The Annals of Statistics
1999, Vol. 27, No. 4, 1368–1384

EMPIRICAL LIKELIHOOD RATIO BASED CONFIDENCE INTERVALS FOR MIXTURE PROPORTIONS

By Jing Qin

University of Maryland and Memorial Sloan–Kettering Cancer Center

We consider the problem of estimating a mixture proportion using data from two different distributions as well as from a mixture of them. Under the model assumption that the log-likelihood ratio of the two densities is linear in the observations, we develop an empirical likelihood ratio based statistic for constructing confidence intervals for the mixture proportion. Under some regularity conditions, it is shown that this statistic converges to a chi-squared random variable. Simulation results indicate that the performance of this statistic is satisfactory. As a by-product, we give estimators for the two distribution functions. Connections with case-control studies and discrimination analysis are pointed out.

1. Introduction. Consider three independent data sets:

$$
\begin{aligned}
&x_1, \dots, x_{n_0}, && \text{iid with distribution } F(x),\\
&y_1, \dots, y_{n_1}, && \text{iid with distribution } G(y),\\
&z_1, \dots, z_{n_2}, && \text{iid with distribution } H(z) = \lambda F(z) + (1-\lambda) G(z).
\end{aligned} \tag{1.1}
$$

Denote the corresponding density functions by $f(x) = dF(x)/dx$, $g(y) = dG(y)/dy$ and $h(z) = dH(z)/dz$, respectively. The goal is to make inference on the mixing parameter λ, treating F and G as nuisance functions. The problem was studied by Hosmer (1973), assuming normality of F and G. He estimated the proportion of male and female fish in a population of halibut from univariate data. Subsequently, maximum likelihood, Bayesian parametric techniques and various approaches using distribution-free kernel methods were summarized in a comprehensive paper by Murray and Titterington (1978). Without any parametric assumptions on F and G, Hall and Titterington (1984) constructed a sequence of multinomial approximations and related maximum likelihood estimators of the mixture proportions by grouping data. They obtained a Cramér–Rao lower bound for their nonparametric estimators.

A different approach was proposed by Anderson (1979). He postulated a semiparametric modeling assumption

$$
\log \frac{g(x)}{f(x)} = \beta_0 + x\beta_1, \tag{1.2}
$$

Received November 1996; revised May 1999.

AMS 1991 *subject classifications*. Primary 62M10; secondary 62E20.

Key words and phrases. Case-control studies, chi-squared distribution, empirical likelihood, exponential tilt models, logistic function, mixture models.

or equivalently

$$g(x) = \exp(\beta_0 + x\beta_1) f(x),$$

where $f(x)$ is an arbitrary density function. The data in (1.1) then come from distributions

$$\begin{aligned} &dF(x), \qquad \exp(\beta_0 + x\beta_1)\, dF(x), \\ &\qquad [\lambda + (1-\lambda)\exp(\beta_0 + x\beta_1)]\, dF(x), \end{aligned} \tag{1.3}$$

respectively.

We can see easily that if $g(x)$ and $f(x)$ are normal densities with common variance or exponential densities, then (1.2) is satisfied. The attractive feature of (1.2), compared with the normal mixture model, is that the distributions are modeled nonparametrically except for a parametric "exponential tilt" that is used to relate one distribution to the other. This is very similar to Cox's proportional hazards models and Lehmann's two sample models in which it is the ratio of two hazard functions that is assumed to have a known parametric form.

The assumption that the ratio of two densities is a known form is very popular in the logistic regression discrimination and case-control studies. Among others, see the papers by Efron (1975), O'Neill (1980), Prentice and Pyke (1979) and books by Breslow and Day (1980) and Cox and Snell (1989). In fact, if we let D be the indicator of data from F ($D = 0$) and G ($D = 1$), respectively, then the assumption (1.2) is equivalent to the logistic model assumption that

$$\begin{aligned} P(D = 1|z) &= \frac{P(D=1)P(z|D=1)}{P(D=1)P(z|D=1) + P(D=0)P(z|D=0)} \\ &= \frac{\exp(\beta_0^* + z\beta_1)}{1 + \exp(\beta_0^* + z\beta_1)}, \end{aligned}$$

where $\beta_0^* = \beta_0 + \log\{(1-\lambda)/\lambda\}$, $P(D = 1) = 1 - \lambda$ and the marginal distribution of z is not specified. When $n_2 = 0$ in (1.1), we have a case-control problem [Prentice and Pyke (1979)]. When $n_0 = 0$ or $n_1 = 0$, we have a so-called case-control problem with contaminated controls discussed recently by Lancaster and Imbens (1996). If either $n_0 = \infty$ or $n_1 = \infty$, that is, we know F or G completely, then we end up with a fully parametric mixture model. Efron and Tibshirani (1996) have used model (1.2) to estimate the underlying densities for F and G. Model (1.3) also can be treated as a biased sampling problem with weights

$$\begin{aligned} &w_0(x) = 1, \qquad w_1(x) = w(x) = \exp(\beta_0 + x\beta_1), \\ &w_2(x) = \lambda + (1-\lambda)w(x). \end{aligned} \tag{1.4}$$

Empirical likelihood is a nonparametric method of inference [Owen (1988, 1990)]. It has sampling properties similar to the bootstrap, but instead of using resampling, empirical likelihood profiles a multinominal likelihood with support on the sample data values. Properties of empirical likelihood have been discussed by Chen and Hall (1993), DiCiccio, Hall and Romano (1989), Hall (1990), Qin (1993) and Qin and Lawless (1994), among others. For more references, see the review paper by Hall and La Scala (1990).

Based on the observed data (1.1) and the semiparametric model (1.2), the likelihood function can be written as

$$\mathscr{L}(\lambda, \beta, F) = \prod_{i=1}^{n_0} dF(x_i) \prod_{j=1}^{n_1} w_1(y_j)\, dF(y_j) \prod_{k=1}^{n_2} w_2(z_k)\, dF(z_k), \tag{1.5}$$

where $\beta = (\beta_0, \beta_1)^\tau$. By maximizing $\mathscr{L}(\lambda, \beta, F)$ under the assumption that F is a discrete distribution function, Anderson (1979) obtained point estimators of (λ, β). The properties of those estimates, however, were not discussed in his paper. He also pointed out the possibility that his method might be extended to the continuous case by subdividing the range of each continuous variate to make it discrete.

Our objective in this paper is fourfold: first, to develop Anderson's (1979) methodology for estimating the parameters (λ, β) without the assumption that F is a discrete distribution function; second, to extend Owen's (1988, 1990) empirical likelihood to the semiparametric model (1.3); third, to estimate the underlying distribution functions and to test the model assumption; finally, to explore the possible applications of the semiparametric model (1.2) to more general exponential tilt models. The organization of this paper is as follows. In Section 2 we present our methodology and main results. We give the asymptotic variance formula for the maximum semiparametric likelihood estimation. We show that the likelihood based test statistic converges in distribution to a chi-squared random variable. Section 3 discusses estimating the underlying distribution functions and model diagnostics. Section 4 presents some simulation results. Concluding remarks are given in Section 5. Proofs are relegated to the Appendix.

