1-2006

# Semi-parametric inference in a bivariate (multivariate) mixture model

Denis H. Y. LEUNG

*Singapore Management University*, denisleung@smu.edu.sg

Jing QIN

*National Institute of Allergy and Infectious Diseases*

Statistica Sinica **16**(2006), 153-163

# SEMI-PARAMETRIC INFERENCE IN A BIVARIATE (MULTIVARIATE) MIXTURE MODEL

Denis Heng-Yan Leung and Jing Qin

*Singapore Management University and*
*National Institute of Allergy and Infectious Diseases*

*Abstract:* We consider estimation in a bivariate mixture model in which the component distributions can be decomposed into identical distributions. Previous approaches to estimation involve parametrizing the distributions. In this paper, we use a semi-parametric approach. The method is based on the exponential tilt model of Anderson (1979), where the log ratio of probability (density) functions from the bivariate components is linear in the observations. The proposed model does not require training samples, i.e., data with confirmed component membership. We show that in bivariate mixture models, parameters are identifiable. This is in contrast to previous works, where parameters are identifiable if and only if each univariate marginal model is identifiable (Teicher (1967)).

*Key words and phrases:* Empirical likelihood, multivariate mixture, semi-parametric, Shannon's mutual information.

## 1. Introduction

Mixture models have been studied by many researchers, dating back to at least the late 1800s (Pearson (1893, 1895)). There is now a vast literature on mixture model inference, we refer readers to the books by Titterington, Smith and Makov (1985) and Lindsay (1995). Statistical analysis of mixture data is not trivial since in general there is no closed form for the maximum likelihood estimators. Empirical methods or an EM algorithm are needed. Furthermore, theoretical results are difficult since the mixture parameter may lie on the boundary of the parameter space. In addition, some nuisance parameters may be absent under the null hypothesis and hence the null distribution of the likelihood ratio test statistic is, in general, unknown even when the sample size is large. Despite these difficulties, mixture models are still popular because of their flexibility in modeling data in a large variety of applications, for example, in social economics (Keane and Wolpin (1997) and Cameron and Heckman (2001)), in marketing (Desarbo, Degrenatu, Wedel and Saxton (2001) and Moe and Fader (2002)), in finance (Zangari (1996), Venkatraman (1997) and Hull and White (1998)), in social psychology (Thomas and Horton (1997)) and in biomedical sciences (Vounatsu, Smith and Smith (1998) and Zou, Fine and Yandell (2002)).

Apart from the technical difficulties discussed above, there is the added concern of robustness when modeling multivariate data using parametric mixture models. As Thomas and Lohaus (1993) and Hettmansperger and Thomas (2000) pointed out, component distributions in a mixture model are often asymmetric, let alone normal. Earlier, MacDonald (1975) also suggested that most parametric estimators are very sensitive to departures from distributional assumptions. Therefore, when modeling multivariate mixture data, it is desirable to make inference under minimal assumptions on the underlying component distributions. To answer this need, we here consider a semi-parametric approach. We assume the component densities are related by Anderson's (1979) exponential tilt model. For ease of illustration, we focus on estimation in a bivariate two-component mixture where the component distributions in the multivariate mixture model can be decomposed into independent identical distributions. However, the method we propose can be generalized to more general situations, such as the decomposition of the components into independent but non-identical distributions. Decomposition of the components into independent identical distributions means that, given we know which component an observation comes from, the within component data are independent and identically distributed (i.i.d.). Such an assumption was used in a study of developmental psychology (Thomas and Horton (1993, 1997) and Hettmansperger and Thomas (2000)). In medical applications, Hall and Zhou (2003) also discussed a problem with $k$-variate data drawn from a mixture of two distributions, each having independent components. They showed that when $k \geq 3$, it is possible to identify the mixture model even without parametrizing the underlying component distributions. Finally, conditional independence of multivariate data can also be seen as a special case of the popular random effects model with clustered data. In such a model, observations from individuals within a cluster are often considered to be independent, when conditioned on an unobserved parameter. The unconditional joint distribution of the data from all the clusters is then a mixture distribution (Liu and Pierce (1994) and Qu and Hadgu (1998)).

