5-1-2005

Testing the Goodness of Fit of Multivariate Multiplicative-intercept Risk Models Based on Case-control Data

Biao Zhang
University of Toledo

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Regular Articles
Testing the Goodness of Fit of Multivariate Multiplicative-intercept Risk Models Based on Case-control Data

Biao Zhang
Department of Mathematics
The University of Toledo

The validity of the multivariate multiplicative-intercept risk model with $I+1$ categories based on case-control data is tested. After reparametrization, the assumed risk model is equivalent to an $(I+1)$-sample semiparametric model in which the I ratios of two unspecified density functions have known parametric forms. By identifying this $(I+1)$-sample semiparametric model, which is of intrinsic interest in general $(I+1)$-sample problems, with an $(I+1)$-sample semiparametric selection bias model, we propose a weighted Kolmogorov-Smirnov-type statistic to test the validity of the multivariate multiplicative-intercept risk model. Established are some asymptotic results associated with the proposed test statistic, also established is an optimal property for the maximum semiparametric likelihood estimator of the parameters in the $(I+1)$-sample semiparametric selection bias model. In addition, a bootstrap procedure along with some results on analysis of two real data sets is proposed.

Key words: Biased sampling problem, bootstrap, Kolmogorov-Smirnov two-sample statistic, logistic regression, mixture sampling, multivariate Gaussian process, semiparametric selection bias model, strong consistency, weak convergence

Introduction

Let Y be a multicategory response variable with $I+1$ categories and X be the associated $p \times 1$ covariate vector. When the possible values of the response variable Y are denoted by $y = 0,1,\ldots,I$ and the first category (0) is the baseline category, Hsieh, Manski, and McFadden (1985) introduced the following multivariate multiplicative-intercept risk model:

$$\frac{P(Y=i \mid X=x)}{P(Y=0 \mid X=x)} = \theta_i^* r_i(x;\beta_i), \quad i=1,\ldots,I, \quad (1)$$

Biao Zhang is a Professor in the Department of Mathematics at the University of Toledo. His research interests include categorical data analysis and empirical likelihood. The author wishes to thank Xin Deng and Shuwen Wan for their help in the manuscript conversion process. Email: bzhang@utnet.utoledo.edu.

where $\theta_1^*,\ldots,\theta_I^*$ are positive scale parameters, $r_1,\ldots,r_I$ are, for fixed x, known functions from R^p to R^+, and $\beta_i = (\beta_i^1,\ldots,\beta_i^p)^\tau$ is a $p \times 1$ vector parameter for $i = 1,\ldots I$. The class of multivariate multiplicative-intercept risk models includes the multivariate logistic regression models and the multivariate odds-linear models discussed by Weinberg and Sandler (1991) and Wacholder and Weinberg (1994). By generalizing earlier works of Anderson (1972, 1979), Farewell (1979), and Prentice and Pyke (1979) in the context of the logistic regression models, Weinberg and Wacholder (1993) and Scott and Wild (1997) showed that under model (1.1), a prospectively derived analysis, including parameter estimates and standard errors for $\beta_1,\ldots,\beta_I$, is asymptotically correct in case-control studies. In this article, testing the validity of model (1) based on case-control data as specified below is considered.

Let $X_{i1}, \ldots, X_{in_i}$ be a random sample from $P(x \mid Y = i)$ for $i = 0, 1, \ldots, I$ and assume that $\{(X_{i1}, \ldots, X_{in_i}) : i = 0, 1, \ldots I\}$ are jointly independent. Let $\pi_i = P(Y = i)$ and $g_i(x) = f(x \mid Y = i)$ be the conditional density or frequency function of X given $Y = i$ for $i = 0, 1, \ldots I$. If $f(x)$ is the marginal distribution of X, then applying Bayes' rule yields

$$f(x \mid Y = i) = \frac{P(Y = i \mid X = x)}{\pi_i} f(x),$$

$$i = 0, 1, \ldots I.$$

It is seen that

$$\frac{f(x \mid Y = i)}{f(x \mid Y = 0)} = \frac{\pi_0}{\pi_i} \frac{P(Y = i \mid X = x)}{P(Y = 0 \mid X = x)}$$

$$= \frac{\pi_0}{\pi_i} \theta_i^* r_i(x; \beta_i), \quad i = 1, \ldots, I.$$

Consequently,

$$g_i(x) = f(x \mid Y = i) = \frac{\pi_0}{\pi_i} \theta_i^* r_i(x; \beta_i) f(x \mid Y = 0)$$

$$= \exp[\theta_i + s_i(x; \beta_i)] g_0(x), \quad i = 1, \ldots, I,$$

where $\theta_i = \log \theta_i^* + \log(\pi_0 / \pi_i)$ and

$s_i(x; \beta_i) = \log r_i(x; \beta_i)$ for $i = 1, \ldots, I$. As a result, the following $(I+1)$-sample semiparametric model is obtained:

$$X_{01}, \ldots, X_{0n_0} \overset{i.i.d.}{\sim} g_0(x),$$

$$X_{i1}, \ldots, X_{in_i} \overset{i.i.d.}{\sim} g_i(x) = \exp[\theta_i + s_i(x; \beta_i)] g_0(x),$$

$$i = 1, \ldots, I. \qquad (2)$$

Throughout this article, let $\theta = (\theta_1, \ldots, \theta_I)^\tau$, $\beta = (\beta_1^\tau, \ldots, \beta_I^\tau)^\tau$, and $G_i(x)$ be the corresponding cumulative

distribution function of $g_i(x)$ for $i = 0, 1, \ldots I$. Note that model (2) is equivalent to an $(I+1)$-sample semiparametric model in which the ith $(i = 1, \ldots, I)$ ratio of a pair of unspecified density functions g_i and g_0 has a known parametric form, and thus is of intrinsic interest in general $(I+1)$-sample problems. Model (2) is equivalent to model (1); it is an $(I+1)$-sample semiparametric selection bias model with weight functions $w_0(x, \theta, \beta) = 1$ and

$$w_i(x, \theta, \beta) = \exp[\theta_i + s(x; \beta_i)]$$

for $i = 1, \ldots, I$ depending on the unknown parameters θ and β. The s-sample semiparametric selection bias model was proposed by Vardi (1985) and was further developed by Gilbert, Lele, and Vardi (1999). Vardi (1982, 1985), Gill, Vardi, and Wellner (1988), and Qin (1993) discussed estimating distribution functions in biased sampling models with known weight functions. Weinberg and Wacholder (1990) considered more flexible design and analysis of case-control studies with biased sampling. Qin and Zhang (1997) and Zhang (2002) considered goodness-of-fit tests for logistic regression models based on case-control data, whereas Zhang (2000) considered testing the validity of model (2) when $I = 1$.

The focus in this article is to test the validity of model (1.2) for $I \geq 1$. Let $\{T_1, \ldots, T_n\}$ denote the pooled sample $\{X_{01}, \ldots, X_{0n_0}; X_{11}, \ldots, X_{1n_1}; \ldots; X_{I1}, \ldots, X_{In_I}\}$ with $n = \sum_{i=0}^{I} n_i$. Furthermore, let

$$\widehat{G}_i(t) = n_i^{-1} \sum_{j=1}^{n_i} I_{[X_{ij} \leq t]}$$

and

$$\overline{G}_0(t) = n^{-1} \sum_{k=1}^{n} I_{[T_k \leq t]}$$

be, respectively, the empirical distribution functions based on the sample $X_{i1}, \ldots, X_{in_i}$ from the ith $(i = 0, 1, \ldots, I)$ category and the pooled sample $T_1, \ldots, T_n$. In the special case of testing

the equality of G_0 and G_1 for which $I = 1$ and $s_1(x; \beta_1) \equiv 0$ in model (2), as argued by (van der Vaart & Wellner, 1996, p. 361; Qin & Zhang, 1997), the Kolmogorov-Smirnov two-sample statistic is equivalent to a statistic based on the discrepancy between the empirical distribution function $\widehat{G_0}$ and the pooled empirical distribution function $\overline{G}_0$. This fact, along with the fact that $\widehat{G_0}$ and $\overline{G}_0$ are, respectively, the nonparametric maximum likelihood estimators of G_0 without and with the assumption of $G_0(t) = G_1(t)$, motivates us to employ a weighted average of the $I + 1$ discrepancies between $\widehat{G_i}$ and $\widetilde{G}_i$ $(i = 0, 1, \ldots, I)$ to assess the validity of model (2), where $\widetilde{G}_i$ is the maximum semiparametric likelihood estimator of G_i under model (2) and is derived by employing the empirical likelihood method developed by Owen (1988, 1990). For a more complete survey of developments in empirical likelihood, see Hall and La Scala (1990) and Owen (1991).

This article is structured as follows: in the method section proposed is a test statistic by deriving the maximum semiparametric likelihood estimator of G_i under model (2). Some asymptotic results are then presented along with an optimal property for the maximum semiparametric likelihood estimator of (θ, β). This is followed by a bootstrap procedure which allows one to find P-values of the proposed test. Also reported are some results on analysis of two real data problems. Finally, proofs of the main theoretical results are offered.

Methodology

Based on the observed data in (2), the likelihood function can be written as

$$L(\theta, \beta, G_0) = \prod_{i=0}^{I} \prod_{j=1}^{n_i} \exp[\theta_i + s_i(X_{ij}; \beta_i)] dG_0(X_{ij})$$

$$= \left(\prod_{k=1}^{n} p_k \right) \left[\exp\left(\sum_{i=1}^{I} \sum_{j=1}^{n_i} [\theta_i + s_i(X_{ij}; \beta_i)] \right) \right],$$

where $\theta_0 = 0$, $s_0(\bullet; \beta_0) \equiv 0$, and $p_k = dG_0(T_k)$, $k = 1, \ldots, n$, are (nonnegative) jumps with total mass unity. Similar to the approach of Owen (1988, 1990) and Qin and Lawless (1994), it can be shown by using the method of Lagrange multipliers that for fixed (θ, β), the maximum value of L, subject to constraints $\sum_{k=1}^{n} p_k = 1$, $p_k \geq 0$ and

$$\sum_{k=1}^{n} p_k \{ \exp[\theta_i + s_i(T_k; \beta_i)] - 1 \} = 0$$

for $i = 1, \ldots, I$, is attained at

$$p_k = \frac{1}{n_0} \frac{1}{1 + \sum_{i=1}^{I} \rho_i \exp[\theta_i + s_i(T_k; \beta_i)]},$$
$$k = 1, \ldots, n,$$

where $\rho_i = n_i / n_0$ for $i = 0, 1, \ldots, I$. Therefore, the (profile) semiparametric log-likelihood function of (θ, β) is given by

$$\ell(\theta, \beta) = -n \log n_0$$
$$- \sum_{k=1}^{n} \log\left[1 + \sum_{i=1}^{I} \rho_i \exp[\theta_i + s_i(T_k; \beta_i)] \right]$$
$$+ \sum_{i=1}^{I} \sum_{j=1}^{n_i} [\theta_i + s_i(X_{ij}; \beta_i)].$$

Next, maximize ℓ over (θ, β). Let $(\tilde{\theta}, \tilde{\beta})$ with $\tilde{\theta} = (\tilde{\theta}_1, \ldots \tilde{\theta}_I)^\tau$ and $\tilde{\beta} = (\tilde{\beta}_1^\tau, \ldots \tilde{\beta}_I^\tau)^\tau$ be the solution to the following system of score equations:

$$\frac{\partial \ell(\theta,\beta)}{\partial \theta_u} = n_u - \sum_{k=1}^{n} \frac{\rho_u \exp[\theta_u + s_u(T_k;\beta_u)]}{1+\sum_{i=1}^{I}\rho_i \exp[\theta_i + s_i(T_k;\beta_i)]}$$

$$= 0, \quad u = 1,\dots,I,$$

$$\frac{\partial \ell(\theta,\beta)}{\partial \beta_u} = \sum_{j=1}^{n_u} d_u(X_{uj};\beta_u)$$

$$-\sum_{k=1}^{n} \frac{\rho_u \exp[\theta_u + s_u(T_k;\beta_u)]}{1+\sum_{m=1}^{I}\rho_m \exp[\theta_m + s_m(T_k;\beta_m)]} d_u(T_k;\beta_u)$$

$$= 0, \quad u = 1,\dots,I, \tag{3}$$

where $d_u(T_k;\beta_u) = \dfrac{\partial s_u(\theta,\beta_u)}{\partial \beta_u}$ for $u = 1,\dots,I$.
That produces the following,

$$\tilde{p}_k = \frac{1}{n_0}\frac{1}{1+\sum_{i=1}^{I}\rho_i \exp[\widetilde{\theta}_i + s_i(T_k;\widetilde{\beta}_i)]},$$

$$k = 1,\dots,n. \tag{4}$$

On the basis of the $\tilde{p}_k$ in (4), it can be proposed to estimate $G_i(t)$, under model (2), by

$$\widetilde{G}_i(t) = \sum_{k=1}^{n} \tilde{p}_k \exp[\widetilde{\theta}_i + s_i(T_k;\widetilde{\beta}_i)]I_{[T_k \le t]}$$

$$= \frac{1}{n_0}\sum_{k=1}^{n} \frac{\exp[\widetilde{\theta}_i + s_i(T_k;\widetilde{\beta}_i)]}{1+\sum_{m=1}^{I}\rho_m \exp[\widetilde{\theta}_m + s_m(T_k;\widetilde{\beta}_m)]}I_{[T_k \le t]},$$

