Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Charnigo et al., J Biomet Biostat 2013, 4:1
DOI: 10.4172/2155-6180.1000157

Research Article Open Access

Contaminated Chi-Square Modeling and Large-Scale ANOVA Testing

Richard Charnigo[1]*, Feng Zhou[1] and Hongying Dai[2]

[1]Department of Statistics, 725 Rose Street, University of Kentucky, Lexington KY, USA
[2]Research Development and Clinical Investigation, 2420 Pershing Road, Children's Mercy Hospital, USA

Abstract

We propose a convenient moment-based procedure for testing the omnibus null hypothesis of no contamination of a central chi-square distribution by a non-central chi-square distribution. In sharp contrast with likelihood ratio tests for mixture models, there is no need for re-sampling or random field theory to obtain critical values. Rather, critical values are available from an asymptotic normal distribution, and there is excellent agreement between nominal and actual significance levels. This procedure may be used to model numerous chi-square statistics, obtained via monotonic transformations of F statistics, from large-scale ANOVA testing, such as that encountered in microarray data analysis. In that context, modeling chi-square statistics instead of p-values may improve detection of differential gene expression, as we demonstrate through simulation studies, while also reducing false declarations of the same, as we illustrate in a case study on aging and cognition. Our procedure may also be incorporated into a gene filtration process, which may reduce type II errors on genewise null hypotheses by justifying lighter controls for Type I errors.

Keywords: Aging; Cognition; Gene expression; Hippocampus; Method of moments; Microarray; Mixture model; Multiple comparisons

Introduction

Consider the mixture model [1-3], with probability density function (pdf)

$$(1-\lambda)\chi^2_\nu(0)+\lambda\,\chi^2_\nu(\mu) \quad (1)$$

where $0 \le \lambda \le 1$, $\chi^2_\nu(0)$ denotes the central chi-square pdf on $\nu>0$ degrees of freedom (df), and $\chi^2_\nu(\mu)$ denotes the chi-square pdf on ν df, with non-centrality parameter $\mu \ge 0$. We assume that ν is known, while λ and μ are unknown. We refer to (1) as the Contaminated Chi-square (CCS) model, since we regard $\chi^2_\nu(0)$ as being contaminated by $\chi^2_\nu(\mu)$.

In this paper, we present a convenient procedure for testing

$$H_0: \lambda\mu=0 \text{ versus } H_1: \lambda\mu>0, \quad (2)$$

we analyze its asymptotic and finite-sample properties, and we propose estimators of these parameters in the event that H_0 is rejected. For a reason that will become apparent later, we refer to H_0 as the omnibus null hypothesis. The CCS model simplifies to $\chi^2_\nu(0)$, if and only if the omnibus null hypothesis is true.

To understand how the CCS model and omnibus null hypothesis relate to large-scale ANOVA testing, suppose that a microarray experiment [4,5] is performed to measure expression levels on each of n genes for subjects in independent samples of sizes $g_1, g_2, \ldots, g_K$ from K populations. For gene i ($1 \le i \le n$), a one-way ANOVA may be conducted to test the genewise null hypothesis of equal mean expression levels across the K populations. This one-way ANOVA yields a test statistic F_i that has a central F distribution on (K-1) numerator and $(g_1+g_2+\ldots+g_K-K)$ denominator df, under the genewise null hypothesis.

Let X_i denote the rescaled test statistic (K-1) F_i. With large $(g_1+g_2+\ldots+g_K-K)$, X_i is distributed approximately $\chi^2_{K-1}(0)$ under the genewise null hypothesis, and approximately $\chi^2_{K-1}(\mu)$, under the genewise alternative hypothesis, for some μ. We explain this approximation in the Appendix. If $g_1, g_2, \ldots, g_K$ are not large enough to warrant this approximation, then a more sophisticated approach may be employed to transform F statistics into chi-square statistics; one such approach is described in and used for our case study.

Letting λ denote the proportion of genes for which mean expression levels are not equal across the K populations, we may regard the collection of rescaled test statistics $X_1, X_2, \ldots, X_n$ as a sample from the CCS model with $\nu=(K-1)$. If mean expression levels are equal across the K populations for all genes, then the CCS model reduces to $\chi^2_{K-1}(0)$. This is why $\lambda\mu=0$ is referred to as the omnibus null hypothesis.

The CCS model may also be applied and the omnibus null hypothesis tested, using subsets of $X_1, X_2 \ldots, X_n$ corresponding to biologically meaningful partitions of the genes. For example, suppose that n=2000 and that the first 1900 genes correspond to autosomes, while the last 100 genes correspond to sex chromosomes [6]. Suppose, moreover, that there are g_1=10 male subjects with a severe form of a disease suspected to be sex-linked, g_2=10 male subjects with a mild form of the same disease, and g_3=10 healthy male subjects. In this case, an investigator may wish to fit the CCS model separately to the first 1900 genes and to the last 100 genes.

If $X_1, X_2 \ldots, X_{1900}$ lead to rejection of the omnibus null hypothesis, then the investigator may question whether the disease is in fact sex-linked.

Otherwise, the investigator may justifiably discard the first 1900 genes and focus attention on the last 100. In particular, multiplicity adjustments for controlling Type I errors on genewise null hypotheses [7-9], can be based on the 100 remaining tests, instead of on the original 2000. Less stringent multiplicity adjustments will reduce Type II errors on the 100 remaining tests. Dai and Charnigo [10,11] have previously referred to this concept as gene filtration, although their earlier work did not consider the CCS model.

