strong law of large numbers, the fact that $||\overline{\boldsymbol{m}}(\overline{\theta})||^2 \leq ||\overline{\boldsymbol{m}}(0)||^2 + ||\overline{\boldsymbol{m}}(1)||^2$, and the order assessment of ε_n in Equation (A6), we have

$$\sum_{j=1}^{n_2} \delta_j^2 = n_2 \left\{ \overline{\boldsymbol{m}}(\overline{\theta}) \right\}^{\tau} \boldsymbol{A} \left\{ \overline{\boldsymbol{m}}(\overline{\theta}) \right\} + o_p(n) \left\{ ||\overline{\boldsymbol{m}}(0)||^2 + ||\overline{\boldsymbol{m}}(1)||^2 \right\}, \tag{A8}$$

$$\sum_{j=1}^{n_2} \delta_j^3 = o_p(n) \left\{ ||\overline{\boldsymbol{m}}(0)||^2 + ||\overline{\boldsymbol{m}}(1)||^2 \right\}. \tag{A9}$$

Combining Equations (A5)–(A9), we get the refined upper bound for $l_{n2}(\overline{\theta}, \overline{\mu}_1, \overline{\mu}_2, \overline{\sigma}_1, \overline{\sigma}_2) - l_{n2}(0.5, 0, 0, 1, 1)$ as follows:

$$\begin{aligned} & l_{n2}(\overline{\theta}, \overline{\mu}_1, \overline{\mu}_2, \overline{\sigma}_1, \overline{\sigma}_2) - l_{n2}(0.5, 0, 0, 1, 1) \\ & \leq \left\{ \overline{\boldsymbol{m}}(\overline{\theta}) \right\}^{\tau} \boldsymbol{a}_2 - \frac{n_2}{2} \left\{ \overline{\boldsymbol{m}}(\overline{\theta}) \right\}^{\tau} \boldsymbol{A} \left\{ \overline{\boldsymbol{m}}(\overline{\theta}) \right\} + o_p(n) \left\{ ||\overline{\boldsymbol{m}}(0)||^2 + ||\overline{\boldsymbol{m}}(1)||^2 \right\}. \end{aligned} \tag{A10}$$

Similarly,

$$\begin{aligned} l_{n3}(\overline{\theta}, \overline{\mu}_1, \overline{\mu}_2, \overline{\sigma}_1, \overline{\sigma}_2) - l_{n3}(0.5, 0, 0, 1, 1) \leq & \left\{ \overline{\boldsymbol{m}}(1 - \overline{\theta}) \right\}^{\tau} \boldsymbol{a}_3 - \frac{n_3}{2} \left\{ \overline{\boldsymbol{m}}(1 - \overline{\theta}) \right\}^{\tau} \boldsymbol{A} \left\{ \overline{\boldsymbol{m}}(1 - \overline{\theta}) \right\} \\ & + o_p(n) \left\{ ||\overline{\boldsymbol{m}}(0)||^2 + ||\overline{\boldsymbol{m}}(1)||^2 \right\}. \end{aligned} \tag{A11}$$

Combining Equations (A3), (A4), (A10), and (A11), we have

$$\begin{aligned} c &\le l_n(\bar{\theta}, \bar{\mu}_1, \bar{\mu}_2, \bar{\sigma}_1, \bar{\sigma}_2) - l_n(0.5, 0, 0, 1, 1) \\ &\le \left\{\overline{\boldsymbol{m}}(1)\right\}^\tau \boldsymbol{a}_1 - \frac{n_1}{2}\left\{\overline{\boldsymbol{m}}(1)\right\}^\tau \boldsymbol{A}\left\{\overline{\boldsymbol{m}}(1)\right\}\{1+o_p(1)\} \\ &+ \left\{\overline{\boldsymbol{m}}(\bar{\theta})\right\}^\tau \boldsymbol{a}_2 - \frac{n_2}{2}\left\{\overline{\boldsymbol{m}}(\bar{\theta})\right\}^\tau \boldsymbol{A}\left\{\overline{\boldsymbol{m}}(\bar{\theta})\right\} \\ &+ \left\{\overline{\boldsymbol{m}}(1-\bar{\theta})\right\}^\tau \boldsymbol{a}_3 - \frac{n_3}{2}\left\{\overline{\boldsymbol{m}}(1-\bar{\theta})\right\}^\tau \boldsymbol{A}\left\{\overline{\boldsymbol{m}}(1-\bar{\theta})\right\} \\ &+ \left\{\overline{\boldsymbol{m}}(0)\right\}^\tau \boldsymbol{a}_4 - \frac{n_4}{2}\left\{\overline{\boldsymbol{m}}(0)\right\}^\tau \boldsymbol{A}\left\{\overline{\boldsymbol{m}}(0)\right\}\{1+o_p(1)\} \end{aligned} \tag{A12}$$

