The Annals of Statistics
1993, Vol. 21, No. 3, 1182–1196

EMPIRICAL LIKELIHOOD IN BIASED SAMPLE PROBLEMS

By Jing Qin

University of Waterloo

It is well known that we can use the likelihood ratio statistic to test hypotheses and to construct confidence intervals in full parametric models. Recently, Owen introduced the empirical likelihood method in nonparametric models. In this paper, we generalize his results to biased sample problems. A Wilks theorem leading to a likelihood ratio confidence interval for the mean is given. Some extensions, discussion and simulations are presented.

1. Introduction. Vardi (1982) discussed a nonparametric two-sample estimation problem in the presence of length bias. Consider one sample $\{x_1, x_2, \ldots, x_m\}$ from F and another sample $\{y_1, y_2, \ldots, y_n\}$ from the length-biased distribution

$$G(y) = \frac{1}{\mu}\int_0^y x\,dF(x), \qquad y \geq 0.$$

Here $\mu = \int_0^\infty x\,dF(x)$, and it is assumed that $\mu < \infty$. The basic idea of a nonparametric maximum likelihood estimate (NPMLE) is to place mass only at the observed points and then maximize the resulting likelihood subject to some constraints. Vardi (1985) and Gill, Vardi and Wellner (1988) considered more general biased sampling problems. They solved the problem of estimating a distribution function based on several independent samples, each subject to a different form of selection bias. It has been shown that the NPMLE for this problem is asymptotically efficient. They also found some applications for this biased sample problem.

The empirical likelihood method for constructing confidence regions was introduced by Owen (1988, 1990). It is a nonparametric method of inference that has sampling properties similar to the bootstrap, but where the bootstrap uses resampling, it amounts to computing the profile likelihood of a general multinomial distribution which has its atoms at data points. Properties of empirical likelihood are described by Owen (1990) and others. Qin (1991) has generalized Owen's empirical likelihood to a two-sample problem, in which one sample comes from a distribution specified up to a parameter, and the other sample comes from a distribution which is unspecified. A likelihood-ratio-based confidence interval is presented for the difference of two sample means. Qin and Lawless (1991) have introduced empirical likelihood into semiparametric models, where the number of estimating equations may be greater than the

Received December 1991; revised August 1992.

AMS 1991 *subject classifications*. Primary 62E20, 62D05.

Key words and phrases. Biased sample, empirical likelihood, test of hypotheses, M-estimator, Wilks' theorem.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

number of parameters. It is shown that the maximized empirical likelihood estimates for both parameters and distribution function are asymptotically efficient. Many theorems for parametric likelihood have analogous versions.

In this paper, we will introduce empirical likelihood in the biased sampling problem. A likelihood-ratio-based confidence interval for the mean is developed. In Section 2 we give our main results. Section 3 gives some simulation results. Some discussion and further problems are presented in Section 4.

2. Main results. In this paper, since we do not wish to involve complicated identification conditions [see, e.g., Vardi (1985)], we assume a two-sample problem. One sample $s_1 = \{x_1, x_2, \dots, x_m\}$ is taken from F and another sample $s_2 = \{y_1, y_2, \dots, y_n\}$ from a weighted distribution

$$G(y) = \frac{1}{w}\int_0^y w(x)\,dF(x), \qquad w(x) \ge 0, y \ge 0,$$

where $w = \int_0^\infty w(x)\,dF(x) < \infty$. Let $N = m + n$, $k = n/N$ and $t_1, t_2, \dots, t_N$ be the combined set of observations. Here for convenience we assume that there are no ties, but an argument similar to that of Owen (1988) shows that the presence of ties does not change our results. To begin with, we are interested in giving a confidence interval for the mean $\theta = E_F(x)$. For the time being, we assume that $w(t)$ is not proportional to t. The case where $w(t)$ is proportional to t is, however, also easily handled by the same development; see the remark after Theorem 1.

