Communications in Statistics—Theory and Methods, 36: 129–141, 2007

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920600966548

Multivariate Analysis

One-Sample Likelihood Ratio Tests for Mixed Data

A. R. DE LEON

Department of Mathematics & Statistics, University of Calgary,
Calgary, Alberta, Canada

We revisit a hypothesis-testing problem recently investigated by de Leon and Carrière (2000). Specifically, we obtain exact likelihood ratio tests of one-sample location hypotheses for multivariate mixed data modeled according to the general location model (Olkin and Tate, 1961). The tests generalize those previously proposed by de Leon and Carrière (2000) for the case of mixed bivariate data. Optimal properties of the tests are briefly studied. Simulations show that the tests are reasonably powerful in detecting differences between the true and hypothesized populations. We illustrate the tests with a few examples, including one concerning data on academic achievement.

Keywords General location model; Level of test; Location hypothesis; Multinomial data; Multivariate normal distribution; Power function; U-distribution; Wishart distribution.

Mathematics Subject Classification Primary 62F03; Secondary 62H15.

1. Introduction

Multivariate data containing mixtures of quantitative and qualitative variables arise frequently in practice. Catalano (1997) gives an example from developmental toxicology where fetal data from laboratory animals include binary, ordered categorical, and continuous outcomes. A model used to study this type of data was introduced by Olkin and Tate (1961), and has since been known as the general location model (GLOM) (Little and Rubin, 1987; Schafer, 1997). It has found numerous applications in multivariate analysis, particularly in mixed data discrimination and classification problems.

Given D categorical $\mathbf{u} = \left(U_1, \ldots, U_D\right)^\top$ and C continuous variables $\mathbf{y} = (Y_1, \ldots, Y_C)^\top$, where the dth categorical variable U_d has s_d categories, so that there are a total of $S = \prod_{d=1}^{D} s_d$ possible patterns of discrete response, or states, for $\mathbf{u}$, a GLOM for the joint distribution of $\mathbf{u}$ and $\mathbf{y}$ assumes that (i) $\mathbf{u}$ falls in state s with probability π_s $\left(\sum_{s=1}^{S} \pi_s = 1\right)$, and (ii) given that $\mathbf{u}$ falls in the sth state, $\mathbf{y}$ has

Received July 6, 2005; Accepted April 7, 2006

Address correspondence to A. R. de Leon, Department of Mathematics & Statistics, University of Calgary, Calgary, AB, Canada T2N 1N4; E-mail: adeleon@math.ucalgary.ca

the multivariate normal distribution with mean $\boldsymbol{\mu}_s$ and covariance matrix $\boldsymbol{\Sigma}$. Recent extensions of the GLOM were given by Liu and Rubin (1998) and Fitzmaurice and Laird (1995).

One particular aspect of mixed data inference that has received little attention so far are the so-called location hypotheses, for which the construction of reasonable statistical tests remains an important and so far unaddressed problem in such applications as quality control (de Leon and Carrière, 2000) and clinical studies (Afifi and Elashoff, 1969). The problem of interest is to test

$$H : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{against } K : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \tag{1}$$

for some specified $\boldsymbol{\theta}_0$, where $\boldsymbol{\theta}^\top = (\boldsymbol{\pi}^\top, \boldsymbol{\mu}^\top)$, with $\boldsymbol{\mu}^\top = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_S^\top)$ the $CS \times 1$ vector of state means. Hypothesis H is referred to in the literature as the one-sample location hypothesis, and much work has been done for the case with continuous data. In practice, the analytic strategy with mixed data has been to perform tests on the parameters separately. This approach entails the problem of multiple significance testing, to which the simplest solution is to adjust the level of each test to control the overall level. de Leon and Carrière (2000) recently showed that such an approach results in substantial loss of power because the correlations between the variables are not utilized explicitly in carrying out the test. They proposed an alternative multivariate approach in the case of mixed bivariate data in which likelihood ratio tests (LRTs) are constructed based on the GLOM. In this article, we extend de Leon and Carrière's (2000) results to the general multivariate case.

The article is organized as follows. We derive the likelihood ratio tests and obtain the exact null and non null distributions of the resulting test statistics in Sec. 2. The consistency and unbiasedness of the test are established in Sec. 3. Section 4 reports the results of a simulation study on the empirical level and power of the LRTs. Examples are provided in Sec. 5 and a brief discussion in Sec. 6 concludes the article.

2. Likelihood Ratio Tests

Suppose $\mathbf{x} = (X_1, \ldots, X_S)^\top$ and $\mathbf{y} = (Y_1, \ldots, Y_C)^\top$ are vectors of binary and continuous variables, respectively, such that X_s is either 1 or 0 and $\sum_{s=1}^S X_s = 1$. The vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be jointly distributed according to the GLOM if and only if $\mathbf{x} \sim p(\mathbf{x}) = \prod_{s=1}^S \pi_s^{x_s}$, and $[\mathbf{y} \mid X_s = 1] \sim \mathcal{N}_C(\boldsymbol{\mu}_s, \Sigma)$, the C-dimensional multivariate normal distribution with mean $\boldsymbol{\mu}_s = (\mu_{1s}, \ldots, \mu_{Cs})^\top$ and covariance matrix Σ. The model is characterized by the vector of state probabilities $\boldsymbol{\pi}^\top = (\pi_1, \ldots, \pi_S)$, the $CS \times 1$ vector of state means $\boldsymbol{\mu}^\top = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_S^\top)$, and the common state covariance matrix Σ. Here, $\pi_s = \Pr(X_s = 1) > 0$ and $\sum_{s=1}^S \pi_s = 1$.

The model above arises for a vector $\mathbf{u} = (U_1, \ldots, U_D)^\top$ consisting of categorical variables where the dth variable U_d has s_d categories, so that there are a total of $S = \prod_{d=1}^D s_d$ possible states for $\mathbf{u}$. In this case, we can define $\mathbf{x}$ as $X_s = 1$ if $\mathbf{u}$ falls in state s and 0 otherwise, and $\sum_{s=1}^S X_s = 1$. For notational convenience, the vector $\mathbf{x}$ for which $X_s = 1$ is denoted by $\mathbf{x}_{(s)}$.