***Corresponding author:** Richard Charnigo, Department of Statistics, 725 Rose Street, University of Kentucky, Lexington KY, USA, Tel: 859.218.2072; Fax: 859.257.6430; E-mail: RJCharn2@aol.com

Received November 26, 2012; **Accepted** December 15, 2012; **Published** December 22, 2012

Citation: Charnigo R, Zhou F, Dai H (2013) Contaminated Chi-Square Modeling and Large-Scale ANOVA Testing. J Biomet Biostat 4:157. doi:10.4172/2155-6180.1000157

The CCS model may potentially be applied in other scenarios involving large numbers of tests. For instance, we envisage that the CCS model may be employed to analyze data on copy number variation [12], or transcript splicing variation [13]. Before presenting our testing and estimation procedures, we briefly review some literature on mixture modeling. This review is not exhaustive but provides some context for this paper, allowing a more explicit articulation of this paper's contributions. The remainder of this paper features empirical investigations, including both simulations, and an application to real data, as well as a discussion highlighting extensions of the ideas contained herein. An appendix explains the rescaling of F statistics into approximate chi-square statistics.

Background on Mixture Modeling

Mixture modeling has been applied to interesting problems in disciplines, as varied as epidemiology [14,15], astronomy [16,17], biochemistry [18,19], and genetics [20,21].

From a technical perspective, mixture modeling is challenging because the usual regularity conditions for likelihood-based inference are not satisfied, when one is testing the number of components in a mixture model [22,23]. In particular, the asymptotic null distribution of a likelihood ratio test statistic for the number of components corresponds, under mild assumptions, to the supremum of a squared truncated Gaussian process defined on a compact parameter space [24-27].

Although likelihood-based inference is still possible via bootstrapping [28], or random field theory [29], more convenient approaches have been developed for many scenarios. These include Modified Likelihood Ratio (MLR) tests and estimators [30,31], Expectation Maximization (EM) tests and estimators [32,33], D tests [34,35] and moment-based tests [36].

Allison et al. [37] proposed applying a beta mixture model to the p-values from genewise hypothesis tests in a microarray experiment. This motivated Dai and Charnigo [10] to present MLR and D tests, for whether a beta mixture model for the p-values could be simplified to a uniform distribution. Subsequently, Dai and Charnigo [11] proposed applying a normal mixture model to the Z scores from genewise hypothesis tests (perhaps obtained by transforming T statistics), and developed tests for whether the normal mixture model could be simplified to a normal distribution. Whether looking at p-values or Z scores, an investigator could incorporate genewise hypothesis tests into a filtration algorithm.

The present work differs from the preceding efforts in that chi-square statistics (perhaps obtained by transforming F statistics) are now the focus, instead of p-values or Z scores. There are two reasons for this focus. First, while some microarray data analyses compare two populations on mean expression levels, other microarray data analyses compare more than two populations. An example, considered in our case study, appears in Blalock et al. [38], who compared three populations based on age strata to identify genes related to aging and cognition. Since ANOVA does not yield a Z score, the methodology of Dai and Charnigo [11] is inapplicable to such a scenario. However, the methodology proposed herein is applicable. In fact, the methodology proposed herein is still applicable when only two populations are compared, since a Z score may be converted to a chi-square statistic via squaring.

Second, a beta mixture model for p-values may differ from a uniform distribution in a way that is not indicative of systematic differential expression. For instance, 0.5 Beta(1,1)+0.5 Beta(2,0.5) corresponds to an excess of large p-values, rather than of small p-values. The tests of Dai and Charnigo [10] will detect an excess in either direction. Thus, the power to detect a specific alternative that is indicative of systematic differential expression may be lower than desired. The test proposed herein overcomes that limitation by rejecting the omnibus null hypothesis in (2), only when there is an excess of large chi-square statistics (or, equivalently, small p-values). Indeed, (2) makes explicit that the alternative to the omnibus null hypothesis is one-sided. As such, the test proposed herein may have better power to detect systematical differential expression than the tests of Dai and Charnigo [10].

Testing and Estimation Procedures

Suppose that X_1, X_2 …, X_n are a random sample from the CCS model (1). Our procedure for testing the omnibus null hypothesis in (2) is an intersection-union test based on the method of moments. More specifically, let

$$S = n^{-1}\sum_{1\le k\le n} X_k - v \quad \text{and}$$

$$W = v^2 + 2v(1-n^{-1}\sum_{1\le k\le n} X_k) + n^{-1}\sum_{1\le k\le n} X_k{}^2 - 4n^{-1}\sum_{1\le k\le n} X_k \quad (3)$$

Then S converges in probability to λμ, and W converges in probability to $\lambda\mu^2$, by the Weak Law of Large Numbers and Slutsky's Theorem. (If one wished to estimate $\lambda\mu^p$ for a generic positive integer p, then one could derive an estimator using the first p moments; or if both S>0 and W>0, then one might estimate $\lambda\mu^p$ by $W^{p-1}S^{2-p}$. However, neither theorem 1 nor theorem 2 below involves estimation of $\lambda\mu^p$, so we do not discuss such estimation further).

The preceding considerations motivate us to reject the omnibus null hypothesis if $S>s_{crit}$ and $W>w_{crit}$, where s_{crit} and w_{crit} are chosen to achieve the desired type I error probability. Theorem 1 below indicates how s_{crit} and w_{crit} may be chosen. Before stating theorem 1, we establish some notation.

Let Φ denote the standard normal cumulative distribution function, and z_c, the c quantile of the same. Let r_j denote the j^{th} moment of $\chi^2_\nu(0)$ for $1 \le j \le 4$, R the 2×2 matrix, whose ij^{th} entry is $r_{i+j}-r_i r_j$, and B the 2×2 matrix, whose first column is (1,0), and whose second column is (-2ν-4,1).