$$i = 0,\dots,I, \tag{5}$$

where $\widetilde{\theta}_0 = 0$ and $s_0(\bullet;\widetilde{\beta}_0) \equiv 0$. Throughout this article, $a \le b$ and $-\infty \le a \le \infty$ with $a = (a_1,\dots,a_p)^\tau$ and $b = (b_1,\dots,b_p)^\tau$ stand for, respectively, $a_i \le b_i$ and $-\infty \le a_i \le \infty$ for $i = 0,1,\dots,p$. Note that $\widetilde{G}_i$ is the maximum semiparametric likelihood estimator of G_i under model (2) for $i = 0,1,\dots,I$. Let $\widehat{G}_i(t) = n_i^{-1}\sum_{j=1}^{n_i} I_{[X_{ij}\le t]}$ be the empirical distribution function based on the sample $X_{i1},\dots,X_{in_i}$ from the i th $(i = 0,1,\dots,I)$ category. Moreover, let

$$\Delta_{ni}(t) = \sqrt{n}\left(\widetilde{G}_i(t) - \widehat{G}_i(t)\right), \quad \Delta_{ni} = \sup_{-\infty \le t \le \infty}\left|\Delta_{ni}(t)\right|,$$

$$i = 0,1,\dots,I.$$

Then, Δ_{ni} is the discrepancy between the two estimators $\widetilde{G}_i(t)$ and $\widehat{G}_i(t)$, and thus measures the departure from the assumption of the multivariate multiplicative-intercept risk model (1) within the i th $(i = 1,\dots,I)$ pair of category i and the baseline category (0). Since $\sum_{i=0}^{I}\rho_i\Delta_{ni}(t) = \sqrt{n}\sum_{i=0}^{I}\rho_i[\widetilde{G}_i(t) - \widehat{G}_i(t)] = 0$, there exists a motivation to employ the weighted average of the Δ_{ni} defined by

$$\Delta_n = \frac{1}{I+1}\sum_{i=0}^{I}\rho_i\Delta_{ni} \tag{6}$$

to assess the validity of model (2). Clearly, the proposed test statistic Δ_n measures the global departure from the assumption of the multivariate multiplicative-intercept risk model (1). Because the same value of Δ_n occurs no matter which category is the baseline category, there is a symmetry among the $I+1$ category designations for such a global test. Thus, the choice of the baseline category in model (1) is arbitrary for testing the validity of model (1) or model (2) based on Δ_n. Note that the test statistic Δ_n reduces to that of Zhang (2000) when $I = 1$ in model (1) since $\Delta_n = 2^{-1}(\Delta_{n0} + \rho_1\Delta_{n1}) = \Delta_{n0}$ for $I = 1$.

Remark 1: The test statistic Δ_n can also be applied to mixture sampling data in which a sample of $n = \sum_{i=0}^{I}n_i$ members is randomly selected from the whole population with $n_0,n_1,\dots,n_I$ being random (Day & Kerridge, 1967). Let (X_k,Y_k), $k = 1,\dots,n$, be a random sample from the joint distribution of (X,Y), then the likelihood has the form of

$$L = \prod_{k=1}^{n} P(Y_k \mid X_k) f(X_k)$$

$$= \prod_{i=0}^{I} \prod_{j=1}^{n_i} [\pi_i f(X_{ij} \mid Y = i)],$$

where $\pi_i = P(Y = i)$ for $i = 1, \ldots, I$. The first expression is a prospective decomposition and the second one is a retrospective decomposition.

Remark 2: In light of Anderson (1972, 1979), the case-control data may be treated as the prospective data to compute the maximum likelihood estimate of (θ^*, β) under model (1), where $\theta^* = (\theta_1^*, \ldots, \theta_I^*)^\tau$. Suppose that the sample data in model (2) are collected prospectively, then the (prospective) likelihood function is, by (1),

$$L(\theta^*, \beta) = \prod_{i=0}^{I} \prod_{j=1}^{n_i} P(Y = i \mid X = X_{ij}) = \prod_{i=0}^{I} \prod_{j=1}^{n_i} \left[\frac{\theta_i^* r_i(X_{ij}; \beta_i)}{1 + \sum_{m=1}^{I} \theta_m^* r_m(X_{ij}; \beta_m)} \right].$$

The log-likelihood function is

$$\ell(\theta^*, \beta) = \sum_{i=0}^{I} \sum_{j=1}^{n_j} [\log \theta_i^* + s_i(X_{ij}; \beta_i)]$$

$$- \sum_{k=1}^{n} \log \left[1 + \sum_{m=1}^{I} \theta_m^* \exp[s_m(T_k; \beta_m)] \right].$$

The system of score equations is given by

$$\frac{\partial \ell(\theta^*, \beta)}{\partial \theta_u^*}$$

$$= \frac{1}{\theta_u^*} \left[n_u - \sum_{k=1}^{n} \frac{\theta_u^* \exp[s_u(T_k; \beta_u)]}{1 + \sum_{m=1}^{I} \theta_m^* \exp[s_m(T_k; \beta_m)]} \right]$$

$$= 0, \quad u = 1, \ldots, I,$$

$$\frac{\partial \ell(\theta^*, \beta)}{\partial \beta_u} = \sum_{j=1}^{n_u} d_u(X_{uj}; \beta_u)$$

$$- \sum_{k=1}^{n} \frac{\theta_u^* \exp[s_u(T_k; \beta_u)]}{1 + \sum_{m=1}^{I} \theta_m^* \exp[s_m(T_k; \beta_m)]} d_u(T_k; \beta_u) = 0,$$

$$u = 1, \ldots, I. \tag{7}$$

Let $(\widehat{\theta^*}, \widehat{\beta})$ with $\widehat{\theta^*} = (\widehat{\theta_1^*}, \ldots, \widehat{\theta_I^*})^\tau$ and $\widehat{\beta} = (\widehat{\beta_1}^\tau, \ldots, \widehat{\beta_I}^\tau)^\tau$ denote the solution to the system of score equations in (7). Then comparing (7) with (3) implies that $\widetilde{\theta_u} = \log \widehat{\theta_u^*} + \log(n_0 / n_u)$ and $\widetilde{\beta_u} = \widehat{\beta_u}$ for $u = 1, \ldots, I$. Thus, the maximum likelihood estimates of are identical under the retrospective sampling scheme and the prospective sampling scheme. In addition, the two estimated asymptotic variance-covariance matrices for $\widehat{\beta}$ and $\widetilde{\beta}$ based on the observed information matrices coincide. See also Remarks 3 and 4 below.

Asymptotic results

In this section, the asymptotic properties of the proposed estimator $\widetilde{G}_i(t)$ $(i = 0, 1, \ldots, I)$ in (5) and the proposed test statistic Δ_n in (6) are studied. To this end, let $(\theta_{(0)}, \beta_{(0)})$ be the true value of (θ, β) under model (2) with

$$\theta_{(0)} = (\theta_{10}, \ldots, \theta_{p0})^\tau$$

and

$$\beta_{(0)} = (\beta_{10}^\tau, \ldots, \beta_{p0}^\tau)^\tau.$$

Throughout this article, it is assumed that $\rho_i = n_i / n_0$ $(i = 0, 1, \ldots, I)$ is positive and finite and remains fixed as $n = \sum_{i=0}^{I} n_i \to \infty$. Write $\rho = \sum_{i=0}^{I} \rho_i$ and

$$d_i(t; \beta_i) = \frac{\partial s_i(t; \beta_i)}{\partial \beta_i},$$

$$D_i(t; \beta_i) = \frac{\partial d_i(t; \beta_i)}{\partial \beta_i^\tau} = \frac{\partial^2 s_i(t; \beta_i)}{\partial \beta_i \partial \beta_i^\tau}$$

$$i = 1, \ldots, I,$$

$$s_{11}^{uv} = -\frac{1}{1+\rho} \times$$

$$\int \frac{\rho_u \exp\left[\theta_{u0}+s_u(y;\beta_{uo})\right]\rho_v \exp\left[\theta_{v0}+s_v(y;\beta_{vo})\right]}{1+\sum_{i=1}^{I}\rho_i \exp[\theta_{i0}+s_i(y;\beta_{i0})]} dG_0(y),$$

$$u \neq v = 0,1,\ldots,I,$$

$$s_{11}^{uu} = -\sum_{v=0,v\neq u}^{I} s_{11}^{uv}, \quad u=0,1,\ldots,I,$$

$$S_{11} = (s_{11}^{uv})_{u,v=1,\ldots,I},$$

$$s_{21}^{uv} = -\frac{1}{1+\rho} \times$$

$$\int \frac{\rho_u \exp\left[\theta_{u0}+s_u(y;\beta_{uo})\right]\rho_v \exp\left[\theta_{v0}+s_v(y;\beta_{vo})\right]}{1+\sum_{i=1}^{I}\rho_i \exp[\theta_{i0}+s_i(y;\beta_{i0})]} d_u(y,\beta_u) dG_0(y),$$

$$u \neq v = 0,1,\ldots,I$$

$$s_{21}^{uu} = -\sum_{v=0,v\neq u}^{I} s_{21}^{uv},$$

$$u=0,1,\ldots,I, \quad S_{21} = (s_{21}^{uv})_{u,v=1,\ldots,I},$$

$$s_{22}^{uv} = -\frac{1}{1+\rho} \times$$

$$\int \frac{\rho_u \exp\left[\theta_{u0}+s_u(y;\beta_{uo})\right]\rho_v \exp\left[\theta_{v0}+s_v(y;\beta_{vo})\right]}{1+\sum_{i=1}^{I}\rho_i \exp[\theta_{i0}+s_i(y;\beta_{i0})]} \times$$

$$d_u(y,\beta_u) d_v^\tau(y,\beta_v)] dG_0(y), \quad u \neq v = 0,1,\ldots,I,$$

$$s_{22}^{uu} = -\sum_{v=0,v\neq u}^{I} s_{22}^{uv}, \quad u=0,1,\ldots,I,$$

$$S_{22} = (s_{22}^{uv})_{u,v=1,\ldots,I}, \quad S = \begin{pmatrix} S_{11} & S_{21}^\tau \\ S_{21} & S_{22} \end{pmatrix}$$

$$\Sigma = S^{-1} - (1+\rho)\begin{pmatrix} D+J & 0 \\ 0 & 0 \end{pmatrix},$$

$$B_{uv}(t) =$$

$$\int_{-\infty}^{t} \frac{\rho_u \exp\left[\theta_{u0}+s_u(y;\beta_{uo})\right]\rho_v \exp\left[\theta_{v0}+s_v(y;\beta_{vo})\right]}{1+\sum_{m=1}^{I}\rho_m \exp[\theta_{m0}+s_m(y;\beta_{m0})]} dG_0(y),$$

$$u,v = 0,1,\ldots,I,$$

$$a_{uv}(t;\theta,\beta) = \rho_u \exp\left[\theta_{u0}+s_u(y;\beta_{uo})\right]\times$$

$$\rho_v \exp\left[\theta_{v0}+s_v(y;\beta_{vo})\right], \quad u \neq v = 0,1,\ldots,I,$$

$$a_{uu}(t) = -\sum_{v=0,v\neq u}^{I} a_{uv}(t;\theta,\beta), \quad u=0,1,\ldots,I,$$

$$C_{1h}(t;\theta,\beta) = (a_{h1}(t;\theta,\beta),\ldots,a_{hI}(t;\theta,\beta))^\tau,$$

$$h = 0,1,\ldots,I,$$

$$b_{uv}(t;\theta,\beta) = \rho_u \exp\left[\theta_{u0}+s_u(y;\beta_{uo})\right]\times$$

$$\rho_v \exp\left[\theta_{v0}+s_v(y;\beta_{vo})\right]d_u(t;\beta_u),$$

$$u \neq v = 0,1,\ldots,I,$$

$$b_{uu}(t) = -\sum_{v=0,v\neq u}^{I} b_{uv}(t;\theta,\beta), \quad u=0,1,\ldots,I,$$

$$C_{2h}(t;\theta,\beta) = (b_{h1}(t;\theta,\beta),\ldots,b_{hI}(t;\theta,\beta))^\tau,$$

$$h = 0,1,\ldots,I,$$

$$A_{kh}(t) =$$

$$\int_{-\infty}^{t} \frac{C_{kh}(y;\theta_{(0)},\beta_{(0)})}{1+\sum_{m=1}^{I}\rho_m \exp[\theta_{m0}+s_m(y;\beta_{m0})]} dG_0(y),$$

$$k = 1,2, \quad h = 0,1,\ldots,I, \tag{8}$$

where J is an $I \times I$ matrix of 1 elements and $D = \text{Diag}(\rho_1^{-1},\ldots,\rho_I^{-1})$ is the $I \times I$ diagonal matrix having elements $\{\rho_1^{-1},\ldots,\rho_I^{-1}\}$ on the main diagonal. In order to formulate the results, the following assumptions are stated.

(A1) There exists a neighborhood Θ_0 of the true parameter point $\beta_{(0)}$ such that for all t the function $r_i(t;\beta_i)$ $(i=1,\ldots,I)$ admits all third derivatives $\frac{\partial^3 r_i(t;\beta_i)}{\partial\beta_i^k \partial\beta_i^l \partial\beta_i^m}$ for all $\beta \in \Theta_0$

(A2) For $i=1,\ldots,I$, there exists a function Q_1 such that $\left|\frac{\partial s_i(t;\beta_i)}{\partial\beta_i^k}\right| \leq Q_1(t)$ for all $\beta \in \Theta_0$ and $k = 1,\ldots,p$, where

$$q_{1j}$$
$$= \int Q_1^j(y)\{1+\rho_i \exp[\theta_{i0}+s_i(y;\beta_{io})]\}dG_0(y) < \infty,$$
$$j = 1,2,3.$$

(A3) For $i=1,\ldots,I$, there exists a function Q_2

such that $\left|\dfrac{\partial^2 s_i(t;\beta_i)}{\partial\beta_i^k\partial\beta_i^l}\right|\le Q_2(t)$ for all $\beta\in\Theta_0$

and $k,l=1,\dots,p$, where

$q_{2j}=$
$\int Q_2^j(y)\{1+\rho_i\exp[\theta_{i0}+s_i(y;\beta_{io})]\}dG_0(y)<\infty,$
$\quad j=1,2.$

(A4) For $i=1,\dots,I$, there exists a function Q_3

such that $\left|\dfrac{\partial^3 s_i(t;\beta_i)}{\partial\beta_i^k\partial\beta_i^l\partial\beta_i^m}\right|\le Q_3(t)$ for all $\beta\in\Theta_0$

and $k,l,m=1,\dots,p$, where

$q_3=$
$\int Q_3(y)\{1+\rho_i\exp[\theta_{i0}+s_i(y;\beta_{io})]\}dG_0(y)<\infty$

First, study the asymptotic behavior of the maximum semiparametric likelihood estimate $(\tilde\theta,\tilde\beta)$ defined in (3). Theorem 8 concerns the strong consistency and the asymptotic distribution of $(\tilde\theta,\tilde\beta)$

Theorem 1: Suppose that model (2) and Assumptions $(A1)-(A4)$ hold. Suppose further that S is positive definite. (a) As $n\to\infty$, with probability 1 there exists a sequence $(\tilde\theta,\tilde\beta)$ of roots of the system of score equations (2.1) such that $(\tilde\theta,\tilde\beta)$ is strongly consistent for estimating $(\theta_{(0)},\beta_{(0)})$, i.e.,

$(\tilde\theta,\tilde\beta)\overset{a.s.}{\to}(\theta_{(0)},\beta_{(0)})$.