The organization of the rest of this paper is as follows. In Section 2, we present the methodology and the main results. By using Vardi's (1985) biased sampling inference technique and Owen's (1988) empirical likelihood, we maximize the empirical likelihood function subject to suitable constraints. The profiled empirical likelihood is related to Shannon's mutual information. It is shown that the maximum semi-parametric likelihood estimate has an asymptotic normal distribution. We also demonstrate that the profiled empirical likelihood behaves like a conventional parametric likelihood, so that the profiled likelihood ratio statistic has an asymptotic chi-squared distribution. Section 3 gives some simulation results.

## 2. Method and Main Results

We assume i.i.d. bivariate data $(x_1, y_1), \ldots, (x_n, y_n)$ from a mixture distribution with density

$$h(x, y) = \lambda f(x) f(y) + (1 - \lambda) g(x) g(y), \tag{1}$$

where $\lambda$ represents the mixing proportion of the two components and $f$ and $g$ are two univariate density functions. We let $H, F$, and $G$ denote the distribution functions corresponding to $h, f$, and $g$, respectively. Throughout the paper, we let $\eta = (\lambda, \alpha, \beta)$, and $\eta_0 = (\lambda_0, \alpha_0, \beta_0)$ be the true values of the parameters.

Assume that $f$ and $g$ are related by an exponential tilt model (Anderson (1979))

$$g(x) = \exp(\alpha + \beta x) f(x), \tag{2}$$

where $\alpha$ and $\beta$ are unknown parameters. If we let $D$ be the indicator for the component membership of an observation ($D = 0$ if it is from $F$ and $D = 1$ if it is from $G$), then the exponential tilt model is equivalent to the logistic regression model

$$P(D = 1|x) = \frac{\exp(\alpha^* + \beta x)}{1 + \exp(\alpha^* + \beta x)},$$

where $\alpha^* = \alpha + \log\{(1 - \lambda)/\lambda\}$ (Qin (1998)). With the exponential tilt model, the marginal distribution of $x$ is

$$h_1(x) = [(1 - \lambda) + \lambda \exp(\alpha + \beta x)] f(x). \tag{3}$$

Similarly, we can define $h_2(y)$ as the marginal distribution of $y$. Model (3) may be considered as a biased sample problem with weight $w(x) = (1 - \lambda) + \lambda \exp(\alpha + \beta x)$, which depends on the parameters $(\alpha, \beta, \lambda)$. Discussions on biased sampling problems can be found in the literature, for example Vardi (1985), Gill, Vardi and Wellner (1988), Qin (1998) and Gilbert, Lele and Vardi (1999). Note that the underlying parameters cannot be identified from the marginal density since the $x_i$'s and $y_i$'s have the same distribution, and $f$ and $(\alpha, \beta, \lambda)$ are unknown. This is in contrast to the situation with a parametric multivariate mixture model with independent components, where the model is identifiable as long as the parameters in the marginal models are identifiable (Teicher (1967)). The non-identifiablity problem can be illustrated by the following example. Suppose the mixture model is used to model brand loyalty in a product. The population of consumers can be considered to be made up of those who are loyal to a particular brand and those who are not brand loyal. A mixture distribution can be used to model their purchase behavior over a number of purchasing periods. In the notation of this paper, the dimension of $F$ (and $G$) represents the number of purchasing periods that consumers' behavior have been recorded. If only one purchase record is recorded per consumer, there is no way to determine whether the customer is brand loyal, hence the model is unidentifiable.

Returning to the estimation problem at (1) and (2), the joint density of $(x, y)$ is

$$h(x,y) = [\lambda + (1-\lambda)\exp\{2\alpha + \beta(x+y)\}]f(x)f(y). \tag{4}$$

It is easy to observe that (4) is identifiable by noting that $f$ is cancelled out in $h(x,y)/\{h_1(x)h_2(y)\}$, and both the joint and marginal distributions can be estimated using the observed data.

