Published in final edited form as:

Biometrika. 2005 June 1; 92(2): 271–282. doi:10.1093/biomet/92.2.271.

Empirical-likelihood-based semiparametric inference for the treatment effect in the two-sample problem with censoring

YONG ZHOU and

Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, 100080, China,
yzhou@amss.ac.cn

HUA LIANG

Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA, hua.liang@stjude.org

Summary

To compare two samples of censored data, we propose a unified semiparametric inference for the parameter of interest when the model for one sample is parametric and that for the other is nonparametric. The parameter of interest may represent, for example, a comparison of means, or survival probabilities. The confidence interval derived from the semiparametric inference, which is based on the empirical likelihood principle, improves its counterpart constructed from the common estimating equation. The empirical likelihood ratio is shown to be asymptotically chi-squared. Simulation experiments illustrate that the method based on the empirical likelihood substantially outperforms the method based on the estimating equation. A real dataset is analysed.

Keywords

Estimating equation; Confidence interval; Coverage; Kaplan-Meier estimation; Empirical likelihood ratio; Empirical likelihood function

1 Introduction

Let X and Y be random variables representing some characteristic of two treatments, where X corresponds to the new treatment and Y to the control treatment. One may wish to compare the average treatment effects, so that $\Delta = E(X) - E(Y)$ is of interest. Alternatively, one may wish to consider the probability of one treatment being better than the other, which is equivalent to considering the survival competition probability, so that $\Delta = \text{pr}(X < Y)$ or $\Delta = \text{pr}(X > Y)$, or to compare the probability of death before a given time t_0 in survival analysis, so that $\Delta = \text{pr}(X < t_0) - \text{pr}(Y < t_0)$. When X and Y are both sampled from parametric families, standard statistical approaches such as maximum likelihood can be used (Brownie et al., 1986; Campbell & Ratnarkhi, 1993; Goddard & Hinberg, 1990; Hsieh & Turnbull, 1996). There are also many nonparametric approaches for comparing the two unknown continuous distributions F and G of X and Y , respectively, based on independent samples; see for example Gastwirth & Wang (1988), Hollander & Korwar (1982), and Li et al. (1996).

Sometimes enough is known about one treatment, usually the control treatment, for a parametric form to be assumed for the distribution G . However, this may not hold for F . This leads to a semiparametric two-sample model. Recently much attention has been paid to semiparametric inference. Li et al. (1999) proposed a semiparametric estimator for quantile comparison. They studied two-sample inference through the quantile comparison function G

$\{F^{-1}(p)\}$ assuming that G is known and F is unknown. A simpler case where Y is normally distributed has been studied by Hsieh & Turnbull (1996).

The contribution of this article is to develop a unified approach to inference about the parameter Δ , introduced in the first paragraph. Although a common estimating equation for Δ and the corresponding confidence interval are presented below, we focus on semiparametric inference about Δ by using empirical likelihood ratio methods; see for example Owen (1988, 1990, 2000), Qin (1994, 1999) and Qin & Lawless (1994). The empirical likelihood ratio has a limiting chi-squared distribution, leading to tests and confidence intervals (Thomas and Grunkemeier, 1975) for a variety of problems, including linear models, generalised linear models and estimating equations. The empirical likelihood method has many advantages over competitors such as the normal-approximation-based method and the bootstrap method (Hall & La Scala, 1990). The appealing features of the empirical likelihood include improvement of the confidence region, increased accuracy of coverage through the use of auxiliary information, and easy implementation. We will derive the asymptotic distribution of the resulting empirical likelihood function in our context, explain how to establish the corresponding confidence interval, and illustrate the advantages numerically.

2 Estimation methods for a two-sample model with censored data

2.1 General framework

Assume that the variables $\{X_1^0, X_2^0, \dots, X_n^0\}$ and $\{Y_1^0, Y_2^0, \dots, Y_m^0\}$ are randomly censored by two sequences of random variables $\{U_1, U_2, \dots, U_n\}$ and $\{V_1, V_2, \dots, V_m\}$, respectively. The members of each sequence of random variables $\{X_i^0\}$, $\{Y_j^0\}$, $\{U_i\}$ and $\{V_j\}$ are independent and identically distributed with survival functions $S(x)$, $D(y)$, $K(u)$, and $Q(v)$, respectively, where $S(x)$ and $K(u)$ are two unknown functions. The distribution function $G(y) = 1 - D(y)$ is of known form and depends on the unknown d -dimensional parameter θ , i.e., $G(y) = G_\theta(y)$. We assume that $G_\theta(y)$ has the density function $g_\theta(y)$. The observed sample are the pairs (X_i, δ_i) , for $i = 1, 2, \dots, n$, and the pairs (Y_j, η_j) , for $j = 1, \dots, m$, where $X_i = \min(X_i^0, U_i)$, $\delta_i = I(X_i^0 \leq U_i)$, $i = 1, 2, \dots, n$ and $Y_j = \min(Y_j^0, V_j)$, $\eta_j = I(Y_j^0 \leq V_j)$, $j = 1, 2, \dots, m$ in which $I(\cdot)$ denotes an indicator function. We assume that the two samples of pairs are independent.