Suppose $(\mathbf{x}_1^\top, \mathbf{y}_1^\top)^\top, \ldots, (\mathbf{x}_N^\top, \mathbf{y}_N^\top)^\top$ are a random sample from the GLOM with parameters $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$. Without loss of generality, we assume $\mathbf{x}_{N_j+1} = \cdots = \mathbf{x}_{N_{j+1}} = \mathbf{x}_{(j+1)}$ so that $\mathbf{y}_{N_j+1}, \ldots, \mathbf{y}_{N_{j+1}}$ are independently and identically distributed as $\mathcal{N}_C(\boldsymbol{\mu}_{j+1}, \boldsymbol{\Sigma})$, $j = 0, 1, \ldots, S-1$, where $N_0 = n_0 = 0$, $N_j = \sum_{s=0}^j n_s$ for

$j = 1, \ldots, S-1$, $N_S = N = \sum_{s=1}^{S} n_s$, and n_s is the number of observations in state $s = 1, \ldots, S$. The likelihood function is then

$$\mathscr{L} = \prod_{s=1}^{S} \pi_s^{n_s} |2\pi\boldsymbol{\Sigma}|^{-N/2} \exp\left[-\frac{1}{2}\sum_{j=0}^{S-1}\sum_{i=N_j+1}^{N_{j+1}} \big(\mathbf{y}_i - \boldsymbol{\mu}_{j+1}\big)^{\top}\boldsymbol{\Sigma}^{-1}\big(\mathbf{y}_i - \boldsymbol{\mu}_{j+1}\big)\right]. \tag{2}$$

Observe that $\mathscr{L}$ consists of two parts, $\mathscr{L}^{(D)}$ and $\mathscr{L}^{(C)}$, the first corresponding to the usual multinomial sample likelihood and the second to that of a multivariate normal sample. Equivalently, the log-likelihood function can be written as

$$\ell = \sum_{s=1}^{S} n_s \log \pi_s - \frac{N}{2}\log|2\pi\boldsymbol{\Sigma}| - \sum_{s=1}^{S}\frac{n_s}{2}\mathrm{tr}\big[\boldsymbol{\Sigma}^{-1}\big(\mathbf{S}_s + \mathbf{d}_s\mathbf{d}_s^{\top}\big)\big], \tag{3}$$

where $\mathrm{tr}(\mathbf{A})$ is the trace of $\mathbf{A}$, $\mathbf{d}_s = \bar{\mathbf{y}}_s - \boldsymbol{\mu}_s$, $\bar{\mathbf{y}}_s$, and $\mathbf{S}_s$ are the sample mean and sample covariance matrix (uncorrected for bias), respectively, of the observations in state $s = 1, \ldots, S$.

Maximum likelihood estimators (MLEs) of unknown parameters of the model are obtained by maximizing either (2) or (3). Because the parameter space is simply the product of the individual spaces of the discrete and continuous parameters, MLEs are obtained by maximizing $\mathscr{L}^{(D)}$ and $\mathscr{L}^{(C)}$ separately. The MLEs of $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ are easily found to be $\hat{\boldsymbol{\pi}}^{\top} = \big(n_1/N, \ldots, n_S/N\big)$, $\hat{\boldsymbol{\mu}}^{\top} = \big(\bar{\mathbf{y}}_1^{\top}, \ldots, \bar{\mathbf{y}}_S^{\top}\big) = \bar{\mathbf{y}}^{\top}$, and $N\widehat{\boldsymbol{\Sigma}} = \sum_{s=1}^{S} n_s\mathbf{S}_s/N = N\mathbf{S}_{pooled}$, provided that $n_s > 0\ \forall s$.

It is easy to see that $\mathrm{E}(\hat{\boldsymbol{\pi}}) = \boldsymbol{\pi}$, $\mathrm{E}(\bar{\mathbf{y}}) = \boldsymbol{\mu}$, and $\mathrm{E}\big(\mathbf{S}_{pooled}\big) = (N-S)\boldsymbol{\Sigma}/N$. Note also that given $\big(n_1, \ldots, n_S\big)^{\top}$, $\bar{\mathbf{y}} \sim \mathscr{N}_{CS}(\boldsymbol{\mu}, \mathbf{D}\otimes\boldsymbol{\Sigma})$, where $\otimes$ is the Kronecker product operator and $\mathbf{D} = \mathrm{diag}\big(1/n_1, \ldots, 1/n_S\big)$. Finally, if $n_s > 0\ \forall s$, then $N\mathbf{S}_{pooled} \sim \mathscr{W}_C(\boldsymbol{\Sigma}, N-S)$ independently of $(n_1, \ldots, n_S)^{\top}$, where $\mathscr{W}_C(\boldsymbol{\Sigma}, N-S)$ is the Wishart distribution with scale matrix $\boldsymbol{\Sigma}$ and $N-S$ degrees of freedom.

2.1. *Case of Known Covariance Matrix*

Consider a GLOM with location parameter $\boldsymbol{\theta}^{\top} = \big(\boldsymbol{\pi}^{\top}, \boldsymbol{\mu}^{\top}\big)$ and known covariance matrix $\boldsymbol{\Sigma}$.

Theorem 2.1. *Consider the hypotheses in* (1).

(i) *The* LRT *is of the form: Reject H if and only if* $\sum_{s=1}^{S}\zeta_s > c$, *for some critical value c, where* $\zeta_s = (-2N/S)\log N - 2n_s\log(\pi_{0s}/n_s) + n_s\mathbf{d}_{0s}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{d}_{0s}$, *with* $\mathbf{d}_{0s} = \bar{\mathbf{y}}_s - \boldsymbol{\mu}_{0s}$.

(ii) *For a level* α *test, the critical value* c_α *is obtained from*

$$\alpha = \sum_{n_1,\ldots,n_S} p\big(n_1, \ldots, n_S \,|\, \boldsymbol{\pi}_0\big)\Pr\big[\chi^2_{d_1} > c\big(n_1, \ldots, n_S\big)\big], \tag{4}$$

where $p(n_1, \ldots, n_S \,|\, \boldsymbol{\pi}_0) = \big(\prod_{s=1}^{S}\pi_{0s}^{n_s}/\prod_{s=1}^{S}n_s!\big)/\sum_{n_1,\ldots,n_S}\big(\prod_{s=1}^{S}\pi_{0s}^{n_s}/\prod_{s=1}^{S}n_s!\big)$ *with the summations taken over all* $\{n_1, \ldots, n_S\}$ *such that* $n_s > 0\ \forall s$ *and* $\sum_{s=1}^{S} n_s = N$, $\chi^2_{d_1}$ *is a* χ^2 *random variable with* $d_1 = CS$ *degrees of freedom, and* $c(n_1, \ldots, n_S) = c_\alpha + 2N\log N + 2\sum_{s=1}^{S} n_s\log\big(\pi_{0s}/n_s\big)$.