$$\begin{aligned} &\le \left\{\overline{\boldsymbol{m}}(1)\right\}^\tau \boldsymbol{a}_1 - \frac{n_1}{2}\left\{\overline{\boldsymbol{m}}(1)\right\}^\tau \boldsymbol{A}\left\{\overline{\boldsymbol{m}}(1)\right\}\{1+o_p(1)\} \\ &+ \left\{\overline{\boldsymbol{m}}(0)\right\}^\tau \boldsymbol{a}_4 - \frac{n_4}{2}\left\{\overline{\boldsymbol{m}}(0)\right\}^\tau \boldsymbol{A}\left\{\overline{\boldsymbol{m}}(0)\right\}\{1+o_p(1)\} + O_p(1). \end{aligned} \tag{A13}$$

By Condition (A5), $\boldsymbol{A}$ is positive definite. Then the upper bound of $l_n(\bar{\theta}, \bar{\mu}_1, \bar{\mu}_2, \bar{\sigma}_1, \bar{\sigma}_2) - l_n(0.5, 0, 0, 1, 1)$ in (A13) has order $O_p(1)$. Together with the lower bound c, this implies that

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

Any values of $(\bar{\mu}_1, \bar{\sigma}_1 - 1, \bar{\mu}_2, \bar{\sigma}_2 - 1)$ outside this range will violate the inequality. This completes the proof. ■

Proof of Theorem 1

Proof of Part (i).

By the definition of the MLE, we have $l_n(\hat{\theta}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2) - l_n(0.5, 0, 0, 1, 1) \ge 0$. Hence, Condition (A1) is satisfied. Then applying the results in Lemma 2, we obtain the results in Part (i).

Proof of Part (ii).

Note that

$$R_n = 2\left\{l_n(\hat{\theta}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2) - l_n(0.5, \hat{\mu}_0, \hat{\mu}_0, \hat{\sigma}_0, \hat{\sigma}_0)\right\} = R_{1n} - R_{2n}, \tag{A14}$$

where

$$R_{1n} = 2\left\{l_n(\hat{\theta}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2) - l_n(0.5, 0, 0, 1, 1)\right\},$$

and

$$R_{2n} = 2\left\{l_n(0.5, \hat{\mu}_0, \hat{\mu}_0, \hat{\sigma}_0, \hat{\sigma}_0) - l_n(0.5, 0, 0, 1, 1)\right\}.$$

Applying some of the classical results for regular models (Serfling, 1980), we have

$$R_{2n} = \boldsymbol{a}^\tau (n\boldsymbol{A})^{-1}\boldsymbol{a} + o_p(1). \tag{A15}$$

Next we use a sandwich method to find the approximation of R_{1n}. We proceed in two steps. In Step 1, we derive an upper bound for R_{1n} and in Step 2, we argue that the upper bound is achievable.