Theorem 1: Let $0<\delta \le 1$ and $0<\varepsilon \le 1$ satisfy $\delta\varepsilon = \alpha$. Under the omnibus null hypothesis,

$$\lim_{n\to\infty} P[S > z_{1-\delta}n^{-1/2}a_{11}^{1/2} \text{ and } W > z_{1-\varepsilon}n^{-1/2}a_{22}^{1/2}] = \alpha \quad (4)$$

w here a_{11} and a_{22} are the diagonal entries of the 2×2 matrix $A=B^TRB$.

Moreover, under any fixed alternative $(\lambda,\mu)=(c_1,c_2)$, with $0<c_1 \le 1$ and $c_2>0$,

$$\lim_{n\to\infty} P[S > z_{1-\delta}n^{-1/2}a_{11}^{1/2} \text{ and } W > z_{1-\varepsilon}n^{-1/2}a_{22}^{1/2}] = 1 \quad (5)$$

Proof: Under the omnibus null hypothesis, $n^{1/2}(n^{-1}\sum_{1\le k\le n} X_k - v, n^{-1}\sum_{1\le k\le n} X_k{}^2 - 2v - v^2)^T$ converges in law to the multivariate normal distribution, with mean vector $(0,0)^T$ and covariance matrix R by the Central Limit Theorem. Then, $(S,W)^T$ converges in law to the multivariate normal distribution, with mean vector $(0,0)^T$ and

covariance matrix A by Cramer's Theorem. The key observation is that the off-diagonal entries of A are 0, hence $P[S > z_{1-\delta} n^{-1/2} a_{11}^{1/2}$ and $W > z_{1-\varepsilon} n^{-1/2} a_{22}^{1/2}]$ converges to $(1-\Phi[z_{1-\delta}])(1-\Phi[z_{1-\varepsilon}])=\delta\varepsilon=\alpha$.

Under the fixed alternative $(\lambda,\mu)=(c_1,c_2)$, S converges in probability to $c_1c_2>0$, and W converges in probability to $c_1c_2^2>0$, so that $P[S \le z_{1-\delta} n^{-1/2} a_{11}^{1/2}]$ and $P[W \le z_{1-\varepsilon} n^{-1/2} a_{22}^{1/2}]$ converge to 0. Since $P[S > z_{1-\delta} n^{-1/2} a_{11}^{1/2} \text{ and } W > z_{1-\varepsilon} n^{-1/2} a_{22}^{1/2}] \ge 1 - P[S \le z_{1-\delta} n^{-1/2} a_{11}^{1/2}] - P[W \le z_{1-\varepsilon} n^{-1/2} a_{22}^{1/2}]$, the former must converge to 1. QED.

A few comments are in order. First, one may choose ε=1 (i.e. choose $w_{crit}=-\infty$), and effectively base the test on only S, rather than on both S and W. In this case, one may replace $z_{1-\delta}\, n^{-1/2}\, a_{11}^{1/2}$ by $n^{-1}\, q_{\nu n,1-\alpha}-\nu$, where $q_{\nu n},1-\alpha$ denotes the 1-α quantile of $\chi^2_{\nu n}(0)$. Then the type I error probability is exactly α, for all finite n, not just converging to α in the limit. However, a potential problem with this choice is that one may reject the omnibus null hypothesis, when W<0. Since W is a moment-based estimator of $\lambda\mu^2$, moment-based estimation of λ and μ, when W<0 leads to the estimator of λ, and/or that of μ, not belonging to the appropriate parameter space. However, a remedy is indicated in the next comment.

Second, choosing $\varepsilon \le ½$ and $\delta \le ½$ (i.e., choosing $w_{crit}>0$ and $s_{crit}>0$) guarantees that λ and μ may be estimated using moments, when the omnibus null hypothesis is rejected. This is described in theorem 2 and its corollary below. More specific choices of ε and δ can be recommended based on power considerations. However, while S and W are asymptotically independent under the omnibus null hypothesis, they may be correlated when the omnibus null hypothesis is false. Thus, analytically evaluating the power, in relation to ε and δ is difficult. However, we can gain some insights from simulation studies, which we pursue later.

Third, in contrast with a likelihood ratio test for the number of components in a mixture model, the testing procedure of theorem 1 does not require a compact parameter space; note that no upper bound for μ was assumed. Moreover, the critical value is known, and thus, need not be estimated *via* resampling or random field theory. On the other hand, the problem in (2) is not, strictly speaking, determining the number of components in a mixture model. This is because, although (1) reduces to one component under the omnibus null hypothesis, (1) also reduces to one component, when λ=1 and μ>0.

Now, we address the estimation of λ and μ. Theorem 2 shows that, when the omnibus null hypothesis is false, S^2/W and W/S are $n^{1/2}$-consistent estimators of λ and μ, respectively. To state theorem 2, we introduce some more notations. Let $m_j=E[X_1^j]$ for $1 \le j \le 4$, M the 2×2 matrix, whose ij^{th} entry is $m_{i+j}-m_i\, m_j$, and D the 2×2 matrix whose first column is

$((m_1-\nu)(2m_2-4m_1-2\nu m_1), -(m_1-\nu)^2)^T/(m_2+2\nu+\nu^2-4m_1-2\nu m_1)^2$ and whose second column is $(-m_2+2\nu+\nu^2, m_1-\nu)^T/(m_1-\nu)^2$.

Theorem 2: Under any fixed alternative $(\lambda,\mu)=(c_1,c_2)$, with $0<c_1 \le 1$ and $c_2>0$,

$n^{1/2}(S^2/W-c_1, W/S-c_2)^T$ converges in law to the multivariate normal distribution, with mean vector $(0,0)^T$ and covariance matrix $D^T M D$.