(iii) *At any* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, *the power of the LRT is*

$$\Pr\left(\sum_{s=1}^{S} \zeta_s > c_\alpha \,\middle|\, \boldsymbol{\theta}\right) = \sum_{n_1,\ldots,n_S} p\left(n_1, \ldots, n_S \,|\, \boldsymbol{\pi}\right) \Pr\left[\chi^2_{d_1,\Delta(n_1,\ldots,n_S)} > c(n_1, \ldots, n_S)\right], \qquad (5)$$

where $p\left(n_1, \ldots, n_S \,|\, \boldsymbol{\pi}\right)$ *is as in* (ii) *with* $\boldsymbol{\pi}_0 = \boldsymbol{\pi}$, *and* $\chi^2_{d_1\Delta(n_1,\ldots,n_S)}$ *is a noncentral* χ^2 *random variable with* $d_1 = CS$ *degrees of freedom and noncentrality parameter* $\Delta\left(n_1, \ldots, n_S\right) = \sum_{s=1}^{S} n_s \left(\boldsymbol{\mu}_s - \boldsymbol{\mu}_{0s}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_s - \boldsymbol{\mu}_{0s}\right)$.

Proof. When $\boldsymbol{\Sigma}$ is known, the maximized log-likelihood under H is

$$\log \widehat{\mathscr{L}}_H = \sum_{s=1}^{S} n_s \log \pi_{0s} - \frac{N}{2} \log|2\pi\boldsymbol{\Sigma}| - \sum_{s=1}^{S} \frac{n_s}{2} \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\mathbf{S}_s + \mathbf{d}_{0s}\mathbf{d}_{0s}^\top\right)\right].$$

Also, since K places no constraints on $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$ is as given above and

$$\log \widehat{\mathscr{L}}_K = \sum_{s=1}^{S} n_s \log\left(\frac{n_s}{N}\right) - \frac{N}{2} \log|2\pi\boldsymbol{\Sigma}| - \sum_{s=1}^{S} \frac{n_s}{2} \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{S}_s\right).$$

Therefore, H is rejected if and only if $-2\log\lambda = \sum_{s=1}^{S} \zeta_s > c$, for some c, which proves (i).

Noting that $n_1\mathbf{d}_{01}^\top\boldsymbol{\Sigma}^{-1}\mathbf{d}_{01}, \ldots, n_S\mathbf{d}_{0S}^\top\boldsymbol{\Sigma}^{-1}\mathbf{d}_{0S}$ are independent and identically distributed χ^2 random variables each with C degrees of freedom, (4) in (ii) is obtained by using Theorem 2.5.2 of Mardia et al. (1979, p. 39) and the fact that $(n_1, \ldots, n_S)^\top$ has a truncated multinomial distribution with parameters N and $\boldsymbol{\pi}_0$ under H, subject to the condition $0 < n_s < N$ $\forall s$.

Part (iii) is proved by noting that $n_1\mathbf{d}_{01}^\top\boldsymbol{\Sigma}^{-1}\mathbf{d}_{01}, \ldots, n_S\mathbf{d}_{0S}^\top\boldsymbol{\Sigma}^{-1}\mathbf{d}_{0S}$ are independently distributed noncentral χ^2 random variables each with C degrees of freedom and respective noncentrality parameters $n_1\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{01}\right)^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{01}\right), \ldots,$ $n_S\left(\boldsymbol{\mu}_S - \boldsymbol{\mu}_{0S}\right)^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_S - \boldsymbol{\mu}_{0S}\right)$, and that for some given $\boldsymbol{\pi}$, $\left(n_1, \ldots, n_s\right)^\top$ has a truncated multinomial distribution with parameters N and $\boldsymbol{\pi}$. ☐

The condition $n_s > 0$ $\forall s$ (i.e., each state has at least one observation) is necessary so that all unknown parameters will be estimable. This results in a truncated multinomial distribution for $\left(n_1, \ldots, n_S\right)^\top$.

Note that if $S = 1$, the LRT statistic in Theorem 2.1 becomes $N\left(\bar{\mathbf{y}} - \boldsymbol{\mu}_0\right)^\top\boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{y}} - \boldsymbol{\mu}_0\right)$, which is the one-sample LRT for the multivariate normal distribution with known covariance matrix (Mardia et al., 1979, p. 124). Thus, Theorem 2.1 generalizes the latter to the case of mixed binary and continuous data.

2.2. *Case of Unknown Covariance Matrix*

Consider now a GLOM with location parameter $\boldsymbol{\theta}^\top = \left(\boldsymbol{\pi}^\top, \boldsymbol{\mu}^\top\right)$ and unknown covariance matrix $\boldsymbol{\Sigma}$. This is usually the case in many applied studies where knowledge about $\boldsymbol{\Sigma}$ is not available.

Theorem 2.2. *Consider the hypotheses in* (1).

(i) *The LRT is of the form: Reject H if and only if* $\sum_{s=1}^{S} \zeta_s > c$, *for some critical value c, where* $\zeta_s = a(n_1, \ldots, n_S; \boldsymbol{\pi}_0)\ \left[1/S + n_S \mathbf{d}_{0s}^{\top} \mathbf{S}_{pooled}^{-1} \mathbf{d}_{0s}/N\right]$ *and* $a(n_1, \ldots, n_S; \boldsymbol{\pi}_0) = N^{-2} \prod_{s=1}^{S} (n_s/\pi_{0s})^{2n_s/N}$.

(ii) *For a level* α *test, the critical value* c_α *is obtained from*

$$\alpha = \sum_{n_1, \ldots, n_s} p(n_1, \ldots, n_S \mid \boldsymbol{\pi}_0) \Pr\left[U^{(M)} > c(n_1, \ldots, n_S)\right], \tag{6}$$

where the summation is taken over all $\{n_1, \ldots, n_s\}$ *such that* $n_s > 0\ \forall s$ *and* $\sum_{s=1}^{S} n_s = N$, $p(n_1, \ldots, n_S \mid \boldsymbol{\pi}_0)$ *is defined in Theorem* 2.1, $c(n_1, \ldots, n_S) = a(n_1, \ldots, n_S; \boldsymbol{\pi}_0)c_\alpha - 1$, *and* $U^{(M)}$ *has the same distribution as that, under H and conditional on* $(n_1, \ldots, n_S)^{\top}$, *of the sum of non zero roots* $\xi_1, \ldots, \xi_M$ *of* $\left|\sum_{s=1}^{S} n_S \mathbf{d}_{0s}\mathbf{d}_{0s}^{\top}/N - \xi \mathbf{S}_{pooled}\right| = 0$.

