A Likelihood Ratio Test for Genome-Wide Association under Genetic Heterogeneity

Meng Qian and Yongzhao Shao

New York University School of Medicine, 650 First Avenue, Rm 538, New York, NY 10016 USA

Summary

Most existing association tests for genome-wide association studies (GWASs) fail to account for genetic heterogeneity. Zhou and Pan proposed a binomial-mixture-model-based association test to account for the possible genetic heterogeneity in case-control studies. The idea is elegant, however, the proposed test requires an expectation-maximization (EM)-type iterative algorithm to identify the penalised maximum likelihood estimates and a permutation method to assess p-values. The intensive computational burden induced by the EM-algorithm and the permutation becomes prohibitive for direct applications to GWASs. This paper develops a likelihood ratio test (LRT) for GWASs under genetic heterogeneity based on a more general alternative mixture model. In particular, a closed-form formula for the LRT statistic is derived to avoid the EM-type iterative numerical evaluation. Moreover, an explicit asymptotic null distribution is also obtained, which avoids using the permutation to obtain p-values. Thus, the proposed LRT is easy to implement for GWASs. Furthermore, numerical studies demonstrate that the LRT has power advantages over the commonly used Armitage trend test and other existing association tests under genetic heterogeneity. A breast cancer GWAS dataset is used to illustrate the newly proposed LRT.

Keywords: Association test, binomial mixture model, complex disease, genetic heterogeneity, genome-wide association study

Introduction

Common and complex diseases (or traits) are often genetically heterogeneous in aetiologies (Lander & Schork, 1994; Zhou & Pan, 2009). Some well-known complex diseases with genetic heterogeneity include asthma, breast cancer (Hall et al., 1990; Wooster et al., 1994; Turnbull et al., 2010), and diabetes (Hattersley, 1998; Sladek et al. 2010). As in Zhou & Pan (2009), this paper considers the situation when a complex disease (or trait) is caused by mutations in multiple unlinked loci, commonly referred to as locus heterogeneity (Ott, 1999; Abreu et al., 2002; Fu et al., 2006). As a consequence of genetic heterogeneity, the population of individuals with disease may be decomposed into various latent sub-populations, each with disease caused by mutations at different loci (or their combinations). Most of the existing association tests for population-based case-control studies, e.g. GWAS, are based on comparing the mean genotype scores (e.g. the Armitage trend test (ATT)) between the case and control groups, which are not efficient in the presence of genetic heterogeneity. Zhou & Pan (2009) showed that it can be beneficial to use methods that account for genetic heterogeneity for testing association in a case-control study.

Similar to admixture mapping in linkage analysis (Smith, 1963; Abreu et al., 2002; Fu et al., 2006), Zhou & Pan (2009) proposed a binomial mixture model to account for genetic heterogeneity and developed a modified likelihood ratio test (MLRT) for a single locus (Fu et al., 2006). They also consider two methods to combine single-locus-based MLRTs across multiple loci in linkage disequilibrium to boost power when causal single nucleotide polymorphisms (SNPs) are not genotyped (Zhou & Pan, 2009). They illustrated, with a wide spectrum of numerical examples, that the proposed MLRT tests are more powerful than some commonly used association tests under genetic heterogeneity. Following Zhou and Pan, we define the genetic score $X$ as the number of the minor
alleles at a single locus for a subject. Zhou and Pan (2009) assumed that the genetic score $X_{ij}$ in a healthy control subject follows a binomial distribution, that is

$$P(X_{ij} = g) = B_2(g, \theta_i), \quad g = 0, 1, 2, \quad 0 < \theta_i < 1,$$  

where $B_2(g, \theta_i) = \binom{2}{g} \theta_i^g (1 - \theta_i)^{2-g}$ and $\theta_i$ represents the minor allele frequency (MAF) on that specific locus of the control subject. On the other hand, under genetic heterogeneity, the genetic score for a diseased subject, $X_D$, follows a simple two-component mixture binomial distribution,

$$P(X_D = g) = \alpha_1 B_2(g, \theta_1) + \alpha_2 B_2(g, \theta_2), \quad g = 0, 1, 2;$$

$$0 \leq \alpha_1, \alpha_2 \leq 1, \quad \alpha_1 + \alpha_2 = 1,$$  

where $\theta$ represents the probability of having the minor allele on one chromosome for a sub-group of cases with disease associated with the minor allele. They adopt a two-step procedure for parameter estimation. First, a maximum likelihood estimate (MLE) of $\theta_j$ is obtained based only on the control sample. Then, fixing the estimated $\theta_j$ at its MLE derived from the control-group data, maximum penalised likelihood estimates of other parameters in the mixture model are obtained using an expectation-maximization (EM)-type algorithm (Li et al., 2009). Subsequently, the penalised MLEs from the EM-step are plugged into a likelihood ratio to form a test statistic to detect the association between the marker genotypes and the disease status. Finally, they proposed a permutation procedure to obtain the $p$-value of the association test.

Zhou and Pan’s idea is applicable to an association study for a limited number of candidate markers; however, there are several challenges in applying their proposed method to genome-wide association studies (GWASs). First, the computation of their proposed MLRT for a vast number of SNPs in a typical GWAS would be very intensive. Since the penalised MLEs are obtained by an EM algorithm for maximization of the penalised mixture likelihood, there are known complexities and caveats associated with the EM or other iterative methods for identifying MLEs and penalised MLEs in mixture models including the challenges in selecting multiple starting points for parameter estimation. Moreover, the $p$-value of the MLRT is proposed to be attained by permutation, which is also difficult to apply directly to detect the SNP-disease association in GWAS with a vast number of SNPs, where the significance level is usually set to be less than $10^{-6}$. In addition, it is widely believed that complex diseases and traits are caused by interplays of a large number of genetic loci and environmental risk factors. The simple binomial mixture model with two components in Equation (2) may be too simple to capture the complex heterogeneity for many complex diseases. A more general form of binomial mixture model can be written as follows:

$$P_\eta(X_D = g) = \sum_{j=1}^{J} \alpha_j B_2(g, \theta_j), \quad J \geq 2,$$

$$1 > \theta_j > 0, \quad \sum_{j=1}^{J} \alpha_j = 1, \quad \alpha_j \geq 0,$$  

where $\eta = (\eta_j)_{j \leq J}, \eta_j = (\theta_j, \alpha_j)^T, \quad j = 1, \ldots, J$, and $\theta_j = \theta_j$ if and only if $i = j$. In particular, for many of the complex diseases with genetic heterogeneity, it is likely that $J$ is quite large. Since it is hard to know the number of the subpopulations $J$ under genetic heterogeneity, it is desirable to have a new test that is applicable without the need to know the exact value of $J$ while allowing $J \geq 2$.