The probability of our data is

$$P(\text{data}) = \prod_{t_i \in s_1} dF(t_i) \prod_{t_j \in s_2} \frac{w(t_j)\,dF(t_j)}{\int w(t)\,dF(t)}.$$

In order to find a cumulative distribution function (cdf) that maximizes P(data), it is not hard to see that we can restrict our search to the class of discrete cdf's which have positive jumps at each of the points in $s_1 \cup s_2$ and only there. Let $p_i = dF(t_i)$, $i = 1, 2, \dots, N$, and

$$L = \left\{\prod_{i=1}^N p_i\right\}\left\{\sum_{i=1}^N p_i w(t_i)\right\}^{-n}. \tag{2.1}$$

Vardi's NPMLE problem is to maximize (2.1) with respect to p_i, $i = 1, 2, \dots, N$, subject to the constraint $\sum_{i=1}^N p_i = 1$. In the following, we use $l(\cdot)$ to represent $\log L$ with various types of constraints. It would be clear from the context which arguments in $\log L$ are allowed to vary. First let $l(\hat{w})$ be defined as the maximum value of

$$\sum_{i=1}^N \log p_i - n \log w \tag{2.2}$$

with respect to p_i, $i = 1, 2, \dots, N$, and w, where p_i and w satisfy constraints

$$(2.3)\quad \sum_{i=1}^{N} w(t_i) p_i = w, \qquad \sum_{i=1}^{N} p_i = 1, \qquad p_i \geq 0, \qquad i = 1, 2, \dots, N.$$

Clearly this is equivalent to Vardi's estimate. An explicit expression for $l(\hat{w})$ can be derived by a Lagrange multiplier argument. It is easy to show that [also see Vardi (1982)]

$$(2.4)\quad l(\hat{w}) = -\sum_{i=1}^{N} \log\left\{ \frac{k w(t_i)}{\hat{w}} + (1-k) \right\} - n \log \hat{w} - N \log N,$$

where $\hat{w}$ satisfies

$$(2.5)\quad \frac{1}{N} \sum_{i=1}^{N} \frac{w(t_i) - \hat{w}}{k w(t_i)/\hat{w} + (1-k)} = 0.$$

To develop an empirical likelihood function for $\theta = E_F(x)$, we let $l(w, \theta)$ be defined as the maximum value of (2.2) with respect to p_i, $i = 1, 2, \dots, N$, where p_i, θ and w now satisfy the constraints

$$(2.6)\quad \sum_{i=1}^{N} w(t_i) p_i = w, \qquad \sum_{i=1}^{N} t_i p_i = \theta, \qquad \sum_{i=1}^{N} p_i = 1,$$
$$p_i \geq 0, \qquad i = 1, 2, \dots, N.$$

Again an explicit expression for $l(w, \theta)$ can be derived by a Lagrange multiplier argument. In order to get $l(w, \theta)$, let

$$\begin{aligned} H(p_1, \dots, p_N, w, \theta) &= \sum_{i=1}^{N} \log p_i - n \log w - N\lambda_1 \sum_{i=1}^{N} \{w(t_i) - w\} p_i \\ &\quad - N\lambda_2 \sum_{i=1}^{N} (t_i - \theta) p_i + a\left(1 - \sum_{i=1}^{N} p_i\right), \end{aligned}$$

where λ_1, λ_2 and a are Lagrange multipliers. Taking derivatives with respect to p_i, we have

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - N\lambda_1\{w(t_i) - w\} - N\lambda_2(t_i - \theta) - a = 0,$$

$$\sum_{i=1}^{N} p_i \frac{\partial H}{\partial p_i} = N - a = 0, \quad \Rightarrow \quad a = N$$

and

$$p_i = \frac{1}{N} \frac{1}{1 + \lambda_1(w(t_i) - w) + \lambda_2(t_i - \theta)}$$

with restrictions

$$(2.7)\qquad \begin{aligned} &\frac{1}{N}\sum_{i=1}^{N}\frac{w(t_i)-w}{1+\lambda_1(w(t_i)-w)+\lambda_2(t_i-\theta)}=0,\\ &\frac{1}{N}\sum_{i=1}^{N}\frac{t_i-\theta}{1+\lambda_1(w(t_i)-w)+\lambda_2(t_i-\theta)}=0. \end{aligned}$$