Proof. The MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are, respectively, $\boldsymbol{\mu}_0$ and $\widehat{\boldsymbol{\Sigma}}_0 = \mathbf{S}_{pooled} + \sum_{s=1}^{S} n_s \mathbf{d}_{0s}\mathbf{d}_{0s}^{\top}/N$ under H, and $\bar{\mathbf{y}}$ and $\mathbf{S}_{pooled}$ under K. Now from (2), it can be deduced that

$$\widehat{\mathscr{L}}_H = \prod_{s=1}^{S} \pi_{0s}^{n_s} |2\pi\widehat{\boldsymbol{\Sigma}}_0|^{-N/2} e^{-CN/2},$$

$$\widehat{\mathscr{L}}_K = \prod_{s=1}^{S} \left(\frac{n_s}{N}\right)^{n_s} \left|2\pi \mathbf{S}_{pooled}\right|^{-N/2} e^{-CN/2}.$$

Thus, $\lambda^{-2/N} = a(n_1, \ldots, n_S; \boldsymbol{\pi}_0)|\widehat{\boldsymbol{\Sigma}}_0|/|\mathbf{S}_{pooled}|$, and H is rejected for large values of $\sum_{s=1}^{S} \zeta_s$. This proves (i).

Now consider (ii), and assume H holds and $(n_1, \ldots, n_S)^{\top}$ is fixed. Noting that $\sqrt{n_s}\mathbf{d}_{0s} \sim \mathscr{N}_C(\mathbf{0}, \boldsymbol{\Sigma})$, it then follows that $\mathbf{M}_1 = \sum_{s=1}^{S} n_s \mathbf{d}_{0s}\mathbf{d}_{0s}^{\top} \sim \mathscr{W}_C(\boldsymbol{\Sigma}, S)$, independently of $\mathbf{M}_2 = N\mathbf{S}_{pooled} \sim \mathscr{W}_C(\boldsymbol{\Sigma}, N - S)$. Rewriting $\lambda^{-2/N} = a(n_1, \ldots, n_S; \boldsymbol{\pi}_0)(1 + U^{(M)})$, with $U^{(M)} = \mathrm{tr}(\mathbf{M}_1\mathbf{M}_2^{-1})$, and using the fact that the trace of a matrix is equal to the sum of its eigenvalues, it is clear that $U^{(M)}$ has the same distribution as $\sum_{m=1}^{M} \xi_m$, where $\xi_1 \neq 0, \ldots, \xi_M \neq 0$ satisfy $0 = \left|\mathbf{M}_1\mathbf{M}_2^{-1} - \xi \mathbf{I}_C\right|$.

Finally, the expression in (6) is obtained by noting that $(n_1, \ldots, n_S)^{\top}$ has a truncated multinomial distribution with parameters N and $\boldsymbol{\pi}_0$ under H, subject to the condition that $0 < n_s < N\ \forall s$. □

It is assumed in Theorem 2.2 that the Wishart distribution is nonsingular, i.e., $N \geq S + C$, so that $\mathbf{S}_{pooled}^{-1}$ exists with probability 1. This holds if $\forall s$, $n_s \geq C$ so that $n_s\mathbf{S}_s$ has a nonsingular Wishart distribution, $s = 1, \ldots, S$. In this case, $M = \text{minimum}(S, N - S)$.

The statistic $U^{(M)}$ is known in the literature as the Lawley–Hotelling trace statistic (Seber, 1984). Approximations to its null, or central, distribution are given by Seber (1984, p. 39). A noncentral distribution arises under K, since $\sum_{s=1}^{S} n_s \mathbf{d}_{0s}\mathbf{d}_{0s}^{\top}$ has a noncentral Wishart distribution in this case.

Note that if $S = 1$, the LRT statistic in Theorem 2.2 becomes $(1 + \mathbf{d}_0^{\top}\mathbf{S}_{pooled}^{-1}\mathbf{d}_0)$, which is the one-sample LRT for a multivariate normal distribution with unknown $\boldsymbol{\Sigma}$ (Mardia et al., 1979, p. 125). Thus, Theorem 2.2 generalizes the latter to the case of mixed binary and continuous data.

Observe that while the LRTs in Theorems 2.1 and 2.2 are derived from the unconditional likelihood function, they are assessed conditionally on the total counts across states for the binary data. This conditioning is computationally convenient because it reduces the burden of getting a critical value or p-value to summing tail probabilities over a set of counts for a contingency table. An important consideration in the unknown covariance case (i.e., Theorem 2.2) is the distribution of $U^{(M)}$. Bilodeau and Brenner (1999) provides an exact and computationally efficient way of evaluating this distribution, which we use in the simulations in Sec. 4.

We establish the consistency and unbiasedness of the LRTs in Theorems 2.1 and 2.2 in the next section.

3. Properties of Likelihood Ratio Tests

Property 3.1. The LRT in Theorem 2.1 is consistent. The same holds for that in Theorem 2.2, provided $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.

Proof. For Theorem 2.1, note that $-2N\log N - 2\sum_{s=1}^{S} n_s \log(\pi_{0s}/n_s) \rightarrow 0$ almost surely as $N \rightarrow \infty$. Therefore it follows that $c(n_1, \ldots, n_S) \rightarrow \infty$, and consistency follows.

For Theorem 2.2, c_α satisfies $\Pr\left(\sum_{s=1}^{S} \xi_s > c_\alpha \,|\, \boldsymbol{\theta}_0\right) = \alpha$. Since $a(n_1, \ldots, n_S; \boldsymbol{\pi}_0) \rightarrow 1$ and $\chi^2_{d_2}/d_2 \rightarrow 1$ almost surely, it follows that $d_2(c_\alpha - 1) \rightarrow c_0$ such that $\Pr(\chi^2_{d_1} \leq c_0) = 1 - \alpha$, where $d_1 = CS$ and $d_2 = N - S - C + 1$. Thus, $c_\alpha \rightarrow 1$ as $N \rightarrow \infty$. By the strong law of large numbers,

$$\frac{d_2}{Nd_1} \sum_{s=1}^{S} n_s \mathbf{d}_{0s}^\top \mathbf{S}_{pooled}^{-1} \mathbf{d}_{0s} \rightarrow \sum_{s=1}^{S} (\boldsymbol{\mu}_s - \boldsymbol{\mu}_{0s})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_s - \boldsymbol{\mu}_{0s}),$$

almost surely. For $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, $\sum_{s=1}^{S} (\boldsymbol{\mu}_s - \boldsymbol{\mu}_{0s})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_s - \boldsymbol{\mu}_{0s}) > 0$ and the LRT is consistent. □

Property 3.2. The LRTs in Theorems 2.1 and 2.2 are both unbiased.