(b) As $n\to\infty$, it may be written

$$\begin{pmatrix}\tilde\theta-\theta_{(0)}\\ \tilde\beta-\beta_{(0)}\end{pmatrix}=\frac{1}{n}S^{-1}\begin{pmatrix}\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}\\ \frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}\end{pmatrix}$$
$$+o_p(n^{-1/2}),\qquad(9)$$

where $\quad\dfrac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}=\dfrac{\partial\ell(\theta,\beta)}{\partial\theta}\Big|_{(\theta,\beta)=(\theta_{(0)},\beta_{(0)})}$ and

$\dfrac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}=\dfrac{\partial\ell(\theta,\beta)}{\partial\beta}\Big|_{(\theta,\beta)=(\theta_{(0)},\beta_{(0)})}$ .As a result,

$$\sqrt{n}\begin{pmatrix}\tilde\theta-\theta_{(0)}\\ \tilde\beta-\beta_{(0)}\end{pmatrix}\overset{d}{\to}N_{(p+1)I}(0,\Sigma).\qquad(10)$$

Remark 3: A consistent estimate of the covariance matrix Σ is given by

$$\tilde\Sigma=\tilde S^{-1}-(1+\rho)\begin{pmatrix}D+J & 0\\ 0 & 0\end{pmatrix}$$

where $\tilde S$ is obtained from S with $(\theta_{(0)},\beta_{(0)})$ replaced by $(\tilde\theta,\tilde\beta)$ and G_0 replaced by $\widetilde{G_0}$.

Remark 4: Because S^{-1} is the prospectively derived asymptotic variance-covariance matrix of $(\widehat{\theta^*},\widehat\beta)$ on the basis of the prospective likelihood function given in Remark 2, it is seen from the expression for the asymptotic variance-covariance matrix Σ of $(\tilde\theta,\tilde\beta)$ that the asymptotic variance-covariance matrices for $\widehat\beta$ and $\tilde\beta$ coincide under the retrospective sampling scheme and the prospective sampling scheme. Consequently, a prospectively derived analysis under model (1.1) on parameter estimates and standard errors for β is asymptotically correct in case-control studies. These results match those of Weinberg and Wacholder (1993) and Scott and Wild (1997).

The two-step profile maximization procedure, by which the maximum semiparametric likelihood estimator $(\tilde\theta,\tilde\beta,\widetilde{G_0})$ is derived, relies on first maximizing the nonparametric part G_0 with (θ,β) fixed and then maximizing $\ell(\theta,\beta)$ with respect to (θ,β). The estimator $(\tilde\theta,\tilde\beta,\widetilde{G_0})$ can also be derived by employing the following "method of moments". Motivated by the work of Gill,

Vardi, and Wellner (1988), let $F = \sum_{i=0}^{I} \frac{n_i}{n} G_i$ be the "average distribution function", then by (2)

$$G_i(t) = \frac{n}{n_0}$$
$$\int \frac{\exp[\theta_i + s_i(y;\beta_i)]}{\sum_{i=0}^{I} \rho_i \exp[\theta_i + s_i(y;\beta_i)]} I_{[y \leq t]} dF(y),$$
$$i = 0, 1, \ldots, I.$$

Let $F_n(t) = \frac{1}{n} \sum_{i=1}^{n} I_{[T_i \leq t]}$ be the empirical distribution function of the pooled sample $\{T_1, \ldots, T_n\}$. Then G_i can be estimated for fixed (θ, β) by

$$\overline{G}_i(t) = \frac{n}{n_0}$$
$$\int \frac{\exp[\theta_i + s_i(y;\beta_i)]}{\sum_{i=0}^{I} \rho_i \exp[\theta_i + s_i(y;\beta_i)]} I_{[y \leq t]} dF_n(y)$$
$$= \frac{1}{n_0} \sum_{k=1}^{n} \frac{\exp[\theta_i + s_i(T_k;\beta_i)]}{\sum_{i=0}^{I} \rho_i \exp[\theta_i + s_i(T_k;\beta_i)]} I_{[T_k \leq t]}$$

for $i = 0, 1, \ldots, I$. Let $\widehat{G}_i(t) = n_i^{-1} \sum_{j=1}^{n_i} I_{[X_{ij} \leq t]}$ be the empirical distribution function based on the sample $X_{i1}, \ldots, X_{in}$ from the i th response category. Let $\psi_i(t; \theta, \beta)$ be a real function from R^p to R^{p+1} for $i = 1, \ldots, I$ and let $\psi(t; \theta, \beta) = (\psi_1^{\tau}(t; \theta, \beta), \ldots, \psi_I^{\tau}(t; \theta, \beta))^{\tau}$. Then, for a particular choice of $\psi(t; \theta, \beta)$, (θ, β) can be estimated by matching the expectation of $n_i \psi_i(t; \theta, \beta)$ under $\overline{G}_i$ with that under $\widehat{G}_i$ for $i = 1, \ldots, I$:

$$E_{\overline{G}_i}[n_i \psi_i(T; \theta, \beta)]$$
$$= \int n_i \psi_i(t; \theta, \beta) d\overline{G}_i(t)$$
$$= \int n_i \psi_i(t; \theta, \beta) d\widehat{G}_i(t) = E_{\widehat{G}_i}[n_i \psi_i(T; \theta, \beta)]$$

for $i = 1, \ldots, I$. In other words, (θ, β) can be estimated by seeking a root to the following system of equations:
$$L_i(\theta, \beta) =$$

$$\sum_{k=1}^{n} \frac{\rho_i \exp[\theta_i + s_i(T_k;\beta_i)]}{\sum_{m=0}^{I} \rho_m \exp[\theta_m + s_m(T_k;\beta_m)]} \psi_i(T_k, \theta, \beta)$$

$$-\sum_{j=1}^{n_i} \psi_i(X_{ij}, \theta, \beta) = 0 \quad i = 1, \ldots, I. \tag{11}$$

It is easy to see that the above system of equations reduces to the system of score equations in (3) if $\psi_i(t; \theta, \beta) = (1, d_i^{\tau}(t; \beta_i))^{\tau}$ is taken for $i = 1, \ldots, I$. Let $(\overline{\theta}, \overline{\beta})$ with $\overline{\theta} = (\overline{\theta}_1, \ldots, \overline{\theta}_I)^{\tau}$ and $\overline{\beta} = (\overline{\beta}_1, \ldots, \overline{\beta}_I)^{\tau}$ be a solution to the system of equations in (11). Note that $(\overline{\theta}, \overline{\beta})$ depends on the choice of $\psi_i(t; \theta, \beta)$ for $i = 1, \ldots, I$. The following theorem demonstrates that the choice of $\psi_i(t; \theta, \beta) = (1, d_i^{\tau}(t; \beta_i))^{\tau}$ for $i = 1, \ldots, I$ is optimal in the sense that the difference between the asymptotic variance-covariance matrices of $(\overline{\theta}, \overline{\beta})$ and $(\widetilde{\theta}, \widetilde{\beta})$ is positive semidefinite for any set of measurable functions $\{\psi_i(t; \theta, \beta): i = 1, \ldots, I\}$. Qin (1998) established this optimal property when $I = 1$.

Theorem 2: Under the conditions of Theorem 1, we have

$$\sqrt{n} \begin{pmatrix} \overline{\theta} - \theta_{(0)} \\ \overline{\beta} - \beta_{(0)} \end{pmatrix} \xrightarrow{d} N_{(p+1)I}(0, \Sigma_{\psi}),$$

where $\Sigma_{\psi} = V^{-1} B_{\psi} (V^{\tau})^{-1}$ with V and B_{ψ} defined in (18) of the proof section. Moreover, the maximum semiparametric likelihood estimator $(\widetilde{\theta}, \widetilde{\beta})$ is optimal in the sense that $\Sigma_{\psi} - \Sigma$ is positive semidefinite for any set of measurable functions $\{\psi_i(t; \theta, \beta): i = 1, \ldots, I\}$.

In the following case, $p = 1$ is considered, although the results can be naturally generalized to the case of $p > 1$. The weak convergence of $\sqrt{n}(\widetilde{G}_0 - \widehat{G}_0, \ldots, \widetilde{G}_I - \widehat{G}_I)^{\tau}$ is

now established to a multivariate Gaussian process by representing $\widetilde{G}_i - \widehat{G}_i$ $(i = 0, 1, \ldots, I)$ as the mean of a sequence of independent and identically distributed stochastic processes with a remainder term of order $o_p(n^{-1/2})$.

Theorem 3: Suppose that model (2) and Assumptions $(A1) - (A4)$ hold. Suppose further that S is positive definite. For $i = 0, 1, \ldots, I$, one can write

$$\widetilde{G}_i(t) - \widehat{G}_i(t) = H_{1i}(t) - \widehat{G}_i(t) - H_{2i}(t) + R_{in}(t),$$
(12)

where

$$H_{1i}(t)$$
$$= \frac{1}{n_0} \sum_{k=1}^{n} \frac{\exp[\theta_{i0} + s_i(T_k; \beta_{i0})]}{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_{m0} + s_m(T_k; \beta_{m0})]} I_{[T_k \leq t]},$$

$$H_{2i}(t) = \frac{1}{n\rho_i}(A_{1i}^\tau(t), A_{2i}^\tau(t))S^{-1} \begin{pmatrix} \frac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \frac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix},$$
(13)

and the remainder term $R_{in}(t)$ satisfies

$$\sup_{-\infty \leq t \leq \infty} |R_{in}(t)| = o_p(n^{-1/2}).$$
(14)

As a result,

$$\sqrt{n} \begin{pmatrix} \widetilde{G}_0 - \widehat{G}_0 \\ \widetilde{G}_1 - \widehat{G}_1 \\ \vdots \\ \widetilde{G}_I - \widehat{G}_I \end{pmatrix} \xrightarrow{D} \begin{pmatrix} W_0 \\ W_1 \\ \vdots \\ W_I \end{pmatrix}$$

$$in \quad D^{I+1}[-\infty, \infty]$$
(15)

where $D^{I+1}[-\infty, \infty]$ is the product space defined by $D[-\infty, \infty] \times \cdots \times D[-\infty, \infty]$ and $(W_0, W_1, \ldots, W_I)^\tau$ is a multivariate Gaussian

process with continuous sample path and satisfies, for $-\infty \leq s \leq t \leq \infty$,

$$EW_i(t) = 0, \quad i = 0, 1, \ldots, I,$$

$$EW_i(s)W_i(t) = \frac{1+\rho}{\rho_i^2}[G_i(s) - B_{ii}(s)]$$

$$- \frac{1}{\rho_i^2}(A_{1i}^\tau(s), A_{2i}^\tau(s))S^{-1} \begin{pmatrix} A_{1i}(t) \\ A_{2i}(t) \end{pmatrix}, \quad i = 0, 1, \ldots, I,$$

$$EW_i(s)W_j(t)$$
$$= -\frac{1+\rho}{\rho_i \rho_j}B_{ij}(s) - \frac{1}{\rho_i \rho_j}(A_{1i}^\tau(s), A_{2i}^\tau(s))S^{-1}$$
$$\begin{pmatrix} A_{1j}(t) \\ A_{2j}(t) \end{pmatrix}, \quad i \neq j = 0, 1, \ldots, I.$$
(16)

Theorem 3 forms the basis for testing the validity of model (2) on the basis of the test statistic Δ_n in (6). Let w_α denote the α-quantile of the distribution of $\frac{1}{I+1}\sum_{i=0}^{I}\rho_i\{\sup_{-\infty \leq t \leq \infty}|W_i(t)|\}$, i.e., w_α satisfies

$$P(\frac{1}{I+1}\sum_{i=0}^{I}\rho_i\{\sup_{-\infty \leq t \leq \infty}|W_i(t)|\} \leq w_\alpha) = \alpha.$$

According to Theorem 3 and the continuous Mapping Theorem (Billingsley, 1968, p. 30):

$$\lim_{n \to \infty} P(\Delta_n \geq w_{1-\alpha})$$

$$= \lim_{n \to \infty} P\left(\frac{1}{I+1}\sum_{i=0}^{I}\rho_i\{\sup_{-\infty \leq t \leq \infty}\sqrt{n}|\widetilde{G}_i(t) - \widehat{G}_i(t)|\} \geq w_{1-\alpha}\right)$$

$$= P\left(\frac{1}{I+1}\sum_{i=0}^{I}\rho_i\left\{\sup_{-\infty \leq t \leq \infty}|W_i(t)|\right\} \geq w_{1-\alpha}\right) = \alpha$$

Thus, the proposed goodness of fit test procedure has the following decision rule: reject model (2) at level α if $\Delta_n > w_{1-\alpha}$. In order for this proposed test procedure to be useful in practice, the distribution of $\frac{1}{I+1}\sum_{i=0}^{I}\rho_i\{\sup_{-\infty \leq t \leq \infty}|W_i(t)|\}$ must be found and the $(1-\alpha)$-quantile $w_{1-\alpha}$ calculated.

