Title	PROPORTIONAL ODDS MODEL WITH LOG-CONCAVE DENSITY ESTIMATION
Manuscript ID	SS-2017-0465
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0465
Complete List of Authors	Jinsong Chen George R. Terrell Inyoung Kim and Martha L. Daviglus
Corresponding Author	Jinsong Chen
E-mail	jinsongc@uic.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

PROPORTIONAL ODDS MODEL WITH LOG-CONCAVE DENSITY ESTIMATION

Jinsong Chen[1], George R. Terrell[2], Inyoung Kim[2] and Martha L. Daviglus[1,3]

[1] *University of Illinois at Chicago,* [2] *Virginia Tech University, and* [3] *Northwestern University*

Abstract: We add a log-concave qualitative constraints on the baseline distribution for proportional odds model. A full maximum likelihood method is developed for joint estimation of both regression parameters and densities. Asymptotic properties of the estimates are established. A likelihood ratio test is constructed to test the significance of regression parameter. We also propose a Kolmogorov-Smirnov type test to assess the log-concavity of the baseline distribution. A simulation study and application to data from the Chicago Healthy Aging Study show usefulness of our method.

Key words and phrases: Density ratio model; Exponential tilting; Semiparametric method; Shape constrained estimation; Survival analysis.

1. Introduction

The density ratio model or exponential tilt model is useful in modeling treatment effects, the biased-sampling problem, and distribution of a mix of discrete and continuous variables (Cheng, Qin, and Zhang, 2009; Qin, 1998; Terrell, 2003; Zou, Fine, and Yandell, 2002; Chen, 2007). Cheng and Chu (2004), and Fokianos (2004) show that the density estimation combining data from all samples under density ratio model is more efficient than traditional estimation based on separated samples. Luo and Tsai (2012), and Diao, Ning, and Qin (2012) generalize the density ratio model to proportional likelihood ratio model with incorporation of covariates:

$$f(y|\mathbf{x}) = \frac{dF_0(y)\exp(y\mathbf{x}^T\boldsymbol{\beta})}{\int \exp(y\mathbf{x}^T\boldsymbol{\beta})dF_0(y)}, \tag{1.1}$$

where $F_0(\cdot)$ is the baseline distribution for response y, and $\mathbf{x}$ and $\boldsymbol{\beta}$ are linear covariates and coefficients vectors respectively. Density estimation for the above models is important, e.g., describing the distributional difference of outcome among groups. To our knowledge, the current literature for model (1.1) treats the baseline distribution as a nuisance, estimated only empirically. We generalize the functional form of regression part in model (1.1), and propose estimation of both baseline density and regression parameters in this paper.

Moreover, we impose a log-concave qualitative constraint on the baseline density for model (1.1). That means the baseline density is $p(y) = \exp\varphi(y)$ for some concave function $\varphi : \mathbb{R} \to [-\infty, \infty)$. It is known the density estimation without any constraint is not efficient since the parameter space is too large. A popular approach is smoothing methods, e.g., kernel density estimation and estimation based on roughness penalization. Estimation of model (1.1) involves iteratively updating estimate of the nonparametric or parametric components conditioning on the others. Using smoothing methods will results in an intensive computational burden since the optimal smoothing parameter needs to be selected at each iteration step. As an useful alternative, the log-concave density estimation is an automatic nonparametric estimation avoiding the problem of selecting tunning parameters. The univariate log-concave density estimation also possess the same minimax rate of order $n^{-4/5}$ as that of estimation of density with two bounded derivatives (Ibragimov and Khas'minskii, 1983; Seregin and Wellner, 2010; Kim and Samworth, 2016). Thus, comparing with traditional approaches involving tunning parameter selection, our proposed approach has computational advantages without the loss of asymptotic efficiency.

The well studied log-concave densities include most of the commonly used parametric distributions (Walther, 2009), e.g., uniform, normal, logistic, chi-square, chi, gamma, beta, and Weibull distributions. Log-concave shape constrained estimation has practical applications in econometric modeling, reliability theory, and estimation of monotonic hazard rate (Bagnoli and Bergstrom, 2005; Barlow and Proschan, 1975; Hall et al., 2001). The nonparametric maximum likelihood estimation of a unimodal density does not exist (Birge, 1997). The nonparametric maximum likelihood estimation of a log-concave density exists, and may be used instead of the larger class of unimodal densities (Dumbgen and Rufibach, 2009).

Maximum likelihood estimation of a multidimensional log-concave density is shown to have smaller mean integrated squared error compared with kernel-based methods for the moderate to large sample sizes (Cule and Samworth, 2010). In addition, we also gain finite sample efficiency for regression parameter estimates by imposing a correct log-concavity constraint on baseline density estimation for model (1.1) comparing with the empirical estimates without shape constraint. An important questions is that: how confident are we that the shape-constraint is correct when we do not priori know the true distribution in practical applications? In other words, it is critical to have a diagnosis of the log-concavity property for baseline distribution. For example, Walther (2002)'s method is equivalent to test whether a parameter c is equal to zero. It is computational expensive since it requires many bootstrap estimates based on a set of values of c. Cule and Samworth (2010) introduce a permutation test, and Hazelton (2010) proposes a test using kernel density estimation. However, theoretical supports are desired for these two proposals. Chen and Samworth (2013) develop a test based on smoothed log-concave density estimates. All these methods only consider the test of log-concavity for marginal density estimation without incorporating of the covariates. We propose a Kolmogorov-Smirnov type test to assess the log-concavity of the baseline distribution, which is shown to be consistent.

The rest of the paper is organized as follows. The model and estimation method are introduced in Section 2. Section 3 describes the asymptotic properties of the estimates and a test of log-concavity of the baseline distribution. The results of simulation studies and an application to data from the Chicago Healthy Aging Study are presented in Section 4. A summary of our conclusions and implications of this work are presented in the final section.

2. Models and Methods

Let the random vector Y follows distribution P_Y on a given set $\mathcal{Y} \subseteq \mathbb{R}$, P_Y has a density p_Y in $\mathcal{Y}$, and $p_Y \in \mathcal{P}_c$ for a log-concave class of probability densities $\mathcal{P}_c$. The random vector $\mathbf{X}$ follows distribution P_X on a given set $\mathcal{X} \subseteq \mathbb{R}^k$. Our conditional model of interest is

$$f(y; \mathbf{x}, \boldsymbol{\beta}, p) = \frac{p(y)e^{\eta(y,\mathbf{x}|\boldsymbol{\beta})}}{\int p(y)e^{\eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta})}dy} \tag{2.1}$$

where $\eta(y, \mathbf{x}|\boldsymbol{\theta})$ is a parametric regression function depending on parameters $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ for $\boldsymbol{\Theta} \subseteq \mathbb{R}^{1\times k}$, and the baseline density p is log-concave. A simple form for $\eta(y, \mathbf{x}|\boldsymbol{\theta})$ is linear $y\mathbf{x}^T\boldsymbol{\beta}$ as in model (1.1). However, it can be specified using other parametric forms to accommodate various applications, e.g., transformed linear form $H_Y(y)H_X(\mathbf{x})^T\boldsymbol{\beta}$ for known functions H_X and H_Y. We name model (2.1) the proportional odds model with log-concave distribution (POML) because the proportionality between conditional odds: $\{f(y_1|\mathbf{x}_1)/f(y_1|\mathbf{x}_2)\}/\{f(y_2|\mathbf{x}_1)/f(y_2|\mathbf{x}_2)\} = \exp\{\eta(y_1,\mathbf{x}_1|\boldsymbol{\beta}) - \eta(y_1,\mathbf{x}_2|\boldsymbol{\beta}) + \eta(y_2,\mathbf{x}_2|\boldsymbol{\beta}) - \eta(y_2,\mathbf{x}_1|\boldsymbol{\beta})\}$.

In addition to the nice properties of model (2.1) described in literature (Rathouz and Gao, 2009; Luo and Tsai, 2012), we will address the relationships between model (2.1) and shape-constrained survival analysis, and generalized models with random component under shape constraint. Distributions under shape constraint on the hazard rate are of considerable practical interest (Hall et al., 2001; Qin et al., 2011). Since it imposes a log-concave constraint, POML might be utilized to model the monotonic hazard rate (Dumbgen and Rufibach, 2009) for complete data using $h(y;\mathbf{x},\boldsymbol{\beta}) = f(y;\mathbf{x},\boldsymbol{\beta})/\{1-F(y;\mathbf{x},\boldsymbol{\beta})\}$. Although challenging, POML for censored data can be estimated via EM type algorithm (Cheng, Qin, and Zhang, 2009; Shen, Jing, and Qin, 2012). Rathouz and Gao (2009) extended generalized linear model with density estimation for categorical response via exponential tilting. POML can be represented as generalized model with canonical link function and an additional log-concave constraint on random component.

Denote $P_{X,Y}$ the joint distribution of $(Y, \mathbf{X})$. The likelihood function for $(\boldsymbol{\beta}, p)$ is

$$L_P(\boldsymbol{\beta}, p) = \int \log p(y) dP_Y + \int \eta(y, \mathbf{x}|\boldsymbol{\beta}) dP_{X,Y} - \int \left[\log \int \exp\{\eta(y, \mathbf{x}|\boldsymbol{\beta})\} p(\mathbf{y}) d\mathbf{y}\right] dP_X. \tag{2.2}$$

The maximum likelihood estimators (MLE) satisfy: $(\hat{\boldsymbol{\beta}}_n, \hat{p}_n) = \underset{\boldsymbol{\beta}, p \in \mathcal{P}_c}{\arg\max} L_{\mathbb{P}}(\boldsymbol{\beta}, p)$, where $\mathbb{P}$ denote the empirical distribution.

Let $(y_{(1)}, \ldots, y_{(k)})$ be the observed ordered distinct response with corresponding observed frequencies $(m_1, \ldots, m_k)$, and vector $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_k)$ for $\varphi_i = \log p(y_{(i)})$. We propose an iterative procedure for simultaneous estimation of parametric and nonparametric components in POML as followed:

Initiation: give initial values $\hat{\boldsymbol{\beta}}$, which may come from an educated guess in practical application. In our study, we choose the initial values from null space, $\hat{\boldsymbol{\beta}} = \mathbf{0}$.

Density Estimation: update φ using

$$\hat{\varphi} = \underset{\exp(\varphi) \in \mathcal{P}_c}{\arg\max} \left[\sum_{i=1}^{n} \eta(y_i, \mathbf{x}_i|\hat{\boldsymbol{\beta}}) + \sum_{l=1}^{k} m_l \varphi_l - \sum_{i=1}^{n} \log \int \exp\{\varphi(y) + \eta(y, \mathbf{x}_i|\hat{\boldsymbol{\beta}})\} dy\right]. \tag{2.3}$$

Regression Parameter Estimation: update $\boldsymbol{\beta}$ using

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \left[\sum_{i=1}^{n} \eta(y_i, \mathbf{x}_i | \boldsymbol{\beta}) + \sum_{l=1}^{k} m_l \hat{\varphi}_l - \sum_{i=1}^{n} \log \int \exp\{\hat{\varphi}(y) + \eta(y, \mathbf{x}_i | \boldsymbol{\beta})\} dy \right].$$

Iteration: iterate for density and regression parameter estimation until convergence.