In this paper, we developed a LRT for GWASs based on the more flexible binomial mixture models in (3). It is widely believed that complex diseases and traits are caused by interplays of a large number of genetic loci and environmental risk factors. Thus, we assume that the genetic score in the case group, $X_D$, follows a general binomial mixture distribution in (3) which allows the possibility of a large and unknown $J$. The proposed LRT overcomes the above-mentioned challenges of using Zhou and Pan’s method for testing association of a vast number of SNPs in a typical GWAS. In particular, we derived the closed-form formula for the LRT statistic even though the MLEs of parameters in the binomial mixture model are non-regular with loss of identifiability (Liu & Shao, 2003). We further derived the simple closed-form asymptotic null distribution of the LRT which avoids the intensive numerical calculations, such as the EM-based iterations for identification of MLEs and the permutations for evaluation of $p$-values. Additionally, the LRT can be implemented without the requirement of knowing the number of components $J$ in the mixture model (3). We conducted extensive simulation studies to show that the LRT has power advantages over Armitage trend test (ATT) and some other association tests under genetic heterogeneity. We applied our test to a real dataset from a breast cancer GWAS to illustrate that it can achieve a much smaller $p$-value than some commonly used tests when there is evidence of genetic heterogeneity. Thus, the proposed LRT might be used to scan SNPs in GWAS to make novel discoveries by taking account of genetic heterogeneity.

**Method**

**Notation and Setup**

We focus on detecting marker-disease association at a single locus with two alleles $A$ and $a$, such as a SNP in a case-control GWAS. Suppose $m_+$ controls and $n_+$ cases are sampled from

---

© 2013 Blackwell Publishing Ltd/University College London

Annals of Human Genetics (2013) 77,174–182 175
the population. For each SNP, the genotype frequencies in a case-control study can be summarised as in the following 2 × 3 table (Table 1).

Let the genetic score $X_H$ and $X_D$ denote the number of minor alleles, say $a$, at a single locus for a healthy control and a diseased case, respectively. It is clear that $\Sigma X_H = 2m_2 + m_1$, $\Sigma X_D = 2n_2 + n_1$. Similar to Zhou and Pan’s setup, we assume that under the null hypothesis, both $X_H$ and $X_D$ have the same binomial distribution $B_2(g, \theta_h)$ as described in Equation 1. As in Zhou & Pan (2009), $X_H$ is assumed to have a binomial distribution under $H_0$. Under the alternative hypothesis of genetic heterogeneity, we assume that $X_D$ has a mixture distribution as described in Equation 3. This last assumption is worthy of further comments. On the one hand, it is possible to have $J > 2$ in Equation 3 under $H_0$ both in practice and in theory; thus, it is conceptually desirable to allow $J > 2$ in Equation 3. On the other hand, for likelihood inference, it is not necessary to have $J > 2$ in the model in order to achieve the maximum of the likelihood because the model is actually saturated with $J = 2$. In other words, for a given dataset, posing a model (3) with $J = 2$ or with $J \geq 2$, the testing results from the LRT are not going to be different. In fact, as will be seen in the next section, our proposed LRT actually has the “non-parametric” nature because it has a closed-form formula, with a simple null distribution shown to be valid.; thus, it will be valid for testing any alternative models including the common models and those under heterogeneity. In this paper, we will establish that the test is actually a LRT under the specified setup motivated by the elegant work of Zhou & Pan (2009) and by the fact that the LRT has well-known optimality in terms of statistical power and efficiency.

### Mixture Binomial and Maximum Likelihood

Assuming the setup in the previous sub-section, under $H_0$, using the notation in Table 1 and denoting the true value of $\theta_h$ as $P_0$, the MLE of $P_0$ for the overall combined case-control data in Table 1 is

$$
\hat{P}_0 = [\Sigma X_D + \Sigma X_H] / (2n_+ + 2m_+)
= [n_2 + m_2 + (n_1 + m_1)/2] / (n_+ + m_+). \quad (4)
$$

Thus, the binomial likelihood function for the overall combined case-control data evaluated at $\hat{P}_0$, $L_0$, is given by

$$
L_0 = \prod_{g=0}^{2} B_2(g, \hat{P}_0)^{n_g + m_g}, \quad (5)
$$

where $B_2(g, \hat{P}_0) = (\hat{P}_0 g (1 - \hat{P}_0))^{2-g}$ and $\hat{P}_0$ is defined in (4). Following Zhou & Pan (2009), in the control group, the genetic score $X_H$ is assumed to follow a binomial distribution under the alternative hypothesis, say

$$
P(X_H = g) = B_2(g, P_{hi}), \quad g = 0, 1, 2. \quad (6)
$$

Using the notation in Table 1, the MLE of $P_{hi}$ within the healthy control group only is given by

$$
\hat{P}_i = \Sigma X_{H_i} / (2m_+) = (m_2 + m_1/2) / m_+. \quad (7)
$$

The binomial likelihood function of the healthy controls data evaluated at $\hat{P}_i$, $L_i$, is given by

$$
L_i = \prod_{g=0}^{2} B_2(g, \hat{P}_i)^{n_g}. \quad (8)
$$

Similarly, in the case group, if the genetic score $X_D$ has the distribution $B_2(g; \hat{P}_D)$, the MLE $\hat{P}_D$ of $P_D$ within the diseased case group only would be

$$
\hat{P}_D = \Sigma X_{D_i} / (2n_+) = (n_2 + n_1/2) / n_. \quad (9)
$$

However, as in Zhou & Pan (2009), we assume that under genetic heterogeneity, the cases can be divided into multiple latent sub-populations. Thus, under the alternative hypothesis of genetic heterogeneity, we assume that $X_D$ has a mixture distribution as described in Equation 3. It can be shown that (see Appendix I), using the above notation, the maximum of the mixture likelihood for $X_D$ has an explicit formula:

$$
L_D = \sup_{\eta} \prod_{g=0}^{2} P_\eta(X_D = g)^{n_g}
= \left\{\begin{array}{ll}
\prod_{g=0}^{2} (n_g / n_+)^{n_g} \quad & \text{if } 4n_0n_2 > n_1^2; \\
\prod_{g=0}^{2} B_2(g; \hat{P}_D)^{n_g} \quad & \text{if } 4n_0n_2 \leq n_1^2.
\end{array}\right. \quad (10)
$$

The derivation of the above equation can be found in Appendix I. It is also clear from the derivation in Appendix I that the mixture likelihood function of the parameter vector $\eta = (\theta_j, \alpha_j)_{1 \leq j \leq J}$ in the mixture model (3) can have many local maxima due to the lack of identifiability in parameters (Liu & Shao, 2003). Nevertheless, the supremum of the mixture likelihood $L_D$ for $X_D$ has a single unique value for each dataset and can be obtained from the explicit formula in Equation (10). In the typical case-control study design, $L_D$ is independent of $L_i$.