Unfortunately, no analytic expressions appear to be available for the distribution function of $\frac{1}{I+1}\sum_{i=0}^{I}\rho_i\{\sup_{-\infty\le t\le\infty}|W_i(t)|\}$ and the quantile function thereof. A way out is to employ a bootstrap procedure as described in the next section.

A Bootstrap Procedure

In this section is presented a bootstrap procedure which can be employed to approximate the quantile $w_{1-\alpha}$ defined at the end of the last section. If model (1) is valid, since $\theta^* = (\theta_1^*,\ldots,\theta_I^*)^\tau$ is not estimable in general on the basis of the case-control data $T_1,\ldots,T_n$, only generated data, respectively, from $\widetilde{G}_0,\widetilde{G}_1,\ldots,\widetilde{G}_I$, where $\widetilde{G}_i$ $(i=0,1,\ldots,I)$ is given by (5). Specifically, let $X_{i1}^*,\ldots,X_{in_i}^*$ be a random sample from $\widetilde{G}_i$ for $i=0,1,\ldots,I$ and assume that $\{(X_{i1}^*,\ldots,X_{in_i}^*): i=0,1,\ldots,I\}$ are jointly independent. Let $\{T_1^*,\ldots,T_n^*\}$ denote the combined bootstrap sample $\{X_{01}^*,\ldots,X_{0n_0}^*;X_{11}^*,\ldots,X_{1n_1}^*;\cdots;X_{I1}^*,\ldots,X_{In_I}^*\}$ and $(\widetilde{\theta}^*,\widetilde{\beta}^*)$ with $\widetilde{\theta}^* = (\widetilde{\theta}_1^*,\ldots,\widetilde{\theta}_I^*)^\tau$ and $\widetilde{\beta}^* = (\widetilde{\beta}_1^{*\tau},\ldots,\widetilde{\beta}_I^{*\tau})^\tau$ be the solution to the system of score equations in (3) with the T_k^* in place of the T_k. Moreover, similar to $(4)-(6)$,

let $\widehat{G}_i^*(t) = \frac{1}{n_i}\sum_{j=1}^{n_i}I_{[X_{ij}^*\le t]}$ for $i=0,1,\ldots,I$ and

$$\widetilde{p}_k^* = \frac{1}{n_0}\frac{1}{1+\sum_{i=1}^{I}\rho_i\exp[\widetilde{\theta}_i^*+s_i(T_k^*;\widetilde{\beta}_i^*)]},$$
$$k=1,\ldots,n.$$

$$\widetilde{G}_i^*(t) = \sum_{k=1}^{n}\widetilde{p}_k^*\exp[\widetilde{\theta}_i^*+s_i(T_k^*;\widetilde{\beta}_i^*)]I_{[T_k^*\le t]}$$

$$= \frac{1}{n_0}\sum_{k=1}^{n}\frac{\exp[\widetilde{\theta}_i^*+s_i(T_k^*;\widetilde{\beta}_i^*)]I_{[T_k^*\le t]}}{1+\sum_{m=1}^{I}\rho_m\exp[\widetilde{\theta}_m^*+s_m(T_k^*;\widetilde{\beta}_m^*)]},$$
$$i=0,\ldots,I,$$

where $\widetilde{\theta}_0^*=0$ and $s_0(\bullet;\widetilde{\beta}_0^*)\equiv0$. Then the corresponding bootstrap version of the test statistic Δ_n in (6) is given by

$$\Delta_n^* = \frac{1}{I+1}\sum_{i=0}^{I}\rho_i\Delta_{ni}^*,$$

where $\Delta_{ni}^* = \sup_{-\infty\le t\le\infty}|\Delta_{ni}^*(t)|$ with

$\Delta_{ni}^*(t) = \sqrt{n}(\widetilde{G}_i^*(t)-\widehat{G}_i^*(t))$ for $i=0,1,\ldots,I$. To see the validity of the proposed bootstrap procedure, the proofs of Theorems 1 and 3 can be mimicked with slight modification to show the following theorem. The details are omitted here.

Theorem 4: Suppose that model (2) and Assumptions $(A1)-(A4)$ hold. Suppose further that S is positive definite and

$$\int_{\infty}^{-\infty}Q_1^2(y)Q_2(y)\{1+\rho_i\exp[\theta_{i0}+s_i(y;\beta_{i0})]\}dG_0(y)<\infty$$

for $i=1,\ldots,I$.

(a) Along almost all sample sequences $T_1,T_2,\ldots$, given $(T_1,\ldots,T_n)$, as $n\to\infty$, we have

$$\sqrt{n}\begin{pmatrix}\widetilde{\theta}^*-\widetilde{\theta}\\\widetilde{\beta}^*-\widetilde{\beta}\end{pmatrix}\xrightarrow{d}N_{(p+1)I}(0,\Sigma).$$

(b) Along almost all sample sequences $T_1,T_2,\ldots$, given $(T_1,\ldots,T_n)$, as $n\to\infty$, we have

$$\sqrt{n}\begin{pmatrix}\widetilde{G}_0^*-\widehat{G}_0^*\\\widetilde{G}_1^*-\widehat{G}_1^*\\\vdots\\\widetilde{G}_I^*-\widehat{G}_I^*\end{pmatrix}\xrightarrow{D}\begin{pmatrix}W_0\\W_1\\\vdots\\W_I\end{pmatrix}\quad\text{in}\quad D^{I+1}[-\infty,\infty],$$

where $(W_0,W_1,\ldots,W_I)^\tau$ is the multivariate Gaussian process defined in Theorem 3.

Theorem 3 and part (b) of Theorem 4 indicate that the limit process of $\sqrt{n}(\widetilde{G}_0^*-\widehat{G}_0^*,\ldots,\widetilde{G}_I^*-\widehat{G}_I^*)^\tau$ agrees with that

of $\sqrt{n}(\widetilde{G}_0 - \widehat{G}_0, \ldots, \widetilde{G}_I - \widehat{G}_I)^\tau$. It follows from the Continuous Mapping Theorem that $\Delta_n^* = \frac{1}{I+1} \sum_{i=0}^{I} \rho_i \Delta_{ni}^*$ has the same limiting behavior as does $\Delta_n = \frac{1}{I+1} \sum_{i=0}^{I} \rho_i \Delta_{ni}$. Thus, the quantiles of the distribution of Δ_n can be approximated by those of Δ_n^*. For $\alpha \in (0,1)$, let $w_{1-\alpha}^n = \inf\{t; \ P^*(\Delta_n^* \le t) \ge 1-\alpha\}$, where P^* stands for the bootstrap probability under $\widetilde{G}_i \ (i = 0,1,\ldots,I)$. Then there is the following bootstrap decision rule: reject model (2) at level α if $\Delta_n > w_{1-\alpha}^n$.

Two real data sets are next considered. Note that the multivariate logistic regression model is a special case of the multivariate multiplicative-intercept risk model (1) with $\theta_i^* = \exp(\alpha_i^*)$ and $r_i(x; \beta_i) = \exp(\beta_i^\tau x)$ for $i = 1,\ldots,I$. In this case, we have $\theta_i = \alpha_i^* + \log(\frac{\pi_0}{\pi_1})$ and $s_i(x; \beta_i) = \beta_i^\tau x$ in model (2) for $i = 1,\ldots,I$.

Example 1: Agresti (1990) analyzed, by employing the continuation-ratio logit model, the relationship between the concentration level of an industrial solvent and the outcome for pregnant mice in a developmental toxicity study. The complete dataset is listed on page 320 in his book. Let X denote "concentration level (in mg/kg per day)" and Y represent "pregnancy outcome", in which $Y = 0,1,$ and 2 stand for three possible outcomes: Normal, Malformation, and Non-live. Here this data set is analyzed on the basis of the multivariate logistic regression model. Because the sample data (X_i, Y_i), $i = 1,\ldots,1435$, can be thought as being drawn independently and identically from the joint distribution of (X,Y), Remark 1 implies that the test statistic Δ_n in (6) can be used to test the validity of the multivariate logistic regression model. Under model (2),

$$(\widetilde{\theta}_1, \widetilde{\beta}_1, \widetilde{\theta}_2, \widetilde{\beta}_2) = (-3.33834, 0.01401,$$
$$-2.52553, 0.01191)$$

and $\Delta_n = 0.49439$ with the observed P-value equal to 0 based on 1000 bootstrap replications of Δ_n^*. Because $n_0 = 1000$, $n_1 = 199$, and $n_2 = 236$, $\alpha_1^* = \log\theta_1^*$ and $\alpha_2^* = \log\theta_2^*$ can be estimated by

$$\widetilde{\alpha}_1^* = -3.33834 + \log(199/1000) = -4.95279$$

and $\widetilde{\alpha}_2^* = -2.52553$
$+ \log(236/1000) = -3.96945$, respectively.

Figure 1 shows the curves of $\widetilde{G}_0$ and $\widehat{G}_0$ (left panel), the curves of $\widetilde{G}_1$ and $\widehat{G}_1$ (middle panel), and the curves of $\widetilde{G}_2$ and $\widehat{G}_2$ (right panel) based on this data set. The middle and right panels indicate strong evidence of the lack of fit of the multivariate logistic regression model to these data within the categories for Malformation and Non-live.

Figure 1. Example 1: Developmental toxicity study with pregnant mice. Left panel: estimated cumulative distribution functions $\tilde{G}_0$ (solid curve) and $\hat{G}_0$ (dashed curve). Middle panel: estimated cumulative distribution functions $\tilde{G}_1$ (solid curve) and $\hat{G}_1$ (dashed curve). Right panel: estimated cumulative distribution functions $\tilde{G}_2$ (solid curve) and $\hat{G}_2$ (dashed curve).
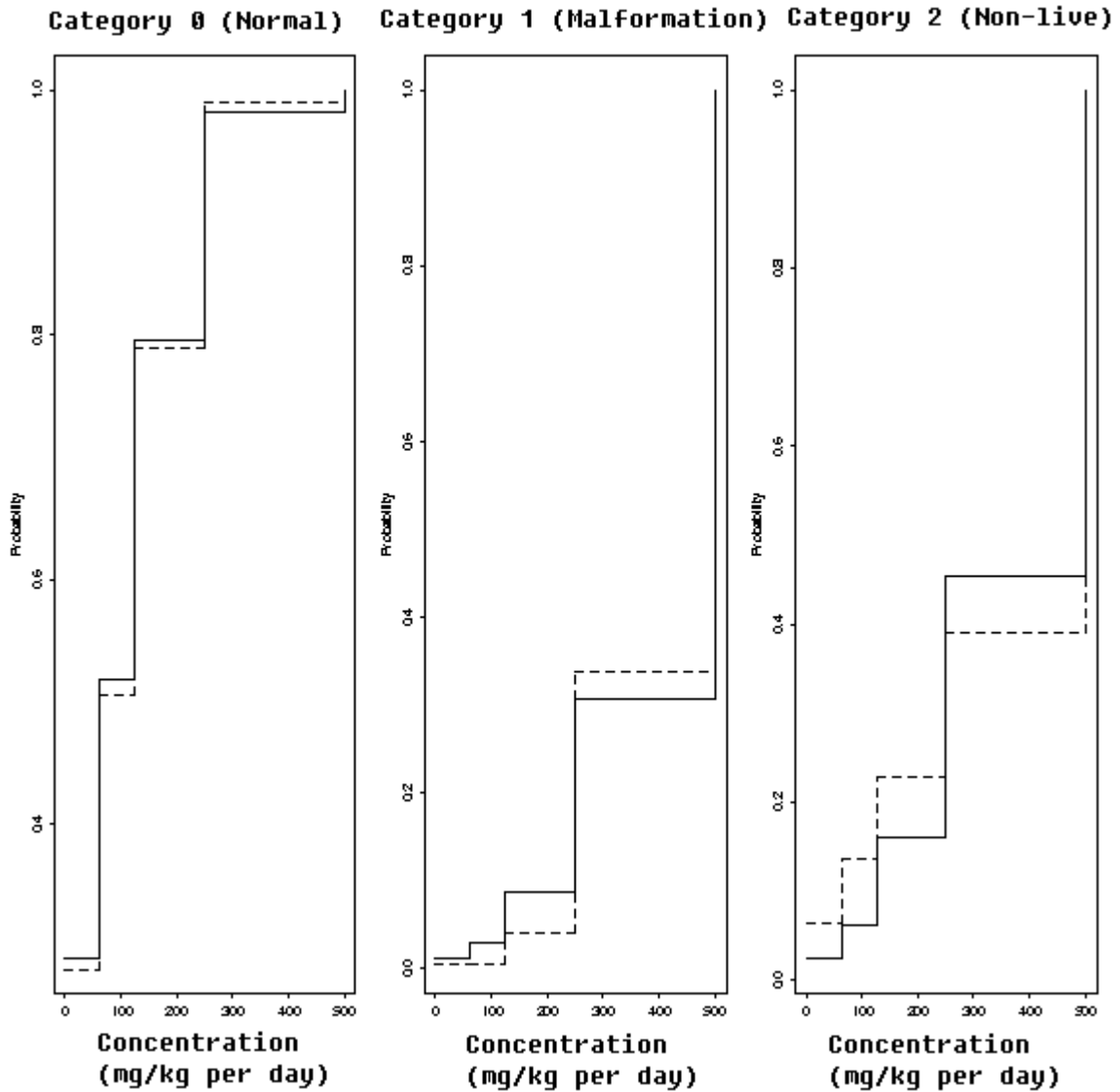
Example 2: Table 9.12 in Agresti (1990, p. 339) contains data for the 63 alligators caught in Lake George. Here the relationship between the alligator length and the primary food choice of alligators is analyzed by employing the multivariate logistic regression model. Let X denote "length of alligator (in meters)" and Y represent "primary food choice" in which $Y = 0, 1,$ and 2 stand for three categories: Other, Fish, and Invertebrate. Since the sample data (X_i, Y_i), $i = 1, \ldots, 63$, can be thought as being drawn independently and identically from the joint distribution of (X, Y), Remark 1 implies that the test statistic Δ_n in (6) can be used to test the validity of the multivariate logistic regression model.