Hence the empirical likelihood function $l(w,\theta)$ is

$$(2.8)\qquad \begin{aligned} l(w,\theta) = &-\sum_{i=1}^{N}\log\{1+\lambda_1(w(t_i)-w)+\lambda_2(t_i-\theta)\}\\ &-n\log w - N\log N, \end{aligned}$$

where λ_1, λ_2, w and θ satisfy the constraint conditions (2.7). Let $l(\tilde{w},\theta)$ be the maximum value of $l(w,\theta)$ when θ is fixed and $l(\hat{w},\hat{\theta})$ be the maximum value of $l(w,\theta)$. It is obvious that $l(\hat{w},\hat{\theta}) = l(\hat{w})$, which is defined by (2.4). We then define the empirical likelihood ratio statistic for θ as

$$(2.9)\qquad R(\theta) = 2\{l(\hat{w}) - l(\tilde{w},\theta)\},$$

and obtain empirical likelihood confidence intervals for θ as

$$\mathscr{R}_u = \{\theta\colon R(\theta) < u\}.$$

We now develop some results that lead to a proof that when $\theta=\theta_0$, $R(\theta_0)$ has a limiting $\chi^2_{(1)}$ distribution. For convenience, we consider minus $l(w,\theta)$ without the constant term, and we let

$$(2.10)\quad l_E(w,\theta) = \sum_{i=1}^{N}\log\{1+\lambda_1(w(t_i)-w)+\lambda_2(t_i-\theta)\} + n\log w.$$

In the following, we reparametrize the parameters. Note that

$$\begin{aligned} &1+\lambda_1(w(t_i)-w)+\lambda_2(t_i-\theta)\\ &\quad= 1+\left(\lambda_1-\frac{k}{w}\right)(w(t_i)-w)+\frac{k}{w}(w(t_i)-w)+\lambda_2(t_i-\theta)\\ &\quad= \frac{k}{w}w(t_i)+(1-k)+\alpha_1(w(t_i)-w)+\alpha_2(t_i-\theta), \end{aligned}$$

where $\alpha_1 = \lambda_1 - k/w$, $\alpha_2=\lambda_2$. Thus (2.7) becomes

$$(2.11)\qquad \begin{aligned} &\frac{1}{N}\sum_{i=1}^{N}\frac{g_1(t_i,w)}{1+\alpha_1 g_1(t_i,w)+\alpha_2 g_2(t_i,w)}=0,\\ &\frac{1}{N}\sum_{i=1}^{N}\frac{g_2(t_i,w)}{1+\alpha_1 g_1(t_i,w)+\alpha_2 g_2(t_i,w)}=0 \end{aligned}$$

and

$$(2.12)\quad p_i = \frac{1}{N}\frac{1}{kw(t_i)/w + (1-k)}\frac{1}{1+\alpha_1 g_1(t_i,w) + \alpha_2 g_2(t_i,w)},$$

where

$$(2.13)\quad \begin{aligned} g_1(t_i,w) &= \frac{w(t_i) - w}{kw(t_i)/w + (1-k)},\\ g_2(t_i,w) &= \frac{t_i - \theta}{kw(t_i)/w + (1-k)}. \end{aligned}$$

We assume that the true values of w and θ are w_0 and θ_0, respectively. We now prove that (2.11) implicitly defines two functions $\alpha_1(w)$ and $\alpha_2(w)$.

LEMMA 1. *Suppose that distribution function F is nondegenerate and $E_F|x|^3 < \infty$, $E_G|x|^3 < \infty$ and $k = n/N \to k_0$, as $N \to \infty$, where $0 < k_0 < 1$. Then at the true value $\theta = \theta_0$, with probability* 1 *in the interior of the interval $|w - w_0| \le N^{-1/3}$ for N large enough, equation* (2.11) *uniquely determines $\alpha_1 = \alpha_1(w)$ and $\alpha_2 = \alpha_2(w)$. Furthermore, $\alpha_1(w)$ and $\alpha_2(w)$ are continuous and differentiable when w belongs to this interval.*

PROOF. Note that when $\theta = \theta_0$ and $w = w_0$,

$$\frac{1}{N}\sum_{i=1}^{N} g_1(t_i,w_0) = \int (w(t) - w_0)\,dF(t) + O_p(N^{-1/2}) = O_p(N^{-1/2}),$$

$$\frac{1}{N}\sum_{i=1}^{N} g_2(t_i,w_0) = \int (t - \theta_0)\,dF(t) + O_p(N^{-1/2}) = O_p(N^{-1/2}).$$