2. Main results. In this section we maximize $\mathscr{L}(\lambda, \beta, F)$, the semiparametric likelihood (1.5), jointly with respect to (λ, β) and F. To do this, we only need to concentrate on those distribution functions with jumps at observed points. Let $n = n_0 + n_1 + n_2$ and t_i, $i = 1, 2, \dots, n$, be the combined sample and let $p_i = dF(t_i)$, $i = 1, 2, \dots, n$, be the jump sizes (nonnegative) such that the total mass is unity. The likelihood $\mathscr{L}(\lambda, \beta, F)$ can be written as

$$\begin{aligned} \mathscr{L}(\lambda, \beta, F) &= \prod_{i=1}^{n_0} dF(x_i) \prod_{j=1}^{n_1} w_1(y_j)\, dF(y_j) \prod_{k=1}^{n_2} w_2(z_k)\, dF(z_k) \\ &= \left\{ \prod_{i=1}^{n} p_i \right\} \left\{ \prod_{j=1}^{n_1} w(y_j) \right\} \left\{ \prod_{k=1}^{n_2} w_2(z_k) \right\}, \end{aligned} \tag{2.1}$$

with w, w_1 and w_2 as in (1.4). We will maximize the likelihood in two steps, as follows:

STEP 1. For fixed (λ, β), maximize

$$\prod_{i=1}^{n} p_i$$

subject to the constraints

$$\sum_{i=1}^{n} p_i = 1, \qquad \sum_{i=1}^{n} p_i\{w(t_i) - 1\} = 0, \qquad p_i \geq 0.$$

Note that the second constraint comes from the fact that $G(t) = \int_{-\infty}^{t} \exp(\beta_0 + x\beta_1)\, dF(x) = \int_{-\infty}^{t} w(x)\, dF(x)$ is a cumulative distribution function. Therefore, $E_F\{w(x)\} = 1$. After maximizing over the p_i's, we have [Qin and Lawless (1994)]

$$p_i = \frac{1}{n} \frac{1}{1 + \nu[w(t_i) - 1]}, \qquad i = 1, 2, \dots, n,$$

where ν is the Lagrange multiplier, which is determined by

$$\frac{1}{n} \sum_{i=1}^{n} \frac{w(t_i) - 1}{1 + \nu[w(t_i) - 1]} = 0.$$

We change the variables (λ, β, ν) to (λ, β, α), where $\alpha = \nu - n_1/n - (n_2/n)(1 - \lambda)$. Then p_i can be written as

$$p_i = \frac{1}{n} \frac{1/\gamma(t_i)}{1 + \alpha(w(t_i) - 1)/\gamma(t_i)},$$

where

$$\gamma(t_i) = \frac{n_0}{n} + \frac{n_2}{n}\lambda + w(t_i)\left[\frac{n_1}{n} + \frac{n_2}{n}(1 - \lambda)\right].$$

The constraint equation becomes

$$\frac{1}{n} \sum_{i=1}^{n} \frac{g(t_i; \lambda, \beta)}{1 + \alpha g(t_i; \lambda, \beta)} = 0, \tag{2.2}$$

where

$$g(t_i; \lambda, \beta) = (w(t_i) - 1)/\gamma(t_i).$$

The advantage of changing variables is the fact that

$$\begin{aligned} E\left\{\frac{1}{n} \sum_{i=1}^{n} g(t_i; \lambda, \beta)\right\} &= \frac{1}{n} \int \{n_0 + n_1 w(t) + n_2[\lambda + (1 - \lambda) w(t)]\} \\ &\qquad \times g(t; \lambda, \beta)\, dF(t) \\ &= \int [w(t) - 1]\, dF(t) = 0 \end{aligned}$$

and the constraint has the same form as in Owen (1988, 1990).

By using Qin and Lawless' (1994) Lemma 1, we know that under certain conditions the constraint equation determines uniquely an implicit function $\alpha = \tilde{\alpha}(\lambda, \beta)$, in a neighborhood of (λ_T, β_T), where (λ_T, β_T) is the true value of (λ, β). Plugging the p_i's in (2.1), therefore, we have the profile log-likelihood

$$l(\lambda, \beta, \tilde{\alpha}(\lambda, \beta)) = l_1(\lambda, \beta, \tilde{\alpha}(\lambda, \beta)) + l_2(\lambda, \beta), \tag{2.3}$$

where

$$l_1(\lambda, \beta, \tilde{\alpha}(\lambda, \beta)) = -\sum_{i=1}^{n} \log\{1 + \tilde{\alpha}(\lambda, \beta) g(t_i; \lambda, \beta)\},$$

$$l_2(\lambda, \beta) = -\sum_{i=1}^{n} \log \gamma(t_i; \lambda, \beta) + \sum_{j=1}^{n_1} \log w(y_j; \beta) + \sum_{k=1}^{n_2} \log w_2(z_k; \beta, \lambda).$$

STEP 2. Maximize $l(\lambda, \beta, \tilde{\alpha}(\lambda, \beta))$, where $\tilde{\alpha}(\lambda, \beta)$ satisfies (2.2), with respect to (λ, β). Differentiating l, we have

$$\frac{\partial l}{\partial \lambda} = \frac{\partial l_2}{\partial \lambda} - \sum_{i=1}^{n} \frac{[\tilde{\alpha}(\lambda, \beta)\, \partial g(t_i; \lambda, \beta)/\partial\lambda + \partial\tilde{\alpha}(\lambda, \beta)/\partial\lambda\, g(t_i; \lambda, \beta)]}{1 + \tilde{\alpha}(\lambda, \beta) g(t_i; \lambda, \beta)} = 0,$$

$$\frac{\partial l}{\partial \beta} = \frac{\partial l_2}{\partial \beta} - \sum_{i=1}^{n} \frac{[\tilde{\alpha}(\lambda, \beta)\, \partial g(t_i; \lambda, \beta)/\partial\beta + \partial\tilde{\alpha}(\lambda, \beta)/\partial\beta\, g(t_i; \lambda, \beta)]}{1 + \tilde{\alpha}(\lambda, \beta) g(t_i; \lambda, \beta)} = 0.$$

Equivalently

$$\frac{\partial l}{\partial \lambda} = \frac{\partial l_2}{\partial \lambda} - \sum_{i=1}^{n} \frac{\tilde{\alpha}(\lambda, \beta)\, \partial g(t_i; \lambda, \beta)/\partial\lambda}{1 + \tilde{\alpha}(\lambda, \beta) g(t_i; \lambda, \beta)} = 0, \tag{2.4}$$

$$\frac{\partial l}{\partial \beta} = \frac{\partial l_2}{\partial \beta} - \sum_{i=1}^{n} \frac{\tilde{\alpha}(\lambda, \beta)\, \partial g(t_i; \lambda, \beta)/\partial\beta}{1 + \tilde{\alpha}(\lambda, \beta) g(t_i; \lambda, \beta)} = 0, \tag{2.5}$$

by using (2.2).

