AISM (2006) 58: 707–719
DOI 10.1007/s10463-006-0043-y

Biao Zhang

A partial empirical likelihood based score test under a semiparametric finite mixture model

Received: 14 January 2005 / Revised: 24 October 2005 / Published online: 11 May 2006

Abstract We propose a score statistic to test the null hypothesis that the two-component density functions are equal under a semiparametric finite mixture model. The proposed score test is based on a partial empirical likelihood function under an I-sample semiparametric model. The proposed score statistic has an asymptotic chi-squared distribution under the null hypothesis and an asymptotic noncentral chi-squared distribution under local alternatives to the null hypothesis. Moreover, we show that the proposed score test is asymptotically equivalent to a partial empirical likelihood ratio test and a Wald test. We present some results on a simulation study.

Keywords Biased sampling problem · Chi-squared · Consistency · Local alternative · Maximum likelihood · Mixture model · Partial empirical likelihood · Power · Score function · Score statistic · Semiparametric selection bias model · Wald test

1 Introduction

Finite mixture models have been used extensively in a wide variety of important practical situations where data can be viewed as arising from two or more populations mixed in varying proportions and have become the most widely used statistical tool in the analysis of heterogeneous data; see for example Titterington et al. (1985), McLachlan and Basford (1988), Titterington (1990), Lindsay (1995), and McLachlan and Peel (2001). In the literature, the component density functions of a finite mixture model are usually assumed either to be completely specified so that mixing proportions are the only parameters to be estimated or to be specified up to a number of unknown parameters that have to be estimated along with mixing

B. Zhang
Department of Mathematics, The University of Toledo, Toledo, OH 43606, USA
E-mail: bzhang@utnet.utoledo.edu

proportions; see for example Hosmer (1973) and Murray and Titterington (1978). In the context of nonparametric estimation of mixture proportions without any parametric assumptions on component density functions, Hall and Titterington (1984) proposed efficient nonparametric estimation of mixture proportions by constructing a sequence of multinomial approximations and related maximum likelihood estimators. In a different approach using the logistic method, Anderson (1979) proposed a semiparametric finite mixture model in which the log ratio of the two-component density functions h and g is linear in data so that

$$h(x) = \exp(\alpha + \beta^{\tau} x) g(x), \tag{1}$$

where α is a scale parameter and β is a $p \times 1$ vector parameter. Here, the component density function g is unspecified. Anderson (1979) and Qin (1999) discussed several attractive features of this model, including connections to the logistic regression discrimination and case-control studies, and also considered the maximum semiparametric likelihood estimation of the underlying parameters and distribution functions. In a more general setup, Zou et al. (2002) and Zou and Fine (2002) considered an I-component semiparametric finite mixture model with known mixing proportions and component density functions g and h satisfying (1). Specifically, let $X_{i1}, \ldots, X_{in_i}$ be independent and identically distributed p-dimensional random vectors from the ith mixture with density function

$$f_i(x) = \pi_i h(x) + (1 - \pi_i) g(x), \qquad i = 1, \ldots, I, \tag{2}$$

where the component density functions g and h satisfy (1) and the mixing proportions π_i are assumed to be known and satisfy $0 \le \pi_i \le 1$, $\pi_1 \ne \cdots \ne \pi_I$. Assume that $\{(X_{i1}, \ldots, X_{in_i}),\ i = 1, \ldots, I\}$ are jointly independent. Nagelkerke et al. (2001) applied model (2) to estimate tuberculous prevalence based on a tuberculosis infection data collected from several populations with different mixes of tuberculosis infection. As pointed out by Zou et al. (2002), model (2) is nonidentifiable when $I = 1$, so we assume that $I \ge 2$ throughout this paper.

Let $w(x) = \exp(\alpha + \beta^{\tau} x)$ and $u_i(x) = \pi_i w(x) + (1 - \pi_i)$ for $i = 1, \ldots, I$. Then the semiparametric finite mixture model (2) is equivalent to the following I-sample semiparametric model in which $\{(X_{i1}, \ldots, X_{in_i}),\ i = 1, \ldots, I\}$ are jointly independent and $X_{i1}, \ldots, X_{in_i}$ are independent with density function

$$f_i(x) = u_i(x) g(x), \quad i = 1, \ldots, I. \tag{3}$$

Throughout this paper, let $G(x)$ be the corresponding cumulative distribution function of $g(x)$. Note that model (3) is an I-sample semiparametric selection bias model with weight functions $u_i(x)$ depending on the unknown vector parameter (α, β). The s-sample semiparametric selection bias model was proposed by Vardi (1985) and was further developed by Gilbert et al. (1999).