The estimation based on conditional density is an optimization problem with a nonlinear objective function and a concave constraint on the result. We aim to estimate the baseline density function $f(\cdot) = dF(\cdot)/d\mu$ nonparametrically for Lebesgue measure μ, or equivalently $\varphi(\cdot) = \log(f)$, and parameters $\boldsymbol{\beta}$. Estimating the vector $\boldsymbol{\varphi}$ is sufficient since the nonparametric maximum likelihood estimate of log-concave density exists and is a piecewise linear continuous function with knots on the observation points (Dumbgen and Rufibach, 2009). The iterative convex minorant algorithm (Groeneboom and Wellner, 1992) and active set algorithm (Fletcher, 1987) have been used to estimate the marginal log-concave density. As discussed by Dumbgen, Husler and Rufibach (2011), the likelihood function in expression (2.3) is an infinitely often differentiable and strictly concave on $\mathbb{R}^k$. We extend the active set algorithm by Dumbgen, Husler and Rufibach (2011) to maximize likelihood (2.3) for conditional density estimation. The term "active set" refer to the set of knots where the slope changes in a continuous piecewise linear function. Essentially, the active set algorithm involve two iterative procedures: updating the active set, and updating the density estimate within the active set. We also tried a gradient method for the conditional log-concave density estimation, and found it to be computationally inefficient for large sample size. The estimation of regression parameters is a nonlinear optimization problem and can be maximized using a Newton-Raphson type algorithm.

Our MLE approach is different from the empirical likelihood approach in literature (Luo and Tsai, 2012; Diao, Ning, and Qin, 2012) in terms of estimation of baseline distribution. The empirical likelihood approach only provides empirical estimate for distribution function, i.e., stepwise function with jump at data points. The likelihood approach for POML provides estimated density function with log-concave shape constraint. Qin and Zhang (2005) develop useful kernel density estimation under density ratio model. However, the kernel density estimation relies on empirical likelihood estimates, i.e., they first have estimates of regression parameters and empirical estimate $\tilde{F}_0$ of baseline distribution function F_0, then the kernel estimator of density $\hat{f}_0$ is obtained by smoothing the increment in $\tilde{F}_0$.

3. Inferential Results

In this section, we consider the asymptotic properties of estimates of regression parameters and baseline distribution. We also propose a log-likelihood ratio test for hypothesis about regression parameters $\boldsymbol{\beta}$, and a Kolmogorov-Smirnov type test for assessing the log-concavity of the baseline density.

To build the theoretical consequence, we make the followed assumptions:

A. The true parameters $(\boldsymbol{\beta}_0, p_0)$ maximize $L_P(\boldsymbol{\beta}, p)$, and the Kullback-Leibler information exists and is finite, i.e,

$$E_0 \left\{ \left| \log \frac{f(y; \mathbf{x}, \boldsymbol{\beta}, p)}{f(y; \mathbf{x}, \boldsymbol{\beta}_0, p_0)} \right| \right\} < \infty$$

where E_0 denote the expectation under $P_{X,Y}$;

B. The domains of P_Y and P_X are compact in Euclidean space;

C. The parameter space $\boldsymbol{\Theta}$ is convex compact. The function $\eta(y, \mathbf{x}|\boldsymbol{\beta})$ is a parametric continuous differentiable function in terms of $\boldsymbol{\beta}$. The parameter $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ is identifiable from $\eta(\mathbf{y}, \mathbf{x}|\boldsymbol{\beta})$;

D. The information matrix $-\frac{\partial^2 E[L_P(\boldsymbol{\beta},p)]}{\partial \boldsymbol{\beta}^2}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ is positive definite;

E. The true log-concave density function p_0 is continuously differentiable.

The assumption A is required for consistency. The condition B is a general regularity condition to apply asymptotic theorem for large sample properties. The identification condition in assumption C is a basic criterion. The theoretical results might not succeed if this condition is weak. The conditions D and E are need for deriving the $\sqrt{n}$ consistency and asymptotic normality of regression parameter estiamtes $\hat{\boldsymbol{\beta}}$.

3.1 Consistency: Multi-dimensional Log-concave Distribution

In this paper, we focus on the case $Y \subseteq \mathbb{R}$. The estimation for POML with multivariate log-concave baseline distribution is challenging, and should be addressed elsewhere (Cule and Samworth, 2010). However, we found that the MLE of POML with $\mathbf{Y} \subseteq \mathbb{R}^d$ is theoretically consistent, and show it here for generality.

Let the random vector $\mathbf{Y}$ follows distribution P_Y on a given set $\mathcal{Y} \subseteq \mathbb{R}^d$, and P_Y has a density p_Y in a log-concave class $\mathcal{P}_c$ of probability density on $\mathcal{Y}$. Let $h(p,q)$ denote the Hellinger distance between two probability measures with densities p and q with respect to Lebesgue measure on $\mathbb{R}^d$: $h^2(p,q) = 1/2 \int (\sqrt{p} - \sqrt{q})^2 d\boldsymbol{\mu} = 1 - \int \sqrt{p.q} d\boldsymbol{\mu}$. Denote the joint density:

$$g_{\boldsymbol{\beta},p} = \frac{p(\mathbf{y}) e^{\eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta})}}{\int p(\mathbf{y}) e^{\eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta})} d\mathbf{y}} p_x(\mathbf{x}).$$

It can be seen that the likelihood (2.2) is $\log g_{\boldsymbol{\beta},p}$ with $p_x(\mathbf{x})$ left out since it does not involve $(\boldsymbol{\beta}, p)$.

Lemma 1: *The Hellinger distance satisfy* $h^2(g_{\boldsymbol{\beta},p}, g_{\boldsymbol{\beta}_0,p_0}) \geq a h^2(p, p_0)$ *for a positive constant* a.

Proof: see Appendix.

The consistency of the MLE of a log-concave density on $\mathbb{R}$ with respect to the Hellinger metric is established by Pal, Woodroofe, and Meyer (2007), and the uniform consistency is shown by Dumbgen and Rufibach (2009). Both the Hellinger consistency (Seregin and Wellner, 2010) and the uniform consistency (Cule and Samworth, 2010; Schuhmacher, Husler, and Duumbgen, 2011) of MLE for multivariate log-concave density on $\mathbb{R}^d$ are established. Dumbgen, Samworth and Schuhmacher (2011) present the consistency of MLE for multivariate log-concave distribution in terms of total variation distance for regression model. We first establish the connection between joint and baseline densities in terms of Hellinger distance under POML in Lemma 1. It implies that $h^2(p, p_0) = 0$ if $h^2(g_{\boldsymbol{\beta},p}, g_{\boldsymbol{\beta}_0,p_0}) = 0$. Then, we will be able to show that the estimates of both baseline density and regression parameters are consistent in the following Theorem. Specifically, the estimate of baseline density is Hellinger consistent.

Theorem 1: *Under assumptions A-C, the sequence of MLEs* $(\hat{\boldsymbol{\beta}}_n, \hat{p}_n) = \arg\max_{\boldsymbol{\beta}, p \in \mathcal{P}_c} L_{\mathbb{P}}(\boldsymbol{\beta}, p)$ *satisfy:* $\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}_0$ *and* $h(\hat{p}_n, p_0) \to 0$ *as* $n \to \infty$. *Moreover,* $\hat{p}_n \to p_0$ *pointwise and the convergence is uniform on compact space.*

Proof: see Appendix.

3.2 **Asymptotic Normality: One-dimensional Log-concave Distribution**

In this section, we establish the asymptotic normality for the estimates of $\boldsymbol{\beta}$ for the case $Y \subseteq \mathbb{R}$. In the first, let us introduce the concept of bracketing number with Hellinger metric specific to our problem (van der Vaart and Wellner, 1996). A ε-bracket is a bracket $[g^L, g^U]$ with $h(g^L, g^U) < \varepsilon$, where the bracket $[g^L, g^U]$ is the set of all functions g with $g^L \leq g \leq g^U$ and $g \in \mathcal{G}$. The bracketing number $N_{[\,]}(\varepsilon, \mathcal{G}, h)$ is the minimum number of ε-brackets needed to cover $\mathcal{G}$. The logarithm of the bracketing number is generally referred as entropy with bracketing.

Lemma 2: *Let* $\mathcal{G}_\delta = \{g_{\boldsymbol{\beta},p} : h(p, p_0) < \delta, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \delta\}$ *and* $h_0^2(p, q) = h^2(p + p_0, q + p_0)$ *for* $\delta > 0$. *There is a constant* $C > 0$ *such that*

$$\log N_{[\,]}(\varepsilon, \mathcal{G}_\delta, h_0) \leq c(\varepsilon^{-1/2}),$$

for ε *small enough, and a constant* c.

Proof: see Appendix.

The bounds for the metric entropy with bracketing for the class of log-concave density determine the global rate of convergence of MLE. Doss and Wellner (2016) obtain the bound of entropy with respect to Hellinger metric to be of the order $O(\varepsilon^{-1/2})$ for MLEs of univariate log-concave densities, and find the rate of convergence to be $O(n^{-2/5})$. Similarly, we establish the entropy of the joint density of our interest in Lemma 2. In the following

Lemma, we establish the rate of convergence of baseline density conditioning on the convergence rate of regression parameters. It implies that the baseline density estimation achieves the optimal rate if the estimates of regression parameters behavior reasonable well.

Lemma 3: *Let* $\hat{p}_{\tilde{\boldsymbol{\beta}}} = \text{argmax}_{p \in \mathcal{P}_c} L_{\mathbb{P}}(\tilde{\boldsymbol{\beta}}, p)$, *and the Assumptions C and D hold, then*

$$h(\hat{p}_{\tilde{\boldsymbol{\beta}}}, p_0) \leq O(n^{-2/5} + ||\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||)$$

Proof: see Appendix.

Utilizing the technique of profile likelihood (Murphy and van der Vaart, 2000) and the results of Lemma 3, we are able to prove the asymptotic normality for the estimates of $\boldsymbol{\beta}$.

Theorem 2: *Under Assumptions A-E,* $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *is asymptotic normal with mean 0 and covariance matrix* $\tilde{I}_0$.

Proof: see Appendix for proof and details of $\tilde{I}_0$.