Let $(\tilde{\lambda}, \tilde{\beta})$ be a solution of (2.4) and (2.5) in the neighborhood of the true value of (λ_T, β_T). We call $(\tilde{\lambda}, \tilde{\beta})$ a maximum semiparametric likelihood estimator of (λ, β).

Now we present our main results; proofs are given in the Appendix.

THEOREM 1. *Suppose that*:

1. *The distribution function F is nondegenerate and $\|g\|^3$, $\|\partial g/\partial\lambda\|$ and $\|\partial g/\partial\beta\|$ are bounded by some integrable function in a neighborhood of the true value of (λ_T, β_T), where*

$$g(t; \lambda, \beta) = (w(t_i) - 1)\left\{\frac{n_0}{n} + \frac{n_2}{n}\lambda + w(t_i)\left[\frac{n_1}{n} + \frac{n_2}{n}(1 - \lambda)\right]\right\}^{-1}$$

and $\|\cdot\|$ denotes Euclidean norm.

2. *As* $n = n_0 + n_1 + n_2 \to \infty$, $n_i/n \to \rho_i > 0$, $i = 0, 1, 2$.
3. $0 < \lambda_T < 1$.

Under assumptions (1)–(3), *with probability* 1, $l(\lambda, \beta, \tilde{\alpha}(\lambda, \beta))$ *has a local maximum in an* $O(n^{-1/3})$ *neighborhood of* (λ_T, β_T). *Moreover, the maximizer* $(\tilde{\lambda}, \tilde{\beta})$ *and* $\tilde{\alpha} = \tilde{\alpha}(\tilde{\lambda}, \tilde{\beta})$ *satisfy the constraint equation* (2.2) *and score equations* (2.4) *and* (2.5), *and*

$$\sqrt{n}\begin{pmatrix} \tilde{\lambda} - \lambda_T \\ \tilde{\beta} - \beta_T \\ \tilde{\alpha} - 0 \end{pmatrix} \to N(0, U), \tag{2.6}$$

where U is defined in (A.5) *in the Appendix.*

Confidence intervals for (λ, β) can be constructed by using the normal approximation theory. A more straightforward method is to use the semiparametric generalized likelihood ratio test statistic

$$\begin{aligned} R(\lambda) &= 2\log\left\{\frac{\sup_{\lambda, \beta, F}\mathscr{L}(\lambda, \beta, F)}{\sup_{\beta, F}\mathscr{L}(\lambda, \beta, F)}\right\} \\ &= 2\left\{\sup_{\lambda, \beta} l(\lambda, \beta, \tilde{\alpha}(\lambda, \beta)) - \sup_{\beta} l(\lambda, \beta, \tilde{\alpha}(\lambda, \beta))\right\}, \end{aligned} \tag{2.7}$$

where $l(\lambda, \beta, \tilde{\alpha}(\lambda, \beta))$ is defined in (2.3).

THEOREM 2. *Under the same regularity conditions specified in Theorem* 1, *if* $H_0 : \lambda = \lambda_T$ *is true, then*

$$R(\lambda_T) \to \chi^2_{(1)}.$$

Therefore the 90 and 95% confidence intervals for λ are

$$\{\lambda \mid R(\lambda) \le 2.706\} \quad \text{and} \quad \{\lambda \mid R(\lambda) \le 3.841\}, \tag{2.8}$$

respectively.

If we are interested in the odds ratio parameter β_1, similarly we can prove the following result:

THEOREM 3. *Under the regularity conditions of Theorem* 1, *for testing* $H_0' : \beta_1 = \beta_{1T}$, *the likelihood ratio statistic*

$$R(\beta_{1T}) = 2\log\left\{\frac{\sup_{\lambda, \beta_0, \beta_1, F}\mathscr{L}(\lambda, \beta_0, \beta_1, F)}{\sup_{\lambda, \beta_0, F}\mathscr{L}(\lambda, \beta_0, \beta_{1T}, F)}\right\} \to \chi^2_{(1)} \tag{2.9}$$

if H_0' *is true.*

3. Distribution function estimations and model diagnostics. Although this paper is mainly concerned with estimation of (λ, β), an auxiliary result is the estimation of the underlying distribution functions.

Based on the observed data sets (1.1) and the model (1.2), the distribution functions $F(x)$, $G(y)$ and $H(z)$ can be estimated by

$$
\begin{aligned}
SF_n(x) &= \sum_{i=1}^{n} \tilde{p}_i I(t_i \le x) \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{1/\gamma\big(t_i; \tilde{\lambda}, \tilde{\beta}\big)}{1 + \tilde{\alpha}\big\{w\big(t_i; \tilde{\beta}\big) - 1\big\}/\gamma\big(t_i; \tilde{\lambda}, \tilde{\beta}\big)} I(t_i \le x), \\
SG_n(y) &= \sum_{i=1}^{n} \tilde{p}_i \exp\big(\tilde{\beta}_0 + \tilde{\beta}_1 t_i\big) I(t_i \le y), \\
SH_n(z) &= \sum_{i=1}^{n} \tilde{p}_i \big\{\tilde{\lambda} + (1 - \tilde{\lambda})\exp\big(\tilde{\beta}_0 + \tilde{\beta}_1 t_i\big)\big\} I(t_i \le z),
\end{aligned} \tag{3.1}
$$

respectively. We will prove the following theorem in the Appendix.

THEOREM 4. *Under the regularity conditions of Theorem* 1,

$$\sqrt{n}\,\{SF_n(x) - F(x)\} \to B(x), \tag{3.2}$$

where $B(x)$ *is a mean zero Gaussian process with continuous paths and covariance structure specified in* (A.7) *in the Appendix.*

If we define the empirical distribution estimators

$$
\begin{aligned}
EF_{n_0}(x) &= \frac{1}{n_0} \sum_{i=1}^{n_0} I(x_i \le x), \\
EG_{n_1}(y) &= \frac{1}{n_1} \sum_{j=1}^{n_1} I(y_j \le y), \\
EH_{n_2}(z) &= \frac{1}{n_2} \sum_{k=1}^{n_2} I(z_k \le z),
\end{aligned} \tag{3.3}
$$

then a natural diagnostic is to plot

$$\big(SF_n(x), EF_{n_0}(x)\big), \qquad \big(SG_n(y), EG_{n_1}(y)\big) \quad \text{and} \quad \big(SH_n(z), EH_{n_2}(z)\big).$$

Substantial differences would indicate that the model (1.2) is inadequate.