Zou et al. (2002) proposed to employ the semiparametric finite mixture model (2) in genetic quantitative trial loci analysis, where the component density functions g and h are associated with the possible genotypes and the mixing proportions are determined by the recombination fractions between a locus and the flanking markers. They showed that a constrained empirical likelihood has an irregularity when $g = h$ or $\beta = 0$ under model (2) and further proposed a partial empirical likelihood which allows for unconstrained estimation of the underlying parameters and distribution functions. To test the null hypothesis that a locus has no genetic influence or $H_0 : \beta = 0$ under model (2), Zou et al. (2002) proposed to use a log partial empirical likelihood ratio statistic and proved that its asymptotic null distribution is chi-squared. Our focus of attention in this paper is to propose a partial empirical likelihood-based score statistic for testing $H_0 : \beta = 0$ under model (2) or (3). We provide a different, yet equivalent construction of the log partial empirical likelihood $\ell(\alpha, \beta)$ of Zou et al. (2002). The proposed score test is based on the slope and expected curvature of $\ell(\alpha, \beta)$ under model (2) at the null value $\beta = 0$. It makes use of the size of the score function $\partial \ell\left(\hat{\alpha}, 0\right)/\partial \beta$ evaluated at $\left(\hat{\alpha}, 0\right)$ with $\hat{\alpha}$ being the maximum partial empirical likelihood estimator of α subject to $\beta = 0$ under model (2). The discrepancy between $\partial \ell\left(\hat{\alpha}, 0\right)/\partial \beta$ and a p-dimensional zero vector indicates evidence against $H_0 : \beta = 0$ under model (2). The proposed score statistic is a quadratic form based on $\partial \ell\left(\hat{\alpha}, 0\right)/\partial \beta$ and is shown to have an asymptotic chi-squared distribution under $H_0 : \beta = 0$ and an asymptotic noncentral chi-squared distribution under local alternatives to $H_0 : \beta = 0$. Compared to the partial empirical likelihood ratio test and the Wald test, one advantage of the proposed score test is that it does not need to calculate the maximum partial empirical likelihood estimator of (α, β) under model (2).

This paper is organized as follows. In Sect. 2, we propose our partial empirical likelihood-based score statistic and establish its asymptotic distribution under model (2). In Sect. 3, we demonstrate the asymptotic equivalence among the score test, the partial empirical likelihood ratio test, and the Wald test under model (2). In Sect. 4, we discuss the consistency of the score test and investigate the power of the score statistic theoretically and via simulation by considering local alternatives to $H_0 : \beta = 0$ under model (2). Proofs of the main theoretical results are provided in the Appendix.

2 Construction of score statistics

We consider the problem of testing $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ under the semiparametric finite mixture model (2). To this end, let $n = \sum_{i=1}^{I} n_i$, $\rho_i = n_i/n$ for $i = 1, \ldots, I$, and $\pi = \sum_{i=1}^{I} \rho_i \pi_i$. We assume that the ρ_i remain fixed as $\min(n_1, \ldots, n_I) \to \infty$. Following Qin (1999) and Zou et al. (2002), the log profile empirical likelihood function of (α, β) under model (2) is given by

$$\begin{aligned}\ell_p(\alpha, \beta) = -n \log n - \sum_{i=1}^{I} \sum_{j=1}^{n_i} \log\left[1 + \lambda\{\exp(\alpha + \beta^\tau x_{ij}) - 1\}\right] \\ + \sum_{i=1}^{I} \sum_{j=1}^{n_i} \log[\pi_i \exp(\alpha + \beta^\tau x_{ij}) + (1 - \pi_i)],\end{aligned} \tag{4}$$

where $\lambda = \lambda(\alpha, \beta)$ is the Lagrange multiplier determined by the constrained equation

$$\frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \frac{\exp(\alpha + \beta^\tau x_{ij}) - 1}{1 + \lambda[\exp(\alpha + \beta^\tau x_{ij}) - 1]}. \tag{5}$$

As shown by Zou et al. (2002), the constrained empirical likelihood $\ell_p(\alpha, \beta)$ has an irregularity under the null hypothesis $H_0 : \beta = 0$. By taking the first-order partial derivative of $\ell_p(\alpha, \beta)$ with respect to α, it follows from (4) and (5) that

$$\frac{\partial \ell_p(\alpha, \beta)}{\partial \alpha} = -n\lambda + \sum_{i=1}^{I} \sum_{j=1}^{n_i} \frac{\pi_i \exp(\alpha + \beta^\tau x_{ij})}{\pi_i \exp(\alpha + \beta^\tau x_{ij}) + (1 - \pi_i)}.$$

Setting $\partial \ell_p(\alpha, \beta)/\partial \alpha = 0$ yields

$$\lambda = \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \frac{\pi_i \exp(\alpha + \beta^\tau x_{ij})}{\pi_i \exp(\alpha + \beta^\tau x_{ij}) + (1 - \pi_i)}. \tag{6}$$

It is seen that the Lagrange multiplier λ is not a constant. However, by taking the expectation on both sides of (6) under model (2), we have that

$$E(\lambda) = \frac{1}{n} \sum_{i=1}^{I} n_i \pi_i E\left[\frac{w(X_{i1})}{u_i(X_{i1})}\right] = \sum_{i=1}^{I} \rho_i \pi_i \int w(t) \mathrm{d}G(t) = \sum_{i=1}^{I} \rho_i \pi_i = \pi.$$

Here, we have used the fact that $\int w(t)\mathrm{d}G(t) = \int \exp(\alpha + \beta^\tau t)\mathrm{d}G(t) = 1$ since $h(t) = w(t)g(t)$ is a density function. Note that $E(\lambda) = \pi$ is a known constant under model (2) with known mixing proportions $\pi_1, \ldots, \pi_I$. Replacing λ with $E(\lambda) = \pi$ in $\ell_p(\alpha, \beta)$ of (4), we obtain the log partial empirical likelihood function of (α, β) Zou et al. (2002):

$$\begin{aligned} \ell(\alpha, \beta) = -n \log n &- \sum_{i=1}^{I} \sum_{j=1}^{n_i} \log[\pi \exp(\alpha + \beta^\tau x_{ij}) + (1 - \pi)] \\ &+ \sum_{i=1}^{I} \sum_{j=1}^{n_i} \log[\pi_i \exp(\alpha + \beta^\tau x_{ij}) + (1 - \pi_i)]. \end{aligned} \tag{7}$$

It is easy to verify that $\ell(\alpha, \beta)$ is equivalent to $\ell_2(\beta)$ of Zou et al. (2002) and Zou and Fine (2002). Throughout this paper, let $(\tilde{\alpha}, \tilde{\beta}) = \arg\max_{(\alpha,\beta)} \ell(\alpha, \beta)$ be the maximum partial empirical likelihood estimator of (α, β) under model (2).