Proof: By the Central Limit Theorem, $n^{1/2}(n^{-1}\sum_{1\le k\le n} X_k-\nu-c_1c_2$, $n^{-1}\sum_{1\le k\le n} X_k^2-(\nu+c_1c_2)^2-c_2^2c_1(1-c_1)-4c_1c_2-2\nu)^T$ converges in law to the multivariate normal distribution, with mean vector $(0,0)^T$ and covariance matrix M. The desired result then follows from Cramer's Theorem. QED.

Although the probability that S<0 or W<0 is nonzero (in which case the estimator of λ, and/or that of μ will not belong to the appropriate parameter space), with $\varepsilon \le ½$ and $\delta \le ½$, this event is a subset of accepting the omnibus null hypothesis. Hence, if one agrees to take $\varepsilon \le ½$ and $\delta \le ½$, as well as to estimate λ and μ, only if the omnibus null hypothesis is rejected, then this event will not be encountered in practice. The following corollary, an immediate consequence of (5) from theorem 1, also demonstrates that such an agreement does not disturb the conclusion of theorem 2.

Corollary: Under any fixed alternative $(\lambda,\mu)=(c_1,c_2)$ with $0<c_1 \le 1$ and $c_2>0$, the conditional distribution of $n^{1/2}(S^2/W-c_1, W/S-c_2)^T$, given that $W>w_{crit}$ and $S>s_{crit}$ converges to the multivariate normal distribution, with mean vector $(0,0)^T$ and covariance matrix $D^T M D$.

Simulation Studies

To assess the type I and type II error rates of our testing procedure in finite samples, we conducted a number of simulation studies. In figure 1 and in the following text, we use this shorthand:

* "CCS 1": The procedure for testing the omnibus null hypothesis in (2) is applied directly to a random sample $X_1, X_2, \ldots, X_n$ from the CCS model (1), with δ=1/2 and ε=1/10. These choices of δ and ε emphasize W over S for rejection of the omnibus null hypothesis, requiring only that the latter be positive.

* "CCS 2": Proceed as above, but with $\delta=\varepsilon=0.05^{1/2}$. These choices emphasize W and S equally.

* "CCS 3": Proceed as above, but with δ=1/10 and ε=1/2. These choices of δ and ε emphasize S over W for rejection of the omnibus null hypothesis, requiring only that the latter be positive.

* "CB": A random sample $X_1, X_2 \ldots, X_n$ from the CCS model

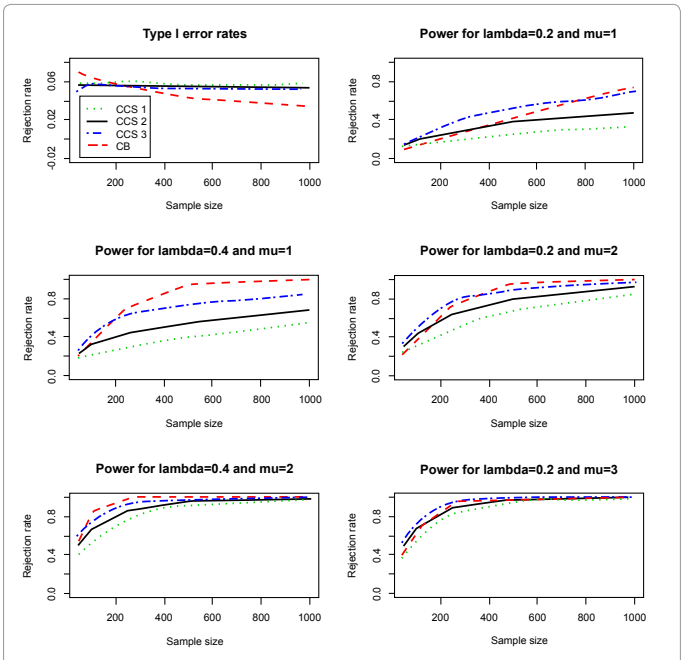

Figure 1: Displayed are rejection rates for omnibus null hypotheses in the simulation studies. Methods CCS1, CCS2, and CCS3, refer to our proposed testing procedure with (δ,ε)=(1/2, 1/10), (δ, ε)=($0.05^{1/2}$,$0.05^{1/2}$), and (δ,ε)=(1/10, 1/2), respectively. Method CB refers to a modified likelihood ratio test applied to p-values, and treated as arising from the contaminated beta model [10].

(1) is transformed by the survival function of the central chi-square distribution on ν df to yield "p-values" $P_1, P_2 \ldots, P_n$. These are treated as if they had arisen from the Contaminated Beta (CB) model with pdf

$$(1-\lambda)1_{0<p<1} + \lambda 1_{0<p<1} p^{\alpha-1}(1-p)^{\beta-1}/B(\alpha,\beta). \quad (6)$$

The MLR test is applied to $P_1, P_2 \ldots, P_n$ to see whether the CB model can be reduced to a uniform distribution [10].

For each n in {50, 100, 250, 500, 1000}, we generated 10,000 random samples $X_1, X_2 \ldots, X_n$ from the CCS model (1) with λμ=0. Each random sample $X_1, X_2 \ldots, X_n$ was meant to mimic a collection of chi-square statistics, corresponding to n genes with no differential expression. We calculated type I error rates as the numbers of omnibus null hypothesis rejections divided by 10,000. The calculated type I error rates are displayed in the top left panel of figure 1. For methods CCS1, CCS2, and CCS3, these are between 0.0504 and 0.0613 at all n. Thus, the critical values for our testing procedure, which were based on the asymptotic result of theorem 1, appear satisfactory for finite samples. For method CB, the calculated type I error rates decrease from 0.0701 at n=50 to 0.0338 at n=1000, indicating that the MLR test applied to p-values is slightly anticonservative for small n.