The parameter of interest, Δ , may be a functional of the functions $S(x)$, $D(y)$, $K(u)$ and $Q(v)$. The estimation of the unknown functions $S(x)$ and $K(u)$ and the parameter θ is a secondary goal; these are considered as nuisance parameters. For any measurement Δ of the difference between the two samples, we assume that the information about θ , Δ and $F(x) = 1 - S(x)$ is available in the known form of an unbiased estimation function, i.e.,

$$E_{\theta_0} \psi(X^0, \theta_0, \Delta) = 0, \quad (2.1)$$

where θ_0 is the true value of θ and ψ , which may be a vector of functions, is assumed to be a real function for the sake of simplicity. As mentioned before plausible candidates for Δ are the following: (i) the difference in the means of the two censored samples, so that $\Delta = E(X^0) - \theta_0$ and $\psi(X^0, \theta_0, \Delta) = X^0 - \theta_0 - \Delta$, where $\theta_0 = EY^0$; (ii) the difference of probabilities for the events $X^0 < t_0$ and $Y^0 < t_0$ for a given constant t_0 , so that $\Delta = F(t_0) - G_{\theta_0}(t_0)$ and $\psi(X^0, \theta_0, \Delta) = I(X^0 < t_0) - G_{\theta_0}(t_0) - \Delta$; or (iii) the probability of the event $I(X^0 > Y^0)$, so that $\Delta = E\{1 - G_{\theta_0}(X^0)\}$ and $\psi(X^0, \theta_0, \Delta) = 1 - G_{\theta_0}(X^0) - \Delta$. Another candidate is (iv) a value on the receiver

operating characteristic curve, so that $\Delta = 1 - F\{G_{\theta_0}^{-1}(1-p)\}$ and $\psi(X^0, \theta_0, \Delta) = 1 - I\{X^0 \leq G_{\theta_0}^{-1}(1-p)\} - \Delta$ for a given $p \in (0, 1)$.

For the case of complete data, Qin (1997) used an empirical likelihood ratio statistic to test the hypothesis that the two populations have the same mean and constructed a confidence interval for the mean difference of two populations in a semiparametric model. This case and Li et al.'s (1999) work on the vertical quantile comparison function for two-sample tests can be viewed as special cases of $\psi(\cdot, \cdot, \cdot)$.

2.2 Estimation of Δ based on the common estimating equation

A naive and direct method of estimating Δ is to solve equation (2.1) with F and θ replaced by estimates, which may be obtained by maximum likelihood estimation.

The likelihood function based on the samples $\{(X_i, \delta_i), i = 1, 2, \dots, n\}$ and $\{(Y_j, \eta_j), j = 1, 2, \dots, m\}$ is

$$L(F, \theta) = \prod_{i=1}^n \{F(X_i) - F(X_i-)\}^{\delta_i} \{1 - F(X_i)\}^{1-\delta_i} \prod_{j=1}^m g_{\theta}^{\eta_j}(Y_j) \{1 - G_{\theta}(Y_j)\}^{1-\eta_j},$$

provided $Q(v)$ does not depend on θ . We assume that $\theta \in \Theta \subset R^d$. It can be shown that $L(F, \theta)$ attains its maximum value at the point $(\widehat{F}, \widehat{\theta}_{MLE})$, where $\widehat{\theta}_{MLE}$ is the maximum likelihood estimator based on the sample (Y_j, η_j) , and $\widehat{F}$ is the Kaplan-Meier estimator:

$$\widehat{F}(t) = 1 - \prod_{u \leq t} \left\{ 1 - \frac{dN(u)}{Y(u)} \right\},$$

where $N(u) = \sum I(X_i \leq u, \delta_i = 1)$ and $Y(u) = \sum I(X_i \geq u)$. The estimator of Δ , denoted by $\widehat{\Delta}$, is defined as the solution of the estimating equation $\int_0^{\tau_1} \psi(x, \widehat{\theta}_{MLE}, \Delta) d\widehat{F}(x) = 0$, which is equivalent to

$$\sum_{i=1}^n \frac{\delta_i}{\widehat{K}(X_i)} \psi(X_i, \widehat{\theta}_{MLE}, \Delta) = 0, \quad (2.2)$$

where $\widehat{K}$ is the Kaplan-Meier estimator of K , i.e.,

$$\widehat{K}(t) = \prod_{u \leq t} \left\{ 1 - \frac{dN^c(u)}{Y(u)} \right\},$$

in which $N^c(u) = \sum I(X_i \leq u, \delta_i = 0)$, and $\tau_1 = \sup\{t : S(t)K(t) > 0\}$.

We assume that $m/n \rightarrow \zeta > 0$ as $n, m \rightarrow \infty$. The asymptotic normality of the estimator $\widehat{\Delta}$ is given in the following theorem.

THEOREM 1—Suppose that Assumptions 1-4 in the Appendix hold. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n W_{ni} \psi(X_i, \widehat{\theta}_{\text{MLE}}, \Delta) \rightarrow N(0, \sigma_\psi^2),$$

in distribution, where $W_{ni} = \delta_i / \widehat{K}(X_i)$, $\sigma_\psi^2 = \sigma^2(\theta, \Delta) + \beta_0^T(\theta, \Delta) \Sigma^{-1} \beta_0(\theta, \Delta)$, $\Sigma = \zeta^{-1} I(\theta)$ with $I(\theta)$

$$\sigma^2(\theta, \Delta) = \text{var}\{\psi(X, \theta, \Delta)\}$$

given in Assumption 2, $+E \int_0^{\tau_1} \{\psi(X, \theta, \Delta) - G(\theta, u)\}^2 I(X \geq u) \lambda^c(u) / K(u) du$, $G(\theta, u)$

$$= S^{-1}(u) E\{\psi(X, \theta, \Delta) I(X \geq u)\}, \beta_0(\theta, \Delta) = \int_0^{\tau_1} \dot{\psi}_\theta(u, \theta, \Delta) dF(u), \text{ and}$$

$\dot{\psi}_\theta(u, \theta, \Delta) = \partial \psi(x, \theta, \Delta) / \partial \theta$, $\lambda^c(t)$ is the hazard function of U . Furthermore, if Assumption 5 holds, then

$$\sqrt{n}(\widehat{\Delta} - \Delta) \rightarrow N(0, \sigma_\Delta^2),$$

in distribution, where $\sigma_\Delta^2 = \sigma_\psi^2 \gamma^{-2}$, and $\gamma(\theta, \Delta) = E\{\dot{\psi}_\Delta(X, \theta, \Delta)\}$, in which

$$\dot{\psi}_\Delta(x, \theta, \Delta) = \partial \psi(x, \theta, \Delta) / \partial \Delta.$$