Based on the observed data, the likelihood is

$$L = \prod_{i=1}^{n}[\lambda + (1-\lambda)\exp\{2\alpha + \beta(x_i+y_i)\}]dF(x_i)dF(y_i).$$

Obviously, $F$ only jumps at each observed $(x_i, y_i), i = 1, 2, \ldots, n$. For convenience, we denote the jump sizes as $p_j, j = 1, 2, \ldots, 2n$. Then the log-likelihood is

$$l = \sum_{i=1}^{n}\log[\lambda + (1-\lambda)\exp\{2\alpha + \beta(x_i+y_i)\}] + \sum_{j=1}^{2n}\log p_j.$$

For fixed $(\lambda, \alpha, \beta)$, we maximize the $p_j$'s subject to the constraints

$$\sum_{j=1}^{2n} p_j = 1, \quad p_j \geq 0, \quad \sum_{j=1}^{2n} p_j\{\exp(\alpha + \beta t_j) - 1\} = 0,$$

where $(t_1, \ldots, t_{2n}) = (x_1, y_1, \ldots, x_n, y_n)$, and the last constraint reflects that $g(x) = \exp(\alpha + \beta x)f(x)$ is a density function with total mass 1. After profiling the $p_j$'s (see, for e.g., Qin (1999)), the log-likelihood can be written as

$$l = \sum_{i=1}^{n}\log[\lambda + (1-\lambda)\exp\{2\alpha + \beta(x_i+y_i)\}] - \sum_{j=1}^{2n}\log[1 + \gamma\{\exp(\alpha + \beta t_j) - 1\}]. \tag{5}$$

Differentiating (5) with respect to $(\lambda, \alpha, \beta)$, we have

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{n}\frac{1 - \exp\{2\alpha + \beta(x_i+y_i)\}}{\lambda + (1-\lambda)\exp\{2\alpha + \beta(x_i+y_i)\}} = 0, \tag{6}$$

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n}\frac{2(1-\lambda)\exp\{2\alpha + \beta(x_i+y_i)\}}{\lambda + (1-\lambda)\exp\{2\alpha + \beta(x_i+y_i)\}} - \sum_{j=1}^{2n}\frac{\gamma\exp(\alpha + \beta t_j)}{1 + \gamma\{\exp(\alpha + \beta t_j) - 1\}} = 0, \tag{7}$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \frac{(x_i+y_i)(1-\lambda)\exp\{2\alpha+\beta(x_i+y_i)\}}{\lambda+(1-\lambda)\exp\{2\alpha+\beta(x_i+y_i)\}} - \sum_{j=1}^{2n} \frac{t_j \gamma \exp(\alpha+\beta t_j)}{1+\gamma\{\exp(\alpha+\beta t_j)-1\}}$$
$$= 0. \tag{8}$$

Multiplying (6) by $(1-\lambda)$, we have

$$\sum_{i=1}^{n} \frac{1}{\lambda+(1-\lambda)\exp(\{2\alpha+\beta(x_i+y_i)\}} = n.$$

Furthermore, using (6) and (7), it can be shown that

$$\gamma = 1-\lambda. \tag{9}$$

*Remark:* Note that since

$$\frac{h(x,y)}{h_1(x)h_2(y)} = \frac{\lambda+(1-\lambda)\exp\{2\alpha+\beta(x+y)\}}{\{\lambda+(1-\lambda)\exp(\alpha+\beta x)\}\{\lambda+(1-\lambda)\exp(\alpha+\beta y)\}},$$

and using $\gamma = 1-\lambda$ in the semi-parametric likelihood (5), we have $l = \sum_{i=1}^{n} \log \{h(x_i,y_i)/h_1(x_i)h_2(y_i)\}$. The quantity $E[\log\{h(x,y)/h_1(x)h_2(y)\}]$ is Shannon's mutual information function (Gourieroux and Monfort (1997, pp.403-404)), and has been used for measuring the partial link between two random variables, conditional on a third variable: it is non-negative, and is zero if and only if $x$ and $y$ are independent. Note that the estimating equation $\sum_{i=1}^{n} \partial \log\{h(x_i,y_i)/h_1(x_i) h_2(y_i)\}/\partial\eta = 0$ is unbiased since $\sum_{i=1}^{n} \partial \log h(x_i,y_i)/\partial\eta$, $\sum_{i=1}^{n} \partial \log h_1(x_i)/\partial\eta$ and $\sum_{i=1}^{n} \partial \log h_2(y_i)/\partial\eta$ are scores based on the joint and marginal distributions, respectively.

Denote the solution of (6) to (8) as $\hat{\eta} = (\hat{\lambda}, \hat{\alpha}, \hat{\beta})$. We have the following results.