For the male data, we find $(\widetilde{\theta}_1, \widetilde{\beta}_1, \widetilde{\theta}_2, \widetilde{\beta}_2) = (0.41781, -0.17678, 4.83809, -2.60093)$ and $\Delta_n = 1.33460$ with the observed P-value identical to 0.389 based on 1000 bootstrap replications of Δ_n^*. For the female data, we find $(\widetilde{\theta}_1, \widetilde{\beta}_1, \widetilde{\theta}_2, \widetilde{\beta}_2) = (-5.58723, 2.57174, 2.70962, -1.50304)$ and $\Delta_n = 1.63346$ with the observed P-value equal to 0.249 based on 1000 bootstrap replications of Δ_n^*. For the combined male and female data, $(\widetilde{\theta}_1, \widetilde{\beta}_1, \widetilde{\theta}_2, \widetilde{\beta}_2) = (-0.19542, 0.08481, 4.48780, -2.38837)$ and $\Delta_n = 1.73676$ is found with the observed P-value identical to 0.225 based on 1000 bootstrap replications of Δ_n^*, indicating that we can ignore the gender effect on primary food choice. Because $n_0 = 10$, $n_1 = 33$, and $n_2 = 20$, $\alpha_1^* = \log \theta_1^*$ and $\alpha_2^* = \log \theta_2^*$ can be estimated by $\widetilde{\alpha}_1^* = -0.19542 + \log(33/10) = 0.99850$ and $\widetilde{\alpha}_2^* = 4.48780 + \log(20/10) = 5.18094$, respectively.

Figures 2-4 display the curves of $\widetilde{G}_0$ and $\widehat{G}_0$ (left panel), the curves of $\widetilde{G}_1$ and $\widehat{G}_1$ (middle panel), and the curves of $\widetilde{G}_2$ and $\widehat{G}_2$ (right panel) based, respectively, on the male, female, and combined data set. For the combined data, the curve of $\widetilde{G}_1(\widetilde{G}_2)$ bears a resemblance to that of $\widehat{G}_1(\widehat{G}_2)$, whereas the dissimilarity between the curves of $\widetilde{G}_0$ and $\widehat{G}_0$ indicates some evidence of lack of fit of the multivariate logistic regression model to these data within the baseline category for Other.

Proofs

First presented are four lemmas, which will be used in the proof of the main results. The proofs of Lemmas 1, 2, and 3 are lengthy yet straightforward and are therefore omitted here. Throughout this section, the norm of a $m_1 \times m_2$ matrix $A = (a_{ij})_{m_1 \times m_2}$ is defined by

$$\| A \| = \left(\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} a_{ij}^2 \right)^{1/2} \quad \text{for} \quad m_1, m_2 \geq 1.$$

Furthermore, in addition to the notation in (8) we introduce some further notation. Write

$$Q_{i11} = (s_{11}^{1i}, \cdots, s_{11}^{Ii})^{\tau}, \quad Q_{i21} = ((s_{21}^{1i})^{\tau}, \cdots, (s_{21}^{Ii})^{\tau})^{\tau},$$

$$Q_i = \begin{pmatrix} Q_{i11} \\ Q_{i21} \end{pmatrix}, \quad i = 0, 1, \cdots, I,$$

$$S_{n11} = -\frac{1}{n} \frac{\partial^2 \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta \, \partial \theta^{\tau}},$$

$$S_{n21} = -\frac{1}{n} \frac{\partial^2 \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta \, \partial \theta^{\tau}},$$

$$S_{n22} = -\frac{1}{n} \frac{\partial^2 \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta \, \partial \beta^{\tau}}, \quad S_n = \begin{pmatrix} S_{n11} & S_{n21}^{\tau} \\ S_{n12} & S_{n22} \end{pmatrix}$$

$$H_{0i}(t) = \frac{1}{n_i}$$

$$\sum_{k=1}^{n} \frac{C_{1i}(T_k; \theta_{(0)}, \beta_{(0)}) I_{[T_k \leq t]}}{\{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_{m0} + s_m(T_k; \beta_{m0})]\}^2},$$
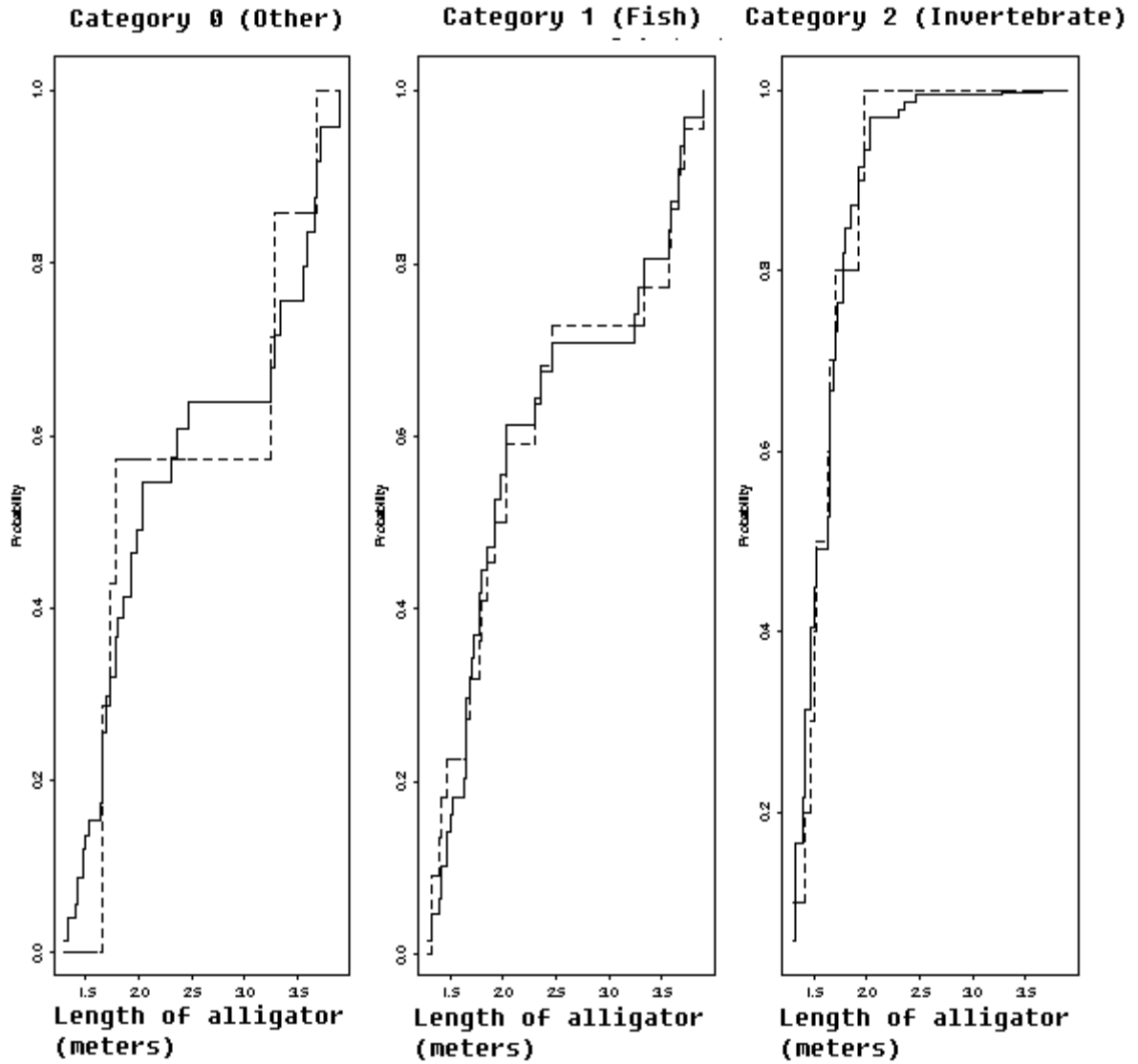
$$i = 0, 1, \cdots, I,$$

Figure 2. Example 2: Primary food choice for 39 male Florida alligators. Left panel: estimated cumulative distribution functions $\widetilde{G}_0$ (solid curve) and $\hat{G}_0$ (dashed curve). Middle panel: estimated cumulative distribution functions $\widetilde{G}_1$ (solid curve) and $\hat{G}_1$ (dashed curve). Right panel: estimated cumulative distribution functions $\widetilde{G}_2$ (solid curve) and $\hat{G}_2$ (dashed curve).
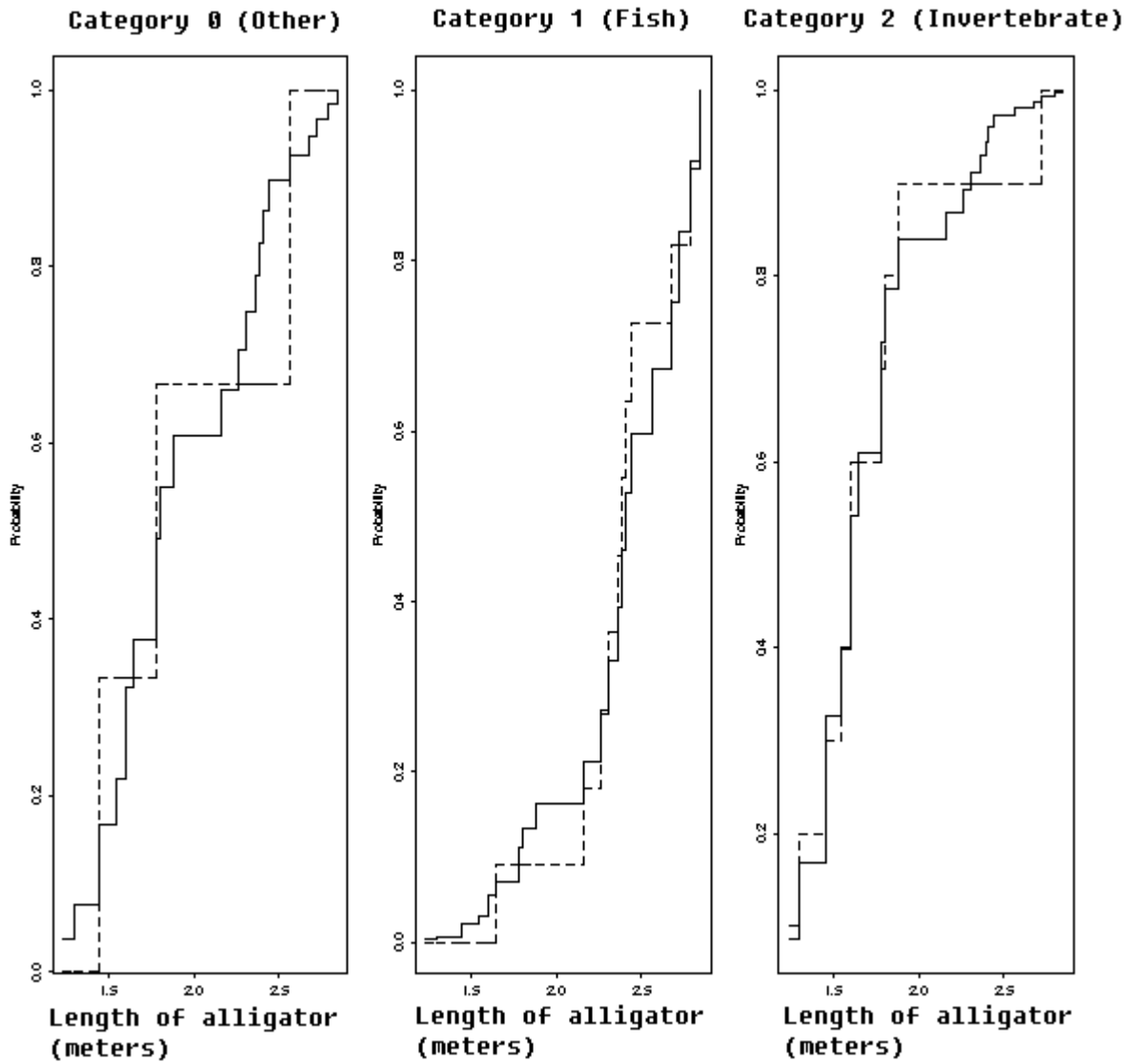
Figure 3. Example 2: Primary food choice for 24 female Florida alligators. Left panel: estimated cumulative distribution functions $\tilde{G}_0$ (solid curve) and $\hat{G}_0$ (dashed curve). Middle panel: estimated cumulative distribution functions $\tilde{G}_1$ (solid curve) and $\hat{G}_1$ (dashed curve). Right panel: estimated cumulative distribution functions $\tilde{G}_2$ (solid curve) and $\hat{G}_2$ (dashed curve).