One application of Theorem 1 is to construct the confidence interval for Δ , for which we need to estimate σ_Δ^2 . This can be done by using a plug-in method. Let

$$\begin{aligned} \widehat{\sigma}_1^2(\widehat{\theta}_{\text{MLE}}, \widehat{\Delta}) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \psi^2(X_i, \widehat{\theta}_{\text{MLE}}, \widehat{\Delta})}{\widehat{K}(X_i)} \\ &\quad + \frac{1}{n} \int_0^\Lambda \frac{dN^c(u)}{\widehat{K}^2(u)} \{ \widehat{G}_1(\widehat{\theta}_{\text{MLE}}, u) - \widehat{G}_2(\widehat{\theta}_{\text{MLE}}, u) \}, \end{aligned}$$

where Λ is the maximum of the uncensored observations,

$$\widehat{G}_1(\theta, u) = \frac{1}{n \widehat{S}(u)} \sum_{i=1}^n \frac{\delta_i \psi^2(X_i, \theta, \widehat{\Delta}) I(X_i \geq u)}{\widehat{K}(X_i)}, \quad (2.3)$$

$$\widehat{G}_2(\theta, u) = \left\{ \frac{1}{n \widehat{S}(u)} \sum_{i=1}^n \frac{\delta_i \psi(X_i, \theta, \widehat{\Delta}) I(X_i \geq u)}{\widehat{K}(X_i)} \right\}^2, \quad (2.4)$$

$$\begin{aligned}
 & \widehat{\beta}_0(\widehat{\theta}_{\text{MLE}}, \widehat{\Delta}) \\
 &= \frac{1}{n} \sum_{i=1}^n W_{ni} \dot{\psi}_{\theta}(X_i, \widehat{\theta}_{\text{MLE}}, \widehat{\Delta}) \\
 & \quad , \widehat{\gamma}(\widehat{\theta}_{\text{MLE}}, \widehat{\Delta}) \\
 &= \frac{1}{n} \sum_{i=1}^n W_{ni} \dot{\psi}_{\Delta}(X_i, \widehat{\theta}_{\text{MLE}}, \widehat{\Delta}).
 \end{aligned}
 \tag{2.5}$$

Replace $\sigma^2(\theta, \Delta)$, $\beta_0(\theta, \Delta)$, $\gamma(\theta, \Delta)$ and Σ in σ_{Δ}^2 by their estimators $\widehat{\sigma}_{\Delta}^2(\widehat{\theta}_{\text{MLE}}, \widehat{\Delta})$, $\widehat{\beta}_0(\widehat{\theta}_{\text{MLE}}, \widehat{\Delta})$, $\widehat{\gamma}(\widehat{\theta}_{\text{MLE}}, \widehat{\Delta})$ and $\widehat{\Sigma} = -\zeta/n\partial^2\ell_g(\widehat{\theta}_{\text{MLE}})/\partial\theta^2$, to obtain an estimator $\widehat{\sigma}_{\Delta}^2$ of σ_{Δ}^2 . It is easy to prove that the estimator is consistent. As a consequence, the approximate $100(1-\alpha)\%$ confidence interval of Δ is $(\widehat{\Delta} - z_{1-\alpha/2}\sqrt{\widehat{\sigma}_{\Delta}^2}, \widehat{\Delta} + z_{1-\alpha/2}\sqrt{\widehat{\sigma}_{\Delta}^2})$, where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution.

2.3 Empirical likelihood inference

In this section, we apply the empirical likelihood principle to obtain a confidence interval for Δ . To deal with the censoring we propose an adjusted empirical likelihood function and prove that the resulting log likelihood ratio is still asymptotically chi-squared under mild conditions.

Let θ_0 be the true value of parameter θ . Note that the auxiliary information depends only on the unknown distributions F and K and the parameter θ_0 , since $E\{\delta\psi(X, \theta_0, \Delta)/K(X-)\} = 0$. If the function $K(\cdot)$ were known, then we could infer about Δ based on the observation sample $\{\delta_i/K(X_i-), i = 1, 2, \dots, n\}$ and the sample $\{(Y_j, \eta_j), j = 1, 2, \dots, m\}$. Let F_p be the distribution function that assigns probabilities p_i at points $\delta_i/K(X_i-)$. Then the adjusted empirical likelihood is $L_{\text{adj}}(\theta, \Delta) = \prod_{i=1}^n p_i \prod_{j=1}^m g_{\theta}^{\eta_j}(Y_j) \{1 - G_{\theta}(Y_j)\}^{1-\eta_j}$, where $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for each i , and the auxiliary information (2.1) is expressed as

$$\sum_{i=1}^n \frac{\delta_i}{K(X_i-)} \psi(X_i, \theta, \Delta) = 0.
 \tag{2.6}$$

Again, we replace $K(\cdot)$ with its Kaplan-Meier estimator $\widehat{K}(\cdot)$, and our adjusted empirical likelihood, evaluated at Δ and θ , is defined by $L_{\text{adj}}(\theta, \Delta)$ subject to the constraints

$$\begin{aligned}
 & \sum_{i=1}^n p_i = 1, p_i \geq 0 \\
 & \quad , \text{ for } i=1, 2, \dots, n,
 \end{aligned}
 \tag{2.7}$$

$$\sum_{i=1}^n \frac{p_i \delta_i}{\widehat{K}(X_i-)} \psi(X_i, \theta, \Delta) = 0.
 \tag{2.8}$$