**Theorem 1.** *Suppose $F$ is non-degenerate, and $|\partial^3 l/\partial\eta_i\partial\eta_j\partial\eta_k|$, $i,j,k = 1,2,3$, are bounded by some integrable functions, that $E_F\{\exp(3\beta x)\} < \infty$ in a neighborhood of the true value of $\beta_0$, and that $0 < \lambda_0 < 1$ and $\beta_0 \neq 0$. Then, as $n \to \infty$,*

$$\sqrt{n}\begin{pmatrix} \hat{\lambda}-\lambda_0 \\ \hat{\alpha}-\alpha_0 \\ \hat{\beta}-\beta_0 \end{pmatrix} \to N(0,\Sigma) \text{ in distribution,}$$

*where $\Sigma$ is defined at* (A.2) *in the Appendix.*

**Theorem 2.** *Let $R(\lambda) = 2\{\max_{\lambda,\alpha,\beta} l(\lambda,\alpha,\beta) - \max_{\alpha,\beta} l(\lambda,\alpha,\beta)\}$. Under the conditions of Theorem 1, $R(\lambda_0) \to \chi^2(1)$ in distribution.*

## 3. Simulation Results

In this section, we give some simulation results. We first present results that illustrate the non-robustness of the parametric maximum likelihood method. Assuming fully parametrized models for $f(x,y) = f(x,y,\theta_1)$ and $g(x,y) = g(x,y,\theta_2)$, inference can be based on the likelihood $l(\lambda,\theta_1,\theta_2) = \sum_{i=1}^{n} \log\{\lambda f(x_i, y_i,\theta_1) + (1-\lambda) g(x_i,y_i,\theta_2)\}$. We considered two different situations of model misspecification. In the first situation, the data came from a mixture distribution with independent identical log-normal components. Specifically, the data distribution was such that $f(x,y) = f(x)f(y)$ and $g(x,y) = g(x)g(y)$, with $f(x)$ being the density for $x$ where $\log(x)$ was $N(0,1)$ and $g(x)$ being the density for $x$ where $\log(x)$ was $N(\mu,1)$. The data were then (mis)modeled using an exponential mixture likelihood. In the second situation, the data came from a log-transformed mixture distribution with independent exponential components. Specifically, the underlying samples were taken from $f(x) = \exp(-x)$ and $g(x) = 1/\mu \exp(-x/\mu)$ and log-transformed. The data were then modeled (incorrectly) using a normal mixture likelihood.

For the log-normal mixture model, the values of $\mu = 1.5, 2, 2.5$ and $\lambda = 0.2, 0.5, 0.8$ were used, and for the transformed exponential mixture model the values of $\mu = 3, 4, 1/3$ and $\lambda = 0.2, 0.5, 0.8$ were used. In both models, for each combination of the parameters $\mu$ and $\lambda$, we generated 1,000 data sets of sample size 200. We considered only the estimation of $\lambda$. The simulation means and variances of the parameter estimates obtained from maximizing the parametric likelihoods are reported in Table 1. As observed from the results, the parameter estimates are grossly biased from the true parameter values. These results illustrate the risks of using a parametric model.

Table 1. Mean (variance) of the parametric estimates of $\lambda$ based on 1,000 simulations, sample size= 200, using an incorrectly specified model.

| Mixture of log-Normal models | | | |
|---|---|---|---|
| $\lambda$ | $\mu = 1.5$ | $\mu = 2.0$ | $\mu = 2.5$ |
| 0.20 | 0.65412 (0.04428) | 0.33248 (0.05012) | 0.17759 (0.00261) |
| 0.50 | 0.58454 (0.00987) | 0.49602 (0.00373) | 0.47465 (0.00204) |
| 0.80 | 0.78703 (0.00265) | 0.77687 (0.00193) | 0.77427 (0.00121) |

| Transformed mixture of exponentials | | | |
|---|---|---|---|
| $\lambda$ | $\mu = 1/3$ | $\mu = 1/4$ | $\mu = 3$ |
| 0.20 | 0.90062 (0.06963) | 0.78115 (0.08675) | 0.83343 (0.11587) |
| 0.50 | 0.87980 (0.06568) | 0.80482 (0.04128) | 0.80186 (0.12021) |
| 0.80 | 0.98741 (0.00794) | 0.94493 (0.03833) | 0.85757 (0.10114) |