We then generated 10,000 random samples, with λ=0.2 and μ=1. Each random sample was meant to mimic a collection of chi-square statistics, corresponding to a mix of differentially expressed genes (20%), with non differentially expressed genes (80%). Power, calculated as the number of omnibus null hypothesis rejections divided by 10,000, is displayed in the top right panel of figure 1. As anticipated, power increases with n for each method. Method CCS3 exhibits better power than method CCS2, which in turn is more powerful than method CCS1. Method CB appears relatively strong for large n, but comparatively weak for small n.

The remaining panels of figure 1 present power for (λ,μ)=(0.4,1), (λ,μ)=(0.2,2), (λ,μ)=(0.4,2), and (λ,μ)=(0.2,3), respectively. All of these scenarios maintain the relative ordering of methods CCS3, CCS2, and CCS1. Roughly speaking, method CB fares well with larger λ, μ, and n, but does not perform as well with smaller λ, μ, and n.

Based on the results of these simulation studies, we recommend taking δ=1/10 and ε=1/2, when applying our testing procedure. If n is large, or if λ and μ are anticipated to be large, then one may also wish to consider transforming chi-square statistics to p-values and then analyzing p-values using the CB model (6). However, the case study will provide an important caveat, namely that a naïve analysis of p-values may lead to an inappropriate declaration of systematic differential expression. Thus, care must be exercised in any decision to transform chi-square statistics to p-values.

We also note that, while convenient to use because no resampling is required to ascertain critical values, our moment-based procedure for testing the omnibus null hypothesis in (2) may be less powerful than other approaches yet to be developed. In particular, we plan to investigate in a future manuscript whether the EM test [32,33], can be adapted to this setting. If so, then transforming chi-square statistics to p-values, and then analyzing p-values using the CB model (6) may become even less appealing.

Case Study

Dai and Charnigo [10] applied the CB model (6) to analyze the p-values generated from a microarray experiment conducted by Blalock et al. [38]. Briefly, gene expression levels were acquired from the hippocampal tissue of 30 male Fischer rats divided into three groups of 10: "old", "middle-aged", and "young". For each of 8799 genes, a one-way ANOVA was conducted to compare expression levels across the three groups. This produced 8799 F statistics, which in turn yielded the p-values. As noted by Dai and Charnigo [10], Blalock et al. [38] employed a three-step process to filter the p-values. In each step, genes were either retained for or eliminated from further consideration.

A major concern emerged when Dai and Charnigo [10] analyzed the p-values and, in particular, employed the MLR test [30], and D test [34], to see whether the CB model could be reduced to a uniform distribution. For the genes eliminated at step 3, the MLR test and D test decisively rejected the omnibus null hypothesis of a uniform distribution. However, the fitted model had λ=0.696, α=1.01, and β=1.28. Since α>1 does not correspond to an excess of small p-values, the departure from a uniform distribution may not indicate differential expression, but rather, as suggested by Allison et al. [37], correlations among the p-values corresponding to different genes. Thus, the alternative to the omnibus null hypothesis of a uniform distribution may be too broad if our main interest is in ascertaining differential expression.

With this concern in mind, we revisited these data. However, instead of analyzing p-values, we examined chi-square statistics. Since the denominator df for the underlying F statistics was not particularly large, we modified the F statistics based on the probability integral transformation [39], a more sophisticated approach than the rescaling described earlier and also consistent with the manner in which Dai and Charnigo [11] transformed T statistics to Z scores. More specifically, we converted the F statistics to chi-square statistics by successively applying the cumulative distribution function (cdf) of the central F distribution on 2 and 27 df, followed by the inverse cdf of the central chi-square distribution on 2 df.

Figure 2 shows histograms of chi-square statistics for all 8799 genes,

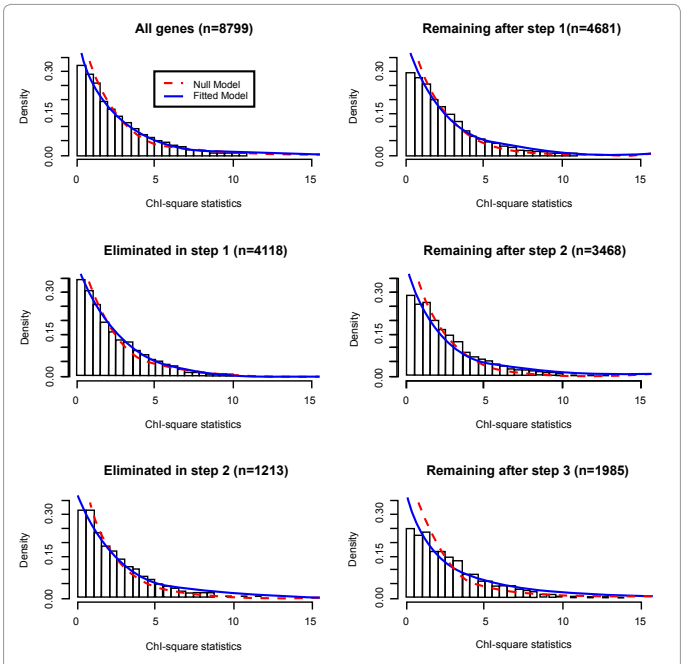

Figure 2: Shown are histograms of chi-square statistics for all 8799 genes in the case study for the genes eliminated in steps 1 and 2 of the filtration process employed by Blalock et al. [38], and for the genes remaining after each of the three steps. Superimposed against each histogram are the fitted CCS model for which parameter estimates are displayed in table 1, and the null model $\chi^2_2(0)$.

for the genes eliminated in steps 1 and 2, and for the genes remaining after each step. Superimposed against each histogram are the fitted CCS model from (1), for which parameter estimates are displayed in table 1, and the null model $\chi^2_2(0)$. In all six panels of figure 2, though most noticeably in the last panel, the fitted model yields a smaller density between 0 and 2, but a larger density between 5 and 10 compared to the null model. Overall, each fitted model is in much better concordance with its respective histogram than the null model, although even the fitted model overstates the number of very small chi-square statistics.