The establishment of the asymptotic normality when $\mathbf{Y} \subseteq \mathbb{R}^d$ presents non-trival technical challenges, and should be studied in future research. The convergence rate shown in Lemma 3 is used to establish the $\sqrt{n}$ asymptotic normality in Theorem 2. If the baseline density is multivariate for

POML (i.e., $\mathbf{Y} \subseteq \mathbb{R}^d$), the derivation of convergence rate require additional technique difficulties. A first step might be to extend Lemma 2 to have an entropy bound for multivariate case. Kim and Samworth (2016) show that the minimax lower bound rate for Hellinger loss is $n^{-1/(d+1)}$ for $d \geq 2$ for log-concave density estimation without covariates involved. If these convergence rates can be established for baseline density estimation in POML, we conjecture that the asymptotic normality for $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ can be achieved for $d = 2, 3$. But for $d > 4$, the $\sqrt{n}$ asymptotic normality might not be satisfied, and we might need to find alternative estimators to achieve the optimal rates of convergence.

3.3 Inference for Regression Parameters

Since the estimates are obtained using the maximum likelihood estimation procedure, it is natural to use the log-likelihood ratio test to test the regression parameters. Likelihood ratio inference proceeds by fitting a series of reduced models which are nested. This means that each reduced model in the sequence is contained within the previous one. Denote the hypothesis of interest as $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_{\o}$. The testing technique can be represented based on profile likelihood since we are interested in testing the low-dimensional parameter $\boldsymbol{\beta}$ instead of the high-dimensional paramter p. Define profile likelihood as $pL(\boldsymbol{\beta}) = L\{\boldsymbol{\beta}, p(\boldsymbol{\beta})\}$, where $p(\boldsymbol{\beta}) = \mathrm{argmax}_p L(\boldsymbol{\beta}, p)$ and $L(\boldsymbol{\beta}, p)$

is the full likelihood. The difference between the reduced model and the full model with no restriction on $\boldsymbol{\beta}$ can be examined by calculating the profile likelihood ratio test statistic $G = 2\{pL(\hat{\boldsymbol{\beta}}) - pL(\boldsymbol{\beta}_{\o})\}$. The asymptotic distribution of this test statistic is presented in the following Lemma.

Lemma 4: *Under Assumption A-E and null hypothesis* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_{\o}$,

$$2\{pL(\hat{\boldsymbol{\beta}}) - pL(\boldsymbol{\beta}_{\o}) \rightarrow \chi^2_u \qquad \text{in } \textit{distribution}$$

where χ^2_u *is a chi-squared distribution with* u *degree of freedom equal to the difference between the number of parameters specified under the reduced model and the full model.*

Proof: This is immediate follow of Corollary 2 in Murphy and van der Vaart (2000).

Substituting the estimator $\hat{\boldsymbol{\beta}}$ into the respective score vector and information matrix of $\boldsymbol{\beta}$, the covariance matrix of $\hat{\boldsymbol{\beta}}$ can be obtained from the sandwich estimator:

$$\left\{ -\frac{\partial^2 \mathrm{pL}_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\bigg|_{\hat{\boldsymbol{\beta}}} \right\}^{-1} \left[\sum_{i=1}^n \left\{ \frac{\partial \mathrm{pL}(y_i, \mathbf{x}_i, \boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big|_{\hat{\boldsymbol{\beta}}} \right\} \left\{ \frac{\partial \mathrm{pL}(y_i, \mathbf{x}_i, \boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big|_{\hat{\boldsymbol{\beta}}} \right\}^T \right] \left\{ -\frac{\partial^2 \mathrm{pL}_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\bigg|_{\hat{\boldsymbol{\beta}}} \right\}^{-1}.$$

The explicit analytical expression of gradient and hessian for the profile likelihood $\mathrm{pL}(\boldsymbol{\beta})$ is very complicated. In practical applications, we can reply on numerical derivatives to have variance estimates.

3.4 Assess Log-concavity of Baseline Density

A essential assumption for POML (2.1) is the log-concave shape constraint on the baseline density, which could be violated in practical application. It is important to have an inferential tool to diagnose the appropriateness of log-concavity. Many tests have been developed in literature. Walther (2002) propose that the mixture of log-concave densities can be represented as $\exp\{\phi(y) + c||y||^2\}$ for a concave function ϕ and constant $c \geq 0$. The test for a log-concave distribution is equivalent to testing whether $c = 0$. A limitation of this approach is that it is only practical for small sample sizes because the computation of the test statistics requires construction of many bootstrap samples. Cule and Samworth (2010) present a permutation test with easy implementation but paid for the price of less power. The test by Hazelton (2010) is based on choosing the smallest bandwidth for kernel density estimate with log-concavity satisfied. An extension of this test to model (2.1) will results in an excessive computational burden since it is nontrival to find optimal kernel estimates, as we discussed in introduction. Moreover, there is a lack of theoretical supports for the tests utilizing kernel density or permutation.

Motivated by the foregoing described works, we aim to develop a test of log-concavity for baseline distribution not only computationally feasible but also theoretically supported. In what follows, we present a Kolmogorov-Smirnov type test to exam the log-concave assumption. The test statistic is essentially a distance between shape constrained and non-constrained MLEs of the baseline distribution in terms of a uniform metric. Denote by $\mathcal{F}_c$ the family of distribution whose density is log-concave, and the MLEs $(\hat{\boldsymbol{\beta}}_n, \hat{F}_n) = \arg\max_{\boldsymbol{\beta}, F \in \mathcal{F}_c} \mathbb{P}_n\{l(\boldsymbol{\beta}, F)\}$, where

$$l(\boldsymbol{\beta}, F) = dF(y) + \eta(\mathbf{y}, \mathbf{x}|\boldsymbol{\beta}) - \log \int \exp(y\mathbf{x}^T\boldsymbol{\beta})dF(y).$$

Let $(\tilde{\boldsymbol{\beta}}, \tilde{\mathbb{F}}_n)$ maximize the empirical likelihood without shape constraint. Luo and Tsai (2012) demonstrate that the empirical likelihood estimates is both computationally and asymptotically efficient. We may test the log-concavity using the test statistic $T_n = \sqrt{n}\|\hat{F}_n(y) - \tilde{\mathbb{F}}_n(y)\|_\infty$, where $\|\cdot\|_\infty$ is the supernorm.

Since the distribution of T_n is very difficulty to derive, we propose a bootstrap testing procedure as follows:

1) Obtain the shape-constraint estimates $(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)$ and empirical estimates $(\tilde{\boldsymbol{\beta}}, \tilde{\mathbb{F}}_n)$ for data $\{Y, \mathbf{X}\}$, and calculate T_n;

2) Using bootstrap to sample data $\{Y^*, \mathbf{X}\}$ from the null distribution $f(y; \mathbf{x}, \hat{\boldsymbol{\beta}}_n, \hat{F}_n)$, obtain shape constrained MLEs $(\hat{\boldsymbol{\beta}}^*, \hat{F}_n^*)$ and empirical estimates $(\tilde{\boldsymbol{\beta}}^*, \tilde{\mathbb{F}}_n^*)$ without shape-constraint, calculate T^*;

3) Repeat the bootstrap process N times;

4) Compute the upper α level critical value ξ_α from T^* and reject the null hypothesis if $T_n > \xi_\alpha$.

We investigate the asymptotic property of our proposed Kolmogorov Smirnov type test under the alternative hypothesis, i.e., the log-concave shape constraint is violated. The results show the test based on proposed bootstrap procedure is consistent.

Theorem 3: *Under Assumption A-C, if the true distribution $F_0 \notin \mathcal{F}_c$, then $P(T_n > \xi_\alpha) \to 1$.*

Proof: see Appendix.

4. Numerical Studies

4.1 Simulation

We conduct simulation study to assess the performance of our methods. The data are generated from the following POML with linear regression function:

$$f(y; x_1, x_2, \beta_1, \beta_2, p) = \frac{p(y)e^{y\beta_1 x_1 + y\beta_2 x_2}}{\int p(y)e^{y\beta_1 x_1 + y\beta_2 x_2} dy} \tag{4.1}$$

with $x_1 \sim \text{Binomial}(1, 0.5)$ under the following four settings: I. $p(y) \sim \text{N}(0,1)$, $x_2 \sim \text{N}(0,1)$, $\beta_1 = 0$, and $\beta_2 = 0$; II. $p(y) \sim \text{N}(0,1)$, $x_2 \sim \text{N}(0,1)$, $\beta_1 = 1$, and $\beta_2 = 0.5$; III. $p(y) \sim \text{Exponential}(1)$, $x_2 \sim \text{Exponential}(1)$, $\beta_1 = 0$, and $\beta_2 = 0$; IV. $p(y) \sim \text{Exponential}(1)$, $x_2 \sim \text{Exponential}(1)$, $\beta_1 = -1$, and $\beta_2 = -0.5$.

For each setting, we generate 500 datasets, and fit data using POML with log-concave density estimation. The bias, standard deviation, meas squared error (MSE), and coverage probability of 95% confidence interval of estimated regression parameters are shown in Table 1. This table also represents the empirical rejection rates for a significance level of 0.05 using likelihood ratio test G. The mean estimated density functions are shown in Figure 1. In summary, our method does a reasonably effective job in providing accurate estimates of both regression parameters and density functions. The proposed likelihood ratio test is an adequate tool for testing the significance of regression parameters.

For the purpose of comparing with existing similar methods, we also fit data using the empirical likelihood approach of Luo and Tsai (2012). Define the relative efficiency as $RE = MSE_{POML}/MSE_{EL}$ where MSE_{POML} is the MSE for estimate using our method and MSE_{EL} is the MSE using empirical likelihood approach. Figure 2 shows the REs for estimating regression parameters (β_1, β_2). We can see that the MLE with log-concave density constraint has smaller MSE comparing with empirical likelihood estimates without shape constraint for moderate to large sample sizes. This might not be surprising that the addition of an appropriate shape constraint on the distribution gains some efficiency in finite sample performance. As discussed by Cule and Samworth (2010), the poor approximation of the convex hull of the data to the support of the underlying density results in relative poor performance of the log-concave maximum likelihood estimator for small sample size. In small sample cases, the less desirable baseline density estimates affect the quality in estimating regression parameter as well.

Another simulation is conduced to evaluate the behavior of our proposed test for the log-concave constraint. The data are generated from model (4.1) with mixture normal baseline distribution $0.5\mathrm{N}(\mu_1, 1)+0.5\mathrm{N}(\mu_2, 1)$. For simplicity, we only consider a binary predictor $x \sim$ binomail$(1, 0.5)$ with regression coefficient $\beta = 1$. The baseline mixture $p(\cdot)$ has three settings: $(\mu_1 = 0, \mu_2 = 0)$, $(\mu_1 = 0, \mu_2 = 2)$, and $(\mu_1 = 0, \mu_2 = 4)$. It is well known that the log-concavity is satisfied when $|\mu_2 - \mu_1| \leq 2$. 200 datasets are generated for each setting with two different sample size (N=100 or N=200). Within each dataset, 100 bootstrap samples are generated to obtain the critical value for test statistics. As shown in Table 2, our proposed testing procedure perform appropriately in terms of type I error rate and power even for relatively small sample size.