4. An example and some simulation results.

To test the performance of the proposed method with small sample sizes, we consider an example used by Anderson (1979) and some simulation results in this section. For comparison, we also consider a fully parametric approach. Assume the density function of x is parametrized as $f(x) = f(x; \eta)$, where η is a $q \times 1$ parameter. Then the log-likelihood based on the data (1.1) is

$$l_F = \sum_{i=1}^{n} \log f(t_i; \eta) + \sum_{j=1}^{n_1} \log w(y_j; \beta) + \sum_{k=1}^{n_2} \log w_2(z_k; \beta, \lambda). \tag{4.1}$$

The resulting likelihood ratio statistic is defined by

$$R_F(\lambda) = 2\left\{ \sup_{\lambda,\beta,\eta} l_F(\lambda,\beta,\eta) - \sup_{\beta,\eta} l_F(\lambda,\beta,\eta) \right\}. \tag{4.2}$$

By the standard large sample theory, we have $R_F(\lambda_T) \to \chi^2_{(1)}$ when $H_0 : \lambda = \lambda_T$ is correct. Hence the 90 and 95% fully parametric likelihood based confidence intervals are

$$\{\lambda | R_F(\lambda) \le 2.706\} \quad \text{and} \quad \{\lambda | R_F(\lambda) \le 3.841\}, \tag{4.3}$$

respectively.

First we consider Anderson's (1979) example. Corresponding to (1.1) we have the following data:

Population X: $1.15, 0.25, 2.31, 2.44, 3.28, 3.34$;

Population Y: $0.74, -0.50, 1.08, 1.34, -0.74, 0.15$;

Population Z: $-0.23, 0.71, 0.92, -0.53, -0.68, 1.04,$
$0.61, -0.88, -0.61, 0.59, 2.96, 2.59.$

Here the random variables x, y and z were randomly generated from $N(2,1)$, $N(0,1)$ and $\lambda N(2,1) + (1-\lambda)N(0,1)$, respectively, with $\lambda = 0.25$, by Anderson (1979). Of course, with these data, (1.1) and (1.2) are satisfied. The maximum semiparametric and fully parametric likelihood estimators of λ are 0.189 and 0.187, respectively. The 90 and 95% semiparametric confidence intervals (2.8) are $(0, 0.569)$ and $(0, 0.658)$ and the 90 and 95% fully parametric confidence intervals (4.3) are $(0, 0.470)$ and $(0, 0.538)$, respectively. Figure 1 is the semiparametric likelihood ratio $R(\lambda)$ and fully parametric likelihood ratio $R_F(\lambda)$ plot based on these data. Figure 2 displays the fitted semiparametric cumulative distribution functions (3.1) for the Anderson data and the corresponding empirical cumulative distribution functions (3.3). They match pretty well.

Next we describe some simulations. We still consider the model used by Anderson (1979). We generated 1000 samples by using Numerical recipes subroutines ran1 and gasdev [Press, Teukólsky, Vetterling and Flannery (1992)]. We assume that x is $N(2,1)$, y is $N(0,1)$ and z is $\lambda N(2,1) + (1-\lambda)N(0,1)$. Therefore, $w(x) = \exp(\beta_0 + \beta_1 x)$ with $\beta_0 = 2$, $\beta_1 = -2$. For comparison we also consider fully parametric likelihood based inferences, where F and G are $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. Confidence intervals were calculated for the proportion parameter λ based on the semiparametric likelihood confidence intervals (2.8) and the fully parametric likelihood confidence intervals (4.3), respectively. In Table 1, we report the estimated true coverage, mean length and mean value of the midpoint of those confidence intervals. Each value in the table is the average of 1000 simulations, and S and P denote the semiparametric and fully parametric approaches, respectively. From the table we can see that the performance of the semiparametric likelihood ratio confidence intervals is satisfactory. All empirical coverage levels are close to the nominal levels. The lengths of

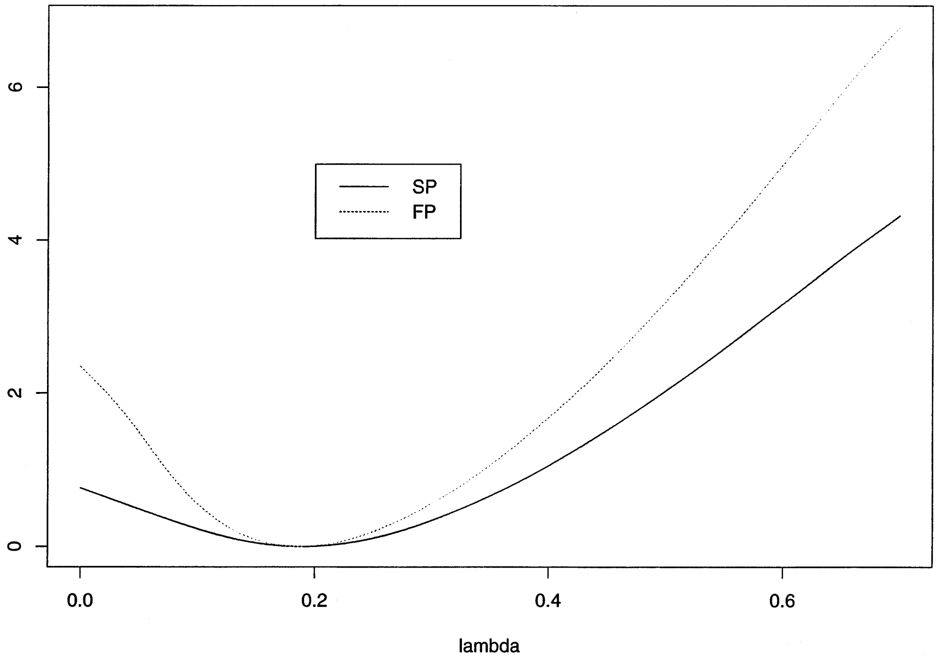

FIG. 1. *Likelihood ratio plot*

confidence intervals are almost the same based on the semiparametric model (1.2) or fully parametric normal mixture model. Therefore, the loss of efficiency when the true distribution is a mean mixture of normals is negligible by using model (1.2). The coverage levels are highest for $\lambda = 0.5$ compared with small and large λ. When λ is small ($\lambda = 0.25$), the mean values of the midpoints of those confidence intervals are a little bit larger than the true values. This is because we considered only λ between 0 and 1 such that $\{\lambda : R(\lambda) < 2.706 \text{ or } 3.841\}$, modifying those confidence intervals with negative left endpoints to have left endpoints 0. Similar changes were made for $\lambda = 0.75$. Figure 3 is the Q-Q plot for the 1000 replications of $R(\lambda)$ at the true value $\lambda = 0.5$ ($n_0 = n_1 = 30$, $n_2 = 60$) versus standard $\chi^2_{(1)}$. The approximation appears satisfactory.