### Table 1: The genotype frequencies for case-control data of a SNP.

|     | $AA$ | $aA$ | $aa$ | Total |
|-----|------|------|------|-------|
| Case | $n_0$ | $n_1$ | $n_2$ | $n_+$ |
| Control | $m_0$ | $m_1$ | $m_2$ | $m_+$ |

$p(X_H = g) = B_2(g, P_{hi}), \quad g = 0, 1, 2.$
The LRT

Using the maximum of the likelihood $L_0$, $L_I$ and $L_D$ in Equations 5, (8) and (10), respectively, we can write down the explicit formula of the log-LRT statistic as follows:

$$2\lambda_N = 2[\log(L_D L_I) - \log L_0].$$  

(11)

No iterative numerical maximization of the mixture likelihood function is needed for the evaluation of the LRT statistic in (11). Thus, the LRT statistic is easy to compute even for GWAS. It is known that the LRT statistics for testing homogeneity in mixture models often have complicated asymptotic distributions that typically lack closed-form representations. However, the above statistic $2\lambda_N$ can be shown to have an explicit form of asymptotic distribution under the null hypothesis. More specifically, under $H_0$, as $n_+ \to \infty$ and $m_+ \to \infty$, we have

$$2\lambda_N \to \frac{1}{2} \chi_1^2 + \frac{1}{2} \chi_3^2,$$  

(12)

where $\chi_d^2$ denotes a chi-squared distribution with $d$ degrees of freedom, $d = 1, 2$. Although the above asymptotic null distribution can be derived from general results such as those in Chernoff & Lander (1995), Chiano & Yates (1995) or Liu & Shao (2003), an elementary and detailed direct derivation of the above asymptotic null distribution is given in Appendix II for readers who are interested in a direct derivation based on first principles.

It is worth pointing out that our extensive numerical simulations discussed in the next section indicate that the simple asymptotic null distribution in (12) approximates the exact finite sample null distribution very well. The asymptotic formula is only slightly conservative. Therefore, the $p$-values of the LRT can be easily read off from the above simple closed-form asymptotic null distribution. For example, given any observed data in Table 1, one can first evaluate the value of $2\lambda_N$ in (11) and then can obtain the $p$-value using the following simple command in the widely used R-platform:

```r
{pchisq(2\lambda_N, 1, lower.tail = F) + pchisq(2\lambda_N, 2, lower.tail = F)}/2
```

Last but not least, it is well known that the LRT generally has better power than other ad hoc tests. Thus, it should not be a surprise to see that the LRT can be more powerful than other commonly used tests which ignore the genetic heterogeneity that exists for many common complex diseases such as breast cancer. Finally, to implement the LRT, there is no need to identify the exact number of mixture components $J$ in (3), which is desirable because $J$ is hard to determine in practice.

### Numerical Results

#### Type I Errors

The LRT has an explicit asymptotic distribution under $H_0$. Consequently, it is convenient to evaluate the $p$-value and type I errors. We conducted comprehensive simulations to compare the empirical type I error of the LRT to the nominal significance level ranging from $10^{-2}$ to $10^{-8}$. In the Monte Carlo simulations, the genotype data for both the control group and the case group were generated from the same binomial distribution $B_s(g; \theta_0)$, where $\theta_0$ takes some fixed value $P_0$, which represents the MAF. A number of simulation setups, which varied over a range of MAF and sample size, were selected. The control and case sample sizes are set to be equal. The nominal significance levels were taken to be $10^{-2}$, $10^{-3}$, $10^{-4}$, $10^{-5}$, $10^{-6}$, $10^{-7}$ and $10^{-8}$, respectively. For each setup, $10^{11}$ samples were generated. We found that the empirical type I error is slightly smaller than the nominal level, but they are extremely close to each other. Thus, using the asymptotic null distribution for the LRT is valid. For illustration, an example with $\theta_0 = 0.4$ and sample size $n_+ = m_+ = 1000$ is shown in Table 2.

#### Power Comparison

The significance level of the association test is usually set very small for GWASs. For example, the genome-wide significance level of $5 \times 10^{-8}$ is being increasingly used for arrays that contain one million SNPs. The most commonly used association tests for GWAS include ATT and the $\chi_2^2$ test, both applicable for testing association in a $2 \times 3$ table between the case-control status and the three genotypes, as illustrated in Table 1. Accordingly, we designed simulation studies to evaluate and compare the powers of the LRT, the ATT and the $\chi_2^2$ test when the significance level is set to be $5 \times 10^{-8}$. Note that, the MLRT of Zhou & Pan (2009) was not included for comparison due to its severe computational challenge when the significance level is very small. In the first set of Monte Carlo simulations, the control sample was generated from a binomial distribution $B_s(g, \theta_0)$; the case sample was generated from a two-component mixture binomial distribution as

| Table 2 | Empirical type I error and nominal significance level at $\theta_0 = 0.4$ and $n_+ = m_+ = 1000$. |
|----------|---------------------------------------------------------------|
| Nominal level | 0.01 | 0.001 | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ | $10^{-7}$ | $10^{-8}$ |
| Empirical level | 0.0098 | 0.00099 | $9.7 \times 10^{-5}$ | $9.8 \times 10^{-6}$ | $9.7 \times 10^{-7}$ | $9.5 \times 10^{-8}$ | $9.8 \times 10^{-9}$ |
Table 3 Empirical power* when $X_D$ has a mixture distribution with $J = 2$.