Using an argument similar to that of Owen (1990), when $w = w_0 + O_p(N^{-1/3})$ and $\theta = \theta_0$, from (2.11) we have $\alpha_1 = O_p(N^{-1/3})$, $\alpha_2 = O_p(N^{-1/3})$. By the implicit function theorem, we easily obtain the lemma's result. □

Since, after the reparametrization, the p_i's are a product of two factors besides a constant factor, we decompose $l_E(w,\theta)$ into two parts. Let

$$(2.14)\qquad l_{1E}(w) = \sum_{i=1}^{N} \log\{kw(t_i)/w + (1-k)\} + n\log w,$$

$$(2.15)\qquad l_{2E}(w,\theta_0) = \sum_{i=1}^{N} \log\{1 + \alpha_1 g_1(t_i,w) + \alpha_2 g_2(t_i,w)\}$$

and

$$(2.16)\qquad l_E(w,\theta_0) = l_{1E}(w) + l_{2E}(w,\theta_0).$$

As in Qin and Lawless (1991), we now prove that $l_E(w,\theta_0)$ has a local minimum in a small neighborhood of w_0.

LEMMA 2. *We assume the conditions in Lemma* 1 *are true. Then for* N *large enough, with probability* 1, $l_E(w, \theta_0)$ *attains a local minimum value at some point* $\tilde{w}$ *in the interior of the interval* $|w - w_0| \le N^{-1/3}$.

PROOF. This proof is similar to the proof of Theorem 4.1 in Lehmann (1983). Let $w = w_0 + uN^{-1/3}$, where $|u| = 1$. First we give a lower bound for $l_E(w, \theta_0)$ at $w = w_0 + uN^{-1/3}$, where $u = 1$ or -1. Note that

$$\frac{\partial l_{1E}(w)}{\partial w} = \sum_{i=1}^{N} \frac{-kw(t_i)/w^2}{kw(t_i)/w + (1-k)} + \frac{n}{w}$$

$$= \frac{n(k-1)}{w^2} \frac{1}{N} \sum_{i=1}^{N} g_1(t_i, w)$$

and

$$\frac{\partial^2 l_{1E}(w)}{\partial w^2} = n(k-1)\left\{ \frac{-2}{w^3} \frac{1}{N} \sum_{i=1}^{N} g_1(t_i, w) + \frac{1}{w^2} \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g_1(t_i, w)}{\partial w} \right\},$$

$$\frac{\partial^2 l_{1E}(w_0)}{\partial w^2} = n(k-1)\frac{1}{w_0^2}\left\{ -\frac{\delta}{w_0} + o(1) \right\} \quad \text{a.s.}$$

where

$$\delta = \int \frac{w(t)}{k_0 w(t)/w_0 + (1-k_0)} dF(t).$$

So

$$l_{1E}(w_0 + uN^{-1/3}) = l_{1E}(w_0) + \frac{\partial l_{1E}(w_0)}{\partial w} uN^{-1/3} + \frac{1}{2} \frac{\partial^2 l_{1E}(w^*)}{\partial w^2} u^2 N^{-2/3}$$

$$= l_{1E}(w_0) + O\big((N \log\log N)^{1/2}\big) uN^{-1/3}$$

$$+ k(1-k)\frac{\delta}{w_0^3} u^2 N^{1/3} + o(N^{1/3}) \quad \text{a.s.}$$

$$> l_{1E}(w_0) \quad \text{a.s.},$$

where w^* falls between w_0 and $w_0 + uN^{-1/3}$. Similarly, expanding $l_{2E}(w_0 + uN^{-1/3}, \theta_0)$ at w_0, we can easily show that [also see Lemma 1 in Qin and Lawless (1991)]

$$l_{2E}(w_0 + uN^{-1/3}, \theta_0) > l_{2E}(w_0, \theta_0), \quad \text{a.s.}$$

when N is large enough. Hence $l_E(w_0 + uN^{-1/3}, \theta_0) > l_E(w_0, \theta_0)$ a.s., and we have proved our claim. □