5. Concluding remarks. Estimating mixture proportions under data structures (1.1) is formulated in the semiparametric model (1.2). The distributions are modeled nonparametrically, but are assumed to be related through an "exponential tilt." Based on our limited simulation studies, the loss of

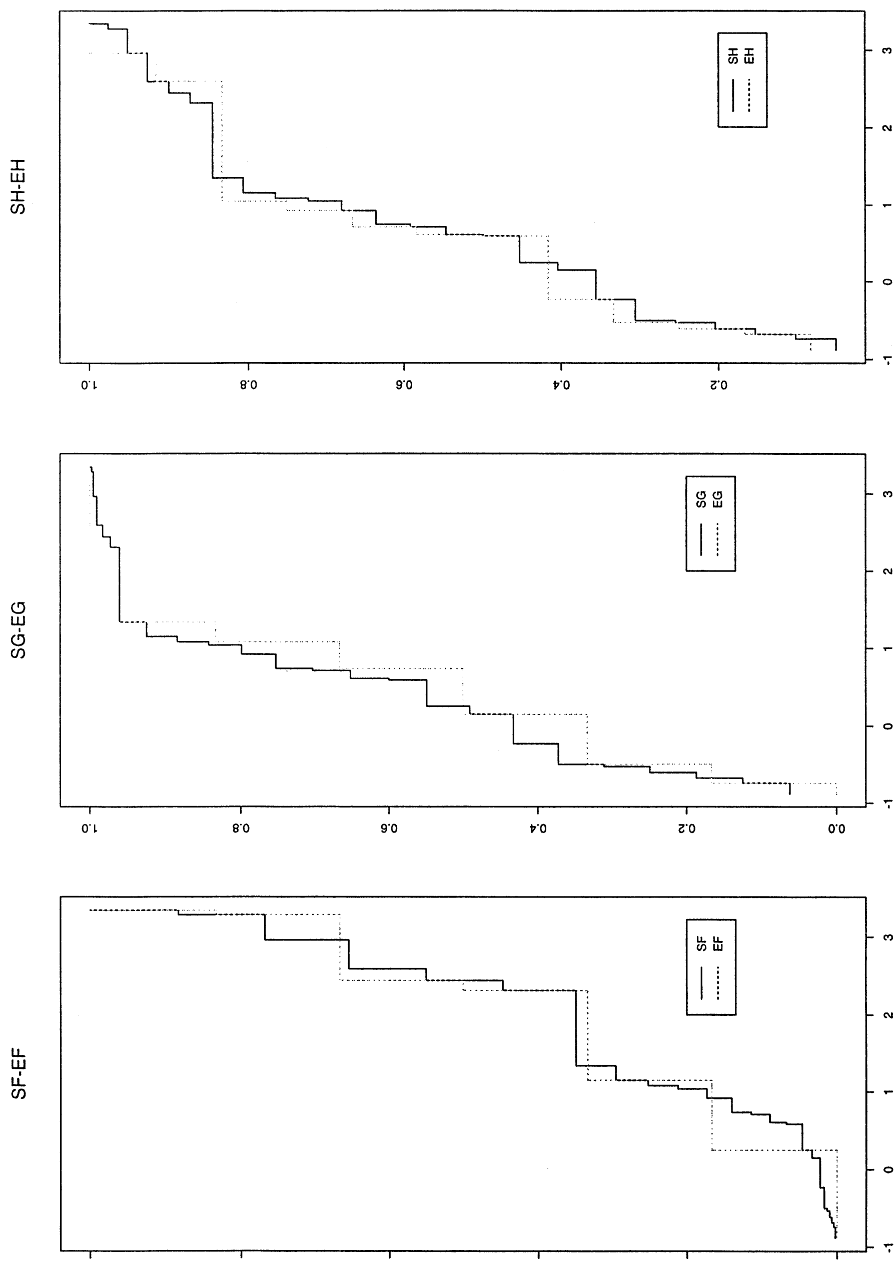

FIG. 2. *Distribution function plots*

TABLE 1
Empirical average length, midpoint and coverage from 1000 samples

Model	λ	90% CD Cov. (%)	90% CD Av. Length	90% CD Av. Midpt.	95% CD Cov. (%)	95% CD Av. Length	95% CD Av. Midpt.
			$n_0 = 30, n_1 = 30, n_2 = 60$				
S	0.25	87.5	0.3063	0.2625	92.8	0.3607	0.2687
P	0.25	90.0	0.2922	0.2629	94.9	0.3456	0.2679
S	0.50	88.7	0.3413	0.4994	93.8	0.4082	0.4954
P	0.50	90.3	0.3258	0.5007	94.9	0.3880	0.5007
			$n_0 = 20, n_1 = 20, n_2 = 40$				
S	0.50	89.6	0.4168	0.4958	94.2	0.4941	0.4959
P	0.50	89.5	0.3980	0.4975	95.3	0.4732	0.4981
S	0.75	87.4	0.3710	0.7224	92.7	0.4345	0.7127
P	0.75	86.8	0.3458	0.7294	93.9	0.4079	0.7205

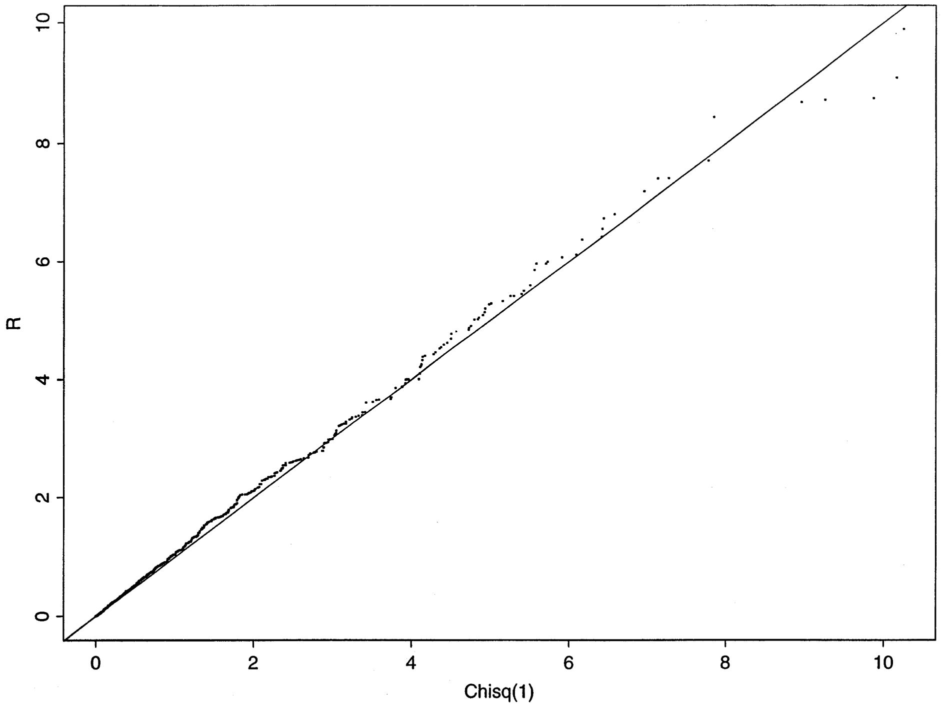

FIG. 3. *Q-Q plot*

information by using a semiparametric approach is insignificant, even when the forms of f and g are available.

The linear log ratio structure (1.2) is not essential to the development given here, which can readily be extended to the more general model

$$\frac{g(x)}{f(x)} = \exp\{\beta_0 + \psi(x, \beta)\},$$

where ψ is a continuous differentiable function. A variety of problems with two or more unknown distribution functions could be tackled with the same essential methodology.

APPENDIX

In this section we give the proofs of the theorems in Sections 2 and 3. Note the log-likelihood (2.3), $l(\lambda, \beta, \tilde{\alpha}(\lambda, \beta))$, contains two terms. The first term, $l_1(\lambda, \beta, \tilde{\alpha}(\lambda, \beta))$, is the empirical likelihood [Qin and Lawless (1994)].