Figure 4. Example 2: Primary food choice for 63 male and female Florida alligators. Left panel: estimated cumulative distribution functions $\tilde{G}_0$ (solid curve) and $\hat{G}_0$ (dashed curve). Middle panel: estimated cumulative distribution functions $\tilde{G}_1$ (solid curve) and $\hat{G}_1$ (dashed curve). Right panel: estimated cumulative distribution functions $\tilde{G}_2$ (solid curve) and $\hat{G}_2$ (dashed curve).
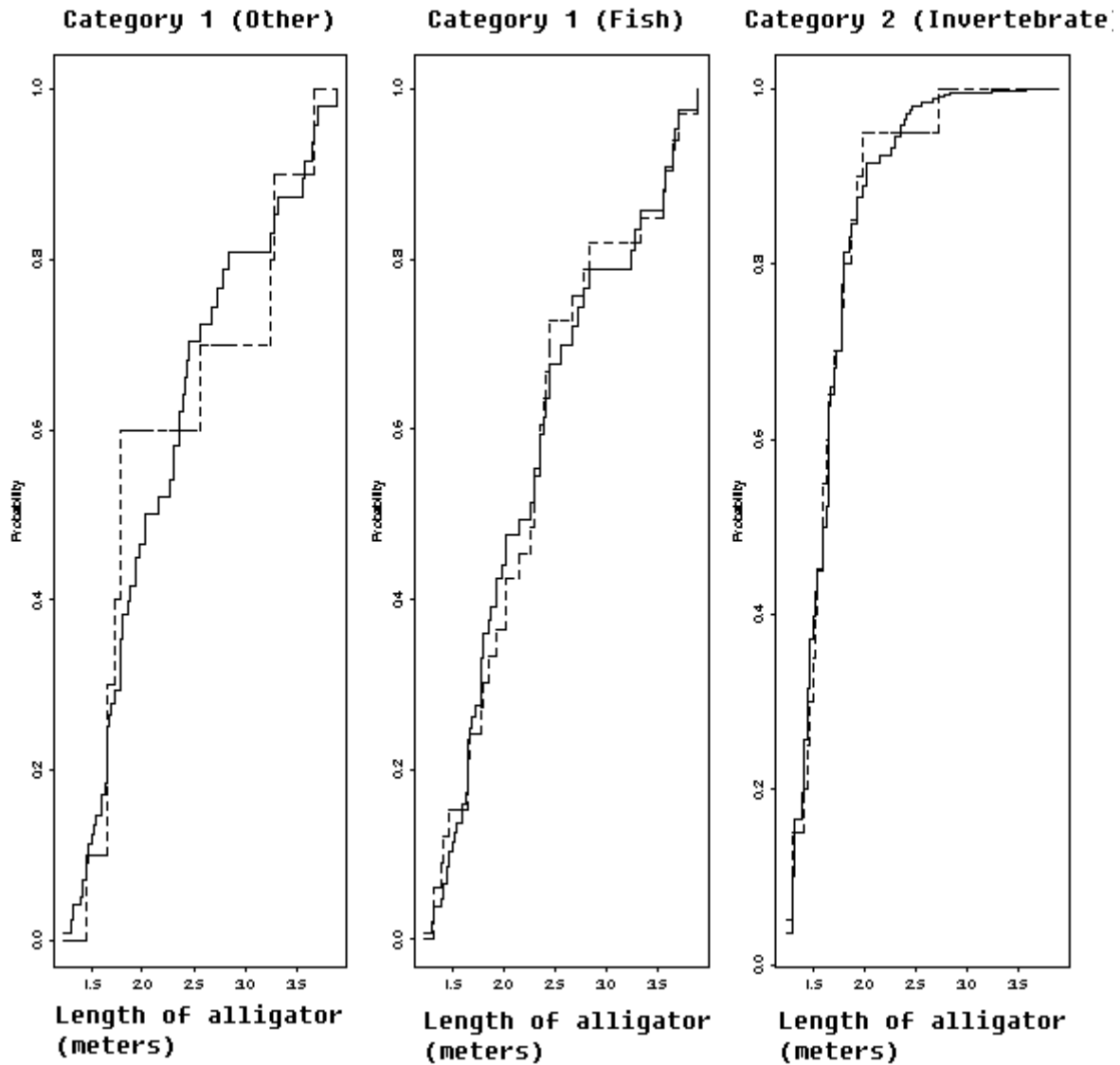
$$H_{3i}(t) = \frac{1}{n_i}$$

$$\sum_{k=1}^{n} \frac{C_{2i}(T_k; \theta_{(0)}, \beta_{(0)}) I_{[T_k \leq t]}}{\{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_{m0} + s_m(T_k; \beta_{m0})]\}^2},$$

$$i = 0, 1, \cdots, I.$$

Lemma 1: Suppose that model (2) holds and S is positive definite. Let J be an $I \times I$ matrix of 1 elements and let $D = \text{Diag}(\rho_1^{-1}, \cdots, \rho_I^{-1})$ denote the $I \times I$ diagonal matrix having elements $\{\rho_1^{-1}, \cdots, \rho_I^{-1}\}$ on the main diagonal, then

$$B \equiv \frac{1}{n} \text{Var} \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix} = S - \sum_{i=0}^{I} \frac{1+\rho}{\rho_i} Q_i Q_i^{\tau},$$

$$S^{-1}BS^{-1} = S^{-1} - (1+\rho) \begin{pmatrix} D+J & 0 \\ 0 & 0 \end{pmatrix} = \Sigma.$$

Lemma 2: Suppose that model (2) holds and S is positive definite. For $-\infty \leq s \leq t \leq \infty$,

$$\text{Cov}(\sqrt{n}[H_{1i}(s) - \hat{G}_i(s)], \sqrt{n}[H_{1i}(t) - \hat{G}_i(t)])$$

$$= \frac{1+\rho}{\rho_i^2}[G_i(s) - B_{ii}(s)]$$

$$- \frac{1+\rho}{\rho_i^2} \sum_{k=0, k \neq i}^{I} \frac{1}{\rho_k} B_{ik}(s) B_{ik}(t)$$

$$- \frac{1+\rho}{\rho_i^3}[G_i(s) - B_{ii}(s)][G_i(t) - B_{ii}(t)],$$

$$i = 0, 1, \cdots, I,$$

$$\text{Cov}(\sqrt{n}[H_{1i}(s) - \hat{G}_i(s)], \sqrt{n}[H_{1j}(t) - \hat{G}_j(t)])$$

$$= -\frac{1+\rho}{\rho_i \rho_j} B_{ij}(s) - \frac{1+\rho}{\rho_i \rho_j} \sum_{k=0}^{I} \frac{1}{\rho_k} B_{ik}(s) B_{jk}(t)$$

$$+ \frac{1+\rho}{\rho_i \rho_j^2} B_{ij}(s) G_j(t) + \frac{1+\rho}{\rho_i^2 \rho_j} G_i(s) B_{ij}(t),$$

$$i \neq j = 0, 1, \cdots, I.$$

Lemma 3: Suppose that model (2) holds and S is positive definite. For $-\infty \leq s \leq t \leq \infty$, we have

$$\text{Cov}(\sqrt{n}[H_{1i}(s) - \hat{G}_i(s)], \sqrt{n}H_{2i}(t))$$

$$= \text{Cov}(\sqrt{n}H_{2i}(s), \sqrt{n}H_{2i}(t))$$

$$= \frac{1}{\rho_i^2}(A_{1i}^{\tau}(t), A_{2i}^{\tau}(t))S^{-1} \begin{pmatrix} A_{1i}(s) \\ A_{2i}(s) \end{pmatrix}$$

$$- \frac{1+\rho}{\rho_i^2} \sum_{k=0, k \neq i}^{I} \frac{1}{\rho_k} B_{ik}(s) B_{ik}(t)$$

$$- \frac{1+\rho}{\rho_i^3}[G_i(s) - B_{ii}(s)][G_i(t) - B_{ii}(t)],$$

$$i = 0, 1, \cdots, I,$$

$$\text{Cov}(\sqrt{n}[H_{1i}(s) - \hat{G}_i(s)], \sqrt{n}H_{2j}(t))$$

$$= \text{Cov}(\sqrt{n}H_{2i}(s), \sqrt{n}H_{2j}(t))$$

$$= \frac{1}{\rho_i \rho_j}(A_{1i}^{\tau}(s), A_{2i}^{\tau}(s))S^{-1} \begin{pmatrix} A_{1j}(t) \\ A_{2j}(t) \end{pmatrix}$$

$$- \frac{1+\rho}{\rho_i \rho_j} \sum_{k=0}^{I} \frac{1}{\rho_k} B_{ik}(s) B_{jk}(t)$$

$$+ \frac{1+\rho}{\rho_i \rho_j^2} B_{ij}(s) G_j(t) + \frac{1+\rho}{\rho_i^2 \rho_j} G_i(s) B_{ij}(t),$$

$$i \neq j = 0, 1, \cdots, I.$$

Lemma 4: Suppose that model (2) and Assumption (A2) hold. If S is positive definite and G_0 is continuous, then the stochastic process $\{\sqrt{n}[H_{1i}(t) - \hat{G}_i(t) - H_{2i}(t)], -\infty \leq t \leq \infty\}$ is tight in $D[-\infty, \infty]$ for $i = 0, 1, \cdots, I$, where $H_{1i}(t)$ and $H_{2i}(t)$ are defined in (13).

Proof: Because $n = \frac{1+\rho}{\rho_i} n_i$ for $i = 0, 1, \cdots, I$, it can be shown after some algebra that

$$\sqrt{n}[H_{1i}(t) - \hat{G}_i(t) - H_{2i}(t)]$$

$$= \frac{1}{\rho_i} \sum_{k=0, k \neq i}^{I} \sqrt{\frac{1+\rho}{\rho_k}} \sqrt{n_k} U_{ik}(t)$$

$$- \frac{1}{\rho_i} \sqrt{\frac{1+\rho}{\rho_i}} \sqrt{n_i} U_{ii}(t) - \sqrt{n} H_{2i}(t), \qquad (17)$$

where

$$U_{ii}(t) = \frac{1}{n_i}$$

$$\sum_{j=1}^{n_i} \frac{\sum_{m=0, m \neq i}^{I} \rho_m \exp[\theta_{m0} + s_m(X_{ij}; \beta_{m0})]}{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_{m0} + s_m(X_{ij}; \beta_{m0})]} \rho_i I_{[X_{ij} \leq t]}$$

$$- [G_i(t) - B_{ii}(t)],$$

$$U_{ik}(t) = \frac{1}{n_k}$$

$$\sum_{j=1}^{n_k} \frac{\rho_i \exp[\theta_{i0} + s_i(X_{kj}; \beta_{i0})] \rho_k I_{[X_{kj} \leq t]}}{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_{m0} + s_m(X_{kj}; \beta_{m0})]}$$

$$- B_{ik}(t), \qquad k \neq i = 0, 1, \cdots, I.$$

Let $\Im = \{I_{(-\infty, t]} : t \in R\}$ be the collection of all indicator functions of cells $(-\infty, t]$ in R. According to the classical empirical process theory, $\Im$ is a $P_{X_{k1}}$-Donsker class for $k = 0, 1, \cdots, I$, where $P_{X_{k1}} = P \circ X_{k1}^{-1}$ is the law of X_{k1} for $k = 0, 1, \cdots, I$. For each $i = 0, 1, \cdots, I$, let us define $I + 1$ fixed functions $f_{i0}, f_{i1}, \cdots, f_{iI}$ by

$$f_{ii}(y) = \frac{\rho_i \sum_{m=0, m \neq i}^{I} \rho_m \exp[\theta_{m0} + s_m(y; \beta_{m0})]}{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_{m0} + s_m(y; \beta_{m0})]},$$

$$f_{ik}(y) = \frac{\rho_k \rho_i \exp[\theta_{i0} + s_i(y; \beta_{i0})]}{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_{m0} + s_m(y; \beta_{m0})]},$$

$$k \neq i = 0, 1, \cdots, I.$$

Then it is seen that $f_{i0}, f_{i1}, \cdots, f_{iI}$ are uniformly bounded functions. According to Example

2.10.10 of van der Vaart and Wellner (1996, p. 192), it can be concluded that $\Im \cdot f_{ik}$ is a $P_{X_{k1}}$-Donsker class for $k = 0, 1, \cdots, I$.

Let $P_{n_k} = \frac{1}{n_k} \sum_{j=1}^{n_k} \delta_{X_{kj}}$ be the empirical measure of $X_{k1}, \cdots, X_{kn_k}$ for $k = 0, 1, \cdots, I$, where δ_x is the measure with mass one at x. Then, it can be shown that

$$\sqrt{n_k}(P_{n_k} - P_{X_{k1}})(I_{(-\infty, t]} f_{ik})$$
$$= \sqrt{n_k} U_{ik}(t), \qquad i, k = 0, 1, \cdots, I.$$

As a result, there exist $I + 1$ zero-mean Gaussian processes $V_{i0}, V_{i1}, \cdots, V_{iI}$ such that

$$\sqrt{n_k} U_{ik} \xrightarrow{D} V_{ik} \quad \text{on } D[-\infty, \infty],$$
$$i, k = 0, 1, \cdots, I$$

Thus, the stochastic process $\{\sqrt{n_k} U_{ik}(t), -\infty \leq t \leq \infty\}$ is tight on $D[-\infty, \infty]$ for $i, k = 0, 1, \cdots, I$. Moreover, it can be shown by using the tightness axiom (Sen & Singer, 1993, p. 330) that the stochastic process $\{\sqrt{n} H_{2i}(t), -\infty \leq t \leq \infty\}$ is tight on $D[-\infty, \infty]$ for $i = 0, 1, \cdots, I$. These results, along with (17), imply that the stochastic process

$$\{\sqrt{n}[H_{1i}(t) - \hat{G}_i(t) - H_{2i}(t)], -\infty \leq t \leq \infty\}$$

is tight in $D[-\infty, \infty]$ for $i = 0, 1, \cdots, I$. The proof is complete.