Furthermore, by (2.7) and (2.10), $\tilde{w}$ satisfies

$$\frac{\partial l_E(w,\theta_0)}{\partial w} = \sum_{i=1}^{N} \frac{-\lambda_1}{1+\lambda_1(w(t_i)-w)+\lambda_2(t_i-\theta_0)} + \frac{n}{w} = 0,$$

that is, $\lambda_1 = k/w$, which implies that $\alpha_1 = 0$, and by (2.11)–(2.16), we have estimating equations

$$Q_{1N}(w,\lambda) = \frac{1}{N}\sum_{i=1}^{N} \frac{g_1(t_i,w)}{1+\lambda g_2(t_i,w)} = 0, \tag{2.17}$$
$$Q_{2N}(w,\lambda) = \frac{1}{N}\sum_{i=1}^{N} \frac{g_2(t_i,w)}{1+\lambda g_2(t_i,w)} = 0,$$

and

$$l_E(\tilde{w}) = l_{1E}(\tilde{w}) + l_{2E}(\tilde{w}), \qquad l_{2E}(\tilde{w}) = \sum_{i=1}^{N} \log\{1+\tilde{\lambda} g_2(t_i,\tilde{w})\}, \tag{2.18}$$

where $\tilde{w}$ and $\tilde{\lambda}$ satisfy (2.17) and $g_2(t_i,w)$ is evaluated at $\theta = \theta_0$. In order to get the asymptotic distribution of the empirical likelihood ratio statistic, we need to know the asymptotic behavior of $\tilde{w}$ and $\tilde{\lambda}$.

LEMMA 3. *Under the conditions of Lemma* 1, *if H_0: $\theta = \theta_0$ is true, then*

$$\begin{pmatrix} \sqrt{N}(\tilde{w}-w_0) \\ \sqrt{N}(\tilde{\lambda}-0) \end{pmatrix} \to N(0,U), \tag{2.19}$$

where

$$U = \begin{pmatrix} \dfrac{w_0^2(w_0-\delta)}{k_0(1-k_0)\delta} - \dfrac{w_0^4\delta_1^2}{k_0^2\delta^2\eta} & 0 \\ 0 & \eta^{-1} \end{pmatrix}, \tag{2.20}$$

δ is defined in the proof of Lemma 2 *and δ_1, δ_2 and η are defined in the following proof.*

PROOF. Using Taylor's expansion for $Q_{iN}(\tilde{w},\tilde{\lambda})$, $i = 1,2$ at $(w_0, 0)$, we have

$$0 = Q_{iN}(w_0,0) + \frac{\partial Q_{iN}(w_0,0)}{\partial w}(\tilde{w}-w_0) + \frac{\partial Q_{iN}(w_0,0)}{\partial \lambda}(\tilde{\lambda}-0) + o_P(\varepsilon_N),$$

where $\varepsilon_N = |\tilde{w}-w_0| + |\tilde{\lambda}-0|$. Hence

$$\begin{pmatrix} \tilde{w}-w_0 \\ \tilde{\lambda}-0 \end{pmatrix} = -S_N^{-1}Q_N(w_0,0) + o_P(\varepsilon_N), \tag{2.21}$$

where

$$S_N = \begin{pmatrix} \dfrac{\partial Q_{1N}(w_0,0)}{\partial w} & \dfrac{\partial Q_{1N}(w_0,0)}{\partial \lambda} \\ \dfrac{\partial Q_{2N}(w_0,0)}{\partial w} & \dfrac{\partial Q_{2N}(w_0,0)}{\partial \lambda} \end{pmatrix}, \qquad Q_N(w,\lambda) = \begin{pmatrix} Q_{1N}(w,\lambda) \\ Q_{2N}(w,\lambda) \end{pmatrix}.$$

From this and $Q_N(w_0, 0) = O_P(N^{-1/2})$, we know that $\varepsilon_N = O_p(N^{-1/2})$.

Note that

$$S_N = \begin{pmatrix} \dfrac{1}{N}\sum_{i=1}^{N} \dfrac{\partial g_1(t_i, w_0)}{\partial w} & -\dfrac{1}{N}\sum_{i=1}^{N} g_1(t_i, w_0) g_2(t_i, w_0) \\ \dfrac{1}{N}\sum_{i=1}^{N} \dfrac{\partial g_2(t_i, w_0)}{\partial w} & -\dfrac{1}{N}\sum_{i=1}^{N} g_2^2(t_i, w_0) \end{pmatrix}$$

$$\to \begin{pmatrix} -\dfrac{\delta}{w_0} & \dfrac{w_0\,\delta_1}{k_0} \\ -\dfrac{(1-k_0)\delta_1}{w_0} & -\delta_2 \end{pmatrix} = S,$$