By Part (i), the results in Equation (A12) are applicable to R_{1n}. Hence,

$$\begin{aligned} R_{1n} &= 2\left\{l_n(\hat\theta,\hat\mu_1,\hat\mu_2,\hat\sigma_1,\hat\sigma_2) - l_n(0.5,0,0,1,1)\right\} \\ &\le 2\{\hat{\boldsymbol m}(1)\}^\tau \boldsymbol a_1 - n_1\{\hat{\boldsymbol m}(1)\}^\tau \boldsymbol A\{\hat{\boldsymbol m}(1)\}\{1+o_p(1)\} \\ &\quad + 2\left\{\hat{\boldsymbol m}(\hat\theta)\right\}^\tau \boldsymbol a_2 - n_2\left\{\hat{\boldsymbol m}(\hat\theta)\right\}^\tau \boldsymbol A\left\{\hat{\boldsymbol m}(\hat\theta)\right\} \\ &\quad + 2\left\{\hat{\boldsymbol m}(1-\hat\theta)\right\}^\tau \boldsymbol a_3 - n_3\left\{\hat{\boldsymbol m}(1-\hat\theta)\right\}^\tau \boldsymbol A\left\{\hat{\boldsymbol m}(1-\hat\theta)\right\} \\ &\quad + 2\{\hat{\boldsymbol m}(0)\}^\tau \boldsymbol a_4 - n_4\{\hat{\boldsymbol m}(0)\}^\tau \boldsymbol A\{\hat{\boldsymbol m}(0)\}\{1+o_p(1)\}. \end{aligned} \tag{A16}$$

Here $\hat{\boldsymbol m}(\theta)$ is defined similarly to Equation (A2) with $(\hat\mu_1,\hat\mu_2,\hat\sigma_1,\hat\sigma_2)$ in place of $(\overline\mu_1,\overline\mu_2,\overline\sigma_1,\overline\sigma_2)$. To simplify the upper bound in Equation (A16), let

$$\hat{\boldsymbol\beta}_1 = \frac{\hat{\boldsymbol m}(1)+\hat{\boldsymbol m}(0)}{2} = \begin{pmatrix} \frac{\hat\mu_1+\hat\mu_2}{2} \\ \frac{\hat\sigma_1+\hat\sigma_2-2}{2} \end{pmatrix} \text{ and } \hat{\boldsymbol\beta}_2 = \frac{\hat{\boldsymbol m}(1)-\hat{\boldsymbol m}(0)}{2} = \begin{pmatrix} \frac{\hat\mu_1-\hat\mu_2}{2} \\ \frac{\hat\sigma_1-\hat\sigma_2}{2} \end{pmatrix}.$$

Then

$$\hat{\boldsymbol m}(\theta) = \hat{\boldsymbol\beta}_1 + 2(\theta-0.5)\hat{\boldsymbol\beta}_2. \tag{A17}$$

Substituting Equation (A17) into Equation (A16) and using Part (i) and the condition that $n_i/n \to p_i$, $i=1,2,3,4$ with

$$(p_1,p_2,p_3,p_4) = \left(\frac{1-r}{2},\frac{r}{2},\frac{r}{2},\frac{1-r}{2}\right),$$

we have

$$\begin{aligned} R_{1n} &= 2\left\{l_n(\hat\theta,\hat\mu_1,\hat\mu_2,\hat\sigma_1,\hat\sigma_2) - l_n(0.5,0,0,1,1)\right\} \\ &\le 2\hat{\boldsymbol\beta}_1^\tau \boldsymbol a - n\hat{\boldsymbol\beta}_1^\tau \boldsymbol A\hat{\boldsymbol\beta}_1\{1+o_p(1)\} + 2\hat{\boldsymbol\beta}_2^\tau \boldsymbol b(\hat\theta) - n\tau(\hat\theta)\hat{\boldsymbol\beta}_2^\tau \boldsymbol A\hat{\boldsymbol\beta}_2 \\ &\le \boldsymbol a^\tau(n\boldsymbol A)^{-1}\boldsymbol a + \left\{\boldsymbol b(\hat\theta)\right\}^\tau\left\{n\tau(\hat\theta)\boldsymbol A\right\}^{-1}\left\{\boldsymbol b(\hat\theta)\right\} + o_p(1) \\ &\le \boldsymbol a^\tau(n\boldsymbol A)^{-1}\boldsymbol a + \sup_{\theta\in[0,1]}\left[\{\boldsymbol b(\theta)\}^\tau\{n\tau(\theta)\boldsymbol A\}^{-1}\{\boldsymbol b(\theta)\}\right] + o_p(1), \end{aligned} \tag{A18}$$

where

$$\boldsymbol b(\theta) = \boldsymbol a_1 - \boldsymbol a_4 + (2\theta-1)(\boldsymbol a_2 - \boldsymbol a_3) \text{ and } \tau(\theta) = 1 + 4r\theta(\theta-1).$$