We now give results for our semi-parametric method. The following algorithm was used to maximize the semi-parametric likelihood. From (9), the Lagrange multiplier $\gamma$ is equal to $1-\lambda$. For fixed $(\alpha,\beta)$, we solved (6) to obtain $\lambda=\lambda(\alpha,\beta)$ by imposing the constraints that $\lambda=0$ if $\partial l(\lambda,\alpha,\beta)/\partial\lambda|_{\lambda=0}\leq 0$ and $\lambda=1$ if $\partial l(\lambda,\alpha,\beta)/\partial\lambda|_{\lambda=1}\geq 0$. Then we employed the downhill simplex algorithm (Press et al. (1992)) to search for $(\alpha,\beta)$ to maximize the semi-parametric likelihood. In the simulations, when the two components were close to each other it was hard to estimate the underlying parameters. A larger sample size was needed when the mixture parameter $\lambda$ was close to 0 or 1 and $\beta$ was close to 0.

For comparison, we also calculated the maximum likelihood estimate of $\lambda$ using the parametric likelihoods. In this comparison, we assumed the models for $f(x,y)$ and $g(x,y)$ were correctly specified.

Similar to a robustness study, two situations were considered. The first was a mixture model with independent identical normal components. Specifically, the underlying densities $f(x)$ and $g(y)$ were $N(0,1)$ and $N(\mu,1)$, respectively. Therefore $g(x)=\exp(\alpha+\beta x)f(x)$, with, $\alpha=-\mu^2/2$; and $\beta=\mu$. The values of $\mu=1.5,2,2.5$ and $\lambda=0.2,0.5,0.8$ were used. We generated 1,000 data sets of sample size 200 for each combination of $\mu$ and $\lambda$. Means and variances of the parameter estimates based on the 1,000 simulations are reported at the top of Table 2. The first and second entries are the mean (variance) of the parametric maximum likelihood estimate and the semi-parametric maximum likelihood estimate, respectively.

From Table 2, we observe that both parametric and semi-parametric likelihood estimations are satisfactory, though there is some bias when $\lambda$ is close to the boundaries. Naturally there is some loss of efficiency in using the semi-parametric likelihood estimation under a correctly specified model.

The second model we used was an exponential mixture model with $f(x)=\exp(-x)$, $g(x)=(1/\mu)\exp(-x/\mu)$. Hence $g(x)=\exp(\alpha+\beta x)f(x)$, $\alpha=-\log\mu$, $\beta=1-1/\mu$. For $\mu=3,4,1/3$ and $\lambda=0.2,0.5,0.8$, the simulation results are reported at the bottom of Table 2. Once again, both models give satisfactory parameter estimates, with the parametric model slightly more efficient.

Our simulation results are similar to those observed by Efron (1975) in comparing the relative efficiency of a logistic regression model to a full parametric model. The similarity in the results between the two studies is not surprising since model (2) is essentially equivalent to a logistic regression model (Qin (1998)). The proposed method offers a few advantages: the semi-parametric nature of the model makes it robust against departures from model assumptions; it is simple to implement; the model allows identification of all model parameters without the requirement of a training sample.

Table 2 Mean (variance) of the parametric and semi-parametric estimates of $\lambda$ based on 1,000 simulations, sample size=200.

| $F \sim N(0,1)$ and $G \sim N(\mu,1)$ | | | |
|---|---|---|---|
| $\lambda$ | $\mu = 1.5$ | $\mu = 2.0$ | $\mu = 2.5$ |
| 0.20 | 0.21099 (0.00512)[1] | 0.20216 (0.00157) | 0.20158 (0.00102) |
| 0.20 | 0.21386 (0.00767)[2] | 0.20165 (0.00198) | 0.20188 (0.00114) |
| 0.50 | 0.50072 (0.00528) | 0.50083 (0.00219) | 0.49842 (0.00152) |
| 0.50 | 0.49815 (0.00798) | 0.49928 (0.00285) | 0.49860 (0.00161) |
| 0.80 | 0.78897 (0.00504) | 0.79698 (0.00165) | 0.79833 (0.00103) |
| 0.80 | 0.78401 (0.00812) | 0.79551 (0.00206) | 0.79901 (0.00111) |