Suppose that $\hat{\alpha}$ maximizes $\ell(\alpha, \beta)$ with respect to α subject to $\beta = 0$. Then it can be shown that $\hat{\alpha}$ satisfies

$$\frac{\partial \ell(\alpha, 0)}{\partial \alpha} = \frac{\exp(\alpha)}{\pi \exp(\alpha) + (1 - \pi)} \sum_{i=1}^{I} \frac{n_i(\pi_i - \pi)}{\pi_i \exp(\alpha) + (1 - \pi_i)} = 0, \tag{8}$$

where the partial derivative notation reflects derivatives with respect to α that are evaluated at $\beta = 0$. It is seen that $\hat{\alpha} = 0$ is a solution to (8). Let

$$\begin{aligned} U_n(\alpha, \beta) = \frac{1}{n} \frac{\partial \ell(\alpha, \beta)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \Bigg[&\frac{\pi_i}{\pi_i \exp(\alpha + \beta^\tau x_{ij}) + (1 - \pi_i)} \\ &- \frac{\pi}{\pi \exp(\alpha + \beta^\tau x_{ij}) + (1 - \pi)} \Bigg] \exp(\alpha + \beta^\tau x_{ij}) x_{ij} \end{aligned}$$

be the partial empirical likelihood-based vector score function corresponding to β and let

$$U_n = U_n(\hat{\alpha}, 0) = U_n(0, 0) = \frac{1}{n}\frac{\partial \ell(0, 0)}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\pi_i - \pi)x_{ij}. \tag{9}$$

The values of the components in U_n tend to be larger in absolute value when $\tilde{\beta}$ is farther from 0. Although the score function U_n is based on a partial empirical likelihood function rather than on a parametric likelihood function under a parametric model, it is the fact that $E(U_n) = 0$ under $H_0 : \beta = 0$ in model (2) that allows construction of a score statistic for testing $H_0 : \beta = 0$ under model (2). As indicated in the following theorem, the asymptotic null distribution of U_n depends on the nonparametric part G in model (2). Write $\eta = \sum_{i=1}^{I} \rho_i(\pi_i - \pi)^2$ and

$$\Gamma = \eta\left[\int tt^{\tau}\mathrm{d}G(t) - \left(\int t\mathrm{d}G(t)\right)\left(\int t\mathrm{d}G(t)\right)^{\tau}\right]. \tag{10}$$

Theorem 1 *Suppose that model (2) holds and that* Γ *is positive definite. Then as* $n \to \infty$, $\sqrt{n}U_n \to N_p(0, \Gamma)$ *in distribution under* $H_0 : \beta = 0$.

The proof of Theorem 1 is given in the Appendix. According to Zou et al. (2002), the maximum partial empirical likelihood estimator of G under model (2) for fixed (α, β) is given by

$$G_{(\alpha,\beta)}(t) = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\frac{I(x_{ij} \le t)}{1 + \pi[\exp(\alpha + \beta^{\tau}x_{ij}) - 1]}.$$

Under $H_0 : \beta = 0$, we propose to estimate G by

$$\hat{G}(t) = \hat{G}_{(\hat{\alpha},0)}(t) = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} I(x_{ij} \le t),$$

the standard empirical distribution function based on $\{(x_{i1}, \ldots, x_{in_i}), i = 1, \ldots, I\}$. Let $\hat{\Gamma}$ be the empirical version of Γ with G replaced by $\hat{G}$, yielding

$$\hat{\Gamma} = \eta\left[\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}x_{ij}x_{ij}^{\tau} - \left(\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}x_{ij}\right)\left(\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}x_{ij}\right)^{\tau}\right]. \tag{11}$$

It can then be shown that the estimated asymptotic covariance matrix $\hat{\Gamma}$ is a consistent estimator of Γ under $H_0 : \beta = 0$ in model (2). According to Theorem 1, a partial empirical likelihood-based score statistic for testing $H_0 : \beta = 0$ under model (2) is given by

$$V_n = nU_n^{\tau}\hat{\Gamma}^{-1}U_n. \tag{12}$$

Clearly, large observed values of V_n indicate evidence against $H_0 : \beta = 0$ under model (2). If Γ^{-1} exists, then it follows from Theorem 1 that $V_n \to \chi_p^2$ in distribution under $H_0 : \beta = 0$ as $n \to \infty$. On the other hand, if $\hat{\Gamma}$ is singular, we can replace $\hat{\Gamma}^{-1}$ by the Moore–Penrose generalized inverse $\hat{\Gamma}^{+}$ of $\hat{\Gamma}$ in (11). According to Theorem 9.2.2 of Rao and Mitra (1971, p 173), $V_n = nU_n^\tau \hat{\Gamma}^{+} U_n \to \chi_r^2$ in distribution under $H_0 : \beta = 0$ as $n \to \infty$, where r is the rank of $\hat{\Gamma}$, whether or not $\hat{\Gamma}^{+}$ converges.