Proof of Theorem 1: Fro part (a), let

$$B_\varepsilon = \{(\theta, \beta) : \| \theta - \theta_{(0)} \|^2 + \| \beta - \beta_{(0)} \|^2 \leq \varepsilon^2\}$$

be the ball with center at the true parameter point $(\theta_{(0)}, \beta_{(0)})$ and radius ε for some $\varepsilon > 0$. For small ε, it can be shown that we can expand $n^{-1}\ell(\theta, \beta)$ on the surface of B_ε about $(\theta_{(0)}, \beta_{(0)})$ to find

$$\frac{1}{n}\ell(\theta,\beta)-\frac{1}{n}\ell(\theta_{(0)},\beta_{(0)})=W_{n1}+W_{n2}+W_{n3},$$

where

$$W_{n1}=(\theta^{\tau}-\theta_{(0)}^{\tau},\beta^{\tau}-\beta_{(0)}^{\tau})\frac{1}{n}\begin{pmatrix}\dfrac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}\\[2mm]\dfrac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}\end{pmatrix},$$

$$W_{n2}=-\frac{1}{2}(\theta^{\tau}-\theta_{(0)}^{\tau},\beta^{\tau}-\beta_{(0)}^{\tau})S_n\begin{pmatrix}\theta-\theta_{(0)}\\\beta-\beta_{(0)}\end{pmatrix},$$

and W_{n3} satisfies $|W_{n3}|\le c_3\varepsilon^3$ for some constant $c_3>0$ and sufficiently large n with probability 1. Because $\frac{1}{n}\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}\xrightarrow{a.s.}0$ and $\frac{1}{n}\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}\xrightarrow{a.s.}0$ by the strong law of large numbers, it follows that for any given $\varepsilon>0$, with probability 1 $|W_{n1}|\le 2\varepsilon^3$ for sufficiently large n. Furthermore, because $S_n\xrightarrow{a.s.}S$ again by the strong law of large numbers, it follows that with probability 1, $\|S_n-S\|<2\varepsilon$ for sufficiently large n. Because S is positive definite, on the surface of B_ε there is,

$$-\frac{1}{2}(\theta^{\tau}-\theta_{(0)}^{\tau},\beta^{\tau}-\beta_{(0)}^{\tau})S\begin{pmatrix}\theta-\theta_{(0)}\\\beta-\beta_{(0)}\end{pmatrix}$$
$$\le-\frac{\varepsilon^2}{2}\inf_{x\ne0}\frac{x^{\tau}Sx}{x^{\tau}x}\le-\frac{\varepsilon^2}{2}\lambda_1,$$

where $\lambda_1>0$ is the smallest eigenvalue of S. As a result, $W_{n2}<-c_2\varepsilon^2$ for sufficiently large n with probability 1 with $c_2=\frac{\lambda_1}{2}-\varepsilon>0$ for sufficiently small $\varepsilon>0$. Consequently, if $\varepsilon<\frac{c_2}{2+c_3}$, then on the surface of B_ε,

$$\frac{1}{n}\ell(\theta,\beta)-\frac{1}{n}\ell(\theta_{(0)},\beta_{(0)})$$
$$\le|W_{n1}|+|W_{n2}|+|W_{n3}|\le2\varepsilon^3-c_2\varepsilon^2+c_3\varepsilon^3<0$$

for sufficiently large n with probability 1. It has been shown that for any sufficiently small $\varepsilon>0$ and sufficiently large n, with probability 1, $\ell(\theta,\beta)<\ell(\theta_{(0)},\beta_{(0)})$ at all points (θ,β) on the surface of B_ε, and hence that $\ell(\theta,\beta)$ has a local maximum in the interior of B_ε. Because at a local maximum the score equations (3) must be satisfied it follows that for any sufficiently small $\varepsilon>0$ and sufficiently large n, with probability 1, the system of score equations (3) has a solution $(\tilde\theta,\tilde\beta)$ within B_ε. Because $\varepsilon>0$ is arbitrary, $(\tilde\theta,\tilde\beta)$ is strongly consistent for estimating $(\theta_{(0)},\beta_{(0)})$, i.e., $(\tilde\theta,\tilde\beta)\xrightarrow{a.s.}(\theta_{(0)},\beta_{(0)})$.

For part (b), since $(\tilde\theta,\tilde\beta)$ is strongly consistent by part (a), expanding $\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}$ and $\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}$ at $(\theta_{(0)},\beta_{(0)})$ gives

$$0=\frac{\partial\ell(\tilde\theta,\tilde\beta)}{\partial\theta}$$
$$=\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}+\frac{\partial\ell^2(\theta_{(0)},\beta_{(0)})}{\partial\theta\,\partial\theta^{\tau}}(\tilde\theta-\theta_{(0)})$$
$$+\frac{\partial\ell^2(\theta_{(0)},\beta_{(0)})}{\partial\theta\,\partial\beta^{\tau}}(\tilde\beta-\beta_{(0)})+o_p(\delta_n),$$

$$0=\frac{\partial\ell(\tilde\theta,\tilde\beta)}{\partial\beta}$$
$$=\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}+\frac{\partial\ell^2(\theta_{(0)},\beta_{(0)})}{\partial\beta\,\partial\theta^{\tau}}(\tilde\theta-\theta_{(0)})$$
$$+\frac{\partial\ell^2(\theta_{(0)},\beta_{(0)})}{\partial\beta\,\partial\beta^{\tau}}(\tilde\beta-\beta_{(0)})+o_p(\delta_n),$$

where

$$\delta_n=\|\tilde\theta-\theta_{(0)}\|+\|\tilde\beta-\beta_{(0)}\|=o_p(1).$$

Thus,

$$nS_n \begin{pmatrix} \tilde{\theta} - \theta_{(0)} \\ \tilde{\beta} - \beta_{(0)} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix} + o_p(\delta_n).$$

Because $S_n = S + o_p(1)$ by the weak law of large numbers and $\dfrac{1}{\sqrt{n}} \begin{vmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{vmatrix} = O_p(1)$ by the central limit theorem, it follows that

$$\begin{pmatrix} \tilde{\theta} - \theta_{(0)} \\ \tilde{\beta} - \beta_{(0)} \end{pmatrix} = \frac{1}{n} S^{-1} \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix}$$

$$+ o_p(n^{-1/2}) \frac{1}{\sqrt{n}} \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix} + o_p(n^{-1} \delta_n)$$

$$= \frac{1}{n} S^{-1} \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix} + o_p(n^{-1/2}),$$

thus establishing (9). To prove (10), it suffices to show that

$$\frac{1}{\sqrt{n}} S^{-1} \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix} \xrightarrow{d} N_{(p+1)I}(0, \Sigma).$$

Because each term in $\dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta}$ and $\dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta}$ has mean 0, it follows from the multivariate central limit theorem that

$$\frac{1}{\sqrt{n}} B^{-1/2} \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix} \xrightarrow{d} N_{(p+1)I}(0, I_{(p+1)I}),$$

where B is defined in Lemma 1. By Slutsky's Theorem and Lemma 1,

$$\frac{1}{\sqrt{n}} S^{-1} \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix}$$

$$= S^{-1} B^{1/2} \frac{1}{\sqrt{n}} B^{-1/2} \begin{pmatrix} \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \theta} \\ \dfrac{\partial \ell(\theta_{(0)}, \beta_{(0)})}{\partial \beta} \end{pmatrix}$$

$$\xrightarrow{d} S^{-1} B^{1/2} N_{(p+1)I}(0, I_{(p+1)I}) \xrightarrow{d} N_{(p+1)I}(0, \Sigma).$$

The proof is complete.

Proof of Theorem 2: Let

$$v_{11}^{ij} = -\frac{1}{1+\rho} \times$$

$$\int \frac{\rho_i \exp[\theta_{i0} + s_i(t; \beta_{i0})] \rho_j \exp[\theta_{j0} + s_j(t; \beta_{j0})]}{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_m + s_m(t; \beta_{m0})]} \times$$

$$\psi_i(t; \theta_{(0)}, \beta_{(0)}) dG_0(t), \quad i \neq j = 0, 1, \cdots, I,$$

$$v_{11}^{ii} = -\sum_{j=0, j \neq i}^{I} v_{11}^{ij},$$

$$i = 0, 1, \cdots, I, \qquad V_{11} = (v_{11}^{ij})_{i,j=1,\cdots,I},$$

$$v_{12}^{ij} = -\frac{1}{1+\rho} \times$$

$$\int \frac{\rho_i \exp[\theta_{i0} + s_i(t; \beta_{i0})] \rho_j \exp[\theta_{j0} + s_j(t; \beta_{j0})]}{1 + \sum_{m=1}^{I} \rho_m \exp[\theta_m + s_m(t; \beta_{m0})]} \times$$

$$\psi_i(t; \theta_{(0)}, \beta_{(0)}) d_j^{\tau}(t; \beta_{(j0)}) dG_0(t), \quad i \neq j = 0, 1, \cdots, I,$$

$$v_{12}^{ii} = -\sum_{j=0, j\neq i}^{I} v_{12}^{ij}, \quad i = 0,1,\cdots,I,$$

$$V_{12} = (v_{12}^{ij})_{i,j=1,\cdots,I}, \quad V = (V_{11}, V_{12}),$$

$$L(\theta,\beta) = (L_1^{\tau}(\theta,\beta),\cdots,L_I^{\tau}(\theta,\beta))^{\tau},$$

$$B_{\psi} = \mathrm{Var}\left[\frac{1}{\sqrt{n}}L(\theta_{(0)},\beta_{(0)})\right]. \tag{18}$$

The proof of the asymptotic normality of $\begin{pmatrix}\bar{\theta}-\theta_{(0)}\\\bar{\beta}-\beta_{(0)}\end{pmatrix}$ is similar to that of $\begin{pmatrix}\tilde{\theta}-\theta_{(0)}\\\tilde{\beta}-\beta_{(0)}\end{pmatrix}$ in Theorem 1 by noting that

$$\begin{pmatrix}\bar{\theta}-\theta_{(0)}\\\bar{\beta}-\beta_{(0)}\end{pmatrix} = \frac{1}{n}V^{-1}L(\theta_{(0)},\beta_{(0)}) + o_p(n^{-1/2}).$$

To prove the second part of Theorem 2, notice that $\begin{pmatrix}\tilde{\theta}-\theta_{(0)}\\\tilde{\beta}-\beta_{(0)}\end{pmatrix}$ and $\begin{pmatrix}\bar{\theta}-\theta_{(0)}\\\bar{\beta}-\beta_{(0)}\end{pmatrix}-\begin{pmatrix}\tilde{\theta}-\theta_{(0)}\\\tilde{\beta}-\beta_{(0)}\end{pmatrix}$ are asymptotically independent because it can be shown after very extensive algebra that

$$\mathrm{Cov}\left[S^{-1}\begin{pmatrix}\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}\\\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}\end{pmatrix}, V^{-1}L(\theta_{(0)},\beta_{(0)})-S^{-1}\begin{pmatrix}\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}\\\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}\end{pmatrix}\right] = 0.$$

Consequently, there is

$$0 \le \mathrm{Var}\left[\frac{1}{\sqrt{n}}S^{-1}\begin{pmatrix}\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}\\\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}\end{pmatrix} - \frac{1}{\sqrt{n}}V^{-1}L(\theta_{(0)},\beta_{(0)})\right] = \Sigma_{\psi} - \Sigma.$$

This completes the proof of Theorem 2.

Proof of Theorem 3: Since $\mathrm{E}H_{0i}(t) = \rho_i^{-1}A_{1i}(t)$ and $\mathrm{E}H_{3i}(t) = \rho_i^{-1}A_{2i}(t)$ for $i = 0,1,\cdots,I$ and $(\tilde{\theta},\tilde{\beta})$ is strongly consistent, applying a first-order Taylor expansion and Theorem 1 gives, uniformly in t,

$$\tilde{G}_i(t) = \frac{1}{n_i}\sum_{k=1}^{n}\frac{\rho_i\exp[\tilde{\theta}_i + s_i(T_k;\tilde{\beta}_i)]I_{[T_k \le t]}}{1 + \sum_{m=1}^{I}\rho_m\exp[\tilde{\theta}_m + s_m(T_k;\tilde{\beta}_m)]}$$

$$= H_{1i}(t) - H_{0i}^{\tau}(t)(\tilde{\theta}-\theta_{(0)}) - H_{3i}^{\tau}(t)(\tilde{\beta}-\beta_{(0)}) + o_p(\delta_n)$$

$$= H_{1i}(t) - (\mathrm{E}H_{0i}^{\tau}(t), \mathrm{E}H_{3i}^{\tau}(t))\begin{pmatrix}(\tilde{\theta}-\theta_{(0)})\\(\tilde{\beta}-\beta_{(0)})\end{pmatrix}$$

$$-r_{in}(t) + o_p(\delta_n)$$

$$= H_{1i}(t) - \frac{1}{n\rho_i}(A_{1i}^{\tau}(t), A_{2i}^{\tau}(t))S^{-1}\begin{pmatrix}\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\theta}\\\frac{\partial\ell(\theta_{(0)},\beta_{(0)})}{\partial\beta}\end{pmatrix}$$

$$+o_p(n^{-1/2}) - r_{in}(t) + o_p(\delta_n)$$

$$= H_{1i}(t) - H_{2i}(t) + R_{in}(t), \qquad i = 0,1,\cdots,I, \tag{19}$$

where $\delta_n = \|\tilde{\theta}-\theta_{(0)}\| + \|\tilde{\beta}-\beta_{(0)}\|$ and for $i = 0,1,\cdots,I$,

$$r_{in}(t) = (H_{0i}^{\tau}(t) - \mathrm{E}H_{0i}^{\tau}(t), H_{3i}^{\tau}(t) - \mathrm{E}H_{3i}^{\tau}(t))\begin{pmatrix}\tilde{\theta}-\theta_{(0)}\\\tilde{\beta}-\beta_{(0)}\end{pmatrix},$$

$$R_{in}(t) = o_p(n^{-1/2}) - r_{in}(t) + o_p(\delta_n).$$

It follows from part (b) of Theorem 3 that $\delta_n = O_p(n^{-1/2})$. Furthermore, it can be shown that $\sup_{-\infty \le t \le \infty}|r_{in}(t)| = o_p(n^{-1/2})$. As a result, $\sup_{-\infty \le t \le \infty}|R_{in}(t)| = o_p(n^{-1/2})$, which along with (19) establishes (12) and (14). To prove (15), according to (12) and (14), it suffices to show that

$$\sqrt{n}\begin{pmatrix}H_{10} - \hat{G}_0 - H_{20}\\H_{11} - \hat{G}_1 - H_{21}\\\vdots\\H_{1I} - \hat{G}_I - H_{2I}\end{pmatrix} \xrightarrow{D} \begin{pmatrix}W_0\\W_1\\\vdots\\W_I\end{pmatrix} \text{ in } D^{I+1}[-\infty,\infty]. \tag{20}$$