| $F \sim \exp(1)$ and $G \sim \exp(\mu)$ | | | |
|---|---|---|---|
| $\lambda$ | $\mu = 1/3$ | $\mu = 1/4$ | $\mu = 3$ |
| 0.20 | 0.25392 (0.03356) | 0.22991 (0.01279) | 0.23532 (0.01606) |
| 0.20 | 0.27510 (0.05946) | 0.22999 (0.02627) | 0.23424 (0.03453) |
| 0.50 | 0.49950 (0.01618) | 0.50026 (0.00655) | 0.50787 (0.01687) |
| 0.50 | 0.50833 (0.03645) | 0.50002 (0.01378) | 0.49267 (0.03870) |
| 0.80 | 0.76769 (0.01348) | 0.79154 (0.00411) | 0.75170 (0.03202) |
| 0.80 | 0.77283 (0.03430) | 0.78301 (0.01271) | 0.73804 (0.05320) |

1. parametric estimate.
2. semi-parametric estimate.

## Appendix. Proofs of Theorem 1 and Theorem 2

Proof of Theorem 1. Let $\phi = (\phi_1, \phi_2, \phi_3)^T$, where

$$\phi_1(\eta) = \frac{1}{n}\sum_{i=1}^{n} \frac{1-\exp[2\alpha+\beta(x_i+y_i)]}{\lambda+(1-\lambda)\exp[2\alpha+\beta(x_i+y_i)]},$$

$$\phi_2(\eta) = \frac{1}{n}\sum_{i=1}^{n} \frac{(x_i+y_i)\exp[2\alpha+\beta(x_i+y_i)]}{\lambda+(1-\lambda)\exp[2\alpha+\beta(x_i+y_i)]} - \frac{1}{n}\sum_{j=1}^{2n} \frac{t_j\exp(\alpha+\beta t_j)}{\lambda+(1-\lambda)\exp(\alpha+\beta t_j)},$$

$$\phi_3(\eta) = \frac{1}{n}\sum_{j=1}^{2n} \frac{\exp(\alpha+\beta t_j)-1}{\lambda+(1-\lambda)\exp(\alpha+\beta t_j)}.$$

It is easy to show that in probability

$$\frac{\partial\phi(\eta_0)}{\partial\eta} \to A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \text{ where}$$

$$a_{11} = -\int \frac{[1-\exp\{2\alpha_0+\beta_0(x+y)\}]^2}{\lambda_0+(1-\lambda_0)\exp\{2\alpha_0+\beta_0(x+y)\}} dF(x)dF(y),$$

$$a_{12} = 2\int \frac{\exp\{2\alpha_0+\beta_0(x+y)\}}{\lambda_0+(1-\lambda_0)\exp\{2\alpha_0+\beta_0(x+y)\}} dF(x)dF(y)$$
$$+4(1-\lambda_0)\int \frac{[1-\exp\{2\alpha_0+\beta_0(x+y)\}]\exp\{2\alpha_0+\beta_0(x+y)\}}{\lambda_0+(1-\lambda_0)\exp\{2\alpha_0+\beta_0(x+y)\}} dF(x)dF(y),$$

$$a_{13} = 2(1-\lambda_0)\int \frac{(x+y)[1-\exp\{2\alpha_0+\beta_0(x+y)\}]\exp\{2\alpha_0+\beta_0(x+y)\}}{\lambda_0+(1-\lambda_0)\exp\{2\alpha_0+\beta_0(x+y)\}} dF(x)dF(y)$$
$$-\int \frac{(x+y)\exp\{2\alpha_0+\beta_0(x+y)\}}{\lambda_0+(1-\lambda_0)\exp\{2\alpha_0+\beta_0(x+y)\}} dF(x)dF(y),$$

$$a_{21} = -\int \frac{(x+y)\exp\{2\alpha_0+\beta_0(x+y)\}[1-\exp\{2\alpha_0+\beta_0(x+y)\}]}{\lambda_0+(1-\lambda_0)\exp\{2\alpha_0+\beta_0(x+y)\}} dF(x)dF(y)$$
$$+2\int \frac{x\exp(\alpha_0+\beta_0 x)\{1-\exp(\alpha_0+\beta_0 x)\}}{\lambda_0+(1-\lambda_0)\exp(\alpha_0+\beta_0 x)} dF(x)dF(y),$$