Remark 2.1 Theorem 1 also forms the basis of constructing a score confidence region for β under model (2). Specifically, let $\chi_p^2(\nu)$ be such that $P\{\chi_p^2 \le \chi_p^2(\nu)\} = \nu$. When Γ is positive definite, $C_{1-\nu} = \{\beta : V_n(\beta) \le \chi_p^2(1-\nu)\}$ is an approximate level $1-\nu$ score confidence region for β under model (2), where $V_n(\beta) = nU_n^\tau\left(\hat{\alpha}(\beta), \beta\right)\hat{\Gamma}^{-1}\left(\hat{\alpha}(\beta), \beta\right)U_n\left(\hat{\alpha}(\beta), \beta\right)$ with $\hat{\alpha}(\beta)$ maximizing $\ell(\alpha, \beta)$ with respect to α for fixed β and $\hat{\Gamma}\left(\hat{\alpha}(\beta), \beta\right)$ is an estimated asymptotic covariance matrix of $U_n\left(\hat{\alpha}(\beta), \beta\right)$ for fixed β. When $p = 1$, a level $1-\nu$ score confidence interval for β can be constructed from $U_n\left(\hat{\alpha}(\beta), \beta\right)$ and is given by $I_{1-\nu} = \{\beta : \sqrt{n}U_n\left(\hat{\alpha}(\beta), \beta\right)\hat{\Gamma}^{-1/2}\left(\hat{\alpha}(\beta), \beta\right) \le z_{1-\nu/2}\}$, where $z_{1-\nu/2}$ satisfies $P(Z \le z_{1-\nu/2}) = 1-\nu/2$ with $Z \sim N(0, 1)$.

Remark 2.2 The proposed construction of the partial empirical likelihood-based score test may be extended to the case of unknown mixing proportions π_i in model (2). Nevertheless, since the π_i are unknown, the partial empirical likelihood function $\ell(\alpha, \beta) = \ell(\alpha, \beta, \pi_1, \ldots, \pi_I)$ in (7) depends on parameters (α, β) as well as $(\pi_1, \ldots, \pi_I)$. Consequently, the partial empirical likelihood-based score statistic for testing $H_0 : \beta = 0$ with unknown mixing proportions π_i in model (2) is based on the partial derivatives $\ell(\alpha, 0, \pi_1, \ldots, \pi_I)/\partial\alpha$ and $\ell(\alpha, 0, \pi_1, \ldots, \pi_I)/\partial\pi_i$ for $i = 1, \ldots, I$ in that we need to maximize $\ell(\alpha, \beta, \pi_1, \ldots, \pi_I)$ with respect to $(\alpha, \pi_1, \ldots, \pi_I)$ subject to $\beta = 0$.

3 Equivalence among the score, empirical likelihood ratio, and Wald tests

According to Theorem 2 of Zou et al. (2002), we have that

$$\sqrt{n}\begin{pmatrix}\tilde{\alpha} - \alpha \\ \tilde{\beta} - \beta\end{pmatrix} \to N_{p+1}(0, \Sigma) \tag{13}$$

in distribution under model (2), where

$$\Sigma = S^{-1} - \frac{\eta}{\pi^2(1-\pi)^2}S^{-1}\begin{pmatrix}S_0 + C_0 \\ S_1 + C_1\end{pmatrix}(S_0 + C_0,\ S_1^\tau + C_1^\tau)S^{-1}, \tag{14}$$

with

$$C_0 = \sum_{i=1}^{I} \rho_i \pi_i (1-\pi_i) \int \frac{\exp(\alpha + \beta^\tau t)}{\pi_i \exp(\alpha + \beta^\tau t) + (1-\pi_i)} \mathrm{d}G(t),$$

$$C_1 = \sum_{i=1}^{I} \rho_i \pi_i (1-\pi_i) \int \frac{\exp(\alpha + \beta^\tau t)}{\pi_i \exp(\alpha + \beta^\tau t) + (1-\pi_i)} t\mathrm{d}G(t),$$

$$S_0 = \int \frac{\pi(1-\pi)\exp(\alpha + \beta^\tau t)}{\pi \exp(\alpha + \beta^\tau t) + (1-\pi)} \mathrm{d}G(t) - C_0,$$

$$S_1 = \int \frac{\pi(1-\pi)\exp(\alpha + \beta^\tau t)}{\pi \exp(\alpha + \beta^\tau t) + (1-\pi)} t\mathrm{d}G(t) - C_1,$$

$$S_2 = \int \frac{\pi(1-\pi)\exp(\alpha+\beta^\tau t)}{\pi\exp(\alpha+\beta^\tau t)+(1-\pi)} tt^\tau \mathrm{d}G(t) - \sum_{i=1}^{I} \int \frac{\rho_i\pi_i(1-\pi_i)\exp(\alpha+\beta^\tau t)}{\pi_i\exp(\alpha+\beta^\tau t)+(1-\pi_i)} tt^\tau \mathrm{d}G(t),$$

$$S = \begin{pmatrix} S_0 & S_1^\tau \\ S_1 & S_2 \end{pmatrix}, \quad S_{22\cdot 1} = S_2 - S_1 S_0^{-1} S_1^\tau. \tag{15}$$

It follows from (13), (14), and (15) that

$$\sqrt{n}(\tilde{\beta}-\beta) \quad \to \quad N_p(0, \Sigma_{\tilde{\beta}}) \tag{16}$$

in distribution under model (2), where

$$\Sigma_{\tilde{\beta}} = S_{22\cdot 1}^{-1} - \frac{\eta}{\pi^2(1-\pi)^2} S_{22\cdot 1}^{-1}(C_1 - S_1 S_0^{-1} C_0)(C_1 - S_1 S_0^{-1} C_0)^\tau S_{22\cdot 1}^{-1}.$$