PROOF OF THEOREM 1. Note that when $\lambda = \lambda_T$ and $\beta = \beta_T$,

$$E\left\{\frac{1}{n}\sum_{i=1}^{n} g(t_i; \lambda_T, \beta_T)\right\} = \int (w(t) - 1)\, dF(t) = 0.$$

As in Lemmas 1 and 2 in Qin (1993), we can prove that $l(\lambda, \beta, \alpha(\lambda, \beta))$ has a local maximum in a $O(n^{-1/3})$ neighborhood of (λ_T, β_T). Also when (λ, β) falls in this neighborhood, $\|\tilde{\alpha}(\lambda, \beta)\| = O_p(n^{-1/3})$.

For notational convenience, we write

$$\frac{\partial l}{\partial \alpha} = -\sum_{i=1}^{n} \frac{g(t_i; \lambda, \beta)}{1 + \alpha g(t_i; \lambda, \beta)}.$$

Then the constraint equation (2.2) is equivalent to $\partial l/\partial\alpha = 0$ and $(\tilde{\lambda}, \tilde{\beta}, \tilde{\alpha})$ satisfies $\partial l/\partial\lambda = 0$, $\partial l/\partial\beta = 0$ and $\partial l/\partial\alpha = 0$. To show the asymptotic normality of $(\tilde{\lambda}, \tilde{\beta}, \tilde{\alpha})$, we expand $\partial l(\tilde{\lambda}, \tilde{\beta}, \tilde{\alpha})/\partial\lambda$, $\partial l(\tilde{\lambda}, \tilde{\beta}, \tilde{\alpha})/\partial\beta$ and $\partial l(\tilde{\lambda}, \tilde{\beta}, \tilde{\alpha})/\partial\alpha$ at $(\lambda_T, \beta_T, 0)$. We can see easily that

$$\begin{pmatrix} \tilde{\lambda} - \lambda_T \\ \tilde{\beta} - \tilde{\beta}_T \\ \tilde{\alpha} - 0 \end{pmatrix} = -S_n^{-1} Q_n + o_p(n^{-1/2}) \tag{A.1}$$

where

$$S_n^{-1} = -\begin{pmatrix} \frac{1}{n}\frac{\partial^2 l}{\partial\lambda\,\partial\lambda} & \frac{1}{n}\frac{\partial^2 l}{\partial\lambda\,\partial\beta^{\tau}} & \frac{1}{n}\frac{\partial^2 l}{\partial\lambda\,\partial\alpha} \\ \frac{1}{n}\frac{\partial^2 l}{\partial\beta\,\partial\lambda} & \frac{1}{n}\frac{\partial^2 l}{\partial\beta\,\partial\beta^{\tau}} & \frac{1}{n}\frac{\partial^2 l}{\partial\beta\,\partial\alpha} \\ \frac{1}{n}\frac{\partial^2 l}{\partial\alpha\,\partial\lambda} & \frac{1}{n}\frac{\partial^2 l}{\partial\alpha\,\partial\beta^{\tau}} & \frac{1}{n}\frac{\partial^2 l}{\partial\alpha\,\partial\alpha} \end{pmatrix}^{-1}_{(\lambda_T, \beta_T, 0)}$$

and

$$Q_n = \begin{pmatrix} \dfrac{1}{n} \dfrac{\partial l(\lambda_T, \beta_T, 0)}{\partial \lambda} \\ \dfrac{1}{n} \dfrac{\partial l(\lambda_T, \beta_T, 0)}{\partial \beta} \\ \dfrac{1}{n} \dfrac{\partial l(\lambda_T, \beta_T, 0)}{\partial \alpha} \end{pmatrix} + o_p(n^{-1/2})$$

After long algebraic calculations, we have

$$\text{(A.2)} \qquad \begin{gathered} S_n \to S = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{12}^\tau & s_{22} & s_{23} \\ s_{13}^\tau & s_{23}^\tau & s_{33} \end{pmatrix} = \begin{pmatrix} e_{11} & e_{12} \\ e_{12}^\tau & s_{33} \end{pmatrix}, \\ e_{11} = \begin{pmatrix} s_{11} & s_{12} \\ s_{12}^\tau & s_{22} \end{pmatrix}, \qquad e_{12} = e_{21}^\tau = (s_{13}^\tau, s_{23}^\tau)^\tau, \end{gathered}$$

where

$$s_{11} = \rho_2^2 \eta_1 - \rho_2 \eta_2, \qquad s_{12} = \rho_2 \phi_1 - \rho_2 \psi_1, \qquad s_{13} = -\rho_2 \eta_1,$$

$$s_{22} = -ab\phi_2 + \lambda_T(1-\lambda_T)\rho_2 \psi_2, \qquad s_{23} = -\phi_1, \qquad s_{33} = \eta_1,$$

$$\gamma(t) = a + bw(t), \qquad a = (\rho_0 + \rho_2 \lambda_T), \qquad b = \rho_1 + \rho_2(1-\lambda_T),$$

$$\delta = \rho_0 b^2 + \rho_1 a^2 + \rho_2(\lambda_T - a)^2,$$

$$\eta_1 = \int \frac{\{1-w(t)\}^2}{\gamma(t)}\, dF(t), \qquad \eta_2 = \int \frac{\{1-w(t)\}^2}{w_2(t)}\, dF(t),$$

$$\phi_1 = \int \frac{\partial w(t)}{\partial \beta} \frac{1}{\gamma(t)}\, dF(t), \qquad \phi_2 = \int \frac{\partial^2 w(t)}{\partial \beta\, \partial \beta^\tau} \frac{1}{\gamma(t)}\, dF(t),$$

$$\psi_1 = \int \frac{\partial w(t)}{\partial \beta} \frac{1}{w_2(t)}\, dF(t), \qquad \psi_2 = \int \frac{\partial^2 w(t)}{\partial \beta\, \partial \beta^\tau} \frac{1}{w_2(t)}\, dF(t).$$

Note that

$$\text{(A.3)} \qquad Q_n = \frac{1}{n} \left\{ \sum_{i=1}^{n_0} q_0(x_i) + \sum_{j=1}^{n_1} q_1(y_j) + \sum_{k=1}^{n_2} q_2(z_k) \right\},$$

where

$$q_0(x) = -\begin{pmatrix} \dfrac{\rho_2(1-w(x))}{\gamma(x)} \\ \dfrac{b\, \partial w(x)/\partial \beta}{\gamma(x)} \\ g(x; \lambda_T, \beta_T) \end{pmatrix}, \qquad q_1(y) = -\begin{pmatrix} \dfrac{\rho_2(1-w(y))}{\gamma(y)} \\ \dfrac{b\, \partial w(y)/\partial \beta}{\gamma(y)} - \dfrac{\partial w(y)/\partial \beta}{w(y)} \\ g(y; \lambda_T, \beta_T) \end{pmatrix}$$

and

$$q_2(z) = -\begin{pmatrix} \dfrac{\rho_2(1-w(z))}{\gamma(z)} - \dfrac{(1-w(z))}{w_2(z)} \\ \dfrac{b\,\partial w(z)/\partial\beta}{\gamma(z)} - \dfrac{(1-\lambda_T)\,\partial w(z)/\partial\beta}{w_2(z)} \\ g(z;\lambda_T,\beta_T) \end{pmatrix}.$$