$$a_{22} = 2\lambda_0\int \frac{(x+y)\exp\{2\alpha_0+\beta_0(x+y)\}}{\lambda_0+(1-\lambda_0)\exp(2\alpha_0+\beta_0(x+y))} dF(x)dF(y)$$
$$-2\lambda_0\int \frac{x\exp(2\alpha_0+2\beta_0 x)}{\lambda_0+(1-\lambda_0)\exp(\alpha_0+\beta_0 x)} dF(x),$$

$$a_{23} = \lambda_0\int \frac{(x+y)^2\exp\{2\alpha_0+\beta_0(x+y)\}}{\lambda_0+(1-\lambda_0)\exp\{2\alpha_0+\beta_0(x+y)\}} dF(x)dF(y)$$
$$-2\lambda_0\int \frac{x^2\exp(2\alpha_0+\beta_0 x)}{\lambda_0+(1-\lambda_0)\exp(\alpha_0+\beta_0 x)} dF(x),$$

$$a_{31} = 2\int \frac{\{\exp(\alpha_0+\beta_0 x)-1\}^2}{\lambda_0+(1-\lambda_0)\exp(\alpha_0+\beta_0 x)} dF(x),$$

$$a_{32} = \int \frac{\exp(\alpha_0+\beta_0 x)}{\lambda_0+(1-\lambda_0)\exp(\alpha_0+\beta_0 x)} dF(x),$$

$$a_{33} = -(1-\lambda_0)\int \frac{x\exp(\alpha_0+\beta_0 x)}{\lambda_0+(1-\lambda_0)\exp(\alpha_0+\beta_0 x)} dF(x).$$

By expanding $\phi(\hat\eta)$ at $\eta_0$, we have

$$\begin{aligned}\sqrt{n}(\hat\eta-\eta_0) &= -(\partial\phi(\eta_0)/\partial\eta)^{-1}\sqrt{n}\phi(\eta_0)+o_p(1) = \sqrt{n}(\hat\eta-\eta_0)\\ &= -A^{-1}\sqrt{n}\phi(\eta_0)+o_p(1).\end{aligned} \tag{A.1}$$

Let $\psi = (\psi_1,\psi_2,\psi_3)^T$. Then $\sqrt{n}\phi(\eta_0) = (1/\sqrt{n})\sum_{i=1}^n \psi(x_i,y_i,\eta_0)$, where

$$\psi_1(x_i,y_i,\eta_0) = \frac{1-\exp\{2\alpha_0+\beta_0(x_i+y_i)\}}{\lambda_0+(1-\lambda_0)\exp\{2\alpha_0+\beta_0(x_i+y_i)\}},$$

$$\psi_2(x_i, y_i, \eta_0) = \frac{(x_i + y_i)\exp[2\alpha_0 + \beta_0(x_i + y_i)]}{\lambda_0 + (1-\lambda_0)\exp[2\alpha_0 + \beta_0(x_i + y_i)]} - \frac{x_i \exp(\alpha_0 + \beta_0 x_i)}{\lambda_0 + (1-\lambda_0)\exp(\alpha_0 + \beta_0 x_i)}$$
$$- \frac{y_i \exp(\alpha_0 + \beta_0 y_i)}{\lambda_0 + (1-\lambda_0)\exp(\alpha_0 + \beta_0 y_i)},$$
$$\psi_3(x_i, y_i, \eta_0) = \frac{\exp(\alpha_0 + \beta_0 x_i) - 1}{\lambda_0 + (1-\lambda_0)\exp(\alpha_0 + \beta_0 x_i)} + \frac{\exp(\alpha_0 + \beta_0 y_i) - 1}{\lambda_0 + (1-\lambda_0) + \exp(\alpha_0 + \beta_0 y_i)}.$$

Hence, in distribution, $\sqrt{n}\phi(\eta_0) \to N(0, V)$, where $V = \mathrm{Cov}\,(\psi, \psi)$. Therefore, by using Slutsky's theorem on (A.1),

$$\sqrt{n}(\hat{\eta} - \eta_0) \to N(0, \Sigma), \qquad \Sigma = A^{-1} V A^{-1}. \tag{A.2}$$

Proof of Theorem 2. The proof is tedious, but similar to the proof of Theorem 2 in Qin (1999).