Correspondingly, our procedure for testing the omnibus null hypothesis in (2) yields a p-value less than 0.0001 for the omnibus null hypothesis, regardless of whether one defines this p-value by taking $\delta=1/2$, $\varepsilon=2\alpha$ (i.e. p-value is half the smallest ε, at which the omnibus null hypothesis is rejected when δ is fixed at 1/2), or $\delta=\varepsilon=\alpha^{1/2}$ (i.e. p-value is the square of the smallest ε, at which the omnibus null hypothesis is rejected when δ and ε are constrained to equality) or $\delta=2\alpha$, $\varepsilon=1/2$ (i.e. p-value is half the smallest δ, at which the omnibus null hypothesis is rejected when ε is fixed at 1/2).

The top panel of figure 3 shows a histogram of chi-square statistics for the 1483 genes eliminated in step 3, along with the null model $\chi^2_2(0)$. No fitted CCS model is shown because we have $W<0$. This precludes valid moment-based estimation of λ and μ. Although a likelihood-based approach to estimating λ and μ could be employed, this is not called for because the omnibus null hypothesis is not rejected at any $\alpha \leq 0.25$, regardless of whether one takes $\delta=1/2$, $\varepsilon=2\alpha$ or $\delta=\varepsilon=\alpha^{1/2}$, or $\delta=2\alpha$, $\varepsilon=1/2$. In fact, the null model is not a bad fit to the histogram, except for overstating the number of very small chi-square statistics. (Recall that the fitted CCS models in figure 2 had the same difficulty.)

Genes	Estimated λ	Estimated μ
all 8799	0.231	3.25
remaining after step 1	0.236	4.13
eliminated in step 1	0.389	1.28
remaining after step 2	0.223	4.54
eliminated in step 2	0.314	2.77
remaining after step 3	0.308	5.19

Note: Shown are parameter estimates for the CCS model as applied to 8799 genes in the Case Study, along with subsets of genes retained or eliminated in the filtration process employed by Blalock et al. [38]. Each of these fitted CCS models is displayed graphically in figure 2.

Table 1: Parameter Estimates for the CCS Model.

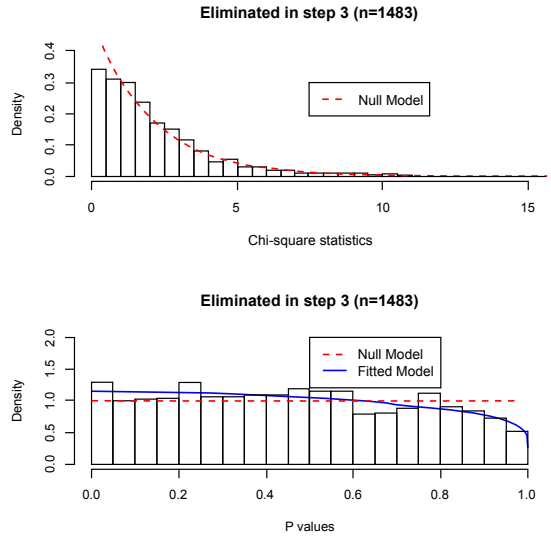

Figure 3: The top panel shows a histogram of chi-square statistics for the 1483 genes eliminated in step 3 of the filtration process employed by Blalock et al. [38], along with the null model $\chi^2_2(0)$. No fitted CCS model is shown, as the omnibus null hypothesis is not rejected at any $\alpha \leq 0.25$. The bottom panel shows a histogram of the p-values for these same 1483 genes, along with the fitted CB model, and the null model of a uniform distribution.

The bottom panel of figure 3 shows a histogram of the p-values for these same 1483 genes, along with the fitted CB model (6), and the null model of a uniform distribution. The fitted CB model is not suggestive of differential expression, as there is no marked surplus of small p-values. However, there are noticeably fewer extremely large p-values than would be compatible with a uniform distribution, and for this reason, both the MLR test and D test decisively reject the omnibus null hypothesis of a uniform distribution. This rejection is inappropriate in so far as one uses it to infer differential expression.

In summary, employing the CCS model to analyze chi-square statistics, instead of the CB model to assess p-values resolves the aforementioned concern, because the omnibus null hypothesis from (2) is not rejected for the genes eliminated in step 3. Thus, using the CCS model avoided an inappropriate declaration of differential expression.

Discussion

We have developed a convenient procedure for testing the omnibus null hypothesis of no contamination of a central chi-square distribution by a non-central chi-square distribution. This procedure is based on the first two sample moments, which permits critical values to be derived from quantiles of the standard normal distribution. Our simulation studies show that, even for small sample sizes, there is excellent agreement between the nominal and actual significance levels. In sharp contrast with likelihood ratio tests for mixture models, the asymptotic null distribution is uncomplicated [24-27], and thus there is no need for re-sampling [28], or random field theory [29], to obtain critical values.