Without the restrictions (2.7) and (2.8), the adjusted empirical likelihood has the maximum value $n^{-n} \left[\prod_{j=1}^m g_{\hat{\theta}_{MLE}}^{\eta_j}(Y_j) \left\{ 1 - G_{\hat{\theta}_{MLE}}(Y_j) \right\}^{1-\eta_j} \right]$, where $\hat{\theta}_{MLE}$ is the maximum likelihood estimate of θ based on the sample $\{(Y_j, \eta_j), j = 1, 2, \dots, m\}$, as in §2.2. Our adjusted semiparametric empirical likelihood ratio statistic is

$$R(\Delta) = \sup_{\mathcal{P}, \theta} \left[\prod_{i=1}^n (np_i) \prod_{j=1}^m g_{\theta}^{\eta_j}(Y_j) \left\{ 1 - G_{\theta}(Y_j) \right\}^{1-\eta_j} \right] \left[\prod_{j=1}^m g_{\hat{\theta}_{MLE}}^{\eta_j}(Y_j) \left\{ 1 - G_{\hat{\theta}_{MLE}}(Y_j) \right\}^{1-\eta_j} \right]^{-1},$$

where $\mathcal{P}$ denotes the region of possible p_i subject to the constraints (2.7) and (2.8). We first investigate $R(\Delta)$ for a given θ , which we denote by $R(\theta, \Delta)$. The empirical log likelihood ratio

$$\mathcal{L}(\theta, \Delta) = \sup_{\mathcal{P}} \sum_{i=1}^n \log(np_i)$$

of $R(\theta, \Delta)$ is described as $+\ell_g(\theta) - \ell_g(\hat{\theta}_{MLE})$, where

$\ell_g(\theta) = \sum_{j=1}^m \log \left[g_{\theta}^{\eta_j}(Y_j) \left\{ 1 - G_{\theta}(Y_j) \right\}^{1-\eta_j} \right]$. Provided that $\sum_{i=1}^n n\psi^2(X_i, \theta, \Delta) W_{ni}^2 > 0$ and the origin is inside the convex hull of $\{\psi(X_i, \theta, \Delta) W_{ni}, i = 1, 2, \dots, n\}$, where $W_{ni} = \delta_i / \widehat{K}(X_i -)$, there exists a unique maximum value of $\mathcal{L}(\theta, \Delta)$ or $R(\theta, \Delta)$ (Qin & Lawless, 1994). By the Lagrange multiplier method, we conclude that $\mathcal{L}(\theta, \Delta)$ attains its maximum value at $\widehat{p}_i = n^{-1} \{1 + \lambda(\theta) W_{ni} \psi(X_i, \theta, \Delta)\}^{-1}$, for $i = 1, 2, \dots, n$. Alternatively, the estimated empirical log likelihood ratio is

$$\begin{aligned} \mathcal{L}(\theta, \Delta) &= \sum_{i=1}^n \log \{1 + \lambda(\theta) W_{ni} \psi(X_i, \theta, \Delta)\} \\ &\quad + \ell_g(\theta) - \ell_g(\hat{\theta}_{MLE}), \end{aligned} \quad (2.9)$$

and the adjusted empirical log likelihood is defined by

$$\ell_{EL}(\theta, \Delta) = \sum_{i=1}^n \log \{1 + \lambda(\theta) W_{ni} \psi(X_i, \theta, \Delta)\} + \ell_g(\theta), \quad (2.10)$$

where $\lambda(\theta)$ is determined by the equation:

$$\frac{1}{n} \sum_{i=1}^n \frac{\psi(X_i, \theta, \Delta) W_{ni}}{1 + \lambda(\theta) W_{ni} \psi(X_i, \theta, \Delta)} = 0. \quad (2.11)$$

In the following, we use $\|\cdot\|$ to denote the Euclidean norm.

THEOREM 2—Suppose that Assumptions 1-4 in the Appendix hold. Then (i) $\ell_{EL}(\theta, \Delta)$ attains its maximum value at some point $\widehat{\theta}_{EL}$ in the interior of the ball $\|\theta - \theta_0\| < n^{-\varrho}$ ($1/3 < \varrho < 1/2$) such that $\widehat{\theta}_{EL}$ and $\widehat{\lambda} = \lambda(\widehat{\theta}_{EL})$ satisfy $\partial \ell_{EL}(\widehat{\theta}_{EL}, \Delta) / \partial \theta = 0$ and (2.11); and (ii) the maximiser $\widehat{\theta}_{EL}$ of $R(\theta, \Delta)$ is strongly consistent and such that $2\rho(\theta, \Delta) \mathcal{L}(\widehat{\theta}_{EL}, \Delta) \rightarrow \chi_{(1)}^2$ in distribution, where

$$\rho(\theta, \Delta) = \frac{\sigma_1^2(\theta, \Delta) + \beta_0^T(\theta, \Delta) \Sigma^{-1} \beta_0(\theta, \Delta)}{\sigma_0^2(\theta, \Delta) + \beta_0^T(\theta, \Delta) \Sigma^{-1} \beta_0(\theta, \Delta)},$$

$$\sigma_0^2(\theta, \Delta) = \int_0^{\tau_1} \psi^2(u, \theta, \Delta)$$

in which $\Sigma = \zeta^{-1}I(\theta)$ with $I(\theta)$ given in Theorem 1, and $\int_0^{\tau_1} \psi^2(u, \theta, \Delta) dF(u)$.

Were it not for the term $\rho(\theta, \Delta)$, the confidence interval could be easily derived without estimating variance as in traditional empirical likelihood (Owen, 2000, p.23). With censorship, the situation becomes complex, and we need to estimate $\rho(\theta, \Delta)$. As in §2.2, we replace $\sigma^2(\theta, \Delta)$, $\beta_0(\theta, \Delta)$ and Σ in $\rho(\theta, \Delta)$ by their estimators $\widehat{\sigma}^2(\widehat{\theta}_{EL}, \widehat{\Delta})$ and $\widehat{\beta}_0(\widehat{\theta}_{EL}, \widehat{\Delta})$, which are defined in a similar way to (2.3)-(2.5), and $\widehat{\Sigma} = -(\zeta/n) \partial^2 \ell_g(\widehat{\theta}_{EL}) / \partial \theta^2$. The quantity $\sigma_0^2(\theta, \Delta)$ is estimated by $\widehat{\sigma}_0^2(\widehat{\theta}_{EL}, \widehat{\Delta})$

by $= n^{-1} \sum_{i=1}^n \{W_{ni} \psi(X_i, \widehat{\theta}_{EL}, \widehat{\Delta})\}^2$, where $\widehat{\Delta}$ is a consistent estimator of Δ . The resulting $\widehat{\rho}(\widehat{\theta}_{EL}, \widehat{\Delta})$ is a consistent estimator of $\rho(\theta, \Delta)$. The confidence interval of Δ based on empirical likelihood can then be established. Another approach is not to estimate Δ in the adjusted factor $\rho(\theta, \Delta)$. A confidence interval can then be constructed on the basis of the quantity $2\rho(\widehat{\theta}_{EL}, \Delta) \mathcal{L}(\widehat{\theta}, \Delta)$.