Proof. Because the distribution function of $\chi^2_{d_1,\Delta(n_1,\ldots,n_s)}$ is decreasing in $\Delta(n_1, \ldots, n_S)$ $\forall n_1, \ldots, n_S$ (Johnson and Kotz, 1970, p. 135), it follows that

$$\Pr\left[\chi^2_{d_1,\Delta(n_1,\ldots,n_S)} \leq c\right] \leq \Pr\left[\chi^2_{d_1} \leq c\right],$$

for any constant c. Therefore,

$$\mathrm{E}_{\boldsymbol{\pi}}\left\{\Pr\left[\chi^2_{d_1,\Delta(n_1,\ldots,n_S)} > c(n_1, \ldots, n_S)\right]\right\} \geq \mathrm{E}_{\boldsymbol{\pi}_0}\left\{\Pr\left[\chi^2_{d_1} > c(n_1, \ldots, n_S)\right]\right\},$$

where the expectations are taken with respect to the truncated multinomial distributions with parameters $\boldsymbol{\pi}$ and $\boldsymbol{\pi}_0$. This implies that the power achieves its minimum at $\boldsymbol{\theta}_0$, and the LRT in Theorem 2.1 is unbiased.

Given $n_1, \ldots, n_S$, Perlman and Olkin (1980) showed that the Lawley–Hotelling trace criterion is unbiased. Using the same argument used above, the unbiasedness of the LRT in Theorem 2.2 now follows from the usual properties of expectations. □

The above results include that for the case $C = 1$ and $S = 2$, details of which are found in de Leon and Carrière (2000).

4. Simulation Study

To evaluate the performance of the LRT in terms of its nominal level and power, we conducted a series of simulation experiments using mixed data generated from the GLOM with $C = 2$ and $S = 4$. This model corresponds to the case of two continuous variables $\mathbf{y}^\top = (Y_1, Y_2)$ and the binary vector $\mathbf{x}^\top = (X_1, \ldots, X_4)$. The parameter is then $\boldsymbol{\theta}^\top = \left(\boldsymbol{\pi}^\top, \boldsymbol{\mu}^\top\right)$, where $\boldsymbol{\pi}^\top = \left(\pi_1, \ldots, \pi_4\right)$, $\boldsymbol{\mu}^\top = \left(\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top\right)$ with $\boldsymbol{\mu}_s^\top = \left(\mu_{1s}, \mu_{2s}\right)$ the sth $(s = 1, \ldots, 4)$ state mean vector of $\mathbf{y}$. Data were simulated from this GLOM with sample sizes $N = 50$, 100, and 160, and this was repeated 10,000 times. The hypothesized value $\boldsymbol{\theta}_0$ under the null hypothesis H in (1) was specified as $\boldsymbol{\pi}_0^\top = (0.25, 0.25, 0.25, 0.25)$, $\boldsymbol{\mu}_{01} = \boldsymbol{\mu}_{02} = \boldsymbol{\mu}_{03} = \boldsymbol{\mu}_{04} = (0, 0)^\top$, and the covariance matrix was taken to be

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

To assess the level and power of the LRT, the following five cases were considered:

(0) no differences between hypothesized and true populations;
(a) there is difference between hypothesized and true populations only with respect to binary vector $\mathbf{x}$;
(b) there is difference between hypothesized and true populations only with respect to continuous vector $\mathbf{y}$;
(c) hypothesized and true populations are different with respect to both variable types.

Note that (0) corresponds to the case where H is true. For (a), we considered the cases (a_1) $\boldsymbol{\pi}^\top = (0.2, 0.3, 0.25, 0.25)$, and (a_2) $\boldsymbol{\pi}^\top = (0.2, 0.3, 0.1, 0.4)$. For (b), we have (b_1) $\boldsymbol{\mu}_1^\top = (0.2, 0.2)$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = (0, 0)^\top$; (b_2) $\boldsymbol{\mu}_1^\top = (0.2, 0.2)$, $\boldsymbol{\mu}_2^\top = (-0.2, -0.2)$, $\boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = (0, 0)^\top$; (b_3) $\boldsymbol{\mu}_1^\top = (0.2, 0.2)$, $\boldsymbol{\mu}_2^\top = (-0.2, -0.2)$, $\boldsymbol{\mu}_3^\top = (0.4, 0.4)$, $\boldsymbol{\mu}_4^\top = (0, 0)$; and (b_4) $\boldsymbol{\mu}_1^\top = (0.2, 0.2)$, $\boldsymbol{\mu}_2^\top = (-0.2, -0.2)$, $\boldsymbol{\mu}_3^\top = (0.4, 0.4)$, $\boldsymbol{\mu}_4^\top = (-0.4, -0.4)^\top$. For (c), we have (c_1) $\boldsymbol{\pi}^\top = (0.2, 0.3, 0.25, 0.25)$ and $\boldsymbol{\mu}_1^\top = (0.2, 0.2)$ $\boldsymbol{\mu}_2^\top = (-0.2, -0.2)$, $\boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = (0, 0)^\top$; (c_2) $\boldsymbol{\pi}^\top = (0.2, 0.3, 0.25, 0.25)$ and $\boldsymbol{\mu}_1^\top = (0.2, 0.2)$, $\boldsymbol{\mu}_2^\top = (-0.2, -0.2)$, $\boldsymbol{\mu}_3^\top = (0.4, 0.4)$, $\boldsymbol{\mu}_4^\top = (-0.4, -0.4)^\top$; (c_3) $\boldsymbol{\pi}^\top = (0.2, 0.3, 0.1, 0.4)$, and $\boldsymbol{\mu}_1^\top = (0.2, 0.2)$, $\boldsymbol{\mu}_2^\top = (-0.2, -0.2)$, $\boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = (0, 0)^\top$; and (c_4) $\boldsymbol{\pi}^\top = (0.2, 0.3, 0.1, 0.4)$ and $\boldsymbol{\mu}_1^\top = (0.2, 0.2)$, $\boldsymbol{\mu}_2^\top = (-0.2, -0.2)$, $\boldsymbol{\mu}_3^\top = (0.4, 0.4)$, $\boldsymbol{\mu}_4^\top = (-0.4, -0.4)^\top$.

Tables 1 and 2 display the simulation results, the former for the case of known covariance matrix and the latter for that of unknown covariance matrix. Critical values were obtained by the bisection method in both cases, and the power function in (5) was used to evaluate the theoretical power of the LRT in the case of known covariance matrix. Observe from Table 1 that the empirical power values are quite close to their theoretical values, as expected.