## References


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

Cameron, S. and Heckman, J. (2001). The dynamics of educational attainment for Black, Hispanic, and White males. *J. Polit. Economy* **109**, 455-499.

Desarbo, W. S., Degrenatu, A. M., Wedel, M. and Saxton, M. K. (2001). The spatial representation of market information. *Marketing Sci.* **20**, 426-441.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *J. Amer. Statist. Assoc.* **70**, 892-898.

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

Gill, R. D., Vardi, Y. and Wellner, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16**, 1069-1112.

Gourieroux, C. and Monfort, A. (1997). *Time Series and Dynamic Models.* Cambridge University Press, Cambridge.

Hall, P. and Zhou, X. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* **31**, 201-224.

Hettmansperger, T. P. and Thomas, H. (2000). Almost nonparametric inference for repeated measure in mixture models. *J. Roy. Statist. Soc.* B **62**, 811-825.

Hull, J. and White, A. (1998). Value-at-risk when daily changes in market variables are not normally distributed. *J. Derivatives* **5**, 9-19.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *J. Polit. Economy* **105**, 473-522.

Lindsay, B. (1995). *Mixture Models: Theory, Geometry and Applications.* Institute of Mathematical Statistics, Haywood, California.

Liu, Q., and Pierce, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika* **81**, 624-629.

MacDonald, P. D. M. (1975). Estimation of finite distribution mixtures. In *Applied Statistics* (Edited by R. P. Gupta), 231-245. North-Holland, Amsterdam.

Moe, W. W. and Fader, P. S. (2002). Using advance purchase orders to forecast new product sales. *Marketing Sci.* **21**, 347-364.

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

Pearson, K. (1893). Contributions to the mathematical theory of evolution. *J. Roy. Statist. Assoc.* **56**, 675-679.

Pearson, K. (1895). Contributions to the mathematical theory of evolution II: Skew variation in homogeneous material. *Philos. Trans. A* **186**, 342-414.

Press, W. H., Teukolsky, S. A., Vettering, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C, The Art of Scientific Computing.* Second edition. Cambridge University Press, Cambridge.

Qin, J. (1998). Inferences for case-control and semi-parametric two-sample density ratio models. *Biometrika* **85**, 619-630.

Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *Ann. Statist.* **27**, 1368-1384.

Qu, Y. S. and Hadgu A. (1998). A model for evaluating sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference test. *J. Amer. Statist. Assoc.* **93**, 920-928.

Teicher, H. (1967). Identifiability of mixtures of product measures. *Ann. Math. Statist.* **38**, 1300-1302.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York.

Thomas, H. and Horton, J. J. (1997). Competency criteria and the class inclusion task: Modeling judgments and justifications, *Develop. Psych.* **33**, 1060-1073.

Thomas, H. and Lohaus, A. (1993). Modeling growth and individual differences in spatial tasks. *Monograph of the Society of Research on Child Development*, Number 237.

Vardi, Y. (1985). Empirical distribution in selection bias models. *Ann. Statist.* **13**, 178-203.

Venkatraman , S. (1997). Value at risk for a mixture of normal distributions: The use of quasi-Bayesian estimation techniques. *Economics Perspective*, March/April, 2-13. Federal Reserve Bank of Chicago.

Vounatsou, P., Smith, T. and Smith, A. F. M. (1998). Bayesian analysis of two-component mixture distributions applied to estimating malaria attributable fractions. *Appl. Statist.* **47**, 575-587.

Zangari, P. (1996). An improved methodology for measuring VaR. *Risk Metrics Monitor*, 7-25. Reuters, J. P. Morgan.

Zou, F., Fine, J. P. and Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika* **89**, 61-75.

School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903.

E-mail: denisleung@smu.edu.sg

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, NIH, 6700B Rockledge Drive MSC 7609, Bethesda, MD 20892, U.S.A.

E-mail: jingqin@niaid.nih.gov

(Received June 2004; accepted October 2004)