Let

$$\tilde{G}(t) = \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \frac{I(x_{ij} \le t)}{\pi\exp(\tilde{\alpha}+\tilde{\beta}^\tau x_{ij})+(1-\pi)}$$

be the maximum partial empirical likelihood estimator of $G(t)$ under model (2). Then the partial empirical likelihood based Wald test of $H_0 : \beta = 0$ under model (2) has test statistic $W_n = n\tilde{\beta}^\tau \tilde{\Sigma}_{\tilde{\beta}}^{-1} \tilde{\beta}$, where $\tilde{\Sigma}_{\tilde{\beta}}$ is the empirical version of $\Sigma_{\tilde{\beta}}$ with (α, β, G) replaced by $(\tilde{\alpha}, \tilde{\beta}, \tilde{G})$. It can be shown that $\tilde{\Sigma}_{\tilde{\beta}} \to \Gamma^{-1}$ in probability under the null hypothesis $H_0 : \beta = 0$ in model (2). The asymptotic multivariate normal distribution for $\tilde{\beta}$ in (16) implies that W_n has the same asymptotic χ_p^2 distribution under $H_0 : \beta = 0$ as does V_n. Moreover, it is seen from Theorem 1 and (16) that the random vectors $\sqrt{n}U_n$ and $\sqrt{n}\Gamma\tilde{\beta}$ have the same asymptotic distribution namely $N_p(0, \Gamma)$ under $H_0 : \beta = 0$. In fact, there holds a stronger relationship between these two random vectors, as described below.

According to the proof of Theorem 2 of Zou et al. (2002), one can write

$$\begin{pmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{pmatrix} = \frac{1}{n} A^{-1} \begin{pmatrix} \frac{\partial \ell(0,0)}{\partial \alpha} \\ \frac{\partial \ell(0,0)}{\partial \beta} \end{pmatrix} + o_p\left(n^{-1/2}\right) \tag{17}$$

under H_0: $\beta = 0$ in model (2), where

$$A = \begin{pmatrix} A_0 & A_1^\tau \\ A_1 & A_2 \end{pmatrix},$$

$$A_0 = \eta, \quad A_1 = \eta \int t \mathrm{d}G(t), \quad A_2 = \eta \int tt^\tau \mathrm{d}G(t). \tag{18}$$

Equation (17) implies the following asymptotic expression for $\sqrt{n}\tilde{\beta}$ under H_0: $\beta = 0$:

$$\sqrt{n}\tilde{\beta} = \sqrt{n}\Gamma^{-1}U_n + o_p(1). \tag{19}$$

It now follows from (19) that $\sqrt{n}(U_n - \Gamma\tilde{\beta}) = o_p(1)$ under $H_0 : \beta = 0$. This asymptotic equivalence between $\sqrt{n}U_n$ and $\sqrt{n}\Gamma\tilde{\beta}$, along with the consistency of $\tilde{\Sigma}_{\tilde{\beta}}$, implies that $V_n - W_n = o_p(1)$ under $H_0 : \beta = 0$ in model (2). Consequently, the score statistic V_n and the Wald statistic W_n are asymptotically equivalent for testing the null hypothesis $H_0 : \beta = 0$ under model (2).

It follows from the proof of Theorem 2 of Zou et al. (2002) and (7) that the partial empirical likelihood ratio statistic for testing $H_0 : \beta = 0$ has the following asymptotic expression:

$$\begin{aligned}
R_n &= 2[\ell(\tilde{\alpha}, \tilde{\beta}) - \ell(0,0)] \\
&= 2\left[\sum_{i=1}^{I}\sum_{j=1}^{n_i} \log[\pi_i \exp(\tilde{\alpha} + \tilde{\beta}^\tau x_{ij}) + (1-\pi_i)]\right. . \\
&\qquad \left. - \sum_{i=1}^{I}\sum_{j=1}^{n_i} \log[\pi \exp(\tilde{\alpha} + \tilde{\beta}^\tau x_{ij}) + (1-\pi)]\right] \\
&= \frac{1}{n}\left(\frac{\partial \ell(0,0)}{\partial \alpha}, \ \frac{\partial \ell(0,0)}{\partial \beta}\right) A^{-1} \begin{pmatrix} \frac{\partial \ell(0,0)}{\partial \alpha} \\ \frac{\partial \ell(0,0)}{\partial \beta} \end{pmatrix} + o_{\mathrm{p}}(1) \\
&= n(0, U_n)A^{-1}\begin{pmatrix} 0 \\ U_n \end{pmatrix} + o_{\mathrm{p}}(1) \\
&= nU_n^\tau \Gamma^{-1} U_n + o_{\mathrm{p}}(1).
\end{aligned} \tag{20}$$