As a follow-up to rejection of the omnibus null hypothesis, we have also proposed moment-based estimators of the contamination fraction and non-centrality parameter of the contaminating distribution. Provided that the quantities in question are both nonzero, our estimators are $n^{1/2}$-consistent. Moreover, with suitable choices of δ and ε in the testing procedure, our estimators have probability 1 of being positive, conditional on rejection of the omnibus null hypothesis. This result is remarkable because moment-based estimators in mixture models ordinarily do not belong to their respective parameter spaces with probability 1, as noted by Charnigo et al. [36] for another type of contamination model.

Our testing and estimation procedures are primarily motivated by the modeling of numerous chi-square statistics arising from microarray data analysis specifically or large-scale testing generally. Such modeling expedites a filtration process, which, if successful, can reduce type II errors by justifying lighter controls for type I errors. While this filtration process was advocated by Dai and Charnigo [10] for the analysis of p-values, our case study provides a clear caveat against naïve analyses of p-values, and illustrates a real-world scenario in which analyzing chi-square statistics avoids an inappropriate declaration of differential expression. Moreover, our simulation studies show that under certain conditions, analysis of chi-square statistics may actually yield better power to detect differential expression than analysis of p-values.

While we have envisaged applying the CCS model to chi-square statistics monotonically related to F statistics from one-way ANOVA,

For example, if the normality and equal variance assumptions underlying one-way ANOVA are untenable, then one may employ the nonparametric Kruskal-Wallis test for equal medians. Since the Kruskal-Wallis test statistic is distributed approximately $\chi^2_{K-1}(0)$ when the medians are equal, the CCS model can be applied in conjunction with chi-square statistics from Kruskal-Wallis tests, as easily as with F statistics from one-way ANOVA.

Moreover, sophisticated experimental designs or sampling schemes may preclude using either one-way ANOVA or Kruskal-Wallis tests. For instance, Mao et al. [40] obtained multiple tissue samples from some of their subjects, so that linear mixed models were required to test genewise null hypotheses. However, as long as genewise null hypotheses are tested using chi-square or F statistics (or even Z or T statistics, since these can be squared), the CCS model remains applicable.

A number of promising avenues exist for future research. One of them is to investigate whether the EM test [32,33], can be profitably employed in the setting of the CCS model, and if so, whether power to reject a false omnibus null hypothesis is improved. Our simulation studies suggest that there may indeed be room for improvement, as the procedure proposed herein was not uniformly more powerful than the MLR test applied to p-values derived from the chi-square statistics.

Another topic for future research is to generalize the CCS model to provide greater flexibility for describing real data. For instance, suppose that each X_i has its own non-centrality parameter μ_i under the genewise alternative hypothesis. Then we may consider a new model,

$$(1-\lambda)\,\chi^2_\nu(0)+\lambda\int\chi^2_\nu(\mu)\,dG(\mu), \qquad (7)$$

where $\int$ denotes integration and G is some cumulative distribution function defined on the nonnegative real numbers. Note that the first sample moment of data from (7) is ν, if and only if (3) reduces to $\chi^2_\nu(0)$, as both are equivalent to $\lambda\{1-G(0)\}=0$. Thus, one obtains a consistent level α test for whether (7) reduces to $\chi^2_\nu(0)$, by asking whether the first sample moment exceeds $n^{-1}\,q_{\nu n,1-\alpha}$. However, the subsequent estimation of λ and G are anticipated to be considerably more delicate.

Appendix

We now explain our earlier statement that a rescaled F statistic may be regarded as an approximate chi-square statistic. Suppose that Y_1 has the central chi-square distribution on ν_1 df, and that independently, Y_2 has the central chi-square distribution on ν_2 df. Then, the quotient $(Y_1/\nu_1)/(Y_2/\nu_2)$ has the central F distribution on ν_1 numerator and ν_2 denominator df [39].

Since Y_2 has mean ν_2 and variance $2\nu_2$, Chebychev's Inequality implies that (Y_2/ν_2) converges in probability to 1 as $\nu_2\to\infty$. Thus, when ν_2 is large, $(Y_2/\nu_2)\approx 1$, and so $(Y_1/\nu_1)/(Y_2/\nu_2) \approx (Y_1/\nu_1)$. In other words, a quantity with a central F distribution on large denominator df resembles a chi-square random variable divided by the numerator df.

To make explicit the connection to our earlier statement let Y_1 be the between sum of squares from a one-way ANOVA divided by the underlying variance of the individual observations, let $\nu_1=K-1$ be the corresponding df, let Y_2 be the within sum of squares divided by the underlying variance, and let $\nu_2=g_1+g_2+\ldots+g_K-K$ be the corresponding df. Put $F=(Y_1/\nu_1)/(Y_2/\nu_2)$. If $g_1+g_2+\ldots+g_K-K$ is sufficiently large, then $(K-1)F \approx Y_1$.

From a practical perspective, one may decide whether $g_1+g_2+\ldots+g_K-K$ is sufficiently large by evaluating whether $P[|\log(Y_2/\nu_2)| \geq tol\,]$ $\leq$ tol, where tol is a specified tolerance. One may interpret tol as the maximum acceptable Levy distance between the cumulative distribution functions of log[(K-1)F] and $\log[Y_1]$. As such, we recommend setting tol no larger than 0.20, and preferably as small as 0.10. Corresponding to these choices, one has $g_1+g_2+\ldots+g_K-K \geq 83$ and $g_1+g_2+\ldots+g_K-K \geq 543$, respectively. Since $g_1+g_2+\ldots+g_K-K=27$ in the case study we did not use rescaling but instead relied on a more sophisticated approach for transforming F statistics into chi-square statistics.