| Setup | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| $\theta_0$ | 0.1 | 0.1 | 0.2 | 0.2 | 0.1 | 0.1 | 0.3 | 0.3 |
| $\theta_1$ | 0.12 | 0.08 | 0.18 | 0.23 | 0.20 | 0.30 | 0.32 | 0.28 |
| $\theta_2$ | 0.50 | 0.50 | 0.60 | 0.60 | 0.70 | 0.70 | 0.70 | 0.70 |
| $\alpha_1$ | 0.90 | 0.85 | 0.80 | 0.9 | 0.8 | 0.9 | 0.9 | 0.8 |
| $\alpha_2$ | 0.10 | 0.15 | 0.20 | 0.1 | 0.2 | 0.1 | 0.1 | 0.2 |
| N    | 1000 | 1000 | 1500 | 1000 | 1500 | 1000 | 1500 | 1500 |
| Power | LRT | 0.717 | 0.840 | 0.987 | 0.833 | 0.997 | 0.829 | 0.673 |
|       | ATT | 0.439 | 0.059 | 0.576 | 0.717 | 0.087 | 0.756 | 0.490 |
|       | $\chi^2_2$ | 0.399 | 0.150 | 0.810 | 0.687 | 0.718 | 0.709 | 0.484 |

*Significance level is set at $5 \times 10^{-8}$.

Table 4 Empirical power* when $X_D$ has a mixture distribution with $J = 3$.

| Setup | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| $\theta_0 = \theta_3$ | 0.1 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 | 0.3 |
| $\theta_1$ | 0.13 | 0.15 | 0.15 | 0.1 | 0.25 | 0.22 | 0.2 | 0.33 |
| $\theta_2$ | 0.5 | 0.5 | 0.4 | 0.6 | 0.6 | 0.6 | 0.7 | 0.7 |
| $\alpha_1$ | 0.35 | 0.4 | 0.3 | 0.2 | 0.3 | 0.35 | 0.4 | 0.4 |
| $\alpha_2$ | 0.15 | 0.1 | 0.2 | 0.2 | 0.1 | 0.15 | 0.2 | 0.15 |
| $\alpha_3$ | 0.5 | 0.5 | 0.5 | 0.6 | 0.6 | 0.5 | 0.4 | 0.45 |
| N    | 800  | 1200 | 800  | 1000 | 2000 | 1500 | 1500 | 1500 |
| Power | LRT | 0.883 | 0.903 | 0.843 | 0.829 | 0.846 | 0.937 | 0.864 |
|       | ATT | 0.554 | 0.707 | 0.713 | 0.125 | 0.612 | 0.696 | 0.011 |
|       | $\chi^2_2$ | 0.528 | 0.683 | 0.641 | 0.32 | 0.64 | 0.757 | 0.274 |

*Significance level is set at $5 \times 10^{-8}$.

described in (3) with $J = 2$:

$$P_q(X_D = k) = \sum_{j=1}^{2} \alpha_j B_2(k, \theta_j), \quad 0 \leq \alpha_1, \alpha_2 \leq 1, \quad \alpha_1 + \alpha_2 = 1.$$ 

Twenty thousand replicate datasets of $n = m = N$ controls and cases were simulated for each of the eight simulation setups and the empirical power for each test is shown in Table 3. The simulation results indicate that the LRT has power advantage over the ATT and the $\chi^2_2$ test under genetic heterogeneity.

Similar power advantages of the LRT over other tests are also observed when the alternative mixture model has three components ($J = 3$), as demonstrated in Table 4, where $\theta_3$ for the cases is set as equal to $\theta_0$ for the control group.

Note that the ATT, also called Cochran–Armitage trend test (CATT) by many researchers, has good power only when the disease risk of the genotypes $AA$, $Aa$ and $aa$ is monotone increasing or decreasing under the alternative hypothesis (Armitage, 1955; Freidlin et al., 2002). Thus, ATT can have very low power when there is a violation of a linear trend in the disease risk across the ordered genotypes $AA$, $Aa$ and $aa$, as in the case of both setups #3 and #5 in Table 3.

It is clear from the power simulations across the multiple simulation setups that the LRT can be much more powerful than the commonly used ATT and the $\chi^2_2$ test in GWAS in the presence of genetic heterogeneity.

A Breast Cancer GWAS

Breast cancer is the most common cancer among women. Many genes on different chromosomes that underlie breast cancer have been identified including many well-known studies conducted two decades ago (Hall et al., 1990; Wooster et al., 1994). Many more genetic variants underlying breast cancer are still being discovered nowadays; thus, there is little doubt about the existence of genetic heterogeneity in the case of breast cancer. For illustration, we applied the newly proposed LRT to a breast cancer GWAS dataset. In particular, Turnbull et al. (2010) conducted a GWAS to identify breast cancer susceptibility alleles. They studied 582,886 SNPs in 3659 breast cancer cases and 4897 controls in the first stage, and evaluated promising SNPs that were identified in stage 1 in a second stage with 12,576 cases and 12,223 controls. In the paper, they reported five new susceptibility SNPs with summary genotype data of the five SNPs made publicly available. A literature search indicates that four of the five SNPs (rs1011970, rs10995190, rs704010 and rs614367) have been independently confirmed by other studies since the publication of their GWAS results in 2010 (Peng et al., 2011; Lambrechts et al., 2012). We evaluated the $p$-values of the LRT, ATT and $\chi^2_2$ test for these four SNPs for comparison. The results are summarised in Table 5.

Note that for the SNP rs10995190 and SNP rs614367, the $p$-values are smaller than the genome-wide significance level $5 \times 10^{-8}$ for the newly proposed LRT and the ATT, and for each of the two stages. The performance of the LRT is as good as or better than the other two tests. In particular, the LRT has an extremely small $p$-value $6 \times 10^{-15}$ for stage 2 data of SNP rs614367 showing statistical significance at even lower levels.

It is thus not surprising that these SNPs are independently replicated by other GWAS. For the SNP rs704010 and SNP 1011970, a simple combined $p$-value (for combining the two stages), e.g. Fisher's $\chi^2$ test, indicates that both SNPs are significant even using the genome-wide significance level $5 \times 10^{-8}$ for all three tests. The newly proposed LRT is also very competitive for these two SNPs. For example, for the stage 1 data of SNP rs704010, only the $p$-value of the LRT is smaller than the genome-wide significance level $5 \times 10^{-8}$.