This implies that $R_n - V_n = o_p(1)$ under $H_0 : \beta = 0$ in model (2). Hence, the score statistic V_n, the Wald statistic W_n, and the partial empirical likelihood ratio statistic R_n are asymptotically equivalent for testing the null hypothesis $H_0 : \beta = 0$ under the semiparametric finite mixture model (2). Note, however, that one advantage of the score statistic V_n over the partial empirical likelihood ratio statistic R_n and the Wald statistic W_n is that according to (9), (11), and (12), the score statistic V_n has a closed-form expression dependent only on the data and the mixing proportions $\pi_1, \ldots, \pi_I$; thus it does not need to calculate the maximum partial empirical likelihood estimator $(\tilde{\alpha}, \tilde{\beta})$ of (α, β) under model (2), making the calculation of V_n straightforward without the need to solve a system of equations. By contrast, both the partial empirical likelihood ratio test and the Wald test depend on $(\tilde{\alpha}, \tilde{\beta})$ and are thus in need of seeking $(\tilde{\alpha}, \tilde{\beta})$ by solving the following system of $p+1$ equations:

$$\begin{aligned}
\begin{pmatrix} \frac{\partial \ell(\alpha,\beta)}{\partial \alpha} \\ \frac{\partial \ell(\alpha,\beta)}{\partial \beta} \end{pmatrix} = \sum_{i=1}^{I}\sum_{j=1}^{n_i} &\left[\frac{\pi_i}{\pi_i \exp(\alpha + \beta^\tau x_{ij}) + (1-\pi_i)}\right. \\
&\left. - \frac{\pi}{\pi \exp(\alpha + \beta^\tau x_{ij}) + (1-\pi)}\right] \exp(\alpha + \beta^\tau x_{ij}) \begin{pmatrix} 1 \\ x_{ij} \end{pmatrix}.
\end{aligned}$$

In all applications, one needs to use the Newton–Raphson method or some variant to obtain R_n and W_n, but this is unnecessary for V_n. Consequently, the proposed score test is computationally easier to implement than the partial empirical likelihood ratio test and the Wald test.

4 Power considerations

In this section, we investigate the power of the proposed score statistic V_n theoretically and via simulation. We first discuss the consistency of V_n. For $\beta \neq 0$ under model (2), let α be the true α-value corresponding to β under model (2). Then it can be shown after some algebra that $U_n \to \eta(\mu_h - \mu_g)$ in probability, where

$$\mu_g = \int t\mathrm{d}G(t), \qquad \mu_h = \int \exp(\alpha + \beta^\tau t)t\mathrm{d}G(t).$$

It then follows that when Γ is positive definite and $\mu_g \neq \mu_h$, the proposed score test based on V_n in (12) is consistent against any fixed alternative $\beta \neq 0$ under model (2).

We now consider the local asymptotic power of the proposed score test under local alternatives $H_1 : \beta_n = n^{-1/2}\gamma\{1 + o(1)\}$ under model (2), where $\gamma \in R^p$. Let α_n be the true value of α when $\beta = \beta_n$ under model (2). Then it can be shown under model (2) that $\alpha_n = n^{-1/2}\lambda\{1+o(1)\}$ with $\lambda = -\int \gamma^\tau t\mathrm{d}G(t) = -A_0^{-1}A_1^\tau\gamma$, where A_0 and A_1 are defined in (18). The following theorem establishes the large-sample distribution of the proposed score statistic V_n for testing $H_0 : \beta = 0$ under the sequence of parameter values (α_n, β_n), where $\beta_n = n^{-1/2}\gamma\{1 + o(1)\}$ and $\alpha_n = n^{-1/2}\lambda\{1 + o(1)\}$ as $n \to \infty$ for fixed $\gamma \in R^p$.

Theorem 2 *Suppose that model* (2) *holds and that* Γ *is positive definite. Then under model* (2) *with* $(\alpha, \beta) = (\alpha_n, \beta_n)$, $V_n \to \chi_p^2(\delta^2)$ *in distribution as* $n \to \infty$, *where* $\delta^2 = \gamma^\tau\Gamma\gamma$ *with* Γ *defined in* (10) *and* $\chi_p^2(\delta^2)$ *is a noncentral chi-squared random variable with p degrees of freedom and noncentrality parameter* δ^2.

The proof of Theorem 2 is given in the Appendix. According to the asymptotic equivalence among the score test, the Wald test, and the empirical likelihood ratio test, it follows from Theorem 2 and Eqs. (9), (19), and (20) that $W_n \to \chi_1^2(\delta^2)$ in distribution and $R_n \to \chi_1^2(\delta^2)$ in distribution under model (2) with $(\alpha, \beta) = (\alpha_n, \beta_n)$ as $n \to \infty$.

The asymptotic null and alternative distributions of V_n, presented in Theorems 1 and 2, can be employed to obtain critical values of the proposed score test and power against various local alternatives $\beta_n \neq 0$ by numerical integration, although explicit computation is unfortunately somewhat complicated.

In the following, we present a small simulation study to compare the performance of the proposed score statistic V_n with those of the partial empirical likelihood ratio statistic R_n and the Wald statistic W_n, by examining their powers against some local alternatives $H_1 : \beta \neq 0$ under the semiparametric finite mixture model (2). We consider the genetic experiment discussed by Zou et al. (2002), in which $(\pi_1, \pi_2, \pi_3) = (0.99, 0.5, 0.01)$ and $(\rho_1, \rho_2, \rho_3) = (0.418, 0.164, 0.418)$. Moreover, we consider both continuous and discrete mixture distributions under model (2) with $p = 1$. Since $p = 1$, all three test statistics—V_n, W_n, and R_n—have an asymptotic chi-squared distribution with one degree of freedom.