By the central limit theorem we have

$$\sqrt{n}\,Q_n \to N(0,V),$$

where

$$V = \begin{pmatrix} -e_{11} - \delta e_{12}e_{12}^{\tau} & -\delta e_{12}s_{33} \\ -\delta e_{12}^{\tau}s_{33} & s_{33} - \delta s_{33}^2 \end{pmatrix}. \tag{A.4}$$

Therefore

$$\sqrt{n}\begin{pmatrix} \tilde\lambda - \lambda_T \\ \tilde\beta - \beta_T \\ \tilde\alpha - 0 \end{pmatrix} \to N(0,U), \qquad U = S^{-1}VS^{-1} = \begin{pmatrix} u_{11} & 0 \\ 0 & u_{22} \end{pmatrix}, \tag{A.5}$$

where

$$u_{11} = -\begin{pmatrix} s_{11} - s_{13}s_{33}^{-1}s_{13}^{\tau} & s_{12} - s_{13}s_{33}^{-1}s_{23}^{\tau} \\ s_{12}^{\tau} - s_{23}s_{33}^{-1}s_{13}^{\tau} & s_{22} - s_{23}s_{33}^{-1}s_{23}^{\tau} \end{pmatrix}^{-1},$$

$$u_{22} = -s_{33}^{-1}e_{12}^{\tau}u_{11}e_{12}s_{33}^{-1} + s_{33}^{-1} - \delta.$$

Because the off-diagonal elements of U are zero, $(\tilde\lambda, \tilde\beta)$ and $\tilde\alpha$ are asymptotically independent. □

PROOF OF THEOREM 2. To prove that $R(\lambda_T)$ converges to a chi-squared variable, we write $\tilde\xi = (\tilde\beta - \beta_T, \tilde\alpha - 0)^{\tau}$. By (A.1) we have

$$\begin{pmatrix} \tilde\lambda - \lambda_T \\ \tilde\xi \end{pmatrix} = -\begin{pmatrix} \dfrac{1}{n}\dfrac{\partial^2 l}{\partial\lambda\,\lambda^{\tau}} & \dfrac{1}{n}\dfrac{\partial^2 l}{\partial\lambda\,\partial\xi^{\tau}} \\ \dfrac{1}{n}\dfrac{\partial^2 l}{\partial\xi\,\partial\lambda^{\tau}} & \dfrac{1}{n}\dfrac{\partial^2 l}{\partial\xi\,\partial\xi^{\tau}} \end{pmatrix}^{-1} \begin{pmatrix} Q_{1n} \\ Q_{2n} \end{pmatrix} + o_p(n^{-1/2})$$

$$= -\begin{pmatrix} s_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1} \begin{pmatrix} Q_{1n} \\ Q_{2n} \end{pmatrix} + o_p(n^{-1/2}),$$

where

$$c_{12} = c_{21}^{\tau} = (s_{12}, s_{13}), \qquad c_{22} = \begin{pmatrix} s_{22} & s_{23} \\ s_{32} & s_{33} \end{pmatrix},$$

$$Q_{1n} = \frac{1}{n}\frac{\partial l(\lambda_T,\beta_T,0)}{\partial\lambda}, \qquad Q_{2n} = \left(\frac{1}{n}\frac{\partial l(\lambda_T,\beta_T,0)}{\partial\beta}, \frac{1}{n}\frac{\partial l(\lambda_T,\beta_T,0)}{\partial\alpha}\right)^{\tau}.$$

Let $\hat{\beta}$ maximize $l(\lambda_T, \beta, \tilde{\alpha}(\lambda_T, \beta))$ with λ fixed at λ_T. It satisfies

$$\text{(A.6)} \qquad \frac{\partial l\left(\lambda_T, \hat{\beta}, \hat{\alpha}\right)}{\partial \beta} = 0, \qquad \frac{\partial l\left(\lambda_T, \hat{\beta}, \hat{\alpha}\right)}{\partial \alpha} = 0,$$

where $\hat{\alpha} = \tilde{\alpha}(\lambda_T, \hat{\beta})$. Expanding $(\partial l(\partial_T, \hat{\beta}, \hat{\alpha}))/\partial\beta$ and $(\partial l(\lambda_T, \hat{\beta}, \hat{\alpha}))/\partial\alpha$ at the point $(\lambda_T, \beta_T, 0)$, we have

$$\hat{\xi} = \begin{pmatrix} \hat{\beta} - \beta_T \\ \hat{\alpha} - 0 \end{pmatrix} = -c_{22}^{-1} Q_{2n} + o_p(n^{-1/2}).$$

Hence

$$\begin{aligned}
\begin{pmatrix} \lambda_T - \tilde{\lambda} \\ \hat{\xi} - \tilde{\xi} \end{pmatrix} &= -\begin{pmatrix} 0 & 0 \\ 0 & c_{22}^{-1} \end{pmatrix}\begin{pmatrix} Q_{1n} \\ Q_{2n} \end{pmatrix} + \begin{pmatrix} s_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1}\begin{pmatrix} Q_{1n} \\ Q_{2n} \end{pmatrix} + o_p(n^{-1/2}) \\
&= \left\{\begin{pmatrix} 0 & 0 \\ 0 & -c_{22}^{-1} \end{pmatrix}\begin{pmatrix} s_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} + \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}\right\}\begin{pmatrix} \tilde{\lambda} - \lambda_T \\ \tilde{\xi} \end{pmatrix} + o_p(n^{-1/2}) \\
&= -\begin{pmatrix} I \\ -c_{22}^{-1} c_{21} \end{pmatrix}\left(\tilde{\lambda} - \lambda_T\right) + o_p(n^{-1/2}).
\end{aligned}$$

Expanding $l(\lambda_T, \hat{\beta}, \hat{\alpha})$ at $(\tilde{\lambda}, \tilde{\beta}, \tilde{\alpha})$, we have

$$\begin{aligned}
l\left(\lambda_T, \hat{\beta}, \hat{\alpha}\right) - l\left(\tilde{\lambda}, \tilde{\beta}, \tilde{\alpha}\right) &= \tfrac{1}{2} n\left(\lambda_T - \tilde{\lambda}, \hat{\xi} - \tilde{\xi}\right)\begin{pmatrix} s_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}\begin{pmatrix} \lambda_T - \tilde{\lambda} \\ \hat{\xi} - \tilde{\xi} \end{pmatrix} + o_p(1) \\
&= \tfrac{1}{2} n\left(\tilde{\lambda} - \lambda_T\right)\left(s_{11} - c_{12} c_{22}^{-1} c_{21}\right)\left(\tilde{\lambda} - \lambda_T\right) + o_p(1) \\
&\to -\tfrac{1}{2}\chi^2_{(1)},
\end{aligned}$$

since by (A.5) the asymptotic variance of $\sqrt{n}(\tilde{\lambda} - \lambda_T)$ is $-(s_{11} - c_{12} c_{22}^{-1} c_{21})^{-1}$. □

The proof of Theorem 3 is similar to the proof of Theorem 2, hence it is omitted.