As an indication of overall strength of the test, Fisher's meta
Discussion

In the analysis of GWAS data, potential latent genetic heterogeneity has largely been ignored by researchers. Zhou & Pan (2009) first proposed mixture models to account for genetic heterogeneity. However, for the analysis of a vast number of SNPs in GWAS, the MLRT of Zhou and Pan has major computational challenges. In this paper, using a more general binomial mixture model, we have derived a LRT for case-control association studies that improve the MLRT by Zhou and Pan on computational efficiency and multiple other aspects. In particular, the LRT statistic has a simple closed-form formula, which could avoid intensive computation, such as the EM algorithm for penalised MLEs. Additionally, we have derived an explicit asymptotic null distribution for the proposed LRT, which is convenient to obtain \( p \)-values even at a small significance level. Moreover, to perform the LRT, there is no need to decide the exact number of mixture components, which is convenient in practice. Therefore, the new LRT has computational advantages over the MLRT proposed by Zhou and Pan and is suitable for scanning SNPs in GWAS data.

As demonstrated by our numerical studies, in the presence of genetic heterogeneity, the LRT can be much more powerful than either Armitage’s trend test or the \( \chi^2 \) test, both of which are among the most widely used tests in GWAS. Given that most complex diseases are widely believed to be polygenic and have environmental components, genetic heterogeneity is a hallmark of complex diseases. As illustrated using the GWAS data for breast cancer, newly proposed LRT can be easily used for any GWAS data; thus, researchers can use the simple algorithm to scan their SNPs as a cost-effective way to potentially make novel and important discoveries using existing data already collected in the large number of GWAS. Given that there are already about 1000 published GWAS, and many more genome-wide studies are being planned and conducted, the new LRT has the potential to become one of the useful tests to scan the SNPs in these GWAS, maybe as a secondary analysis to account for genetic heterogeneity. Thus, the new user-friendly LRT can potentially be used to increase the impact of existing and future GWASs.

Acknowledgements

This research is partially supported by the NIH Cancer Center Supporting Grant to NYU (2P30 CA16087), and the NIEHS Center Grant to NYU (5P30 ES00260), as well as a Stony Wold-Herbert Foundation grant to YS. The authors have no conflict of interest. The authors would like to thank the reviewers for insightful suggestions that lead to improvement of the paper.

References

Abreu, P. C., Hodge, S. E. & Greenberg, D. A. (2002) Quantification of type I error probabilities for heterogeneity lod scores. *Genet Epidemiol* 22, 156–169.

Armitage, P. (1955) Tests for linear trends in proportions and frequencies. *Biometrics* 11, 375–386.

Bickel, P. J. & Doksum, K. A. (2000) *Mathematical Statistics: Basic Ideas and Selected Topics*, Vol I. New Jersey: Prentice Hall.

Chernoff, H. & Lander, E. (1995) Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *J Statist Plann Inference* 43, 19–40.

Chiano, M. N. & Yates, J. R. W. (1995) Linkage detection under heterogeneity and the mixture problem. *Ann Hum Genet* 59, 83–95.

Emigh, T. H. (1980) A comparison of tests for Hardy–Weinberg equilibrium. *Biometrics* 36, 627–642.

Freidlin, B., Zheng, G., Li, Z. & Gastwirth, J. L. (2002) Trend tests for case-control studies of genetic markers: Power, sample size and robustness. *Hum Hered* 53, 146–152.

Fu, Y., Chen, J. & Kalbfleisch, J. D. (2006) Testing for homogeneity in genetic linkage analysis. *Stat Sinica* 16, 805–823.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B. & Kang, M. C. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250, 1684–1689.

Table 5 Comparison of \( p \)-values for the four SNPs reported in Turnbull et al. (2010).

| SNP       | Stage | \( p \)-values | ATT \( p \)-values | \( \chi^2 \) test \( p \)-values |
|-----------|-------|----------------|-------------------|---------------------------------|
| rs10995190| Stage 1 | 7 \times 10^{-9} | 5 \times 10^{-9} | 3 \times 10^{-8} |
|           | Stage 2 | 2 \times 10^{-8} | 10^{-8} | 2 \times 10^{-8} |
| Fisher \( p \)-value | 5 \times 10^{-15} | 2 \times 10^{-15} | 2 \times 10^{-14} |
| rs614367  | Stage 1 | 6 \times 10^{-10} | 2 \times 10^{-10} | 3 \times 10^{-10} |
|           | Stage 2 | 6 \times 10^{-15} | 10^{-8} | 6 \times 10^{-8} |
| Fisher \( p \)-value | 2 \times 10^{-22} | 10^{-16} | 7 \times 10^{-16} |
| rs704010  | Stage 1 | 3 \times 10^{-8} | 3 \times 10^{-6} | 7 \times 10^{-7} |
|           | Stage 2 | 7 \times 10^{-4} | 3 \times 10^{-4} | 4 \times 10^{-4} |
| Fisher \( p \)-value | 5 \times 10^{-10} | 2 \times 10^{-8} | 6 \times 10^{-9} |
| rs1011970 | Stage 1 | 3 \times 10^{-5} | 9 \times 10^{-6} | 5 \times 10^{-5} |
|           | Stage 2 | 7 \times 10^{-5} | 2 \times 10^{-4} | 4 \times 10^{-4} |
| Fisher \( p \)-value | 4 \times 10^{-8} | 4 \times 10^{-8} | 4 \times 10^{-7} |
Hattersley, A. T. (1998) Maturity-onset diabetes of the young: Clinical heterogeneity explained by genetic heterogeneity. Diabet Med 15, 15–24.