Table 1
Empirical level and power of LRT in Theorem 2.1 for $C = 2$ and $S = 4$ at 5% level, based on 10,000 Monte Carlo samples

		Sample size (N)		
Source of difference	Case	50	100	160
No difference	(0)	0.0506	0.0511	0.0503
Difference with respect only to **x**	(a_1)	0.0774 (0.0800)	0.1190 (0.1168)	0.1685 (0.1685)
	(a_2)	0.5511 (0.5599)	0.9204 (0.9160)	0.9943 (0.9946)
Difference with respect only to **y**	(b_1)	0.0709 (0.0693)	0.0947 (0.0917)	0.1167 (0.1222)
	(b_2)	0.0914 (0.0914)	0.1470 (0.1441)	0.2169 (0.2189)
	(b_3)	0.2085 (0.2048)	0.4188 (0.4155)	0.6454 (0.6528)
	(b_4)	0.3456 (0.3427)	0.6731 (0.6753)	0.8973 (0.9010)
Differences with respect to **x** & **y**	(c_1)	0.1355 (0.1299)	0.2348 (0.2391)	0.3942 (0.3890)
	(c_2)	0.3953 (0.3973)	0.7557 (0.7523)	0.9404 (0.9440)
	(c_3)	0.6304 (0.6261)	0.9542 (0.9490)	0.9986 (0.9980)
	(c_4)	0.8178 (0.8199)	0.9934 (0.9940)	1.0000 (1.0000)

Figures in parentheses are theoretical power values, calculated from (5).

From Theorem 2.2, note that the LRT statistic can be written as

$$\begin{aligned}\sum_{s=1}^{S}\zeta_s &= a(n_1,\ldots,n_S;\boldsymbol{\pi}_0)\frac{|\mathbf{M}_1+\mathbf{M}_2|}{|\mathbf{M}_2|}\\ &= a(n_1,\ldots,n_S;\boldsymbol{\pi}_0)U^{-1},\end{aligned}$$

where $U \sim U(C; S, N-S)$, the U-distribution of dimension C with degrees of freedom C and $N-S$ (Seber, 1984). Bilodeau and Brenner (1999) give an **S**-plus program for calculating exact probabilities of U via the fast Fourier transform. We used this to obtain critical values from (6) for the case of unknown covariance matrix.

The samples were checked to ensure that $n_s \geq 1$ for every state $s = 1, \ldots, 4$. This is necessary so that all the parameters are estimable and should work well when the sample size N is appreciably larger than the total number of states S. When this is not the case, a few states may be collapsed to reduce the number of parameters. Alternatively, linear restrictions may be imposed on the model as in Schafer (1997), and the hypotheses in (1) are then expressed in terms of the restricted parameters.

Table 2

Empirical level and power of LRT in Theorem 2.2 for $C = 2$ and $S = 4$ at 5% level, based on 10,000 Monte Carlo samples

		Sample size (N)		
Source of difference	Case	50	100	160
No difference	(0)	0.0488	0.0489	0.0493
Difference with respect only to **x**	(a_1)	0.0643	0.1088	0.1552
	(a_2)	0.5288	0.9071	0.9943
Difference with respect only to **y**	(b_1)	0.0571	0.0878	0.1142
	(b_2)	0.0717	0.1268	0.2005
	(b_3)	0.1590	0.3814	0.6172
	(b_4)	0.2662	0.6113	0.8832
Differences with respect to **x** & **y**	(c_1)	0.10865	0.2205	0.3652
	(c_2)	0.3243	0.7096	0.9323
	(c_3)	0.5935	0.9434	0.9979
	(c_4)	0.7743	0.9936	0.9999

Four observations are apparent from the tables. First, the power of the test increases with the sample size N. Second, while the empirical levels are slightly different from the nominal 5% level, 95% confidence intervals based on the binomial distribution indicate they are all within the margin of error. Third, the LRT for the unknown covariance matrix case is less powerful than that for the case of known covariance matrix. Last, the power of the test is higher when differences exist with respect to both continuous and discrete variables than when the difference is only with respect to just one variable type; moreover, the power increases with the distance between the null and true values of the parameters.

5. Examples

Consider the simplest GLOM with $S = 2$ and $C = 1$. Suppose $\mathbf{x} = \mathbf{x}_{(1)}$ and $\mathbf{x} = \mathbf{x}_{(2)}$ have respective probabilities p and $q = 1 - p$, and the conditional distributions of Y for $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$ are assumed to be $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$, respectively. Given a random sample of size N, the likelihood function is given by $\mathcal{L} = p^n q^{N-n} (2\pi\sigma^2)^{-N/2} \exp[-Q(\mu_1, \mu_2)/(2\sigma^2)]$, where $Q(\mu_1, \mu_2) = \sum_{i=1}^{n} (Y_i - \mu_1)^2 + \sum_{i=n+1}^{N} (Y_i - \mu_2)^2$, and it is assumed that the first n observations have $\mathbf{x} = \mathbf{x}_{(1)}$. The following two examples apply Theorems 2.1 and 2.2 to this model.

Example 5.1. Suppose that σ^2 is known. Taking $S = 2$ and $C = 1$ in Theorem 2.1, the LRT of $H : (p, \mu_1, \mu_2)^\top = (p_0, \mu_{01}, \mu_{02})^\top$ against $K : (p, \mu_1, \mu_2)^\top \neq (p_0, \mu_{01}, \mu_{02})^\top$ in this case rejects H if and only if

$$-N \log N - n \log\left(\frac{p_0}{n}\right) - (N - n) \log\left(\frac{q_0}{N - n}\right) + n \frac{(\overline{Y}_1 - \mu_{01})^2}{\sigma^2} + (N - n) \frac{(\overline{Y}_2 - \mu_{02})^2}{\sigma^2} > \frac{c_\alpha}{2},$$

for some α-critical value c_α obtained from

$$\alpha = \frac{1}{1 - p_0^N - q_0^N} \sum_{n=1}^{N-1} \binom{N}{n} p_0^n q_0^{N-n} \Pr\left[\chi_2^2 > c(n)\right], \tag{7}$$

where $c(n) = c_\alpha / [-2N \log N - 2n \log(p_0/n) - 2(N-n) \log\{q_0/(N-n)\}]$.

The power of the test at $(p, \mu_1, \mu_2)^\top \neq (p, \mu_{01}, \mu_{02})^\top$ is

$$\frac{1}{1 - p^N - q^N} \sum_{n=1}^{N-1} \binom{N}{n} p^n q^{N-n} \Pr\left[\chi_{2,\Delta(n)}^2 > c(n)\right], \tag{8}$$

where $\Delta(n) = \left[n(\mu_1 - \mu_{01})^2 + (N-n)(\mu_2 - \mu_{02})^2\right] / \sigma^2$.