4.2 **Chicago Healthy Aging Study**

As described in Introduction Section, the Chicago Healthy Aging Study (CHAS) is a re-examination of a sample of 1395 surviving participants (ages 65-84, 28% female) from the Chicago Heart Association Detection Project in Industry 1967-1973 cohort (CHA) (Pirzada et al., 2013). Their cardiovascular disease (CVD) risk profiles were originally ascertained at ages 25-44. This study re-examined 421 participants who were low-risk (LR) and 974 participants who were not-LR at baseline. LR is defined as having favorable levels of five major CVD risk factors: serum total cholesterol <200 mg/dL and not taking cholesterol-lowering medication; blood pressures $\leq 120/\leq 80$ mmHg and not taking anti-hypertensive medication; BMI <25 (mass(kg)/$\{(\text{height(m)}\}^2$); not smoking; and no history of diabetes or heart attack. In CHAS study, LR and not-LR CHA participants were randomly selected from the 12,119 surviving original CHA participant, in which there are 1034 LR and 11085 not-LR individuals at baseline. An issue of biased sampling arose since baseline LR participants were over-sampled to obtain adequate samples for between-group comparisons. In addition, the CHAS participants tended to be in healther status compared to the CHA participants who are not selected for CHAS.

Although the importance of the LR status in overcoming the CVD epidemic is often recognized, the long-term association of LR status at a younger age with objectively measured health in older age has not been examined (Daviglus et al., 2004). We divide the CHAS participants into four groups: LR, 0 RF, 1 RF, 2+ RF. The 0 RF, 1 RF, 2+ RF refer to have 0, 1, and ≥ 2 of the five adverse CVD risk factors respectively (Daviglus et al., 2016). POML is applied to study the distributional difference of six minutes walking distance (in feet) between LR and not-LR participants defined at baseline. For illustration purpose, we only consider four discrete predictors, binary indicators of risk group (0 RF, 1 RF, and 2+ RF), gender, and a continuous predictor age. The estimated regression coefficients for 0 RF, 1 RF, and 2+ R groups are -0.056 (pvalue=0.583), -0.339 (pvalue$<.001$), and -0.627 (pvalue$<.001$) respectively. The estimated coefficients for male and age are 0.848 (pvalue$<.001$) and -0.368 (pvalue$<.001$) respectively. The test statistics for testing log-concavity is $T_n = 0.088$ with $\#\{b : T_n > T_b^*\}/100 = 0.4$, where T_b^* for $b = 1, \ldots, 99$ is calculated from 99 bootstrap samples following our proposed bootstrap procedure. Consequently, we fail to reject the hull hypothesis that the baseline distribution is log-concave. The plots of estimated densities and cumulative functions for four risk groups are shown in Figure 3. The estimated six minutes walking distance of participants in LR and 0 RF groups cluster around 1760 feet, while those of participants in 1 RF and 2+ RF groups cluster around 1640 feet. The estimated densities using our proposed method clearly capture the left skewness, and provide insightful information about the distributional difference of six minutes walking distance for individuals in each groups. Comparing with individuals without risk factor in younger age, the results imply that the individuals who have risk factors in younger age and have survived to an older age will have shorter six minutes walking distance after adjustment for gender and age.

5. Discussion

We propose a log-concave shape constraint on the baseline density function for POML. It allows for modeling a variety of distributions. We present a method of maximum likelihood estimation to jointly estimate both regression parameters and densities. The asymptotic properties including consistency and normality of the estimates are explored. The inferential tests are also developed: log-likelihood ratio test for significance of regression parameter, and a Kolmogorov-Smirnov type test to assess the log-concavity. The simulation study and application of CHAS study show the usefulness of our method.

To improve the small sample performance, a smoothed log-concave estimate of the baseline density in POML might helps. Denote s^2 the sample variance of observed Y, and $\sigma^2_{\hat{p}}$ the variance from estimated log-concave density. A smoothed version of $\hat{p}$ can be derived via convolution as $\tilde{p}(z) = \int \phi_{\hat{\gamma}}(z-y)\hat{p}(y)dy$, where $\phi_{\hat{\gamma}}$ is the density for $N(0,\hat{\gamma})$. For observation Y generated from marginal log-concave density, the non-negativity of $\hat{\gamma}$ is ensured by the fact $\sigma^2_{\hat{p}} \leq s^2$ (corollary 2.3, Dumbgen and Rufibach (2009)). If $\sigma^2_{\hat{p}}$ is the variance from estimated baseline density in POML, then the criterion $\sigma^2_{\hat{p}} \leq s^2$ is not always satisfied since the observed Y is generated from a distribution conditioning on various values of $\mathbf{x}$. Thus, it is difficult but promising to develop smoothed baseline log-concave density estimates for POML in future research.

Another interesting further topic is to extend POML to multi-dimensional response which allows the joint modeling of association among muti-response and multi-covaraites. This is stimulated by the work of Cule and Samworth (2010), where the existence, uniqueness, and computation of nonparametric maximum likelihood estimator of multi-dimensional log-concave density were established. An advantage of multi-dimensional log-concave density estimation is that this is a fully automatic non-parametric density estimator.

In general kernel estimation for multi-dimensional density, the specification of a symmetric, positive definite bandwidth matrix is challenging.

Acknowledgements

We thank two referees and associate editor for their careful reading of manuscripts and constructive comments. We would like to thank Mr. Daniel B. Garside for insightful suggestions and support in improving this paper. We also appreciate that Dr. Hua Yun Chen provides helpful comments.

Appendix: Proof of Theorems

Proof of Lemma 1

Let assume that

$$h^2(g_{\boldsymbol{\beta},p}, g_{\boldsymbol{\beta}_0,p_0}) < ah^2(p, p_0) \tag{A.1}$$

for any $a > 0$. Since $0 \leq h^2(g_{\boldsymbol{\beta},p}, g_{\boldsymbol{\beta}_0,p_0}) \leq 1$ and $0 \leq h^2(p, p_0) \leq 1$, both $h^2(g_{\boldsymbol{\beta},p}, g_{\boldsymbol{\beta}_0,p_0}) = 0$ and $h^2(p, p_0) > 0$ have to be satisfied to meet the inequality (A.1).

If $h^2(g_{\boldsymbol{\beta},p}, g_{\boldsymbol{\beta}_0,p_0}) = 0$, then $\int \sqrt{g_{\boldsymbol{\beta},p} g_{\boldsymbol{\beta}_0,p_0}} d\mathbf{y} dx = 1$, and $g_{\boldsymbol{\beta},p} = g_{\boldsymbol{\beta}_0,p_0}$ follows by Cauchy–Schwarz inequality. We have $p = p_0$ and $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ by the identifiable property of the parameter $(p_0, \boldsymbol{\beta}_0)$ (Lemma 1 of Luo and Tsai (2012)). This contradict $h^2(p, p_0) > 0$. $\square$

Proof of Theorem 1

Denote $\phi = \log p$, and $f_{\boldsymbol{\beta},p} = p(\mathbf{y})\exp\{\eta(\mathbf{y}, \mathbf{x}|\boldsymbol{\beta})\}/Q_p(\boldsymbol{\beta})$ where $Q_p(\boldsymbol{\beta}) = \int p(\mathbf{y})\exp\{\eta(\mathbf{y}, \mathbf{x}|\boldsymbol{\beta})\}d\mathbf{y}$. For $\varepsilon > 0$, we have

$$0 \leq L_{\mathbb{P}}(\hat{\boldsymbol{\beta}}_n, \hat{p}_n) - L_{\mathbb{P}}(\boldsymbol{\beta}_0, p_0) = \int \log g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n} d\mathbb{P} - \int \log g_{\boldsymbol{\beta}_0,p_0} d\mathbb{P} \leq \int \log(\varepsilon + g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}) d\mathbb{P}$$
$$- \int \log g_{\boldsymbol{\beta}_0,p_0} d\mathbb{P} = \int \log\{\varepsilon + g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}\} d(\mathbb{P} - P_{X,Y}) + \int \log\left\{\frac{\varepsilon + g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}}{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}}\right\} dP_{X,Y}$$
$$+ \int \log\{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}\} dP_{X,Y} - \int \log g_{\boldsymbol{\beta}_0,p_0} d\mathbb{P}.$$

By assumption B and Lemma 3.2 in Seregin and Wellner (2010), it is not difficulty to show that $\int \log\{\varepsilon + g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}\}\}d(\mathbb{P} - P_{X,Y}) \to 0$ almost surely for ε small enough. Following Lemma 1 in Pal, Woodroofe, and Meyer (2007), it can be derived:

$$\int \log\left(\frac{\varepsilon + g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}}{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}}\right) dP_{X,Y} \leq 2\int \sqrt{\frac{\varepsilon}{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}}} dP_{X,Y} - 2h^2(g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}, g_{\boldsymbol{\beta}_0,p_0})$$

By the strong law of large numbers:

$$\int \log\{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}\} dP_{X,Y} - \int \log g_{\boldsymbol{\beta}_0,p_0} d\mathbb{P} \to \int \log\left\{\frac{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}}{g_{\boldsymbol{\beta}_0,p_0}}\right\} dP_{X,Y} \to 0 \text{ as } \varepsilon \to 0.$$

Consequently, we have

$$0 \leq \liminf \left[\int \log\{\varepsilon + g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}\} d(\mathbb{P} - P_{X,Y}) + \int \log\left\{\frac{\varepsilon + g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}}{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}}\right\} dP_{X,Y}\right.$$
$$\left. + \int \log\{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}\} dP_{X,Y} - \int \log g_{\boldsymbol{\beta}_0,p_0} d\mathbb{P}\right] \leq 2\int \sqrt{\frac{\varepsilon}{\varepsilon + g_{\boldsymbol{\beta}_0,p_0}}} dP_{X,Y} - 2\limsup\{h^2(\hat{p}_n, p_0)\}$$

As $\varepsilon \to 0$, we have lim sup$\{h^2(g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n}, g_{\boldsymbol{\beta}_0,p_0})\} \to 0$, and lim sup$\{h^2(\hat{p}_n, p_0)\} \to 0$ follows from Lemma 1. Following the same arguments of Lemma 3.14 in Seregin and Wellner (2010), $\hat{p}_n \to p_0$ pointwise and the convergence is uniform on compact space.