Corollary 1: Suppose that the Assumptions of Theorem 2 hold. Then

$\lim_{n \rightarrow \infty} \text{pr}(\Delta \in I_\alpha) = 1 - \alpha$, where $I_\alpha = \{\Delta: 2\rho(\widehat{\theta}_{EL}, \Delta) \mathcal{L}(\widehat{\theta}, \Delta) \leq c_\alpha\}$, and $\text{pr}(\chi_{(1)}^2 \leq c_\alpha) = 1 - \alpha$.

Theorem 2 showed that the maximum point is close to the true value, and the adjusted empirical log likelihood ratio is distributed approximately as chi-squared at the maximum point. One way of obtaining this maximum point is to consider $\partial \ell_{EL}(\theta, \Delta) / \partial \theta = 0$, which results in the score equation of semiparametric empirical likelihood, i.e.,

$$\begin{aligned} \frac{1}{n} \lambda(\theta) \sum_{i=1}^n \frac{\dot{\psi}_\theta(X_i, \theta, \Delta) W_{ni}}{1 + \lambda(\theta) \psi(X_i, \theta, \Delta) W_{ni}} \\ = \frac{1}{m} \sum_{j=1}^m \frac{\partial \log [g_\theta^{\eta_j}(Y_j) \{1 - G_\theta(Y_j)\}^{\eta_j}]}{\partial \theta}. \end{aligned} \quad (2.12)$$

There may be multiple solutions of (2.12) but, if (2.12) has a unique root, this root is $\widehat{\theta}_{EL}$.

An obvious result holds when we use $\widehat{\theta}_{MLE}$ in (2.9), as stated in the following theorem.

THEOREM 3—Suppose that Assumptions 1-4 hold. Then $2r(\theta, \Delta) \mathcal{L}(\widehat{\theta}_{MLE}, \Delta) \rightarrow \chi_{(1)}^2$, in distribution, where $r(\theta, \Delta) = \sigma_1^2(\theta, \Delta) / \sigma_0^2(\theta, \Delta)$.

Theorem 3 can be proved in a similar way to Theorem 2 by using Lemma A.3 and Taylor expansion. We omit the details.

Theorem 3 indicates that the empirical likelihood function based on the estimator $\widehat{\theta}_{MLE}$ is still asymptotically chi-squared, but its asymptotic variance is bigger than that of $\widehat{\theta}_{EL}$ (see Lemma A.5). A confidence interval is available, based on the result of Theorem 3.

3 Simulation Study

In our simulation experiments, we consider only one-dimensional θ for the sake of simplicity. We mainly compare the lengths and coverages of confidence intervals based on $\widehat{\theta}_{EL}$ and the estimating equation. Note that closed-form solutions of (2.11) and (2.12) rarely exist, and we use the Newton-Raphson algorithm for implementation. Codes in S-plus are developed which are modified from El.s, a function written by A.B.Owen.

Example 1

Let X^0 and Y^0 be independent $Un(0, 6)$ and $Un(0, 10)$ random variables. The censored random variables U and V are generated by the uniform distributions $Un(0, c_x)$ and $Un(0, c_y)$, respectively. The parameters c_x and c_y are used to control the rate of censored data; the larger the c_x and c_y , the lighter the censorship. The four choices used for (c_x, c_y) are a combination of $c_x = 12.5$ and $c_x = 15$ with $c_y = 15$ and $c_y = 20$. We are concerned with semiparametric inference for $\Delta = E(X) - \theta_0 = -2$. The simulation study is conducted as follows. For each of the 4 censoring configurations, sample sizes $(n, m) = (30, 30)$, $(50, 50)$ and $(100, 100)$ are considered. For each censoring configuration and sample size, 2000 independent sets of data are generated. The confidence interval is based on the 95% level, and Table 1 summarises the results. The confidence interval for Δ based on the adjusted empirical likelihood is almost symmetric about the true value. Coverages based on the empirical likelihood method are consistently close to the nominal level, and the coverages of the normal approximation are all larger than that of the empirical likelihood method and the nominal level. The lengths of the confidence intervals are inversely proportional to square root of the sample size and censoring degree of sample X , and the confidence interval based on the empirical likelihood method is consistently and substantially shorter than its normal counterpart.

Example 2

We use a scenario similar to that in Example 1, except that X^0 and Y^0 are exponential random variables with density functions $1/6\exp(-x/6)$ and $1/10\exp(-x/10)$, respectively. Here, we consider $\Delta = \text{pr}(X^0 > Y^0)$. Consequently, $\theta_0 = 10$, $\psi(x, \theta_0, \Delta) = \exp(-x/\theta_0) - \Delta$, and $\Delta = 0.625$. The simulation results are given in Table 2. The conclusions are the same as those in Example 1, in general. However, the confidence interval based on the empirical likelihood method is not symmetric about the true value in this example. These confidence intervals also depend more upon the censorship of X than did those in Example 1.