In our simulation study, we first assume that $g(x)$ is the standard normal density function, $h(x) = g(x - \mu)$ is the density function of a $N(\mu, 1)$ distribution, and $f_i(x) = \pi_i h(x) + (1 - \pi_i)g(x)$ is the density function of a $\pi_i N(\mu, 1) + (1 - \pi_i)$ $N(0, 1)$ distribution for $i = 1, 2, 3$. Then model (2) holds with $\alpha = -\mu^2/2$ and

Table 1 Achieved significance levels and powers of W_n, V_n, and R_n in the case of normal mixture distributions

γ	n	β_n	Nominal level	Power		
				W_n	V_n	R_n
0	100	0	0.10	0.095	0.095	0.100
0	100	0	0.05	0.049	0.048	0.050
0	100	0	0.01	0.009	0.008	0.012
2	100	0.2	0.10	0.224	0.230	0.234
2	100	0.2	0.05	0.135	0.137	0.139
2	100	0.2	0.01	0.046	0.046	0.051
4	100	0.4	0.10	0.580	0.583	0.585
4	100	0.4	0.05	0.441	0.445	0.451
4	100	0.4	0.01	0.192	0.198	0.211
0	250	0	0.10	0.105	0.110	0.107
0	250	0	0.05	0.050	0.051	0.050
0	250	0	0.01	0.007	0.007	0.007
2	250	0.126	0.10	0.230	0.232	0.230
2	250	0.126	0.05	0.152	0.152	0.154
2	250	0.126	0.01	0.046	0.046	0.048
4	250	0.253	0.10	0.570	0.571	0.574
4	250	0.253	0.05	0.455	0.461	0.460
4	250	0.253	0.01	0.213	0.214	0.218

$\beta = \mu$. Let $\beta_n = n^{-1/2}\gamma$. Then it is easy to see that under model (2), $\alpha_n = 0$ for all $n \geq 1$. Our aim is to compare the performance of V_n with those of W_n and R_n by examining their powers against some local alternatives $H_1 : \beta = \beta_n$ under model (2). In our simulations, we considered $\gamma = 0, 2.0, 4.0$ and $n = 100, 250$. Note that for $\gamma = 0, 2.0, 4.0$, we have that $\beta_n = 0, 0.2, 0.4$ when $n = 100$ and $\beta_n = 0, 0.126, 0.253$ when $n = 250$. For each value of n and γ, we generated 1,000 independent sets of combined random samples from the $0.99N(\beta_n, 1) + 0.01N(0, 1)$, $0.5N(\beta_n, 1)+0.5N(0, 1)$, and $0.01N(\beta_n, 1)+0.99N(0, 1)$ distributions. The simulation results are summarized in Table 1.

Next, we study how the proposed score statistic V_n performs when the data is discrete under the same sample sizes considered in Table 1. To this end, we assume that $g(x) = (\mathrm{e}^{-1}/x!)I(x \in \mathcal{N})$ is the standard Poisson $P(1)$ frequency function with mean 1, $h(x) = (\mu^x/x!)\mathrm{e}^{-\mu}I(x \in \mathcal{N})$ is the frequency function of a Poisson $P(\mu)$ distribution with mean μ, and $f_i(x) = \pi_i h(x) + (1 - \pi_i)g(x)$ is the density function of a $\pi_i P(\mu) + (1 - \pi_i)P(1)$ distribution for $i = 1, 2, 3$, where $\mathcal{N} = \{0, 1, \ldots, \}$. Then model (2) holds with $\alpha = 1 - \mu$ and $\beta = \log \mu$. Let $\beta_n = n^{-1/2}\gamma$ and $\mu_n = \exp(\beta_n) = \exp\left(n^{-1/2}\gamma\right)$. Then it is easy to see that under model (2), $\alpha_n = -\beta_n = -n^{-1/2}\gamma$ for $n \geq 1$. Again, our objective is to compare the performances of V_n, W_n, and R_n by examining their powers against some local alternatives $H_1 : \beta = \beta_n$ under model (2). In our simulations, we considered $\gamma = 0, 2.0, 4.0$ and $n = 100, 250$. Note that for $\gamma = 0, 2.0, 4.0$, we have that $\mu_n = 1, 1.221, 1.492$ when $n = 100$ and $\mu_n = 1, 1.135, 1.288$ when $n = 250$. For each value of n and γ, we generated 1,000 independent sets of combined random samples from the $0.99P(\mu_n) + 0.01P(1)$, $0.5P(\mu_n) + 0.5P(1)$,

Table 2 Achieved significance levels and powers of W_n, V_n, and R_n in the case of Poisson mixture distributions

γ	n	β_n	μ_n	Nominal level	Power		
					W_n	V_n	R_n
0	100	0	1	0.10	0.101	0.100	0.103
0	100	0	1	0.05	0.053	0.053	0.059
0	100	0	1	0.01	0.010	0.009	0.011
2	100	0.2	1.221	0.10	0.252	0.255	0.256
2	100	0.2	1.221	0.05	0.160	0.162	0.165
2	100	0.2	1.221	0.01	0.053	0.054	0.060
4	100	0.4	1.492	0.10	0.600	0.608	0.605
4	100	0.4	1.492	0.05	0.477	0.480	0.487
4	100	0.4	1.492	0.01	0.239	0.242	0.252
0	250	0	1	0.10	0.103	0.101	0.106
0	250	0	1	0.05	0.046	0.048	0.048
0	250	0	1	0.01	0.009	0.010	0.009
2	250	0.126	1.135	0.10	0.215	0.219	0.217
2	250	0.126	1.135	0.05	0.143	0.146	0.147
2	250	0.126	1.135	0.01	0.046	0.047	0.047
4	250	0.253	1.288	0.10	0.618	0.617	0.620
4	250	0.253	1.288	0.05	0.496	0.501	0.498
4	250	0.253	1.288	0.01	0.241	0.254	0.250

and $0.01P(\mu_n) + 0.99P(1)$ distributions. The simulation results are summarized in Table 2.