The $\lim_{n\to\infty}\{\int \log g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n} d\mathbb{P} - \int \log g_{\boldsymbol{\beta}_0,p_0} d\mathbb{P}\} \geq 0$ yields

$$\int \log g_{\hat{\boldsymbol{\beta}}_n,\hat{p}_n} dP_{X,Y} - \int \log g_{\boldsymbol{\beta}_0,p_0} dP_{X,Y} \geq 0,$$

and

$$\int(\hat{\phi}_n - \phi_0)dP_Y + \int\{\eta(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_n) - \eta(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_0)\}dP_{X,Y} \geq \int\{\log Q_{\hat{p}_n}(\hat{\boldsymbol{\beta}}_n) - \log Q_p(\boldsymbol{\beta}_0)\}dP_X. \tag{A.2}$$

Let $\hat{\phi}(\mathbf{y}) - \phi_0(\mathbf{y}) = c(\mathbf{y})$ and $\eta(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_n) - \eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta}_0) = b(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_n,\boldsymbol{\beta}_0)$, we have

$$\frac{Q_{\hat{p}_n}(\hat{\boldsymbol{\beta}}_n)}{Q_{p_0}(\boldsymbol{\beta}_0)} = \int e^{c(\mathbf{y})+b(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_n,\boldsymbol{\beta}_0)} f_{\boldsymbol{\beta}_0,p_0} d\mathbf{y} \geq e^{\int\{c(\mathbf{y})+b(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_n,\boldsymbol{\beta}_0)\}f_{\boldsymbol{\beta}_0,p_0}d\mathbf{y}}$$

with the equality hold when $c(\mathbf{y}) + b(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_n,\boldsymbol{\beta}_0) = c_0$ for a constant c_0 by Jensen's inequality. Furthermore,

$$\begin{aligned}\int \log\left\{\frac{Q_{\hat{p}_n}(\hat{\boldsymbol{\beta}}_n)}{Q_{p_0}(\boldsymbol{\beta}_0)}\right\}dP_X &\geq \int\int\{c(\mathbf{y}) + b(y,\mathbf{x}|\hat{\boldsymbol{\beta}}_n,\boldsymbol{\beta}_0)\}f_{\boldsymbol{\beta}_0}(p_0)d\mathbf{y}dP_X \\ &= \int c(\mathbf{y})dP_Y + \int b(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_n,\boldsymbol{\beta}_0)dP_{X,Y},\end{aligned}$$

by the law of total expectation. Combining with (A.2), we have

$$\int(\hat{\phi}_n - \phi_0)dP_Y + \int\{\eta(\mathbf{y},\mathbf{x}|\hat{\boldsymbol{\beta}}_n) - \eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta}_0)\}dP_{X,Y} = \int\{\log Q_{\hat{p}}(\boldsymbol{\beta}_0) - \log Q_p(\boldsymbol{\beta}_0)\}dP_X,$$

and it has to be $c(\mathbf{y}) + b(\mathbf{y}, \mathbf{x}|\hat{\boldsymbol{\beta}}_n, \boldsymbol{\beta}_0) = c_0$. Based on the facts that both $\exp(\hat{\phi}_n)$ and $\exp(\phi_0)$ need to be a density function, and both $\mathbf{Y}$ and $\mathbf{X}$ are not degenerate, we can deduce $c_0 = 0$. Furthermore, it implies $\eta(\mathbf{y}, \mathbf{x}|\hat{\boldsymbol{\beta}}_n) \rightarrow \eta(\mathbf{y}, \mathbf{x}|\boldsymbol{\beta}_0)$ because again both Y and $\mathbf{X}$ are not degenerate, followed by $\hat{\boldsymbol{\beta}}_n \rightarrow \boldsymbol{\beta}_0$ based on assumption C. $\square$

Proof of Lemma 2

By Theorem 4.2 of Doss and Wellner (2016), we know that $N_{[\,]}(\varepsilon, \sqrt{\mathcal{P}_c}, L_2) \lesssim \exp\{\varepsilon^{-1/2}\}$ for $L_2(p, q) = (\int |p - g|^2 d\lambda)^{1/2}$, where $\lesssim$ means the left side bounded by a constant times the right side. It imply that there is a set of functions $\{(p_1^l, p_1^u, \ldots, p_s^l, p_s^u) : L_2(\sqrt{p_i^l}, \sqrt{p_i^u}) < \varepsilon, i \in (1, \ldots, s)\}$ such that, for each $p \in \mathcal{P}$, $p_i^l \leq p \leq p_i^u$ for some i. Furthermore, let $p_i^L = p_i^l - \varepsilon$ and $p_i^U = p_i^u + \varepsilon$, which satisfy $p_i^L + \varepsilon \leq p \leq p_i^U - \varepsilon$.

Consider t points $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_t$ in the neighborhood $B(\boldsymbol{\beta}_0, \delta)$. By our model assumptions B and C, the $\exp\{\eta(y, \mathbf{x}|\boldsymbol{\beta})\}$ is bounded for $\boldsymbol{\beta} \in B(\boldsymbol{\beta}_0, \delta)$., Following the arguments in the proof of Lemma 3.1 by Huang (1996), by choosing appropriate δ and $t \lesssim 1/\varepsilon$, we have

$$\exp\{y, \mathbf{x}|\boldsymbol{\beta}_j)\}p_i^L(y) \leq \exp\{\eta(y, \mathbf{x}|\boldsymbol{\beta})\}p(y) \leq \exp\{\eta(y, \mathbf{x}|\boldsymbol{\beta}_j)\}p_i^U(y), \quad (A.3)$$

for $j \in (1, \ldots, t)$.

For each $(\boldsymbol{\beta}, p) \in B(\boldsymbol{\beta}_0, \delta) \times \mathcal{P}_c$, $i \in (1, \ldots, s)$, and $j \in (1, \ldots, t)$, inequalities (A.3) imply

$$g_{ij}^L \leq g_{\boldsymbol{\beta},p} \leq g_{ij}^U,$$

where

$$g_{ij}^L = \frac{p_i^L(y)e^{\eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta}_j)}}{\int p_i^U(y)e^{\eta(y,\mathbf{x}|\boldsymbol{\beta}_j)}dy}p_x(\mathbf{x}) \quad \text{and} \quad g_{ij}^U = \frac{p_i^U(y)e^{\eta(y,\mathbf{x}|\boldsymbol{\beta}_j)}}{\int p_i^L(y)e^{\eta(y,\mathbf{x}|\boldsymbol{\beta}_j)}dy}p_x(\mathbf{x}).$$

By Assumption B, we can see that $L_2(g_{ij}^L, g_{ij}^U) \lesssim L_2(p_i^L, p_i^U)$. It implies, there exist $\{g_{ij}^L, g_{ij}^U : i = 1, \ldots, s, j = 1, \ldots, t\}$ such that $g_{ij}^L \leq g \leq g_{ij}^U$ for any $g \in \mathcal{G}_\delta$ and some $i \in (1, \ldots, s)$, $j \in (1, \ldots, t)$. That is, $N_{[\,]}(\varepsilon, \sqrt{\mathcal{G}_\delta}, L_2) \lesssim \varepsilon^{-1}\exp(\varepsilon^{-1/2})$. For small enough ε, the claim of the theorem is followed since $N_{[\,]}(\varepsilon, \mathcal{G}_\delta, h_0) \leq N_{[\,]}(\varepsilon, \mathcal{G}_{4\delta}, h) \leq N_{[\,]}(\varepsilon/\sqrt{2}, \sqrt{\mathcal{G}_\delta}, L_2)$. $\square$

Proof of Lemma 3

Define $m_{\boldsymbol{\beta},p} = \log\{(g_{\boldsymbol{\beta},p}+g_{\boldsymbol{\beta}_0,p_0})/2g_{\boldsymbol{\beta}_0,p_0}\}$. Utilizing the relation $P\{\log(p/q)\} \lesssim -h^2(p,q)$ and the arguments in Theorem 3.4.4 of van der Vaart and Wellner (1996), it can be shown that $P_0(m_{\boldsymbol{\beta},p} - m_{\boldsymbol{\beta}_0,p_0}) \lesssim -h^2(g_{\boldsymbol{\beta},p}, g_{\boldsymbol{\beta}_0,p_0})$. Lemma 1 leads to $P_0(m_{\boldsymbol{\beta},p} - m_{\boldsymbol{\beta}_0,p_0}) \lesssim -h^2(p, p_0)$. By Taylor series expansion in $\boldsymbol{\beta}$, we have $P_0(m_{\boldsymbol{\beta},p_0} - m_{\boldsymbol{\beta}_0,p_0}) \gtrsim -\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2$. Thus decomposing $P_0(m_{\boldsymbol{\beta},p} - m_{\boldsymbol{\beta},p_0})$ as $P_0(m_{\boldsymbol{\beta},p} - m_{\boldsymbol{\beta}_0,p_0}) - P_0(m_{\boldsymbol{\beta},p_0} - m_{\boldsymbol{\beta}_0,p_0})$ yeilds

$$P_0(m_{\boldsymbol{\beta},p} - m_{\boldsymbol{\beta},p_0}) \lesssim -h^2(p, p_0) + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2. \tag{A.4}$$

Denote the empirical process $\mathbb{G}_n f = \sqrt{n}(\mathbb{P} - P)f$. Following Lemma 3.4.2

and Theorem 3.4.4 of van der Vaart and Wellner (1996), we have

$$E^*_{\mathcal{G}_\delta}|\mathbb{G}(m_{\boldsymbol{\beta},p} - m_{\boldsymbol{\beta},p_0})| \lesssim \zeta(\delta) = J_{[\ ]}(\delta, \mathcal{G}_\delta, h_0)\left\{1 + \frac{J_{[\ ]}(\delta, \mathcal{G}_\delta, h_0)}{\delta^2\sqrt{n}}\right\}, \quad \text{(A.5)}$$

where $J_{[\ ]}(\delta, \mathcal{G}_\delta, h_0) = \int_0^\delta \sqrt{1 + \log N_{[\ ]}(\varepsilon, \mathcal{G}_\delta, h_0)}d\varepsilon$.

The inequalities (A.4) and (A.5) correspond to expressions (3.5) and (3.6) of Murphy and van der Vaart (1999). The entropy in Lemmea 2 imply $J_{[\ ]}(\delta, \mathcal{G}_\delta, h_0) \lesssim \delta^{3/4}$. If a sequence δ_n satisfy $\delta_n \lesssim n^{-2/5}$, we have $\delta_n^{-2} J_{[\ ]}(\delta_n, \mathcal{G}_\delta, h_0) \lesssim \sqrt{n}$, which is equivalent to $\zeta_n(\delta_n) \leq \sqrt{n}\delta_n^2$. Then $h(\hat{p}_{\tilde{\boldsymbol{\beta}}}, p_0) \leq O(n^{-2/5} + ||\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||)$ follows from Theorem 3.2 of Murphy and van der Vaart (1999) and Theorem 3.4.1 van der Vaart and Wellner (1996). $\square$

Proof of Theorem 2

In the context of least favorable model, we assume that for each $(\boldsymbol{\beta}, p)$, there exist a map $\mathbf{t} \to p_{\mathbf{t}}(\boldsymbol{\beta}, p) = p + (\boldsymbol{\beta} - \mathbf{t})h_0 p$, where h_0 is the least favorable direction at the true parameter $(\boldsymbol{\beta}_0, p_0)$. We then form the map $\mathbf{t} \to l(\mathbf{t}, \boldsymbol{\beta}, p)(y)$ by $l(\mathbf{t}, \boldsymbol{\beta}, p)(y) = l\{\mathbf{t}, p_{\mathbf{t}}(\boldsymbol{\beta}, p)\}(y)$, where $l(\boldsymbol{\beta}, p) = \log p(y) + \eta(y, \mathbf{x}|\boldsymbol{\beta}) - \log\int \exp(\eta(y, \mathbf{x}|\boldsymbol{\beta}))p(y)dy$ is twice continuously differentiable for all y. The corresponding derivatives are denoted as $\dot{l}(\mathbf{t}, \boldsymbol{\beta}, p)(y)$ and $\ddot{l}(\mathbf{t}, \boldsymbol{\beta}, p)(y)$.