4 Real Data Analysis

Mantel et al. (1977) reported a litter-matched study of the tumourigenesis of a drug. One rat was randomly selected from each of 50 litters and given the drug. Two rats from each litter were selected as controls and given a placebo; see Mantel et al. (1977) for the details of the study. We are interested in comparing the average survival times of the treatment and control groups. Let X and Y be the survival times of the treatment and control groups. We first calculate the Kaplan-Meier estimate of the survival distribution of Y . A Q-Q plot then shows that the exponential-distribution assumption for Y is appropriate. Consequently, the estimated average survival time of the control group is 476, and the estimate of the difference of the averages, $\Delta = E(X) - E(Y)$, of the survival times is -240. The 95% confidence intervals for Δ based on the normal approximation and our empirical likelihood method are $(-676.5, 195.4)$ and $(-258.8, -68.7)$, respectively. The conclusions drawn from these two confidence intervals are puzzling

because the first interval suggests that there is no significant difference, but the second interval implies the contrary. Recalling the performance of the two methods in the simulation experiment, we prefer the conclusion based on the empirical likelihood method. For comparison, we also conducted a log-rank test on the dataset and obtained a p -value less than 10^{-3} , which reinforces our conclusion based on the empirical likelihood method. Moreover, the 99% confidence interval (empirical likelihood) is $(-304.7, -47.4)$, which indicates a significant difference at the 1% significance level.

5 Conclusion

The finite-sample behaviour of our method seems superior to that of the common estimating equation. Furthermore implementation of the normal-approximation method requires $\psi(x, \theta_0, \Delta)$ to be continuously differentiable in Δ , whereas, this assumption is unnecessary for constructing the confidence interval in the empirical likelihood framework. In addition, in absence of censoring, the factors $\rho(\theta, \Delta)$ and $r(\theta, \Delta)$ equal 1, so that the confidence interval based on empirical likelihood does not require the calculation of any covariance or variance. The factors $\rho(\theta, \Delta)$ and $r(\theta, \Delta)$ can therefore be viewed as the cost of censoring.

If the function $\psi(x, \theta_0, \Delta)$ in equation (2.1) is discontinuous at θ_0 , then neither the normal-approximation-based method nor the empirical likelihood method is applicable even in the absence of censoring. In particular, this happens if $\psi(x, \theta_0, \Delta)$ includes the term $I\{X^0 < G_{\theta_0}(p)\}$. Hence, the current version is not applicable to the problems of two-sample quantile comparison, and inference for the ROC curve. In principle, the method proposed in this paper can be extended to cover these cases by incorporating smoothing techniques. Detailed investigation of these issues will be discussed in further research.

Acknowledgments

The authors thank the Editor, an associate editor and a referee for their helpful suggestions and constructive comments. Zhou's research was partially supported by the National Natural Science Foundation of China. Liang's research was partially supported by the American Lebanese Syrian Associated Charities and a grant from the National Institute of Allergy and Infectious Diseases.

APPENDIX

Theory

Assumption 1

We assume that $\tau_S \leq \tau_K$, where $\tau_S = \sup\{x : S(x) > 0\}$ and $\tau_K = \sup\{x : K(x) > 0\}$, and

$$\int_0^{\tau_1} \frac{\psi^2(u, \theta, \Delta)}{K(u)} dF(u) < \infty.$$

Assumption 2

The functions $\psi(x, \theta, \Delta)$ and $\dot{\psi}_\theta(x, \theta, \Delta)$ are continuous and bounded by some function $M(x)$ in

a neighbourhood of the true value θ_0 such that, $\int_0^{\tau_1} \frac{M(u)}{K^2(u)} dF(u) < \infty$, and $E_\theta \dot{\psi}_\theta(X, \theta, \Delta) \neq 0$.

Assumption 3

The density function $g_\theta(y)$ is three times differentiable with respect to θ on $A = \{y : g_\theta(y) > 0\}$, which is assumed to be independent of θ . There exists a function $M_1(y)$ such that, for any $y \in A$ and θ in the neighbourhood of θ_0 , $E_\theta |M_1(Y)| < \infty$.

Assumption 4

The information matrix $I(\theta)$ with entries

$$I_{ij} = - \int_0^{\tau_2} \frac{\partial^2 \log g_\theta(y)}{\partial \theta_i \partial \theta_j} \{1 - Q(y)\} g_\theta(y) dy \\ - \int_0^{\tau_2} \frac{\partial^2 \log \{1 - G_\theta(y)\}}{\partial \theta_i \partial \theta_j} \{1 - G_\theta(y)\} dQ(y),$$

for $i, j = 1, \dots, d$, which are continuous, is positive definite. Here $\tau_2 = \sup\{t : D(t)Q(t) > 0\}$.

Assumption 5

The function $\psi(x, \theta, \Delta)$ is continuously differentiable in Δ , and $E\{\dot{\psi}_\Delta(X, \theta, \Delta)\} \neq 0$.

We next state preliminary lemmas. Their detailed proofs are referred to Zhou and Liang (2004).

Lemma A.1—Under Assumptions 1 and 2, we have that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n W_{ni} \psi(X_i, \theta_0, \Delta) \rightarrow N\{0, \sigma^2(\theta_0, \Delta)\},$$

in distribution, where $\sigma^2(\theta_0, \Delta)$ is defined in Theorem 2.

Lemma A.2—Suppose that Assumptions 1, 2 and 5 hold. Then

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\psi(X_i, \theta, \Delta) \delta_i}{\bar{K}(X_{i-})} \right\}^2 = \sigma_0^2(\theta, \Delta) + o_p(1), \\ \frac{1}{n} \sum_{i=1}^n \frac{\dot{\psi}_\theta(X_i, \theta, \Delta) \delta_i}{\bar{K}(X_{i-})} = \beta_0(\theta, \Delta) + o_p(1), \text{ and } \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \dot{\psi}_\Delta(X_i, \theta, \Delta)}{\bar{K}(X_{i-})} = \gamma(\theta, \Delta) + o_p(1).$$

Lemma A.3—Suppose that Assumptions 1-4 hold. Then $\lambda(\theta) = O_p(n^{-\varrho})$ uniformly on $\{\theta : \|\theta - \theta_0\| \leq cn^{-\varrho}\}$, where $1/3 < \varrho < 1/2$, c is some constant, and

$$\lambda(\theta) = \frac{\sum_{i=1}^n \psi(X_i, \theta, \Delta) W_{ni}}{\sum_{i=1}^n \{\psi(X_i, \theta, \Delta) W_{ni}\}^2} \\ + o_p(n^{-\varrho})$$

uniformly on $\{\theta : \|\theta - \theta_0\| \leq cn^{-\varrho}\}$, where $\lambda(\theta)$ satisfies (2.11).