Next, we show that the upper bound in Equation (A18) for R_{1n} is achievable. Let

$$\tilde\theta = \arg\max_{\theta\in[0,1]}\left[\{\boldsymbol b(\theta)\}^\tau\{n\tau(\theta)\boldsymbol A\}^{-1}\{\boldsymbol b(\theta)\}\right],$$

and $(\tilde\mu_1,\tilde\mu_2,\tilde\sigma_1,\tilde\sigma_2)$ be determined by

$$\begin{pmatrix} \frac{\tilde\mu_1+\tilde\mu_2}{2} \\ \frac{\tilde\sigma_1+\tilde\sigma_2-2}{2} \end{pmatrix} = (n\boldsymbol A)^{-1}\boldsymbol a \quad\text{and}\quad \begin{pmatrix} \frac{\tilde\mu_1-\tilde\mu_2}{2} \\ \frac{\tilde\sigma_1-\tilde\sigma_2}{2} \end{pmatrix} = \left\{n\tau(\tilde\theta)\boldsymbol A\right\}^{-1}\left\{\boldsymbol b(\tilde\theta)\right\}.$$

Note that it is easy to verify that $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma}_1, \tilde{\sigma}_2)$ exists and

$$\tilde{\mu}_h = O_p(n^{-1/2}), \quad \tilde{\sigma}_h - 1 = O_p(n^{-1/2}), \quad h = 1, 2.$$

With this order assessment and applying a second-order Taylor expansion, we have

$$\begin{aligned} R_{1n} &\geq 2\left\{l_n(\tilde{\theta}, \tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma}_1, \tilde{\sigma}_2) - l_n(0.5, 0, 0, 1, 1)\right\} \\ &= \boldsymbol{a}^\tau (n\boldsymbol{A})^{-1}\boldsymbol{a} + \sup_{\theta\in[0,1]} \left[\{\boldsymbol{b}(\theta)\}^\tau \{n\tau(\theta)\boldsymbol{A}\}^{-1} \{\boldsymbol{b}(\theta)\}\right] + o_p(1). \end{aligned} \tag{A19}$$

Combining Equations (A18) and (A19) leads to

$$R_{1n} = \boldsymbol{a}^\tau (n\boldsymbol{A})^{-1}\boldsymbol{a} + \sup_{\theta\in[0,1]} \left[\{\boldsymbol{b}(\theta)\}^\tau \{n\tau(\theta)\boldsymbol{A}\}^{-1} \{\boldsymbol{b}(\theta)\}\right] + o_p(1).$$

With Equations (A14) and (A15), we further have that

$$R_n = \sup_{\theta\in[0,1]} \left[\{\boldsymbol{b}(\theta)\}^\tau \{n\tau(\theta)\boldsymbol{A}\}^{-1} \{\boldsymbol{b}(\theta)\}\right] + o_p(1). \tag{A20}$$

It can be verified that the process $\{\boldsymbol{b}(\theta)\}^\tau \{n\tau(\theta)\boldsymbol{A}\}^{-1} \{\boldsymbol{b}(\theta)\}$ converges weakly to the process $Z_1^2(\theta) + Z_2^2(\theta)$, where $Z_1(\theta)$ and $Z_2(\theta)$ are defined in Equation (4) (see Kim et al., 2013). Hence,

$$R_n \to R = \sup_{0\leq\theta\leq 1} \{Z_1^2(\theta) + Z_2^2(\theta)\},$$

in distribution, as $n \to \infty$. This completes the proof of Part (ii).