It is seen from Tables 1 and 2 that the achieved significance levels of V_n, W_n, and R_n are quite close to the corresponding nominal significance levels, and the powers of V_n, W_n, and R_n are getting larger as γ moves away from 0. Our simulation results also reveal that the powers of V_n are all greater than or equal to those of W_n except for the case in Table 2 with $\gamma = 4$, $n = 250$, and the nominal significance level equal to 0.10, and that the powers of R_n are slightly larger than those of V_n in most of the cases. In summary, our simulation study indicates that the proposed score statistic V_n is superior to the Wald statistic W_n and is quite comparable to the partial empirical likelihood ratio statistic R_n, in terms of their power performances.

Appendix

Proof of Theorems 1 and 2

Under model (2) with $(\alpha, \beta) = (\alpha_n, \beta_n)$ as $n \to \infty$, we have

$$
\begin{aligned}
E(\sqrt{n}U_n) &= \sqrt{n}\sum_{i=1}^{I}\rho_i(\pi_i - \pi)\int[\pi_i \exp(\alpha_n + \beta_n^\tau t) + (1-\pi_i)]t\mathrm{d}G(t) \\
&= \sqrt{n}\sum_{i=1}^{I}\rho_i(\pi_i - \pi)\int\left[\pi_i(1 + \alpha_n + t^\tau\beta_n) + (1-\pi_i)\right]t\mathrm{d}G(t) + o(1)
\end{aligned}
$$

$$
\begin{aligned}
&= \sqrt{n}\sum_{i=1}^{I}\rho_i(\pi_i - \pi)\int\left[\pi_i\left(1 + \frac{\lambda}{\sqrt{n}} + t^\tau\frac{\gamma}{\sqrt{n}}\right) + (1-\pi_i)\right]t\mathrm{d}G(t)\\
&\quad + o(1)\\
&= \sum_{i=1}^{I}\rho_i(\pi_i - \pi)\pi_i\int(\lambda + t^\tau\gamma)t\mathrm{d}G(t) + o(1)\\
&= \eta\lambda\int t\mathrm{d}G(t) + \eta\left(\int tt^\tau\mathrm{d}G(t)\right)\gamma + o(1)\\
&= A_1\lambda + A_2\gamma + o(1) = (A_2 - A_1A_0^{-1}A_1^\tau)\gamma + o(1)\\
&= \Gamma\gamma + o(1). \qquad (21)
\end{aligned}
$$

Moreover, it can be shown that

$$\mathrm{Var}(\sqrt{n}U_n) = \Gamma + o(1) \qquad (22)$$

under model (2) with $(\alpha, \beta) = (\alpha_n, \beta_n)$ as $n \to \infty$. It now follows from (21), (22), the multivariate central limit theorem, and Slutsky's theorem that under (α_n, β_n),

$$\sqrt{n}U_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\pi_i - \pi)x_{ij} \quad\to\quad N_p(\Gamma\gamma, \Gamma)$$

in distribution as $n \to \infty$. Taking $\gamma = 0$ so that $(\alpha_n, \beta_n) = (0, 0)$ yields Theorem 1. Since it can be shown that $\hat{\Gamma}$ is a consistent estimator of Γ under (α_n, β_n) as $n \to \infty$, it follows from Slutsky's theorem and the well-known results on the distribution of quadratic forms of normal random variables that $V_n = nU_n^\tau\hat{\Gamma}^{-1}U_n \to \chi_p^2(\delta^2)$ in distribution under (α_n, β_n) as $n \to \infty$, thus establishing Theorem 2. The proof is completed. □

Acknowledgements I am grateful to the editor, an associate editor, and two referees for their very careful reading and a number of helpful comments and suggestions that have greatly improved my original submission. This research was supported by a grant from the National Security Agency.

References

Anderson, J. A. (1979). Multivariate logistic compounds. *Biometrika 66*, 17–26.

Hall, P., Titterington, D. M. (1984). Efficient nonparametric estimation of mixture proportions. *Journal of the Royal Statistical Society, Series B 46*, 465–473.

Hosmer, D. W. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three types of sample. *Biometrics 29*, 761–770.

Gilbert, P., Lele, S., Vardi, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika 86*, 27–43.

Lindsay, B. (1995). *Mixture models: theory, geometry and applications*. Haywood, CA: Institute of Mathematical Statistics.

McLachlan, G. J., Basford, K. E. (1988). *Mixture models: inference and applications to clustering*. New York: Marcel Dekker.

McLachlan, G. J., Peel, D. (2001). *Finite mixture models*. New York: Wiley.

Murray, G. D., Titterington, D. M. (1978). Estimation problems with data from a mixture. *Applied Statistics 27*, 325–334.

Nagelkerke, N. J. D., Borgdroff, M. W., Kim, S. J. (2001). Logistic discrimination of mixtures of M. tuberculosis and non-specific tuberculin reactions. *Statistics in Medicine 20*, 1113–1124.
Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *Annals of Statistics 27*, 1368–1384.
Rao, C. R., Mitra, S. K. (1971). *Generalized inverse of matrices and its applications*. New York: Wiley.
Titterington, D. M. (1990). Some recent research in the analysis of mixture distributions. *Statistics 21*, 619–641.
Titterington, D. M., Smith, A. F. M., Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
Vardi, Y. (1985). Empirical distribution in selection bias models. *Annals of Statistics 13*, 178–203.
Zou, F., Fine, J. P. (2002). A note on a partial empirical likelihood. *Biometrika 89*, 958–961.
Zou, F., Fine, J. P., Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika 89*, 61–75.