where

$$\delta_1 = \int \frac{t - \theta_0}{k_0 w(t)/w_0 + (1-k_0)}\, dF(t), \qquad \delta_2 = \int \frac{(t-\theta_0)^2}{k_0 w(t)/w_0 + (1-k_0)}\, dF(t)$$

and

$$\sqrt{N}\, Q_N(w_0, 0) \to N(0, V),$$

where

$$V = \begin{pmatrix} \dfrac{\delta(w_0 - \delta)}{k_0(1-k_0)} & -\dfrac{\delta\delta_1}{k_0} \\ -\dfrac{\delta\delta_1}{k_0} & \delta_2 - \dfrac{(1-k_0)\delta_1^2}{k_0} \end{pmatrix}.$$

Let $\eta = \delta_2 + (1-k_0)\delta_1^2 w_0/(k_0\,\delta) > 0$. Then

$$U = S^{-1}V(S^{-1})^{\tau} = \begin{pmatrix} \dfrac{w_0^2(w_0-\delta)}{k_0(1-k_0)\delta} - \dfrac{w_0^4\,\delta_1^2}{k_0^2\,\delta^2\eta} & 0 \\ 0 & \eta^{-1} \end{pmatrix}. \qquad \square$$

Now we are ready to prove our main result.

THEOREM 1. *If the conditions in Lemma 1 are true, under H_0: $\theta = \theta_0$, the empirical likelihood statistic* (2.9) *satisfies*

$$R(\theta_0) \to \chi^2_{(1)}. \tag{2.22}$$

PROOF. From (2.4), (2.5) and (2.14) it is easy to see that $l(\hat{w}) = -l_{1E}(\hat{w}) - N \log N$, where $\hat{w}$ satisfies $Q_{1N}(\hat{w}, 0) = 0$. Hence

$$\begin{aligned}\hat{w} - w_0 &= -\left(\frac{\partial Q_{1N}(w_0,0)}{\partial w}\right)^{-1} Q_{1N}(w_0,0) + o_P(N^{-1/2})\\ &= \frac{w_0}{\delta}(1,0)Q_N(w_0,0) + o_p(N^{-1/2}).\end{aligned} \tag{2.23}$$

So by (2.23) and (2.21) we have

$$\begin{aligned}\tilde{w} - \hat{w} &= -(1,0)\left(\frac{w_0}{\delta}I + S^{-1}\right)Q_N(w_0,0) + o_p(N^{-1/2})\\ &= -(1,0)\left[\begin{pmatrix} -1 & \dfrac{w_0^2\delta_1}{k\delta} \\ -\dfrac{(1-k)\delta_1}{\delta} & -\dfrac{w_0\delta_2}{\delta}\end{pmatrix} + \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right]S^{-1}Q_N(w_0,0)\\ &\qquad + o_p(N^{-1/2})\\ &= -\frac{w_0^2\delta_1}{k\delta}(0,1)S^{-1}Q_N(w_0,0) + o_p(N^{-1/2})\\ &= \frac{w_0^2\delta_1}{k\delta}\tilde{\lambda} + o_p(N^{-1/2}),\end{aligned} \tag{2.24}$$

where I is a 2×2 identity matrix. Expanding $l_{1E}(\tilde{w})$ at $\hat{w}$ and noting that

$$\frac{\partial l_{1E}(\hat{w})}{\partial w} = -\frac{n(1-k)}{\hat{w}^2}Q_{1N}(\hat{w},0) = 0,$$

we have

$$\begin{aligned}l_{1E}(\tilde{w}) - l_{1E}(\hat{w}) &= \frac{\partial l_{1E}(\hat{w})}{\partial w}(\tilde{w} - \hat{w}) + \frac{1}{2}\frac{\partial^2 l_{1E}(\hat{w})}{\partial w^2}(\tilde{w} - \hat{w})^2 + o_p(1)\\ &= \frac{1}{2}\frac{k(1-k)\delta}{w_0^3}N(\tilde{w} - \hat{w})^2 + o_p(1).\end{aligned}$$