PROOF OF THEOREM 4. Denote

$$r\left(t_i; \tilde{\theta}\right) = \frac{1/\gamma\left(t_i; \tilde{\lambda}, \tilde{\beta}\right)}{1 + \tilde{\alpha}\left\{w\left(t_i; \tilde{\beta}\right) - 1\right\}/\gamma\left(t_i; \tilde{\lambda}, \tilde{\beta}\right)},$$

where $\tilde{\theta} = (\tilde{\lambda}, \tilde{\beta}, \tilde{\alpha})$ and denote $\theta_T = (\lambda_T, \beta_T, 0)$. Then

$$SF_n(t) = \frac{1}{n}\sum_{i=1}^{n} r\left(t_i, \tilde{\theta}\right) I(t_i \le t).$$

By expanding $r(t_i;\tilde{\theta})$ at θ_T and using (A.1) and (A.2), we can prove that

$$\begin{aligned} SF_n(t) - F(t) &= \frac{1}{n}\sum_{i=1}^{n} r(t_i;\theta_T)I(x_i \le t) - F(t) \\ &\quad + \frac{\partial r(t_i;\theta_T)}{\partial\theta} I(t_i \le t)\big(\tilde{\theta} - \theta_T\big) + R_{1n}(t) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{I(t_i \le t)}{\gamma(t_i;\lambda_T,\beta_T)} - F(x) - r_1(t)S^{-1}Q_n + R_{2n}(t), \end{aligned}$$

where $R_{in}(t)$, $i = 1,2$, satisfy $\sup_{-\infty<t<\infty}|R_{in}(t)| = o_p(n^{-1/2})$ and

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \frac{\partial r(t_i;\theta_T)}{\partial\theta} I(t_i \le t) \to r_1(t) \\ &\qquad = \int \{\rho_0 + \rho_1 w_1(u) + \rho_2 w_2(u)\} \frac{\partial r(u;\theta_T)}{\partial\theta} I(u \le t)\, du, \quad \text{a.s.} \end{aligned}$$

Let

$$r_2(t) = -r_1(t)S^{-1}, \qquad \varepsilon_0(x_i,t) = \frac{I(x_i \le t)}{\gamma(x_i)} - \int_{-\infty}^{t} \frac{dF(x)}{\gamma(x)},$$

$$\varepsilon_1(y_j,t) = \frac{I(y_j \le t)}{\gamma(y_j)} - \int_{-\infty}^{t} \frac{w_1(y)\,dF(y)}{\gamma(y)},$$

$$\varepsilon_2(z_k,t) = \frac{I(z_k \le t)}{\gamma(z_k)} - \int_{-\infty}^{t} \frac{w_2(z)\,dF(z)}{\gamma(z)}.$$

Therefore, by using (A.3) we have

$$\begin{aligned} \sqrt{n}\,\{SF_n(t) - F(t)\} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n_0} \{\varepsilon_0(x_i;t) + r_2(t)q_0(x_i)\} \\ &\quad + \sum_{j=1}^{n_1} \big\{\varepsilon_1(y_j;t) + r_2(t)q_1(y_j)\big\} \\ &\quad + \sum_{k=1}^{n_2} \{\varepsilon_2(z_k;t) + r_2(t)q_2(z_k)\} + o_p(1). \end{aligned}$$

By using the criteria of Billingsley [(1968), page 128], we can show that

$$\sqrt{n}\,\{SF_n(t) - F(t)\} \to B(t) \quad \text{in distribution,}$$

where $B(t)$ is a mean zero Gaussian process with continuous paths and covariance structure

$$\begin{aligned} \Sigma(t_1,t_2) &= \rho_0 \operatorname{cov}\{\varepsilon_0(x_i,t_1) + r_2(t_1)q_0(x_i), \varepsilon_0(x_i,t_2) + r_2(t_2)q_0(x_i)\} \\ &\quad + \rho_1 \operatorname{cov}\big\{\varepsilon_1(y_j,t_1) + r_2(t_1)q_1(y_j), \varepsilon_1(y_j,t_2) + r_2(t_2)q_1(y_j)\big\} \qquad \text{(A.7)} \\ &\quad + \rho_2 \operatorname{cov}\{\varepsilon_2(z_k,t_1) + r_2(t_1)q_2(z_k), \varepsilon_2(z_k,t_2) + r_2(t_2)q_2(z_k)\}, \\ &\qquad\qquad t_1 \le t_2. \quad \square \end{aligned}$$

Acknowledgments. The author wishes to thank an Associate Editor, referees and Professor Paul Smith for their insightful comments that led to a substantially improved version of this paper.

REFERENCES

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

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BRESLOW, N. and DAY, N. E. (1980). *Statistical Methods in Cancer Research: I. The Analysis of Case-Control Studies*. IARC, Lyon.

CHEN, S. X. and HALL, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21** 1166–1181.

COX, D. R. and SNELL, E. J. (1989). *Analysis of Binary Data*, 2nd ed. Chapman and Hall, London.

DICICCIO, T. J., HALL, P. and ROMANO, J. P. (1989). Comparison of parametric and empirical likelihood. *Biometrika* **76** 447–456.

EFRON, B. (1975). The efficiency of logistic regression compared to discrimination analysis. *J. Amer. Statist. Assoc.* **70** 892–897.

EFRON, B. and TIBSHIRANI, R. (1996). Using specially designed exponential families for density estimation. *Ann. Statist.* **24** 2431–2461.

HALL, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.* **18** 121–140.

HALL, P. and LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *I. S. Review* **58** 109–127.

HALL, P. and TITTERINGTON, D. M. (1984). Efficient nonparametric estimation of mixture proportions. *J. Roy. Statist. Soc. Ser. 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.

LANCASTER, T. and IMBENS, G. (1996). Case-control studies with contaminated controls. *J. Econometrics* **71** 145–160.

MURRAY, G. D. and TITTERINGTON, D. M. (1978). Estimation problems with data from a mixture. *J. Roy. Statist. Soc. Ser. C* **27** 325–334.

O'NEILL, T. J. (1980). The general distribution of the error rate of a classification procedure with applications to the logistic regression discrimination. *J. Amer. Statist. Assoc.* **75** 154–160.

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249.

OWEN, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* **18** 90–120.

PRENTICE, R. L. and PYKE, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66** 403–411.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. and FLANNERY, B. P. (1992). *Numerical Recipes in C*, 2nd ed. Cambridge Univ. Press.

QIN, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.* **21** 1182–1196.

QIN, J. and LAWLESS, J. F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742
AND
DEPARTMENT OF EPIDEMIOLOGY AND BIOSTATISTICS
MEMORIAL SLOAN–KETTERING CANCER CENTER
1275 YORK AVENUE, NEW YORK, NEW YORK 10021