Lemma A.4—Assume that the conditions of Theorem 2 hold, and that ϱ is defined as in Lemma A.1. Then $R(\theta, \Delta)$ attains its local maximum value at $\widehat{\theta}_{\text{EL}}$ such that $\|\widehat{\theta}_{\text{EL}} - \theta_0\| = O(n^{-\varrho})$, almost surely and $\widehat{\theta}_{\text{EL}}$ is a root of (2.12).

Lemma A.5—Suppose that Assumptions 1-4 hold. Then $\sqrt{n} \begin{pmatrix} \widehat{\theta}_{\text{EL}} - \theta_0 \\ \widehat{\lambda} \end{pmatrix} \rightarrow N(0, \Sigma_1)$, in distribution, where

$$\Sigma_1 = \begin{pmatrix} C_\theta \sigma^2 \Sigma^{-1} \beta_0 \beta_0^T \Sigma^{-1} + (\Sigma - \beta_0 \sigma_0^2 \beta_0^T)^{-1} \Sigma (\Sigma - \beta_0 \sigma_0^2 \beta_0^T)^{-1} & * \\ C_\theta \beta_0^T (\Sigma - \beta_0 \sigma_0^2 \beta_0^T)^{-1} - C_\theta^2 \sigma^2 \beta_0^T \Sigma^{-1} & C_\theta^2 \sigma^2 + C_\theta^2 \beta_0^T \sigma_0^2 \beta_0 \end{pmatrix},$$

$C_\theta = (\sigma_0^2 + \beta_0^T \Sigma^{-1} \beta_0)^{-1}$, $\sigma_0^2 = \sigma_0^2(\theta, \Delta)$ and $\beta_0^2 = \beta_0^2(\theta, \Delta)$ as given in Theorem 2.

Proof of Theorem 1: By Assumption 5, we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i, \widehat{\theta}_{\text{MLE}}, \Delta) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{ni} \psi(X_i, \theta, \Delta) \\ &+ \gamma(\theta, \Delta) \sqrt{n} (\widehat{\theta}_{\text{MLE}} - \theta) + o_p(1). \end{aligned}$$

Theorem 1 follows from Lemmas A.1 and A.2 .

Proof of Theorem 2: The result in (i) is a direct consequence of Lemma A.3 . We now prove result (ii). By using Lemma A.4 and Taylor expansion, we can show that

$$\begin{aligned} 2\mathcal{L}(\widehat{\theta}_{\text{EL}}, \Delta) &= -2\widehat{\lambda} \sum_{i=1}^n \psi(X_i, \widehat{\theta}_{\text{EL}}, \Delta) W_{ni} + \widehat{\lambda}^2 \sum_{i=1}^n \{\psi(X_i, \widehat{\theta}_{\text{EL}}, \Delta) W_{ni}\}^2 \\ &- 2 \frac{\partial \ell_g^T(\widehat{\theta}_{\text{MLE}})}{\partial \theta} (\widehat{\theta}_{\text{EL}} - \widehat{\theta}_{\text{MLE}}) + (\widehat{\theta}_{\text{EL}} - \widehat{\theta}_{\text{MLE}})^T \frac{\partial^2 \ell_g(\widehat{\theta}_{\text{MLE}})}{\partial \theta \partial \theta^T} (\widehat{\theta}_{\text{EL}} - \widehat{\theta}_{\text{MLE}}) + o_p(1). \end{aligned} \quad (\text{A.1})$$

Note that (2.11) implies that

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \psi(X_i, \widehat{\theta}_{\text{EL}}, \Delta) W_{ni} \\ &= \frac{\widehat{\lambda}}{n} \sum_{i=1}^n \{\psi(X_i, \widehat{\theta}_{\text{EL}}, \Delta) W_{ni}\}^2 \\ &+ o_p(n^{-\varrho}); \end{aligned} \quad (\text{A.2})$$

see Lemma A.3 . Recall that $\widehat{\theta}_{\text{MLE}}$ is the maximiser of $\ell_g(\theta)$ and that $\partial \ell_g(\widehat{\theta}_{\text{MLE}}) / \partial \theta = 0$. This implies that

$$\begin{aligned} & \frac{1}{n} \frac{\partial \ell_g^T(\widehat{\theta}_{EL})}{\partial \theta} \\ &= \frac{1}{n} \frac{\partial^2 \ell_g^T(\widehat{\theta}_{MLE})}{\partial \theta \partial \theta^T} (\widehat{\theta}_{EL} - \widehat{\theta}_{MLE}) \\ &+ o_p(n^{-\varrho}). \end{aligned}$$

On the other hand, (2.12) and Lemma A.2 yield that

$$\begin{aligned} & \frac{1}{n} \frac{\partial \ell_g(\widehat{\theta}_{EL})}{\partial \theta} \\ &= \frac{\widehat{\lambda}}{n} \sum_{i=1}^n \frac{\dot{\psi}_\theta(X_i, \widehat{\theta}_{EL}, \Delta) W_{ni}}{1 + \widehat{\lambda} \psi(X_i, \widehat{\theta}_{EL}, \Delta) W_{ni}} \\ &= \widehat{\lambda} \beta_0 + o_p(1). \end{aligned}$$

As a result,

$$\widehat{\theta}_{EL} - \widehat{\theta}_{MLE} = \widehat{\lambda} \left\{ \frac{1}{n} \frac{\partial^2 \ell_g(\widehat{\theta}_{MLE})}{\partial \theta \partial \theta^T} \right\}^{-1} \beta_0 + o_p(1). \quad (\text{A.3})$$

$$\begin{aligned} & 1/n \partial^2 \ell_g(\widehat{\theta}_{MLE}) \\ & / \partial \theta \partial \theta^T \rightarrow \end{aligned}$$

Noting the fact that $-\zeta^{-1}I(\theta)$, together with (A.1)-(A.3), we obtain that

$$\begin{aligned} & 2\mathcal{L}(\widehat{\theta}_{EL}, \Delta) \\ &= (\sqrt{n\lambda})^2 \left[\frac{1}{n} \sum_{i=1}^n \left\{ \psi(X_i, \widehat{\theta}_{EL}, \Delta) W_{ni} \right\}^2 + \beta_0^T \Sigma^{-1} \beta_0 \right] + o_p(1). \end{aligned}$$

Using Lemmas A.2 and A.5, we complete the proof of Theorem 2.