In the followed, we will establish four conditions.

Condition 1. $p_{\boldsymbol{\beta}}(\boldsymbol{\beta}, p) = p$ for every $(\boldsymbol{\beta}, p)$.

It is satisfied for $p_{\mathbf{t}}(\boldsymbol{\beta}, p) = p + (\boldsymbol{\beta} - \mathbf{t})h_0 p$.

Condition 2. $\dot{l}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, p_0) = \dot{l}_{\boldsymbol{\beta}_0, p_0}$.

The score function for $\boldsymbol{\beta}$ is $l'_{\boldsymbol{\beta},p} = \eta' - \int \eta' e^{\eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta})} p(y)dy / \int e^{\eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta})} p(y)dy$. Let $\mathcal{H}$ be the set of measurable function $h : \mathcal{Y} \to [0, 1]$, given a fixed p, let $p_{\mathbf{t}}(\boldsymbol{\beta}, p) = p + thp$. When p is log-concave, $p_{\mathbf{t}}$ is log-concave for $th > 0$. Differentiating at $t = 0$, we have the score for p as $A_{\boldsymbol{\beta},p}h = h - B_{\boldsymbol{\beta},p}h = h - \int h e^{\eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta})} p(y)dy / \int e^{\eta(\mathbf{y},\mathbf{x}|\boldsymbol{\beta})} p(y)dy$. The efficient score function for $\boldsymbol{\beta}$ at $(\boldsymbol{\beta}_0, p_0)$ is defined as $\dot{l}_{\boldsymbol{\beta}_0,p_0} = l'_{\boldsymbol{\beta}_0,p_0} + A_{\boldsymbol{\beta}_0,p_0} h_{\boldsymbol{\beta}_0,p_0}$. Substituting $\boldsymbol{\beta} = \mathbf{t}$ and $p_{\mathbf{t}} = p$ in $l(\boldsymbol{\beta}, p)$ and differentiating with respect to $\mathbf{t}$, it is straight forward to show that $\dot{l}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, p_0) = \dot{l}_{\boldsymbol{\beta}_0,p_0}$.

Condition 3. For any $\tilde{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}_0$, $\hat{p}_{\tilde{\boldsymbol{\beta}}_n} \xrightarrow{P} p_0$.

It is followed by Lemma 3.

Condition 4. For any $\tilde{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}_0$, $P_0 \dot{l}(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\beta}}_n, \hat{p}_{\tilde{\boldsymbol{\beta}}_n}) = o_P(||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0|| + n^{-1/2})$.

As shown in expression (17) of Murphy and van der Vaart (2000), $P_0 \dot{l}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, p)$ is of order $O_p\{h^2(p, p_0)\}$ since $p \to f_{\boldsymbol{\beta},p}$ is twice differentiable and $p \to \dot{l}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, p)$ is differentiable at p_0. By Lemma 3, we can see that $P_0 \dot{l}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, \hat{p}_{\tilde{\boldsymbol{\beta}}_n}) = o_P(||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0|| + n^{-1/2})$ is satisfied, which is equivalent to

Condition 4.

The conditions 1-4 correspond to conditions (8)-(11) in Murphy and van der Vaart (2000), in addition, we need to prove the invertibility of the information matrix to build the asymptotic property. Let $\mathbf{D}_n(\boldsymbol{\beta}, p) = \{\mathbf{D}_n1(\boldsymbol{\beta}, p), \mathbf{D}_n2(\boldsymbol{\beta}, p)\}$ be the element of $\mathbb{R}^g \times l^\infty(\mathcal{H})$ given by

$$\mathbf{D}_n1(\boldsymbol{\beta}, p) = \mathbb{P}l'_{\boldsymbol{\beta},p}, \quad \mathbf{D}_n2(\boldsymbol{\beta}, p) = \mathbb{P}A_{\boldsymbol{\beta},p}h - P_{\boldsymbol{\beta},p}A_{\boldsymbol{\beta},p}h.$$

The expectation of $\mathbf{D}_n(\boldsymbol{\beta}, p)$ under the true distribution $P_0 = P_{\boldsymbol{\beta}_0,p_0}$ is the element $\mathbf{D}(\boldsymbol{\beta}, p) = \{\mathbf{D}_1(\boldsymbol{\beta}, p), \mathbf{D}_2(\boldsymbol{\beta}, p)\}$ of $\mathbb{R}^k \times l^\infty(\mathcal{H})$ given by

$$\mathbf{D}_1(\boldsymbol{\beta}, p) = \mathbb{P}_0 l'_{\boldsymbol{\beta},p}, \quad \mathbf{D}_2(\boldsymbol{\beta}, p) = P_0 A_{\boldsymbol{\beta},p}h - P_{\boldsymbol{\beta},p}A_{\boldsymbol{\beta},\phi}h.$$

A Hilbert-space adjoint $B^*_{\boldsymbol{\beta},p}$ of $B_{\boldsymbol{\beta},p}$ is given by $B^*_{\boldsymbol{\beta},p}q = \int q(\mathbf{x})e^{\eta(y,\mathbf{x}|\boldsymbol{\beta})}dP_X(\mathbf{x})$. The least favourable direction, h_0, for the estimation of $\boldsymbol{\beta}$ in the presence of p is given by $(A^*_{\boldsymbol{\beta}_0,p_0}A_{\boldsymbol{\beta}_0,p_0})^{-1}A^*_{\boldsymbol{\beta}_0,p_0}l'_{\boldsymbol{\beta}_0,p_0}$, and it can be shown that $A^*_{\boldsymbol{\beta}_0,p_0}l'_{\boldsymbol{\beta}_0,p_0} = -B^*_{\boldsymbol{\beta}_0,p_0}l'_{\boldsymbol{\beta}_0,p_0}$ and $A^*_{\boldsymbol{\beta}_0,p_0}A_{\boldsymbol{\beta}_0,p_0} = I - B^*_{\boldsymbol{\beta}_0,p_0}B_{\boldsymbol{\beta}_0,p_0}$.

The derivative of D at $(\boldsymbol{\beta}_0, p_0)$ is given by the map:

$$\dot{\mathbf{D}} : (\boldsymbol{\beta} - \boldsymbol{\beta}_0, p - p_0) \to \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} - \boldsymbol{\beta}_0 \\ p - p_0 \end{pmatrix},$$

where $\mathbf{H}_{11}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) = P_0 l''_0(\boldsymbol{\beta}-\boldsymbol{\beta}_0) = -I_0(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$, $\mathbf{H}_{12}(p-p_0) = \int B^*_0 l'_0(p-p_0)dy$, $\mathbf{H}_{21}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)h = P_0 A_0 h(l'_0-\eta'_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$, and $\mathbf{H}_{22}(p-p_0)h = -\int(I-$

$B_0^*B_0)h(p-p_0)dy$. Since

$$\dot{\mathbf{D}}^{-1} = \begin{pmatrix} \mathbf{H}_{11}^{-1}(\mathbf{H}_{11}+\mathbf{H}_{12}\Lambda^{-1}\mathbf{H}_{21})\mathbf{H}_{11}^{-1} & -\mathbf{H}_{11}^{-1}\mathbf{H}_{12}\Lambda^{-1} \\ -\Lambda^{-1}\mathbf{H}_{21}\mathbf{H}_{11}^{-1} & \Lambda^{-1} \end{pmatrix},$$

the continuous invertibility of $\dot{\mathbf{D}}$ can be verified by continuous invertibility of $\mathbf{H}_{11}$ and $\Lambda = \mathbf{H}_{22} - \mathbf{H}_{21}\mathbf{H}_{11}^{-1}\mathbf{H}_{12}$. By Assumption D, the matrix $\mathbf{H}_{11}$ is continuous invertible. The operator

$$\Lambda = -\int \{I+P_0A_0(l_0'-\eta_0')I_0^{-1}B_0^*l_0'-B_0^*B_0\}h(p-p_0)dy = -\int (I+K)h(p-p_0)dy$$

is continuous invertible if K is compact and $I+K$ is one-to-one using the theory of Fredholm operator. Since $e^{\eta(y,\mathbf{x}|\boldsymbol{\beta})}$ is sufficiently smooth, the operator B_0^* is compact by Arzelà-Ascoli theorem. The operator $P_0A_0(l_0' - \eta_0')I_0^{-1}B_0^*l_0'$ is compact because it has a one-dimensional range. Thus K is compact. Now it suffices to show that $I+K$ is one-to-one. The spectrum of the self-adjoint operator $I - B_0^*B_0 : L_2(p_0) \to L_2(p_0)$ is contained in $[1, \infty)$. Finally, this operator is continuously invertible in the Hilbert-space sense.

Following Corollary 1 and Theorem 1 in Murphy and van der Vaart (2000), the conditions 1-4, the invertibility of information matrix, and the consistence of estimators imply $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ is asymptotic normal with mean 0 and covariance matrix $\tilde{I}_0 = \mathbf{H}_{11} - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21}$. $\square$

Proof of Theorem 3

Let

$$(\boldsymbol{\beta}_1, F_1) = \underset{\boldsymbol{\beta}, F \in \mathcal{F}_c}{\arg\max}\ E_0 \left\{ \log \frac{f(\mathbf{y}; \mathbf{x}, \boldsymbol{\beta}, F)}{f(\mathbf{y}; \mathbf{x}, \boldsymbol{\beta}_0, F_0)} \right\},$$

where E_0 denote the expectation under $P_{X,Y}$.

Helly's lemma (van der Vaart, 1998) implies that, there exist a subsequence of $\hat{F}_n$ which converges to a distribution F_2 on the continuous points of F_2. There is also a subsequence of $\hat{\boldsymbol{\beta}}_n$ converging to $\boldsymbol{\beta}_2$ because $\boldsymbol{\Theta}$ is a compact set. It follows that

$$0 \le \mathbb{P}_n\{l(\hat{\boldsymbol{\beta}}_n, \hat{F}_n) - l(\boldsymbol{\beta}_0, F_0)\} \to E_0\{l(\boldsymbol{\beta}_2, F_2) - l(\boldsymbol{\beta}_0, F_0)\}.$$

Since $(\boldsymbol{\beta}_1, F_1)$ is the unique miaximizer of the Kullback-Leibler information by assumption A, we conclude that $(\boldsymbol{\beta}_2, F_2) = (\boldsymbol{\beta}_1, F_1)$. That is, $\hat{F}_n$ converge weakly to F_1 whose density is log-concave when $F_0 \notin \mathcal{F}_c$. Since both $\hat{F}_n$ and F_1 are continuous probability distribution function, the weak convergence of $\hat{F}_n$ implies its uniform convergence, i.e., $\|\hat{F}_n(y) - F_1(y)\|_\infty \to 0$ (Chow and Teicher, 1978).