Proof of Equation (6)

Let $c_1(\theta) = \sqrt{1-r}\sqrt{1+4r\theta(\theta-1)}$ and $c_2(\theta) = \sqrt{r}(2\theta-1)\sqrt{1+4r\theta(\theta-1)}$. Recall the forms of $Z_1(\theta)$ and $Z_2(\theta)$ defined in Equation (4). Then R can be written as

$$\begin{aligned} R &= Z_1^2(\theta) + Z_2^2(\theta) = \{c_1(\theta)z_{11} + c_2(\theta)z_{12}\}^2 + \{c_1(\theta)z_{21} + c_2(\theta)z_{22}\}^2 \\ &= \left(c_1(\theta),\ c_2(\theta)\right) \begin{pmatrix} z_{11}^2 + z_{21}^2 & z_{11}z_{12} + z_{21}z_{22} \\ z_{11}z_{12} + z_{21}z_{22} & z_{12}^2 + z_{22}^2 \end{pmatrix} \begin{pmatrix} c_1(\theta) \\ c_2(\theta) \end{pmatrix}, \end{aligned} \tag{A21}$$

where z_{11}, z_{12}, z_{21}, and z_{22} are four independent standard normal variables.

Similarly to Lemma 3 of Zhang et al. (2008), we have

$$\{(c_1(\theta), c_2(\theta)) : 0 \leq \theta \leq 1\} = \{(x_1, x_2) : x_1^2 + x_2^2 = 1, \quad x_1 \geq \sqrt{1-r}\}.$$

Let

$$\boldsymbol{W} = \begin{pmatrix} z_{11}^2 + z_{21}^2 & z_{11}z_{12} + z_{21}z_{22} \\ z_{11}z_{12} + z_{21}z_{22} & z_{12}^2 + z_{22}^2 \end{pmatrix}, \quad \mathscr{B} = \{(x_1, x_2) : x_1^2 + x_2^2 = 1, \quad x_1 \geq \sqrt{1-r}\}.$$

Then, from Equation (A21) we have

$$R = \sup_{(x_1,x_2)\in\mathscr{B}} (x_1, x_2)\boldsymbol{W}(x_1, x_2)^\tau. \tag{A22}$$

To help us find the maximum of $(x_1, x_2)\boldsymbol{W}(x_1, x_2)^\tau$, we make the following polar transformations. Let

$$\begin{cases} (z_{11} - z_{22})/\sqrt{2} = \rho_1 \sin U_1 \\ (z_{12} + z_{21})/\sqrt{2} = \rho_1 \cos U_1 \end{cases} \quad \text{and} \quad \begin{cases} (z_{11} + z_{22})/\sqrt{2} = \rho_2 \cos U_2 \\ (z_{21} - z_{12})/\sqrt{2} = \rho_2 \sin U_2 \end{cases},$$

where ρ_1^2 with $\rho_1 > 0$, ρ_2^2 with $\rho_2 > 0$, U_1, and U_2 are four independent random variables with ρ_1^2 and ρ_2^2 from a χ_2^2 distribution, and U_1 and U_2 from a uniform distribution on $[-3\pi/4, 5\pi/4]$.

It can be verified that the two eigenvalues of $\boldsymbol{W}$ are $\lambda_1 = (1/2)(\rho_1 + \rho_2)^2$ and $\lambda_2 = (1/2)(\rho_1 - \rho_2)^2$, respectively. Further, $\boldsymbol{W}$ can be decomposed as

$$\boldsymbol{W} = \boldsymbol{P}^\tau \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \boldsymbol{P}, \tag{A23}$$

where

$$\boldsymbol{P} = \begin{pmatrix} \cos\{(U_1 + U_2)/2 - \pi/4\} & -\sin\{(U_1 + U_2)/2 - \pi/4\} \\ \sin\{(U_1 + U_2)/2 - \pi/4\} & \cos\{(U_1 + U_2)/2 - \pi/4\} \end{pmatrix}.$$