Lambrechts, D., Truong, T., Justenhoven, C., Humphreys, M. K., Wang, J., Hopper, J. L., Dite, G. S., Apicella, C., Southey, M. C., Schmidt, M. K., Broeks, A., Cornelissen, S., Hien, R. V., Sawyer, E., Tomlinson, I. P., Kerin, M., Miller, N., Milne, R. L., Zamora, M. P., Perez, J. I., Benitez, J., Hamann, U., Ko, Y. D., Bruning, T., Chang-Claude, J., Elipher, U., Hein, R., Nicholas, S., Flesch-Jansky, D., Wang-Gohrke, S., John, E. M., Miron, A., Wingquist, R., Pulkas, J., Jukkola-Vuorinen, A., Grip, M., Chenexiv-Trench, G., Beesley, J., Chen, X. Q., Menegaux, F., Cordina-Duverger, E., Shen, C. Y., Yu, J. C., Wu, P. E., Hou, M. F., Andrilis, I. L., Selander, T., Glendon, G., Mulligan, A. M., Anton-Culver, H., Ziogas, A., Muir, K. R., Lophathanonan, A., Rattanamongkong, S., Puttawibil, P., Jones, J. M., Orr, N., Ashworth, A., Swedlow, A., Severi, G., Baglatto, L., Giles, G. G., Marne, E., Schneweiss, A., Sohn, C., Burwinkel, B., Yusuf, B. T., Neven, P., Paridaens, R., Wilders, H., Brenner, H., Muller, H., Arndt, V., Stegmaier, C., Meindl, A., Schott, S., Barrett, C. R., Schnurzler, R. K., Cox, A., Brock, I. W., Elliott, G., Cross, S. S., Fasching, P. A., Schulz-Wendland, R., Eki, A. B., Beckmann, M. W., Fletcher, O., Johnson, N., dos Santos Silva, I., Petu, J., Nevanlinna, H. A., Muranen, T. A., Attmaki, K., Blomqvist, C., Dork, T., Schurrman, P., Bremer, M., Hilleman, P., Bogdanova, N. V., Antononkova, N. N., Rogov, Y. I., Karstens, J. J., Khusnardinova, E., Bershshireva, M., Prokofeva, D., Gancev, S., Jakubowska, A., Lubinski, J., Jaworska, K., Durda, K., Nordestgaard, B. G., Bojesen, S. E., Lang, C., Manneuoa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Radice, P., Petterlono, P., Manoukian, S., Bernard, L., Lee, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Radice, P., Peterlongo, P., Manoukian, S., Bernard, L., Couch, F. J., Olson, J. E., Wang, X., Fredericken, Z. S., Aalnæs, G. G., Kristensen, V., Borresen-Dale, A. L., Devilee, P., Tollenraa, R. A., Seynaeve, C. M., Hoonging, M. J., Garcia-Closas, M., Chanocek, S. J., Lisowski, J., Sherman, M. E., Hall, P., Liu, J. J., Czene, K., Kang, D., Yoo, K. Y., Noh, D. Y., Lindblom, A., Margolin, S., Dunning, A. M., Pharoah, P. D., Easton, D. F., Guenel, P. & Brauch, H. (2012) 11q13 is a susceptibility locus for hormone receptor positive breast cancer. Hum Mutat 33, 1123–1132.

Lander, E. S. & Schork, N. J. (1994) Genetic dissection of complex traits. Science 265, 2037-2048.

Li, P., Chen, J. H. & Marriott, P. (2009) Non-finite Fisher information and homogeneity: An EM approach. Biometrika 96, 411-426.

Liu, X. & Shao, Y. (2003) Asymptotics for likelihood ratio tests under loss of identifiability. Ann Stat 31, 807–832.

Ott, J. (1999) Analysis of Human Genetic Linkage, 3rd edition. Baltimore: The John Hopkins University Press.

Peng, S., Li, B., Ruan, W., Zhu, Y., Sheng, H. & Li, M. (2011) Genetic polymorphisms and breast cancer risk: Evidence from meta-analyses, pooled analyses, and genome-wide association studies. Breast Cancer Res Treat 127, 309–324.

Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hajdaj, S., Balkau, B., Heude, B., Charpentier, G., Hudson, T. J., Montpetit, A., Pfehthersky, A. V., Prenki, M., Posner, B. F., Balding, D. J., Meyer, D., Polychronkous, C. & Frugel, P. (2010) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445, 881–885.

Smith, C. A. (1963) Testing for heterogeneity of recombination fraction values in Human Genetics. Ann Hum Genet 27, 175–182.

Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maraman, M., Seal, S., Ghoussamon, M., Hines, S., Healey, C. S., Hughes, D., Warren-Perry, M., Tapper, W., Eccles, D., Evans, D. G., Breast Cancer Susceptibility Collaboration (UK), Hoonging, M., Schutte, M., van den Ouweland, A., Houlston, R., Ross, G., Langford, C., Pharoah, P. D., Stratton, M. R., Dunning, A. M., Rahman, N. & Easton, D. F. (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 42, 504–507.

Wooster, R., Neuhausen, S. L., Mangion, J., Quark, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., Fields, P., Marshall, G., Narod, S., Lennor, G. M., Lynch, H., Feunteun, J., Devilee, P., Cornelisse, C. J., Menko, F. H., Daly, P. A., Ormiston, W., McManus, R., Pyy, C., Lewis, C. M., Cannon-Albright, L. A., Peto, J., Ponder, B. A. J., Skolnick, M. H., Easton, D. F., Goldgar, D. E. & Stratton, M. R. (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 265, 2088–2090.

Zhou, H. & Pan, W. (2009) Binomial mixture model-based association tests under genetic heterogeneity. Ann Hum Genet 73, 614–630.

Appendix I

Derivation of the Test Statistic of the LRT

To prove our proposed association test is indeed a LRT under the given setup, we just need to establish Equation 10, that is, when \( X_D \) follows the mixture distribution in (3),

\[
L_D = \sup_{\eta} \left\{ \prod_{g=0}^{2} P_{\eta}(X_D = g)^{\nu} \right\} = \left\{ \prod_{g=0}^{2} \left( \frac{n_g}{n_+} \right)^{\nu} \right\} \text{ if } 4n_0n_2 > n_1^2;
\]

\[
B_2(\hat{p}_D; \nu) \text{ if } 4n_0n_2 \leq n_1^2;
\]

where \( \hat{p}_D \) is defined as in Equation 9. First, we want to show that when \( 4n_0n_2 > n_1^2 \), the MLEs \( \hat{\eta} \) of the \( \eta \) in (3) satisfy

\[
L_D = \prod_{g=0}^{2} P_{\hat{\eta}}(X_D = g)^{\nu} = \prod_{g=0}^{2} \left( \frac{n_g}{n_+} \right)^{\nu}.
\]

A simple application of Jensen’s inequality yields that, for any \( \eta \),

\[
\log \prod_{g=0}^{2} P_{\eta}(X_D = g)^{\nu} \leq \log \prod_{g=0}^{2} \left( \frac{n_g}{n_+} \right)^{\nu}.
\]