Example 5.2 (de Leon and Carrière, 2000). Suppose now that σ^2 is unknown. The LRT of $H: (p, \mu_1, \mu_2)^\top = (p_0, \mu_{01}, \mu_{02})^\top$ against $K: (p, \mu_1, \mu_2)^\top \neq (p_0, \mu_{01}, \mu_{02})^\top$ in this case rejects H if and only if

$$a(n; p_0)\left[1 + \frac{n(\overline{Y}_1 - \mu_{01})^2 + (N-n)(\overline{Y}_2 - \mu_{02})^2}{Q(\overline{Y}_1, \overline{Y}_2)}\right] > c_\alpha,$$

for some α-critical value c_α, obtained as the solution to

$$\alpha = \frac{1}{1 - p_0^N - q_0^N} \sum_{n=1}^{N-1} \binom{N}{n} p_0^n q_0^{N-n} \Pr[F_{2,N-2} > c(n)], \tag{9}$$

where $a(n; p_0) = (1/N^2)(n/p_0)^{2n/N}[(N-n)/q_0]^{2-2n/N}$, $c(n) = (N-2)[c_\alpha / a(n; p_0) - 1]/2$, and $F_{2,N-2}$ is an F random variable with $(2, N-2)$ degrees of freedom.

The power of the test at $(p, \mu_1, \mu_2)^\top \neq (p_0, \mu_{01}, \mu_{02})^\top$ is given by

$$\frac{1}{1 - p^N - q^N} \sum_{n=1}^{N-1} \binom{N}{n} p^n q^{N-n} \Pr\left[F_{2,N-2}^{\Delta(n)} > c(n)\right], \tag{10}$$

where $F_{2,N-2}^{\Delta(n)}$ is a noncentral $F_{2,N-2}$ random variable with noncentrality parameter $\Delta(n) = [n(\mu_1 - \mu_{01})^2 + (N-n)(\mu_2 - \mu_{02})^2]/\sigma^2$.

Figure 1 displays several plots of the power function in (10) for various fixed true values (μ_1, μ_2) and null values (μ_{01}, μ_{02}) of the state means, with $p_0 = 0.5$. Plots (a) and (b) have $N = 25$ while (c) and (d) have $N = 50$. It is clear from Fig. 1 that the power of the LRT increases with N as well as with the distance between the null and true values of the state means.

Similarly, contour plots of (10) for a range of values of (μ_1, μ_2) are shown in Fig. 2, with $p = 0.5$, $\sigma^2 = 1$, and $N = 25$. The null values considered are $\mu_{01} = 0$, $\mu_{02} = 0.5$, and $p_0 = 0.25, 0.5$. The contour levels are generally high (especially for the top plot in Fig. 2) around, but decreasing as they approach, the point $(\mu_1, \mu_2) = (0, 0.5)$. This indicates a funnel-like shape for the power surface, with saddle point at $(\mu_1, \mu_2) = (0, 0.5)$.

A simple illustration of the LRT is provided by data previously analyzed by Mardia et al. (1979), Krzanowski (1983), and Morales et al. (1998). We apply the test to the data in the following example.

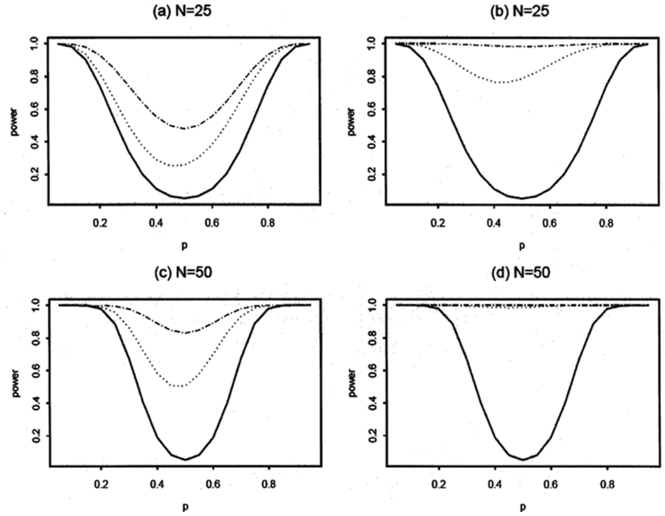

Figure 1. Plots of power function of LRT with $C = 1$, $S = 2$ and unknown $\sigma^2 = 1$, for $p_0 = 0.5$, $\alpha = 0.05$, and fixed state means. For (a) and (c), the solid lines correspond to $\mu_1 = \mu_{01} = 0$, $\mu_2 = \mu_{02} = 0.5$; dotted lines to $\mu_1 = 0$, $\mu_{01} = 0.5$, $\mu_2 = \mu_{02} = 0.5$; dashed lines to $\mu_1 = \mu_{02} = 0$, $\mu_2 = \mu_{02} = 0.5$. For (b) and (d), the solid lines correspond to $\mu_1 = \mu_{01} = 0$, $\mu_2 = \mu_{02} = 1$; dotted lines to $\mu_1 = 0$, $\mu_{01} = 1$, $\mu_2 = \mu_{02} = 1$; dashed lines to $\mu_1 = \mu_{02} = 0$, $\mu_2 = \mu_{02} = 1$.

Example 5.3 (Mardia et al., 1979)**.** The data were collected from 382 university students on the number of GCE A-levels taken and the students' average grades. The average A-level grade obtained is a continuous variable and the number of A-levels taken is a categorical variable with $S = 3$ states (i.e., 2, 3, or 4 A-levels). The students were also grouped according to their final degree classification into seven groups presented in Table 1 of Krzanowski (1983). Let Y denote the average A-level grade and let $\mathbf{x}^\top = (X_1, X_2, X_3)$ represent the number of A-levels taken, with $X_1 = 1$ if 4 A-levels taken and 0 otherwise, $X_2 = 1$ if 3 A-levels taken and 0 otherwise, and $X_3 = 1$ if 2 A-levels taken and 0 otherwise. We focus on group L made up of $N = 22$ students who took four years to complete a three-year degree and test the null hypothesis $H: (\pi_1, \pi_2, \pi_3, \mu_1, \mu_2, \mu_3)^\top = (0.03, 0.896, 0.075, 3.625, 3.555, 3.8)^\top$ against a two-sided alternative. Note that the null parameter values are those for group II(i) made up of students who finished Upper Second Class degrees. Our interest is in determining whether group L is significantly different from group II(i). Repeated one-sample t-tests on the state means lead to acceptance of the hypotheses of no difference yielding p-values of 0.0708 for $\mu_1 = 3.625$, 0.0186 for $\mu_2 = 3.555$, 0.046 for $\mu_3 = 3.8$; an exact χ^2-goodness-of-fit test for the multinomial probabilities yielded a p-value of 0.01354. At a Bonferonni-adjusted overall 5% level (i.e., each test has level 0.0125), these p-values indicate acceptance of the hypotheses of no difference between the true and hypothesized populations.