Based on $\|\hat{F}_n(y) - \tilde{\mathbb{F}}_n(y)\|_\infty \ge \|\hat{F}_n(y) - F_0(y)\|_\infty - \|\tilde{\mathbb{F}}_n(y) - F_0(y)\|_\infty$ and $\|F_1(y) - F_0(y)\|_\infty \ge C$ for a positive constant C, we have

$$\lim_{n\to\infty} \inf_{\hat{F}_n(y) \in \mathcal{F}_c, F_0 \notin \mathcal{F}_c} \|\hat{F}_n(y) - \tilde{\mathbb{F}}_n(y)\|_\infty \ge C$$

because $\|\tilde{\mathbb{F}}_n(y) - F_0(y)\|_\infty \to 0$ by Luo and Tsai (2012). Consequently, we have

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\inf_{\hat{F}_n(y)\in\mathcal{F}_c, F_0\notin\mathcal{F}_c} P(T_n \geq \sqrt{n}\epsilon) = 1. \tag{A.6}$$

When the true distribution satisfy the shape constraint, we have $\lim_{n\to\infty: F_0\in\mathcal{F}_c} P(T_n \leq \sqrt{n}\epsilon) = 1$ for all $\epsilon > 0$ since $\|\hat{F}_n(y) - F_0(y)\|_\infty \to 0$ by Theorem 1. It follows that $\lim_{n\to\infty} P(T^* \leq \sqrt{n}\epsilon) = 1$ because the shape constraint estimate $(\boldsymbol{\beta}^*, F^*)$ is based on data sampled from the null distribution $\mathcal{F}_c$ in the bootstrap procedure. It implies that

$$\lim_{n\to\infty} P(\xi_\alpha \leq \sqrt{n}\epsilon) = 1 \text{ for all } \epsilon > 0, \tag{A.7}$$

where ξ_α is the critical values in the bootstrap procedure. Combining (A.6) and (A.7) together, we complete the proof. □

References

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

Bagnoli, M., and Bergstrom, T. (2005). Log-concave probability and its applications. *Econometric 26*, pp. 445–469.

Barlow, R.E., and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Reinhart and Winston: New York.

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Annal of Statistics 25*, 970–981.

Carroll, R. J., Delaigle, A., and Hall, P. (2011). Testing and estimating shape-constrained nonparametric density and regression in the presence of measurement error. *Journal of the American Statistical Association 106*, 191–202.

Chan, K. C. J., Chen, Y. Q., and Di, C. Z. (2012). Proportinal mean residual life model for right-censored length-biased data. *Biometrika 99*, 995–1000.

Chang, G., and Walther, G. (2007). Clustering with mixtures of log-concave distributions. *Computational Statistics and Data Analysis 51*, 6242–6251.

Chen, H.Y. (2007). A semiparametric odds ratio model for measuring association. *Biometrics 63*, 413–421.

Chen, J., Liu, L., Zhang, D., and Shih, Y.T. (2013). A flexible model for the mean and variance functions, with application to medical cost data. *Statistics in Medicine 32*, 4306–4318.

Chen, Y., and Samworth, R.J. (2013). Smoothed log-concave maximum likelihood estimation with application. *Statistica Sinica 23*, 1373–1398.

Cheng, K.F., and Chu, C.K. (2004). Semiparametric density estimation under a two-sample density ratio model. *Bernoulli 10*, 583–604.

Cheng, J., Qin, J., and Zhang, B. (2009) Semiparametric estimation and inference for distributional and general treatment effects. *Journal of the Royal Statistical Society, Series B 71*, 881–904.

Chow, Y.SH. and Teicher, H. (1978). *Probability Theory. Independence, Interchangeability, Martingales.* Springer-Verlag: New York.

Cule, M., and Samworth, R. (2010). Maximum likelihood estimation for multi-dimensional log-concave density. *Journal of the Royal Statistical Society, Series B 72*, 545–607.

Davidov, O., and Iliopoulos, G. (2013). Convergence of Luo and Tsai's iterative algorithm for estiamtion in proportional likelihood ratio model. *Biometrika 100*, 778–780.

Daviglus, M.L., Stamler, J., Pirzada, A., Yan, L.L., Garside, D.B., Liu, K., Wang, R., Dyer, A.R., Lloyd-Jones, D.M., and Greenland, P. (2004). Favorable cardiovascular risk profile in young women and long-term risk of cardiovascular and all-cause mortality. *Journal of the American Medical Association 13*, 1588–1592.

Daviglus, M.L., Pirzada, A., Durazo-Arvizu, R., Chen, J.S., Allison, M., Aviles-Santa, L., Cai, J.W., Gonzalez, H.M., Kaplan, R.C., and Schneiderman, N. (2016). Prevalence of low cardiovascular risk profile among diverse Hispanic/Latino adults in the United States by age, sex, and level of acculturation: the Hispanic Community Health Study/Study of Latinos. *Journal of the American Heart Association 5*, DOI: 10.1161/JAHA.116.003929.

Dharmadhikari, S., and Joag-Dev, L. (1988). *Unimodality, Convexity, and Application*. Boston: Academic Press.

Diao, G., Ning, J., and Qin, J. (2012). Maximum likelihood estimation for semiparametric density ratio model. *The International Journal of Biostatistics 8*, 1–27.

Doss, C.R., and Wellner J.A. (2016). Global rates of convergence of the MLES of log-concave and s-concave densities. *The Annals of Statistics 44*, 954–981.

Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. *Bernoulli 15*, 40–68.

Dümbgen, L., Hüsler, A., and Rufibach, K. (2011). Active set and EM algorithms for log-concave densities based on complete and censored data. *arXiv* 0707.4643v4.

Dümbgen, L., Samworth, R., and Schuhmacher, D. (2011). Approximation by log-concave distributions, with applications to regression. *The Annals of Statistics 39*, 702–730.

Efron, B., and Tibshirani, R. (1996). Using specially designed exponential families for density estimation. *The Annals of Statistics 24*, 2431–2461.

Fletcher, R. (1987). *Practical Methods for Optimization*. New York: Wiley.

Fokianos, K. (2004). Merging information for semiparametric density estimation. *Journal of the Royal Statistical Society, Series B 66*, 941–958.

Gilbert, P.B. (2000). Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. *The Annals of Statistics 28*, 151–194.

Gu, C., and Wang, J.Y. (2003). Penalized likelihood density estimation: Direct cross-validation and scalable approximation. *Statistica Sinica 13*, 811–826.

Groeneboom, P., and Wellner, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser: Basel.

Hall, P., Huang, L.S., Gifford, J.A., and Gijbels, I. (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *Journal of Computational and Graphical Statitics 10*, 592–614.

Hall, P., and Keilegom, I.V. (2005). Testing for monotone increasing hazard rate. *Annals of Statistics 33*, 1109–1137.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics 24*, 540–568.

Huang, A., and Rathouz, P. (2012). Proportinal likelihood ratio models for mean regression. *Biometrika 99*, 223–229.

Hazelton, M.L. (2010). Assessing log-concavity of multivariate densities. *Statistics and Probability Letters 81*, 121–125.

Ibragimov, I. A. and Khas'minskii, R. Z. (1983). Estimation of distribution density. *Journal of Soviet Mathematics 25*, 40–57.

Kalbfleisch, J.D, and Prentice R.L. (2002). *The Statistical Analysi of Failure Time Data*. Wiley: Hoboken.

Kedem, B., Kim, E., Voulgaraki, A., and Graubard, B.I. (2009). Two-dimensional semiparametric density ratio modeling of testicular germ cell data. *Statistic in Medicine 28*, 2147–2159.

Kim, A.K.H., and Samworth, R.J. (2016). Global rates of convergence in log-concave density estimation. *Annals of Statistics 44*, 2756–2779.

Luo, X., and Tsai, W.Y. (2012). A proportional likelihood ratio model. *Biometrika 99*, 211–222.

Müller S., and Rufibach, L. (2009). Smooth tail-index estimation. *Journal of Statistical Computation and Simulation* **79**, 1155–1167.

Murphy, S.A., and van der Vaart, A.W. (1999). Observed information in semi-parametric models. *Bernoulli 5*, 381–412.

Murphy, S.A., and van der Vaart, A.W. (2000). On profile likelihood. *Journal of the American Statistical Association 95*, 449–485.

Nan, B., and Wellner, J.A. (2013). A general semiparametric Z-estimation approach for case-cohort studies. *Statistica Sinica* **23**, 1155–1180.

Osius, G. (2004). The association between two random elements: A complete characterization and odds ratio models. *Metrika 60*, 261–277.

Owen, A.B. (2001). *Empirical Likelihood.* Chapman & Hall/CRC: New York.

Pal, J.K., Woodroofe, M., and Meyer, M. (2007). Estimating a polya frequency function. *Complex Datasets and Inverse Problems: Tomography, Networks, and Beyond, IMS Lecture Notes-Mongraph Series 54*, 239–249.

Pirzada, A., Reid, K., Kim, D., Garside, D.B., Lu, B., Vu, T.T., Lloyd-Jones, D.M., Zee, P., Liu, K., Stamler, J., and Daviglus, M.L. (2013). Chicago healthy aging study (CHAS): objectives and design. *American Journal of Epidemiology 178*, 635–644.

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

Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *Annals of Statistics 27*, 1368–1384.

Qin, J., Berwich, M., Ashbolt, R., and Dwyer, T. (2002). Quantifying the change of melanoma incidence by Breslow thickness. *Biometrics 58*, 665–670.

Qin, J., Ning, J., Liu, H., and Shen, Y. (2011). Maximum likelihood estimations and EM algorithms with length-biased data. *Journal of the American Statistical Association 106*, 1434–1449.

Qin, J., and Zhang, B. (2005). Density estimation under a two-sample semiparametric model. *Journal of Nonparametric Statistics 17*, 665–683.

Rathouz, P.J., and Gao, L. (2009). Generalized linear models with unspecified reference distribution. *Biostatistics 10*, 205–218.

Seregin, A., and Wellner, J.A. (2010) Nonparametric Estimation of Multivariate Convex-Transformed Densities. *The Annals of Statistics 38*, 3751–3781.

Schuhmacher, D., Hüsler, A. and Düumbgen, L. (2011). Multivariate log-concave distributions as a nealy parametric model. *Statistics and Risk Modeling 28*, 277–295.

Shen, Y., Jing, J. and Qin, J. (2012). Likelihood approaches for the invariant density ratio model with biased-sampling data. *Biometrika 99*, 363–378.

Terrell, G.R. (1990). The maximal smoothing principle in density estimation. *Journal of the American Statistical Association 85*, 470–477.