The right-hand side of the above inequality is an upper bound which may not be achievable in general. However, when \( 4n_0n_2 > n_1^2 \), we can show that the equality in (13) is achievable. In fact, when \( 4n_0n_2 > n_1^2 \), there are infinitely many values of the MLE \( \hat{\eta} \) that can make (13) an equality. It is straightforward and elementary to verify that one set of
solutions for MLE is given as follows:
\[
\hat{\theta}_1 \in \left( \frac{2n_2}{2n_2 + n_1}, 1 \right),
\]
\[
\hat{\theta}_2 = \frac{(n_1 + 2n_2)\hat{\theta}_1 - 2n_2}{2n_2\hat{\theta}_1 - n_1 - 2n_2},
\]
\[
\hat{\alpha}_1 = \frac{4n_0n_2 - n_1^2}{2n_2(2n_1\hat{\theta}_1^2 - 2(n_1 + 2n_2)\hat{\theta}_1 + 2n_2)},
\]
\[
\hat{\alpha}_2 = 1 - \hat{\alpha}_1,
\]
\[
\hat{\alpha}_i = 0, \quad \hat{\theta}_i = 1/2, \quad \text{for}\ i \neq 1, 2.
\]
As indicated above, \(\hat{\theta}_1\) can take any values in an interval; thus, there are infinitely many sets of solutions for the MLE. Thus, Equation 13 is proved. Next, we show that when \(4n_0n_2 \leq n_1^2\),
\[
L_D = \sup_{\eta} \prod_{g=0}^{2} P_\eta(X_D = g)^{\eta_g} = \prod_{g=0}^{2} B_2(g; \hat{p}_D)^{\eta_g}.
\]
(15)
First, we show that for any fixed \(\eta\),
\[
\prod_{g=0}^{2} P_\eta(X_D = g)^{\eta_g} \leq \prod_{g=0}^{2} B_2(g; \hat{p}_D)^{\eta_g}.
\]
Using the inequality \(\log x \leq x - 1\), we get
\[
\sum_{g=0}^{2} \eta_g \log P_\eta(X_D = g) \leq \sum_{g=0}^{2} \eta_g \left( \frac{P_\eta(X_D = g)}{B_2(g; \hat{p}_D)} - 1 \right).
\]
It is straightforward to verify that
\[
\sum_{g=0}^{2} \eta_g \left( \frac{P_\eta(X_D = g)}{B_2(g; \hat{p}_D)} - 1 \right) = \frac{n_1(4n_0n_2 - n_1^2)}{(2n_0 + n_1)(2n_2 + n_1)^2} + \sum_{j=0}^{1} \hat{\alpha}_j(2n_2 + n_1 - 2n_2\hat{\theta}_j)^2.
\]
Therefore, when \(4n_0n_2 \leq n_1^2\), and for any \(\eta\)
\[
\log \prod_{g=0}^{2} P_\eta(X_D = g)^{\eta_g} \leq \log \prod_{g=0}^{2} B_2(g; \hat{p}_D)^{\eta_g}.
\]
Finally, it is obvious that
\[
L_D = \sup_{\eta} \prod_{g=0}^{2} P_\eta(X_D = g)^{\eta_g} \geq \prod_{g=0}^{2} B_2(g; \hat{p}_D)^{\eta_g}.
\]
This finishes the proof (15), thus also (10).

**Appendix II**

**The Asymptotic Null Distribution of the LRT**

Under \(H_0\), both \(X_D\) and \(X_H\) have the same binomial distribution \(B_2(g; \theta_0)\). We denote the true null value for \(\theta_0\) as \(P_0\). Without loss of generality, we assume \(0 < P_0 < 1\) to avoid \(P_0(1 - P_0) = 0\) appearing in any denominator. First, we may consider testing \(H_0 : B_2(g; P_0)\) against \(H_1 : B_2(g; \theta_0), \theta_0 \in (0, 1)\), using only the healthy controls. This is a classic problem, the LRT statistic is well known to have a \(\chi^2\) distribution. It is well known that under \(H_0 : B_2(g; P_0)\), the LRT statistic can be written as
\[
2 \sum_{g=0}^{2} m_2 \log \frac{B_2(g; \hat{p}_H)}{B_2(g; P_0)} = 2m_2 \frac{(\hat{p}_H - P_0)^2}{P_0(1 - P_0)} + o_\rho(1),
\]
(16)
where \(\hat{p}_H = (n_2 + m_1 + 2)/n_2 + m_1\) is the MLE of \(\theta_0\) using only the healthy controls. Similarly, we may consider testing \(H_0 : B_2(g; P_0)\) against \(H_1 : B_2(g; \theta_0), \theta_0 \in (0, 1)\), using only the diseased cases. Then, under \(H_0 : B_2(g; P_0)\), the LRT statistic has \(\chi^2\) distribution and can be written as
\[
2 \sum_{g=0}^{2} m_2 \log \frac{B_2(g; \hat{p}_D)}{B_2(g; P_0)} = 2m_2 \frac{(\hat{p}_D - P_0)^2}{P_0(1 - P_0)} + o_\rho(1),
\]
(17)
where \(\hat{p}_D = (n_2 + n_1 + 2)/n_2 + n_1\) is the MLE of \(\theta_0\) using only the diseased cases. Similarly, we may consider testing \(H_0 : B_2(g; P_0)\) against \(H_1 : B_2(g; \theta_0), \theta_0 \in (0, 1)\), using the overall sample combining both the diseased cases and health controls. Then, the MLE for \(\theta_0 = P_0\) from the combined sample is \(\hat{p}_0\) as defined in (4). The LRT statistic can be written as
\[
2 \sum_{g=0}^{2} \left[ m_2 \log \frac{B_2(g; \hat{p}_H)}{B_2(g; P_0)} + m_2 \log \frac{B_2(g; \hat{p}_D)}{B_2(g; P_0)} \right] = 2(m_2 + n_2) \frac{(\hat{p}_0 - P_0)^2}{P_0(1 - P_0)} + o_\rho(1).
\]
(18)
From the above three equations, and the Equations (5), (8) and (10), we have, when \(4n_0n_2 < n_1^2\),
\[
2 \log \frac{L_DL_H}{L_0} = 2m_2 \frac{(\hat{p}_H - P_0)^2}{P_0(1 - P_0)} + 2n_2 \frac{(\hat{p}_D - P_0)^2}{P_0(1 - P_0)} - 2(m_2 + n_2) \frac{(\hat{p}_0 - P_0)^2}{P_0(1 - P_0)} + o_\rho(1).
\]
(19)
Denote \(\rho = n_2/(m_2 + n_2)\). Then, it is straightforward to verify that
\[
\hat{p}_0 - P_0 = \rho(\hat{p}_D - P_0) + (1 - \rho)(\hat{p}_H - P_0).
\]
and
\[ 2 \log \frac{L_D L_H}{L_0} = (1 - \rho) \frac{2n_+ (\hat{p}_D - P_0)^2}{P_0(1 - P_0)} + \rho \frac{2m_+ (\hat{p}_H - P_0)^2}{P_0(1 - P_0)} \]
\[ -2\sqrt{\rho(1 - \rho)} \sqrt{2n_+ (\hat{p}_D - P_0)} \frac{\sqrt{2m_+ (\hat{p}_H - P_0)}}{\sqrt{P_0(1 - P_0)}} + o_\rho(1). \]