References

1. Titterington DM, Makov UE, Smith AF (1986) Statistical analysis of finite mixture distributions. Wiley, John & Sons.
2. Lindsay BG (1995) Mixture models: Theory, geometry and applications. Institute of Mathematical Statistics, Hayward, California.
3. McLachlan G, Peel D (2000) Finite mixture models. (1st edition), Wiley-Interscience.
4. Leung YF, Cavalieri D (2003) Fundamentals of cDNA microarray data analysis. Trends Genet 19: 649-659.
5. Berrar DP, Dubitzky W, Granzow M (2009) A practical approach to microarray data analysis. Springer.
6. Sumner AT (2003) Chromosomes: organization and function. Blackwell Publishing, USA.
7. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57: 289-300.
8. Shaffer JP (1995) Multiple hypothesis testing. Annu Rev Psychol 46: 561-584.
9. Storey JD (2003) The positive false discovery rate: a Bayesian interpretation and the q-value. Ann Stat 31: 2013-2035.
10. Dai H, Charnigo R (2008) Omnibus testing and gene filtration in microarray data analysis. J Appl Stat 35: 31-47.
11. Dai H, Charnigo R (2010) Contaminated normal modeling with application to microarray data analysis. Can J Stat 38: 315-332.
12. Breheny P, Chalise P, Batzler A, Wang L, Fridley BL (2012) Genetic association studies of copy-number variation: should assignment of copy number states precede testing ? PLoS One 7: 34262.
13. Vandiedonck C, Taylor MS, Lockstone HE, Plant K, Taylor JM, et al. (2011) Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex. Genome Res 21: 1042-1054.
14. Charnigo R, Sun J (2008) Testing homogeneity in discrete mixtures. J Stat Plan Inference 138: 1368-1388.
15. Deng W, Charnigo R, Dai H, Kirby R (2011) Characterizing components in a mixture model for birthweight distribution. J Biom Biostat 2: 118.
16. Roeder K (1990) Density estimation with confidence sets exemplified by superclusters and voids in the Galaxies. J Am Stat Assoc 85: 617-624.
17. Morrison HL, Helmi A, Sun J, Liu P, Gu R, et al. (2009) Fashionably late? Building up the Milky Way's inner halo. Astrophys J 694: 130-143.
18. Bechtel YC, Bonaiti-Pelliee C, Poisson N, Magnette J, Bechtel PR (1993) A population and family study of N-acetyltransferase using caffeine urinary metabolites. Clin Pharmacol Ther 54: 134-141.
19. Roeder K (1994) A graphical technique for determining the number of components in a mixture of normals. J Am Stat Assoc 89: 487-495.
20. Newton MA, Kendziorski CM, Richmond CS, Blattner FR, Tsui KW (2001) On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. J Comput Biol 8: 37-52.
21. Kendziorski CM, Newton MA, Lan H, Gould MN (2003) On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. Stat Med 22: 3899-3914.
22. Ghosh JK, Sen PK (1985) On the asymptotic performance of the log likelihood ratio statistics for the mixture model and related results. In: Proceedings of the Berkeley Conference in Honor of J Neyman and J Kiefer (LM Le Cam, RA Olshen, edn), Wadsworth.

23. Hartigan J (1985) A failure of likelihood asymptotics for normal mixtures. In: Proceedings of the Berkeley Conference in Honor of J Neyman and J Kiefer (LM Le Cam, RA Olshen, edn), Wadsworth.

24. Dacunha-Castelle D, Gassiat E (1999) Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. Ann Stat 27: 1178-1209.

25. Chen H, Chen J (2001) The likelihood ratio test for homogeneity in finite mixture models. Can J Stat 29: 201-215.

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

27. Chambaz A (2006) Testing the order of a model. Ann Stat 34: 1166-1203.

28. McLachlan G (1987) On bootstrapping the likelihood ratio test stastistic for the number of components in a normal mixture. Appl Stat 36: 318-324.

29. Sun J (1993) Tail probabilities of the maxima of Gaussian random fields. Ann Probab 21: 34-71.

30. Chen H, Chen J, Kalbfleisch J (2001) A modified likelihood ratio test for homogeneity in finite mixture models. J R Statist Soc B 63: 19-29.

31. Zhu HT, Zhang H (2004) Hypothesis testing in mixture regression models. J R Statist Soc B 66: 3-16.

32. Chen J, Li P (2009) Hypothesis test for normal mixture models: the EM approach. Ann Stat 37: 2523-2542.

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

34. Charnigo R, Sun J (2004) Testing homogeneity in a mixture distribution via the L^2 distance between competing models. J Am Stat Assoc 99: 488-498.

35. Charnigo R, Sun J (2010) Asymptotic relationships between the D-test and likelihood ratio-type tests for homogeneity. Stat Sin 20: 497-512.

36. Charnigo R, Fan Q, Bittel D, Dai H (2013) Testing unilateral versus bilateral normal contamination. Stat Probab Lett 83: 163-167.

37. Allison DB, Gadbury GL, Heo M, Fernandez JR, Lee CK, et al. (2002) A mixture model approach for the analysis of microarray gene expression data. Comput Stat Data Anal 39: 1-20.

38. Blalock EM, Chen KC, Sharrow K, Herman JP, Porter NM, et al. (2003) Gene microarrays in hippocampal aging: statistical profiling identifies novel processes correlated with cognitive impairment. J Neurosci 23: 3807-3819.

39. Casella G, Berger RL (2002) Statistical Inference. (2^{nd} edn), Duxbury, USA.

40. Mao R, Wang X, Spitznagel EL Jr, Frelin LP, Ting JC, et al. (2005) Primary and secondary transcriptional effects in the developing human Down syndrome brain and heart. Genome Biol 6: R107.