Under the assumption that the underlying distribution function G_0 is continuous (20) is proven. According to (16) and Lemmas 2 and 3, we have for $-\infty \le s \le t \le \infty$,

$$E\{\sqrt{n}[H_{1i}(t) - \hat{G}_i(t) - H_{2i}(t)]\} = 0 = EW_i(t),$$
$$i = 0,1,\cdots,I,$$

$$\text{Cov}(\sqrt{n}[H_{1i}(s) - \hat{G}_i(s)] - \sqrt{n}H_{2i}(s),$$
$$\sqrt{n}[H_{1i}(t) - \hat{G}_i(t)] - \sqrt{n}H_{2i}(t))$$
$$= \text{Cov}(\sqrt{n}[H_{1i}(s) - \hat{G}_i(s)], \sqrt{n}[H_{1i}(t) - \hat{G}_i(t)])$$
$$- \text{Cov}(\sqrt{n}H_{2i}(s), \quad \sqrt{n}H_{2i}(t))$$

$$= \frac{1+\rho}{\rho_i^2}[G_i(s) - B_{ii}(s)]$$

$$- \frac{1+\rho}{\rho_i^2} \sum_{k=0,k\ne i}^{I} \frac{1}{\rho_k} B_{ik}(s)B_{ik}(t)$$

$$- \frac{1+\rho}{\rho_i^3}[G_i(s) - B_{ii}(s)][G_i(t) - B_{ii}(t)]$$

$$- \frac{1}{\rho_i^2}(A_{1i}^\tau(s), A_{2i}^\tau(s))S^{-1}\begin{pmatrix} A_{1i}(t) \\ A_{2i}(t) \end{pmatrix}$$

$$+ \frac{1+\rho}{\rho_i^2} \sum_{k=0,k\ne i}^{I} \frac{1}{\rho_k} B_{ik}(s)B_{ik}(t)$$

$$+ \frac{1+\rho}{\rho_i^3}[G_i(s) - B_{ii}(s)][G_i(t) - B_{ii}(t)]$$

$$= \frac{1+\rho}{\rho_i^2}[G_i(s) - B_{ii}(s)]$$

$$- \frac{1}{\rho_i^2}(A_{1i}^\tau(s), A_{2i}^\tau(s))S^{-1}\begin{pmatrix} A_{1i}(t) \\ A_{2i}(t) \end{pmatrix} = EW_i(s)W_i(t),$$

$$i = 0,1,\cdots,I,$$

$$\text{Cov}(\sqrt{n}[H_{1i}(s) - \hat{G}_i(s)] - \sqrt{n}H_{2i}(s),$$
$$\sqrt{n}[H_{1j}(t) - \hat{G}_j(t)] - \sqrt{n}H_{2j}(t))$$
$$= \text{Cov}(\sqrt{n}[H_{1i}(s) - \hat{G}_i(s)], \sqrt{n}[H_{1j}(t) - \hat{G}_j(t)])$$
$$- \text{Cov}(\sqrt{n}H_{2i}(s), \quad \sqrt{n}H_{2j}(t))$$

$$= -\frac{1+\rho}{\rho_i \rho_j} B_{ij}(s) - \frac{1+\rho}{\rho_i \rho_j} \sum_{k=0}^{I} \frac{1}{\rho_k} B_{ik}(s)B_{jk}(t)$$

$$+ \frac{1+\rho}{\rho_i \rho_j^2} B_{ij}(s)G_j(t) + \frac{1+\rho}{\rho_i^2 \rho_j} G_i(s)B_{ij}(t)$$

$$- \frac{1}{\rho_i \rho_j}(A_{1i}^\tau(s), A_{2i}^\tau(s))S^{-1}\begin{pmatrix} A_{1j}(t) \\ A_{2j}(t) \end{pmatrix}$$

$$+ \frac{1+\rho}{\rho_i \rho_j} \sum_{k=0}^{I} \frac{1}{\rho_k} B_{ik}(s)B_{jk}(t)$$

$$- \frac{1+\rho}{\rho_i \rho_j^2} B_{ij}(s)G_j(t) - \frac{1+\rho}{\rho_i^2 \rho_j} G_i(s)B_{ij}(t)$$

$$= -\frac{1+\rho}{\rho_i \rho_j} B_{ij}(s) - \frac{1}{\rho_i \rho_j}(A_{1i}^\tau(s), A_{2i}^\tau(s))S^{-1}\begin{pmatrix} A_{1j}(t) \\ A_{2j}(t) \end{pmatrix}$$

$$= EW_i(s)W_i(t), \qquad i \ne j = 0, 1, \cdots, I.$$

It then follows from the multivariate central limit theorem for sample means and the Cramer-Wold device that the finite-dimensional distributions of

$$\sqrt{n}(H_{10} - \hat{G}_0 - H_{20}, \cdots, H_{1I} - \hat{G}_I - H_{2I})^\tau$$

converge weakly to those of $(W_0, \cdots, W_I)^\tau$.
Thus, in order to prove (20), it is enough to show that the process

$$\{\sqrt{n}(H_{10}(t) - \hat{G}_0(t) - H_{20}(t), \cdots,$$
$$H_{1I}(t) - \hat{G}_I(t) - H_{2I}(t))^\tau, \quad -\infty \le t \le \infty\}$$

is tight in $D^{I+1}[-\infty, \infty]$. But this has been established by Lemma 4 for continuous G_0. Thus, (20) has been proven when G_0 is continuous.

Suppose now that G_0 is an arbitrary distribution function over $[-\infty, \infty]$. Define the inverse of G_0, or quantile function associated with G_0, by $G_0^{-1}(x) = \inf\{t : G_0(t) \ge x\}$, $x \in (0,1)$. Let $\xi_{i1}, \cdots, \xi_{in_i}$ be independent random variables having the same density function $h_i(x) = \exp[\theta_i + s_i(G_0^{-1}(x); \beta_i)]$ on $(0,1)$ for $i = 0,1,\cdots,I$ and assume that

$\{(\xi_{i1}, \cdots, \xi_{in_i}) : i = 0,1, \cdots, I\}$ are jointly independent. Thus, we have the following $(I+1)$-sample semiparametric model analogous to (2):

$$\xi_{01}, \cdots, \xi_{0n_0} \overset{i.i.d.}{\sim} h_0(x) = I_{(0,1)}(x),$$

$$\xi_{i1}, \cdots, \xi_{in_i} \overset{i.i.d.}{\sim} h_i(x) = \exp[\theta_i + s_i(G_0^{-1}(x); \beta_i)]h_0(x),$$
$$i = 0,1, \cdots, I. \tag{21}$$

Then, it is easy to see that $(X_{i1}, \cdots, X_{in_i})$ and $(G_0^{-1}(\xi_{i1}), \cdots, G_0^{-1}(\xi_{in_i}))$ have the same distribution, i.e.,

$(X_{i1}, \cdots, X_{in_i}) \overset{d}{=} (G_0^{-1}(\xi_{i1}), \cdots, G_0^{-1}(\xi_{in_i}))$ for $i = 0,1, \cdots, I$. Let $\{\psi_1, \cdots, \psi_n\}$ denote the pooled random variables $\{\xi_{01}, \cdots, \xi_{0n_0}; \xi_{11}, \cdots, \xi_{1n_1}; \cdots; \xi_{I1}, \cdots, \xi_{In_I}\}$, then

$(T_1, \cdots, T_n) \overset{d}{=} (G_0^{-1}(\psi_1), \cdots, G_0^{-1}(\psi_n))$. For $u \in (0,1)$ and $m = 0,1, \cdots, I$, let $\tilde{H}_{1m}(u), \tilde{H}_{2m}(u)$, and $\hat{\tilde{G}}_m(u)$ be the corresponding counterparts of $\tilde{H}_{1m}(t)$, $\tilde{H}_{2m}(t)$, and $\hat{\tilde{G}}_m(t)$ under model (21). Now define $\phi : D^{I+1}[-\infty, \infty] \to D^{I+1}[-\infty, \infty]$ by $(\phi K)(t) = K(G_0(t))$, then it can be shown that

$$\sqrt{n}(\tilde{H}_{10}[G_0(t)] - \hat{\tilde{G}}_0[G_0(t)] - \tilde{H}_{20}[G_0(t)], \cdots,$$

$$\tilde{H}_{1I}[G_0(t)] - \hat{\tilde{G}}_I[G_0(t)] - \tilde{H}_{2I}[G_0(t)])^\tau$$
$$\overset{d}{=} \sqrt{n}(H_{10}(t) - \hat{G}_0(t) - H_{20}, \cdots, H_{1I}(t) - \hat{G}_I(t)]$$
$$-H_{2I}(t))^\tau \quad \text{and}$$

$$\sqrt{n}(\tilde{H}_{10} - \hat{\tilde{G}}_0 - \tilde{H}_{20}, \cdots, \tilde{H}_{1I} - \hat{\tilde{G}}_I - \tilde{H}_{2I})^\tau$$
$$\overset{D}{\to} (\tilde{W}_0, \cdots \tilde{W}_I)^\tau$$

in $D^{I+1}[0,1]$, where $(\tilde{W}_0, \cdots \tilde{W}_I)^\tau$ is a multivariate Gaussian process satisfying $\phi[(\tilde{W}_0, \cdots \tilde{W}_I)^\tau] \overset{d}{=} (W_0, \cdots W_I)^\tau$. If K_n converges to K in the Skorohod topology and

$K \in C^{I+1}[-\infty, \infty]$, then the convergence is uniform, so that ϕK_n converges to ϕK uniformly and hence in the Skorohod topology. As a result, Theorem 5.1 of Billingsley (1968, page 30) implies that

$$\sqrt{n}(H_{10} - \hat{G}_0 - H_{20}, \cdots, H_{1I} - \hat{G}_I - H_{2I})^\tau$$
$$\overset{d}{=} \phi[\sqrt{n}(\tilde{H}_{10} - \hat{\tilde{G}}_0 - \tilde{H}_{20}, \cdots, \tilde{H}_{1I} - \hat{\tilde{G}}_I - \tilde{H}_{2I})^\tau]$$
$$\overset{D}{\to} \phi[(\tilde{W}_0, \cdots \tilde{W}_I)^\tau] \overset{d}{=} (W_0, \cdots W_I)^\tau.$$

Therefore, (20) holds for general G_0, and this completes the proof of Theorem 3.

References

Agresti, A. (1990). *Categorical data analysis.* NY: John Wiley & Sons.

Anderson, J. A. (1972). Separate sample logistic discrimination. *Biometrika*, *59*, 19-35.

Anderson, J. A. (1979). Robust inference using logistic models. *International Statistical Institute Bulletin*, 48, 35-53.

Billingsley, P. (1968). *Convergence of probability measures.* NY: John Wiley & Sons.

Day, N. E., & Kerridge, D. F. (1967). A general maximum likelihood discriminant. *Biometrics*, *23*, 313-323.

Farewell, V. (1979). Some results on the estimation of logistic models based on retrospective data. *Biometrika*, *66*, 27-32.

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.

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

Hall, P., & La Scala, B. (1990). Methodology and algorithms of empirical likelihood, *International Statistical Review,58*, 109-127.

Hsieh, D. A., Manski, C. F., & McFadden, D. (1985). Estimation of response probabilities from augmented retrospective observations. *Journal of the American Statistical Association*, 80, 651-662.

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. *Annals of Statistics,* 18, 90-120.

Owen, A. B. (1991). Empirical likelihood for linear models. *Annals of Statistics,* 19, 1725-1747.

Prentice, R. L., & Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66, 403-411.

Qin, J. (1993). Empirical likelihood in biased sample problems. *Annals of Statistics*, 21, 1182-96.

Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika* 85, 619-30.

Qin, J., & Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, 22, 300-325.

Qin, J., & Zhang, B. (1997). A goodness of fit test for logistic regression models based on case-control data. *Biometrika*, *84*, 609-618.

Scott, A. J., & Wild, C. J. (1997). Fitting regression models to case-control data by maximum likelihood. *Biometrika*, *84*, 57-71.

Sen, P. K., & Singer, J. M. (1993). *Large sample methods in statistics: an introduction with applications*. NY: Chapman & Hall.

van der Vaart, A. W., & Wellner, J. A. (1996). *Weak convergence and empirical processes with applications to statistics.* NY: Springer.

Vardi, Y. (1982). Nonparametric estimation in presence of length bias. *Annals of Statistics*, 10, 616-620.

Vardi, Y. (1985). Empirical distribution in selection bias models, *Annals of Statistics*, *13*, 178-203.

Wacholder, S., & Weinberg, C. R. (1994). Flexible maximum likelihood methods for assessing joint effects in case-control studies with complex sampling. *Biometrics*, *50*, 350-357.

Weinberg, C. R., & Wacholder, S. (1990). The design and analysis of case-control studies with biased sampling. *Biometrics*, *46*, 963-975.

Weinberg, C. R., & Sandler, D. P. (1991). Randomized recruitment in case-control studies. *American Journal of Epidemiology*, *134*, 421-433.

Weinberg, C. R., & Wacholder, S. (1993). Prospective analysis of case-control data under general multiplicative-intercept risk models. *Biometrika*, 80, 461-465.

Zhang, B. (2000). A goodness of fit test for multiplicative-intercept risk models based on case-control data. *Statistica Sinica*, *10*, 839-865.

Zhang, B. (2002). Assessing goodness-of-fit of generalized logit models based on case-control data. *Journal of Multivariate Analysis*, *82*, 17-38.