Applying now the LRT in Theorem 2.2 to group L, we obtained a test statistic value of 2.7655 and a p-value of 0.00064, leading to rejection of H. We thus conclude that group L is significantly different from the reference group II(i). If group II(ii), made up of students who finished Lower Second Class degrees, is made

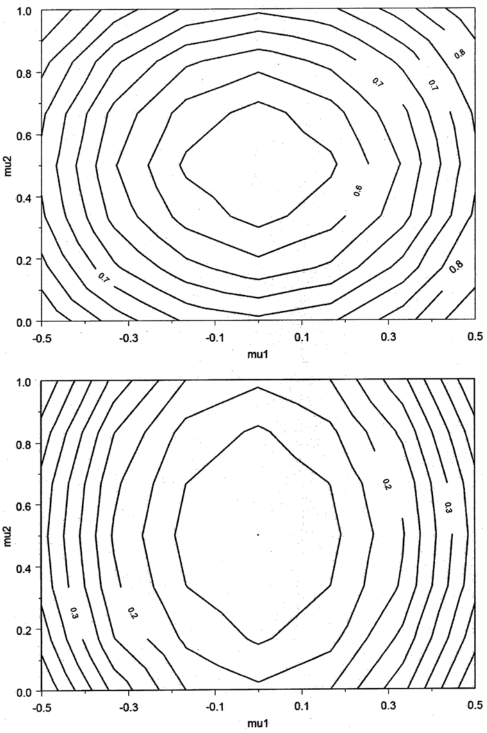

Figure 2. Contour plots of power function of LRT in Example 2 with $C = 1$, $S = 2$ and unknown $\sigma^2 = 1$, for $N = 25$ and $p = 0.5$ at $\alpha = 0.05$, where $\left(\mu_{01}, \mu_{02}\right) = (0, 0.5)$ and $p_0 = 0.25$ (top), 0.5 (bottom). Note that H is false for $p_0 = 0.25$ (top) while it is true for $p_0 = 0.5$ (bottom) at $(\mu_1, \mu_2) = (0, 0.5)$.

the reference group, both the LRT and multiple testing lead to acceptance of the hypothesis of no difference.

The results obtained in this example illustrate the usefulness of the LRT in unifying similar results from a more traditional analysis via repeated multiple tests, by testing only one statistical hypothesis. In addition, by modeling the mixed variables using the GLOM, the LRT is able to explicitly account for correlation between the number of A-levels taken and the average A-level grade, resulting in a more powerful analysis.

6. Discussion

This article was concerned with one-sample tests of location for mixed multivariate data distributed according to the GLOM. Likelihood ratio tests were obtained and their exact distributions were derived. By modeling the joint distribution of the mixed variables by the GLOM, the resulting LRTs can be viewed as extensions of classical LRTs in the one-sample problem based on normal distributions. These LRTs provide global tests of location hypotheses and thus avoid the problem of multiple testing. Simulations show them to have reasonably high power in various cases.

The likelihood ratio approach was employed to construct global tests of mixed data location hypotheses because it allows for a general non-ad hoc approach of simultaneously accounting for both the discrete (i.e., multinomial) and continuous variables in the data. The approach parallels that of Afifi and Elashoff (1969) and is an alternative to the dissimilarity-based tests proposed by Morales et al. (1998). These tests, it should be noted, are all asymptotic, unlike the exact LRTs derived in the article.

The test statistics are similar to their continuous case counterparts and are simple and easy to calculate. In addition, critical values of the tests can be readily obtained by conventional methods. Alternatively, one may resort to using asymptotic distributions to carry out the tests. Another alternative is suggested by the decomposition of the likelihood function into discrete and continuous components. An overall test can then be constructed by adding, say, a χ^2-goodness-of-fit test statistic to a suitably scaled version of Hotelling's T^2 statistic. The size and power of such a test should be explored vis-a-vis those of the LRTs proposed in the article.

Acknowledgments

The author was supported by a Studentship Award from the Alberta Heritage Foundation for Medical Research (AHFMR) and by a grant from the Natural Sciences and Engineering Research Council of Canada. He is grateful to Yongtao Zhu for computational assistance and to the referee for very helpful comments which greatly improved the form and content of the article.

References

Afifi, A. A., Elashoff, R. M. (1969). Multivariate two sample tests with dichotomous and continuous variables. I. The location model. *Ann. Math. Stat.* 40:290–298.

Bilodeau, M., Brenner, D. (1999). *Theory of Multivariate Statistics*. New York: Wiley & Sons.

Catalano, P. J. (1997). Bivariate modeling of clustered continuous and ordered categorical outcomes. *Statist. Med.* 16:883–900.

de Leon, A. R., Carrière, K. C. (2000). On the one-sample location hypothesis for mixed bivariate data. *Commun. Statist. Theor. Meth.* 29(11):2573–2581.

Fitzmaurice, G. M., Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *J. Amer. Statist. Assoc.* 90:845–852.

Johnson, N. L., Kotz, S. (1970). *Continuous Univariate Distributions*. Vol. 2. New York: Houghton Mifflin.

Krzanowski, W. J. (1983). Distance between populations using mixed continuous and categorical variables. *Biometrika* 70:235–243.

Little, R. J., Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley & Sons.

Liu, C., Rubin, D. B. (1998). Ellipsoidally symmetric extensions of the general location model for mixed categorical and continuous data. *Biometrika* 85:673–688.

Mardia, K. V., Kent, J. T., Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

Morales, D., Pardo, L., Zografos, K. (1998). Informational distances and related statistics in mixed continuous and categorical variables. *J. Statist. Plann. Infer.* 75:47–63.

Olkin, I., Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Stat.* 32:448–465 (correction in 36:343–344).

Perlman, M. D., Olkin, I. (1980). Unbiasedness of invariant tests for MANOVA and other multivariate problems. *Ann. Statist* 8:1326–1341.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. New York: Chapman & Hall.

Seber, G. A. F. (1984). *Multivariate Observations*. New York: Wiley & Sons.