Denote
\[ Z_H = \frac{\sqrt{2m_+ (\hat{p}_H - P_0)}}{\sqrt{P_0(1 - P_0)}} \]
and
\[ Z_D = \frac{\sqrt{2n_+ (\hat{p}_D - P_0)}}{\sqrt{P_0(1 - P_0)}}. \]

Then,
\[ 2 \log \frac{L_D L_H}{L_0} = (\sqrt{\rho} Z_H + \sqrt{1 - \rho} Z_D)^2 + o_\rho(1). \]

(20)

Note that \( Z_H \sim N(0, 1) \) and \( Z_D \sim N(0, 1) \) and \( Z_H \) and \( Z_D \) are independent. Thus,
\[ \sqrt{\rho} Z_H + \sqrt{1 - \rho} Z_D \sim N(0, 1). \]

Therefore, when \( 4n_0n_2 \leq n_1^2 \), we have \( 2\lambda_N = 2(\log L_D + \log L_H - \log L_0) \sim \chi^2_2. \)

On the other hand, under \( H_0 \), when \( 4n_0n_2 > n_1^2 \), we can first consider testing goodness-of-fit of \( H_0 : B_2(g; P_0) \) using only the diseased cases. The LRT statistic has a \( \chi^2_2 \) asymptotic distribution and can be written as
\[ \frac{2}{z = 0} n_z \log \frac{n_z / n_+}{B_2(g; P_0)} = \frac{2}{z = 0} n_z \log \frac{n_z / n_+}{B_2(g; \hat{P}_D)} \]
\[ + 2 \sum_{z = 0} n_z \log \frac{B_2(g; \hat{P}_D)}{B_2(g; P_0)} = 2n_+ \frac{(n_z / n_+ - \hat{p}_D)^2}{\hat{P}_D(1 - \hat{P}_D)} \]
\[ + 2n_+ \frac{(\hat{p}_D - P_0)^2}{P_0(1 - P_0)} + o_\rho(1). \]

(21)

The first term at the right-hand side of the last equality is equivalent to the Pearson's classic \( \chi^2 \) statistic (via comparing observed to expected cell frequencies) for testing Hardy–Weinberg equilibrium which is well known to have the \( \chi^2_1 \) distribution (Emigh, 1980). Using the above equations, when \( 4n_0n_2 > n_1^2 \), we have
\[ 2 \log \frac{L_D L_H}{L_0} = 2n_+ \frac{(\hat{p}_H - P_0)^2}{P_0(1 - P_0)} + 2n_+ \frac{(\hat{p}_D - P_0)^2}{P_0(1 - P_0)} \]
\[ -2\sqrt{\rho(1 - \rho)} \sqrt{2n_+ (\hat{p}_H - P_0)} \frac{\sqrt{2m_+ (\hat{p}_H - P_0)}}{\sqrt{P_0(1 - P_0)}} + o_\rho(1). \]

(22)

By Equations (19) and (20), from the above equation, we have
\[ 2 \log \frac{L_D L_H}{L_0} = (\sqrt{\rho} Z_H + \sqrt{1 - \rho} Z_D)^2 \]
\[ + 2n_+ \frac{(n_z / n_+ - \hat{p}_D)^2}{\hat{P}_D(1 - \hat{P}_D)} + o_\rho(1). \]

(23)

Note that the two terms in the right-hand side of (21) are well known to be asymptotically independent which, in turn, implies asymptotic independence of the two terms at the right-hand side of (23). Therefore, when \( 4n_0n_2 > n_1^2 \), we have
\[ 2\lambda_N = 2 \log \frac{L_D L_H}{L_0} = 2(\log L_D + \log L_H - \log L_0) \sim \chi^2_2. \]

Finally, it suffices to show that \( P(4n_0n_2 > n_1^2 | H_0) \rightarrow 1/2 \) as \( n_2 \rightarrow \infty \). Note that under \( H_0 \), \( (n_0, n_1, n_2) \) follow a multinomial distribution \( (n_0, \pi_0, \pi_1, \pi_2) \), where \( \pi_0 = P(X_D = g) \), for \( g = 0, 1, 2 \). Let \( U^T \) be the random vector \( (\frac{n_0}{n_0 + n_2}, \frac{n_1}{n_1 + n_2}, \frac{n_2}{n_2}) \). Then, we have (Bickel & Docksum, 2000)
\[ E(U^T) = \Pi = (\pi_0, \pi_1, \pi_2), \]
\[ \text{Var}(U) = \Sigma / n_+ = \begin{pmatrix} \pi_0(1 - \pi_0) & -\pi_0 \pi_1 & -\pi_0 \pi_2 \\ -\pi_0 \pi_1 & \pi_1(1 - \pi_1) & -\pi_1 \pi_2 \\ -\pi_0 \pi_2 & -\pi_1 \pi_2 & \pi_2(1 - \pi_2) \end{pmatrix}. \]

(24)

Let \( G(U) \) denote \( \frac{4n_2n_0 - n_1^2}{n_0 + n_2} \) and \( G(\Pi) \) denote \( 4\pi_0 \pi_2 - \pi_1^2 \). Under \( H_0 \), then
\[ \pi_0 = 4(1 - P_0)^2, \quad \pi_1 = 2P_0(1 - P_0), \quad \pi_2 = P_0^2. \]

\[ G(\Pi) = 4\pi_0 \pi_2 - \pi_1^2 = 0. \]

By the central limit theorem and the multivariate delta method, \( G(U) \) has an asymptotic normal distribution with mean 0. That is
\[ \sqrt{N}(G(U) - G(\Pi)) \rightarrow N(0, G'(U) \text{Var}(U) G'(U)^T). \]

(25)

Thus, under \( H_0 \), as \( n_2 \rightarrow \infty \),
\[ P(4n_0n_2 - n_1^2 < 0) \rightarrow P(G(U) < 0) \rightarrow 1/2. \]

This finishes the proof of the following convergence in distribution, under \( H_0 \),
\[ 2\lambda_N \rightarrow \frac{1}{2} \chi^2_1 + \frac{1}{2} \chi^2_1. \]

(26)