References

- Brownie C, Habicht JP, Cogill B. Comparing indicators of health or nutritional status. *Am. J. Epidemiol* 1986;124:1031–44. [PubMed: 3776968]
- Campbell G, Ratnaparkhi MV. An application of Lomax distributions in receiver operating characteristic (ROC) curve analysis. *Comp. Statist* 1993;22:1681–97.
- Gastwirth JL, Wang JL. Control percentile test for censored data. *J. Statist. Plan. Infer* 1988;18:267–76.
- Goddard MJ, Hineberg I. Receiver operating characteristic (ROC) curves and non-normal data: an empirical study. *Statist. Med* 1990;9:325–37.
- Hall P, La Scala B. Methodology and algorithms of empirical likelihood. *Int. Statist. Rev* 1990;58:109–27.
- Hsieh F, Turnbull BW. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist* 1996;24:25–40.

- Hollander, M.; Korwar, RW. Nonparametric Bayesian estimation of the horizontal distance between two populations. In: Gnedenko, BV.; Puri, ML.; Vincze, I., editors. *Nonparametric Statistical Inference I*. North Holland; New York: 1982. p. 409-16.
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc* 1958;53:457–81.
- Li G, Tiwari RC, Wells MT. Quantile comparison function in two-sample problems: with applications to comparisons of diagnostic markers. *J. Am. Statist. Assoc* 1996;91:689–98.
- Li G, Tiwari RC, Wells MT. Semiparametric inference for a quantile comparison function with application to receiver operating characteristic curves. *Biometrika* 1999;86:487–502.
- Mantel N, Bohidar NR, Ciminera JL. Mantel-Haenszel analysis of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Res* 1977;37:3863–8. [PubMed: 561656]
- Owen AB. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 1988;75:237–49.
- Owen AB. Empirical likelihood confidence regions. *Ann. Statist* 1990;18:90–120.
- Owen, AB. *Empirical Likelihood*. Chapman and Hall; London: 2000.
- Qin J. Semi-empirical likelihood ratio confidence intervals for the difference of two sample means. *Ann. Statist* 1994;46:117–26.
- Qin J. Empirical likelihood ratio based confidence intervals for mixture proportions. *Ann. Statist* 1999;27:1368–84.
- Qin J, Lawless J. Empirical likelihood and general estimating equations. *Ann. Statist* 1994;22:300–25.
- Thomas DR, Grunkemeier GL. Confidence interval estimation of survival probability for censored data. *J. Am. Statist. Assoc* 1975;70:865–71.
- Zhou, Y.; Liang, H. Discussion paper. Department of Biostatistics, St Jude Children's Research Hospital; Memphis, USA: 2004. Empirical-likelihood-based semiparametric inference for the treatment effect in the two-sample problem with censoring.

Table 1

Example 1. Confidence intervals and the corresponding coverage based on the empirical likelihood (EL) and normal approximation (Norm) methods

Sample size			CI(95%)		coverage	
<i>n</i>	<i>m</i>	ϵ_x ϵ_y	EL	Norm	EL	Norm
30	30	12.5 15	(-2.686, -1.311)	(-3.153, -0.847)	94.46	97.39
		20	(-2.682, -1.311)	(-3.354, -0.646)	94.13	98.60
	15	15	(-2.652, -1.313)	(-3.106, -0.894)	94.66	97.72
50	30	20	(-2.688, -1.347)	(-3.399, -0.601)	95.28	98.15
		15	(-2.531, -1.457)	(-2.753, -1.247)	95.02	97.84
	15	20	(-2.534, -1.46)	(-2.904, -1.096)	95.68	98.54
100	30	15	(-2.522, -1.469)	(-2.751, -1.249)	95.07	97.53
		20	(-2.530, -1.477)	(-2.926, -1.074)	94.9	98.34
	15	15	(-2.386, -1.619)	(-2.442, -1.558)	94.35	96.60
	100	20	(-2.379, -1.613)	(-2.519, -1.481)	93.78	97.64
		15	(-2.374, -1.625)	(-2.433, -1.567)	94.45	97.35
		20	(-2.369, -1.619)	(-2.498, -1.502)	94.98	97.69

CI(95%), approximate 95% confidence interval.

Table 2

Example 2. Confidence intervals and the corresponding coverage based on the empirical likelihood (EL) and normal approximation (Norm) methods

Sample size		CI(95%)			coverage	
		ϵ_x	ϵ_y	EL	Norm	EL
30	30	12.5	15	(0.403, 0.777)	(0.245, 1.005)	96.90
			20	(0.411, 0.785)	(0.274, 0.976)	97.20
	15	15	15	(0.46, 0.792)	(0.254, 0.996)	98.05
50	50	12.5	15	(0.455, 0.789)	(0.282, 0.968)	98.35
			20	(0.460, 0.755)	(0.333, 0.917)	96.55
	15	15	15	(0.461, 0.758)	(0.355, 0.895)	97.40
100	100	12.5	15	(0.506, 0.769)	(0.341, 0.909)	98.15
			20	(0.506, 0.768)	(0.363, 0.887)	98.35
	15	15	15	(0.507, 0.723)	(0.420, 0.830)	96.05
			20	(0.508, 0.722)	(0.435, 0.815)	96.65
			15	(0.553, 0.742)	(0.426, 0.824)	98.10
			20	(0.551, 0.740)	(0.440, 0.810)	98.30

CI(95%), approximate 95% confidence interval.