Since $\boldsymbol{P}$ is an orthogonal transformation and $x_1^2 + x_2^2 = 1$, it follows that $(x_1, x_2)\boldsymbol{P}^\tau\boldsymbol{P}(x_1, x_2)^\tau = x_1^2 + x_2^2 = 1$. Therefore, we can write $(x_1, x_2)\boldsymbol{P}^\tau = (\cos\alpha, \sin\alpha)$ with $\alpha \in [-\pi, \pi]$. From Equations (A22) and (A23), we have

$$R = \sup_{\alpha \in \mathscr{F}} (\lambda_1 \cos^2\alpha + \lambda_2 \sin^2\alpha) = \sup_{\alpha \in \mathscr{F}} \{(\lambda_1 - \lambda_2)\cos^2\alpha + \lambda_2\},$$

where

$$\mathscr{F} = \{\alpha : \cos\{(U_1 + U_2)/2 - \pi/4\}\cos\alpha + \sin\{(U_1 + U_2)/2 - \pi/4\}\sin\alpha \geq \sqrt{1 - r}\}.$$

Recall that $\eta = (U_1 + U_2)/2 - \pi/4$, which satisfies $\eta \in [-\pi, \pi]$. Hence,

$$\mathscr{F} = \{\alpha : \cos(\alpha - \eta) \geq \sqrt{1 - r}\}.$$

Therefore, after some simple analysis, Equation (6) follows. This completes the proof of Theorem 1.

Proof of Theorem 3

Proof of Part (i).

Assume $(\mu_0, \sigma_0) = (0, 1)$, then the local alternative (8) is equivalent to

$$H_A^n : \theta = \theta_0, \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} -n^{-1/2}\delta_\mu/\sigma_0 \\ n^{-1/2}\delta_\mu/\sigma_0 \end{pmatrix}, \quad \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix} = \begin{pmatrix} 1 - n^{-1/2}\delta_\sigma/\sigma_0 \\ 1 + n^{-1/2}\delta_\sigma/\sigma_0 \end{pmatrix}.$$

By Equation (A20), $R_n = \sup_{\theta \in [0,1]}\{R_n(\theta)\} + o_p(1)$ under the null model, where

$$R_n(\theta) = \{\boldsymbol{b}(\theta)\}^\tau \{n\tau(\theta)\boldsymbol{A}\}^{-1} \{\boldsymbol{b}(\theta)\}.$$

Next we derive the limiting distribution of R_n under H_A^n.

Let $\Lambda_n = \Lambda_{n1} + \Lambda_{n2} + \Lambda_{n3} + \Lambda_{n4}$, where

$$\Lambda_{n1} = l_{n1}(\mu_1, \sigma_1) - l_{n1}(0,1), \Lambda_{n2} = l_{n2}(\theta_0, \mu_1, \mu_2, \sigma_1, \sigma_2) - l_{n3}(0.5, 0, 0, 1, 1),$$

$$\Lambda_{n3} = l_{n3}(\theta_0, \mu_1, \mu_2, \sigma_1, \sigma_2) - l_{n3}(0.5, 0, 0, 1, 1), \text{and } \Lambda_{n4} = l_{n4}(\mu_1, \sigma_1) - l_{n4}(0,1).$$

Using the second Taylor expansion, under the null, we have

$$\Lambda_{n1} = -n^{-1/2} \sum_{j=1}^{n_1} (\Delta_\mu T_{1j} + \Delta_\sigma U_{1j}) - 0.5 p_1 (\Delta_\mu, \Delta_\sigma) \boldsymbol{A} \begin{pmatrix} \Delta_\mu \\ \Delta_\sigma \end{pmatrix} + o_p(1),$$

$$\Lambda_{n4} = n^{-1/2} \sum_{j=1}^{n_4} (\Delta_\mu T_{4j} + \Delta_\sigma U_{4j}) - 0.5 p_4 (\Delta_\mu, \Delta_\sigma) \boldsymbol{A} \begin{pmatrix} \Delta_\mu \\ \Delta_\sigma \end{pmatrix} + o_p(1),$$

where $\Delta_\mu = \delta_\mu / \sigma_0$ and $\Delta_\sigma = \delta_\sigma / \sigma_0$. Similarly,

$$\Lambda_{n2} = n^{-1/2} \sum_{j=1}^{n_2} \{ m_1(\theta_0) T_{2j} + m_2(\theta_0) U_{2j} \} - 0.5 p_2 \left(m_1(\theta_0), m_2(\theta_0) \right) \boldsymbol{A} \begin{pmatrix} m_1(\theta_0) \\ m_2(\theta_0) \end{pmatrix} + o_p(1),$$