Terrell, G.R. (2003). The wilson-hilferty transformation is locally saddle-point. *Biometrika 90*, 445–453.

Tricomi, F.G. (1985). *Integral Equations.* Dover: New York.

van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge: New York.

van der Vaart, A.W., and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer: New York.

Wang, C., Tan, Z., and Louis, T.A. (2011). Exponential tilt models for two-group comparison with censored data. *Journal of Statistical Planning and Inference 141*, 1102–1117.

Walther, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *Journal of the American Statistical Association 97*, 508–513.

Walther, G. (2009). Inference and modeling with log-concave distributions. *Statistical Science 24*, 319–327.

Yu, S., Jing, J., and Qin, J. (2012). Likelihood approaches for the invariant density ratio model with biased-sampling data. *Biometrika 99*, 363–378.

Zou, F., Fine, J.P., and Yandell, B.S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika 89*, 661–75.

College of Applied Health Sciences, University of Illinois at Chicago, Chicago, IL, 60612, U.S.A.

E-mail: jinsongc@uic.edu

Department of Statistics, Virginia Tech University, Blacksburg, VA, 24061, U.S.A.

E-mail: terrell@vt.edu

Department of Statistics, Virginia Tech University, Blacksburg, VA, 24061, U.S.A.

E-mail: inyoungk@vt.edu

College of Medicine, University of Illinois at Chicago, Chicago, IL, 60612; Feinberg School of Medicine, Northwestern University, Chicago, IL, 60611, U.S.A.

E-mail: daviglus@uic.edu; daviglus@northwestern.edu

Table 1: Estimates of Regression Parameters in the Simulation Studies Using POML with Log-concave Density Estimation. The data are generated from followed ORM: $f(y; x_1, x_2, \beta_1, \beta_2, p) = \frac{p(y)e^{y\beta_1 x_1 + y\beta_2 x_2}}{\int p(y)e^{y\beta_1 x_1 + y\beta_2 x_2} dy}$ with $x_1 \sim$ Binomial$(1, 0.5)$ and followed four settings: I. $p(y) \sim \mathrm{N}(0,1)$, $x_2 \sim \mathrm{N}(0,1)$, $\beta_1 = 0$, and $\beta_2 = 0$; II. $p(y) \sim \mathrm{N}(0,1)$, $x_2 \sim \mathrm{N}(0,1)$, $\beta_1 = 1$, and $\beta_2 = 0.5$; III. $p(y) \sim$ Exponential(1), $x_2 \sim$ Exponential(1), $\beta_1 = 0$, and $\beta_2 = 0$; IV. $p(y) \sim$ Exponential(1), $x_2 \sim$ Exponential(1), $\beta_1 = -1$, and $\beta_2 = -0.5$. Bias: estimated regression parameters minus true values; Est.: estimates; sd: sampling standard deviation of estimates; mse: average of estimated mean squared error; CP: coverage proability of 95% confidence interval; RR: empirical rejection rate of a nominal 0.05 level using log-likelihood ratio test.

	n	β_1					β_2				
		Bias	sd.	mse	CP	RR	Est.	sd.	mse	CP	RR
I	200	0.009	0.148	0.022	0.920	0.058	0.006	0.072	0.005	0.964	0.038
	500	0.001	0.090	0.008	0.956	0.062	-0.00005	0.043	0.002	0.940	0.046
II	200	0.013	0.186	0.035	0.960	0.992	0.021	0.090	0.009	0.960	0.966
	500	-0.005	0.111	0.012	0.944	1.000	-0.005	0.054	0.003	0.966	0.996
III	200	0.006	0.151	0.023	0.952	0.066	0.016	0.081	0.007	0.960	0.052
	500	0.005	0.085	0.007	0.952	0.044	0.009	0.047	0.002	0.954	0.046
IV	200	-0.031	0.320	0.103	0.950	0.984	-0.044	0.210	0.046	0.955	0.880
	500	-0.015	0.190	0.036	0.944	1.00	-0.001	0.122	0.015	0.958	0.994

Table 2: Results of Power Simulation. The data are generated from model (4.1) with mixture normal baseline distribution $0.5\mathrm{N}(\mu_1, 1) + 0.5\mathrm{N}(\mu_2, 1)$. There is only a binary predictor $x \sim \text{binomail}(1, 0.5)$ with regression coefficient equal to 1. The significance level is $\alpha = 0.05$.

n	Type I Error Rate		Power
	$\mu_1 = 0$, $\mu_2 = 0$	$\mu_1 = 0$, $\mu_2 = 2$	$\mu_1 = 0$, $\mu_2 = 4$
100	0.070	0.105	0.580
200	0.055	0.055	0.900

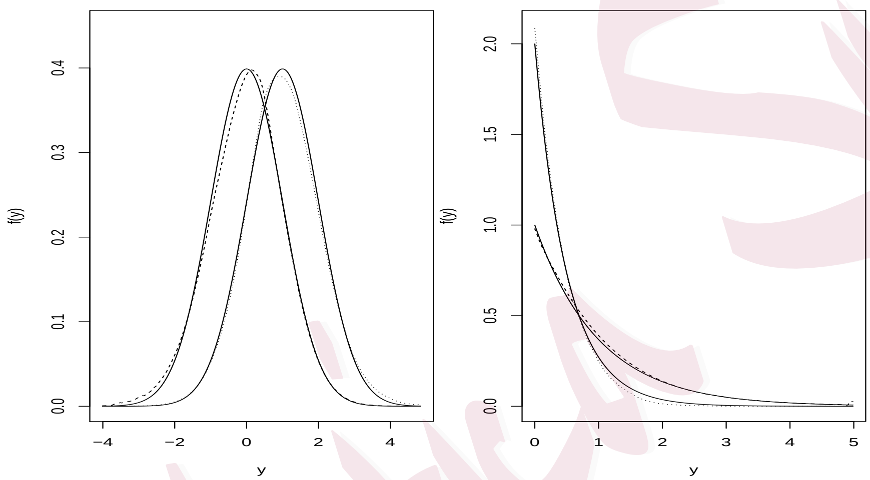

Figure 1: Plots of Mean Estimated Densities and Distributions for Control and Treatment Groups in Simulation. Left: normal distribution in setting II with N=200; right: exponential distribution in setting IV with N=200. Dashed line: estimated density for control group ($\beta_1 = 0$); dotted line: estimate density for treatment group ($\beta_1 = 1$); solid line: true densities.

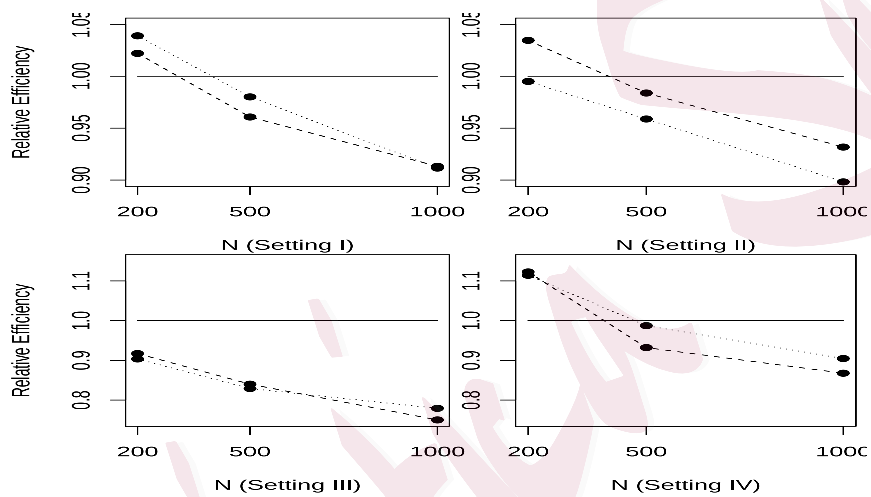

Figure 2: Relative Efficiency for Regression Parameters (β_1, β_2) between Estimates using POML and Empirical Likelihood (EL) Estiamtes. Dashed line: estimated relative efficiency MSE_{POML}/MSE_{EL} for $\hat{\beta}_1$; dotted line: estimated relative efficiency MSE_{POML}/MSE_{EL} for $\hat{\beta}_2$.

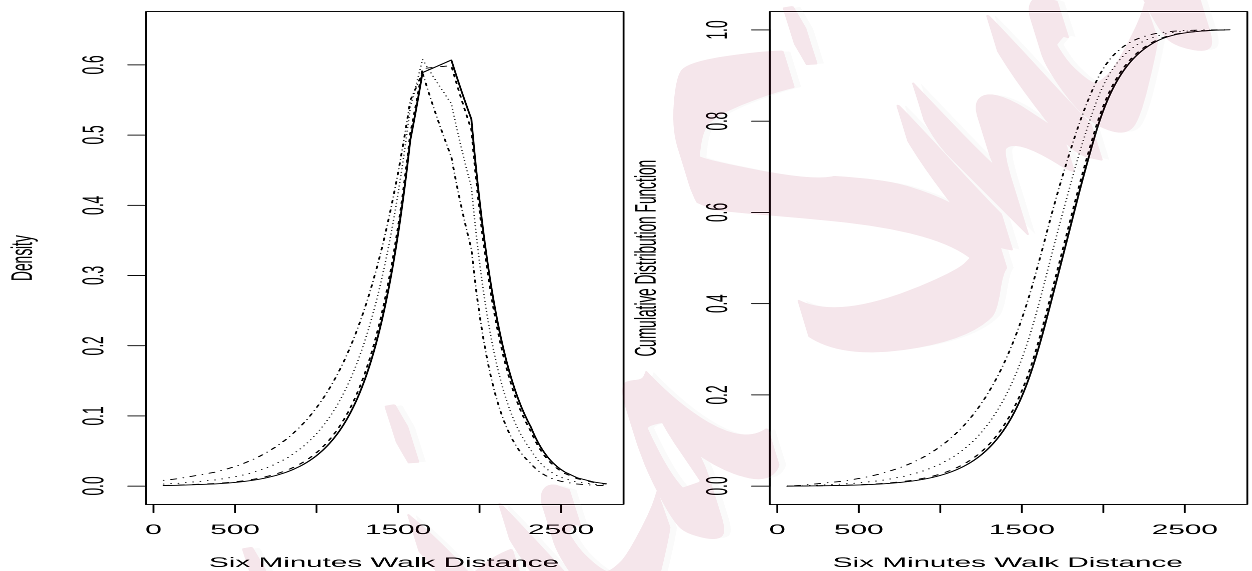

Figure 3: Estimated Densities and Cumulative Distribution Functions for Six Minutes Walking Distance of Participants in CHAS Study using POML by Risk Groups Adjusted for Gender and Age. Left: estimated density functions; right: estimated cumulative distribution functions. Solid Line: LR; dashed line: 0 RF; dotted line: 1 RF; dash-dotted line: 2+ RF. The unit for waking distance is foot.