and

$$\Lambda_{n3} = n^{-1/2} \sum_{j=1}^{n_3} \{ m_1(1-\theta_0) T_{3j} + m_2(1-\theta_0) U_{3j} \} - 0.5 p_3 \left(m_1(1-\theta_0), m_2(1-\theta_0) \right) \boldsymbol{A} \begin{pmatrix} m_1(1-\theta_0) \\ m_2(1-\theta_0) \end{pmatrix} + o_p(1),$$

where $m_1(\theta) = (1-2\theta)\Delta_\mu$, and $m_2(\theta) = (1-2\theta)\Delta_\sigma$.

By the central limit theorem, we get $\Lambda_n \xrightarrow{d} N(-0.5c_0, c_0)$ under the null, where

$$c_0 = 2p_1 \left(\Delta_\mu, \Delta_\sigma \right) \boldsymbol{A} \begin{pmatrix} \Delta_\mu \\ \Delta_\sigma \end{pmatrix} + 2p_2 \left(m_1(\theta_0), m_2(\theta_0) \right) \boldsymbol{A} \begin{pmatrix} m_1(\theta_0) \\ m_2(\theta_0) \end{pmatrix}.$$

Therefore, the local alternative H_A^n is contiguous to the null distribution (Le Cam & Yang, 1990 and example 6.5 of van der Vaart, 2000). By Le Cam's contiguity theory, the limiting distribution of $R_n(\theta)$ under H_A^n is determined by the joint limiting distribution of $\{n\tau(\theta)\boldsymbol{A}\}^{-1/2}\boldsymbol{b}(\theta)$ and Λ_n under the null model.

By the central limit theorem and Slutsky's theorem, the joint limiting distribution of $\{n\tau(\theta)\boldsymbol{A}\}^{-1/2}\boldsymbol{b}(\theta)$ and Λ_n under the null model is multivariate normal

$$\mathcal{N}_3 \left(\begin{pmatrix} \boldsymbol{0} \\ -0.5c_0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{I}_2 & \boldsymbol{\rho}_{\theta_0}(\theta) \\ \boldsymbol{\rho}_{\theta_0}^\tau(\theta) & c_0 \end{pmatrix} \right) \text{ with } \boldsymbol{\rho}_{\theta_0}(\theta) = \boldsymbol{\Delta}_{\theta_0}(\theta) \begin{pmatrix} \Delta_\mu \\ \Delta_\sigma \end{pmatrix},$$

where $\boldsymbol{\Delta}_{\theta_0}(\theta) = -\{1 + 2r(2\theta_0\theta - \theta_0 - \theta)\}\{\tau(\theta)\}^{-\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}$. By Le Cam's third lemma (van der Vaart, 2000), we have under H_A^n,

$$\{n\tau(\theta)\boldsymbol{A}\}^{-1/2}\boldsymbol{b}(\theta) \to^d \mathcal{N}_2\left(\boldsymbol{\rho}_{\theta_0}(\theta), \boldsymbol{I}_2\right).$$

Further, we have $R_n(\theta) \xrightarrow{d} \chi_2^2\left(\boldsymbol{\rho}_{\theta_0}^{\tau}(\theta)\boldsymbol{\rho}_{\theta_0}(\theta)\right)$ under H_A^n. Because $R_n = \sup_{\theta\in[0,1]}\{R_n(\theta)\} + o_p(1)$ under the null model, by Le Cam's first lemma (van der Vaart, 2000), $R_n = \sup_{\theta\in[0,1]}\{R_n(\theta)\} + o_p(1)$ still holds under the local alternative H_A^n. Therefore, the limiting distribution of R_n under H_A^n is $\sup_{\theta\in[0,1]}\left\{\chi_2^2\left(\boldsymbol{\rho}_{\theta_0}^{\tau}(\theta)\boldsymbol{\rho}_{\theta_0}(\theta)\right)\right\}$.

Proof of Part (ii).

The proof of this part is similar to that of Part (i) and hence is omitted.