From $Q_{2N}(\tilde{w}, \tilde{\lambda}) = 0$, we have

$$\begin{aligned}\tilde{\lambda} &= \left\{\frac{1}{N}\sum_{i=1}^{N} g_2^2(t_i, \tilde{w})\right\}^{-1}\left\{\frac{1}{N}\sum_{i=1}^{N} g_2(t_i, \tilde{w})\right\} + o_P(N^{-1/2})\\ &= \delta_2^{-1}Q_{2E}(\tilde{w}, 0) + o_p(N^{-1/2}).\end{aligned}$$

Hence

$$
\begin{aligned}
l_{2E}(\tilde{w}, \theta_0) &= \sum_{i=1}^{N} \log\{1 + \tilde{\lambda} g_2(t_i, \tilde{w})\} \\
&= \sum_{i=1}^{N} \tilde{\lambda} g_2(t_i, \tilde{w}) - \frac{1}{2} \sum_{i=1}^{N} \lambda^2 g_2^2(t_i, \tilde{w}) + o_p(1) \\
&= \frac{N}{2} \tilde{\lambda}^2 \frac{1}{N} \sum_{i=1}^{N} g_2^2(t_i, \tilde{w}) + o_p(1) \\
&= \frac{N}{2} \tilde{\lambda}^2 \delta_2 + o_p(1).
\end{aligned}
$$

So by (2.24)

$$
\begin{aligned}
l_E(\tilde{w}, \theta_0) - l_{1E}(\hat{w}) &= l_{1E}(\tilde{w}) - l_{1E}(\hat{w}) + l_{2E}(\tilde{w}, \theta_0) \\
&= \frac{1}{2} \frac{k(1-k)\delta}{w_0^3} N(\tilde{w} - \hat{w})^2 + \frac{N}{2} \tilde{\lambda}^2 \delta_2 + o_p(1) \\
&= \frac{N}{2} \left\{ \frac{k(1-k)\delta}{w_0^3} \frac{w_0^4 \delta_1^2}{k^2 \delta^2} + \delta_2 \right\} \tilde{\lambda}^2 + o_p(1) \\
&= \frac{N}{2} \eta \tilde{\lambda}^2 + o_P(1).
\end{aligned}
$$

Noting that $\sqrt{N}\tilde{\lambda} \to N(0, \eta^{-1})$, we have

$$
R(\theta_0) = 2\{l(\hat{w}) - l(\tilde{w}, \theta_0)\} = 2\{l_E(\tilde{w}, \theta_0) - l_{1E}(\hat{w})\} \to \chi^2_{(1)}. \qquad \square
$$

REMARK. When $w(t) = t$, $w = \theta$, we replace the constraints (2.6) by

$$
\sum_{i=1}^{N} t_i p_i = w, \qquad \sum_{i=1}^{N} p_i = 1, \qquad p_i \geq 0, \qquad i = 1, 2, \ldots, N.
$$

Using the same procedure as for the case of $w(t) \neq t$, we can check that the empirical likelihood statistic $R(w_0)$ also asymptotically follows a $\chi^2_{(1)}$ distribution. We omit the details.

3. Some simulation results. Empirical likelihood ratio confidence intervals make very weak distributional assumptions and are justified by having asymptotically correct coverage levels. This method can be easily extended to other functionals θ. We consider an example for median estimation based on biased samples. It is a special case of the general M-estimate, which estimates a functional θ defined by the solution to

$$
\int \psi(x, \theta)\, dF(x) = 0. \tag{3.1}
$$

Conditions must be imposed on $\psi(x, \theta)$ to guarantee existence of a solution to (3.1). Using empirical likelihood in the biased sample situation, we would then

TABLE 1

m	n	90% confidence interval			95% confidence interval		
		Cov.	Average midpoint	Average length	Cov.	Average midpoint	Average length
(a) Exp(1), Gamma(2), true value = 0.69315							
15	15	89.4	0.73023	0.68240	95.0	0.74658	0.82685
15	20	90.2	0.72214	0.65422	95.6	0.73370	0.79189
20	20	89.1	0.73335	0.58319	94.3	0.74344	0.69292
20	30	90.2	0.72910	0.56504	95.0	0.74291	0.66371
30	20	88.4	0.72486	0.51708	92.8	0.73394	0.60948
30	30	91.8	0.71772	0.49809	95.6	0.72574	0.59252
(b) Exp(1), Gamma(3), true value = 0.69315							
15	15	92.1	0.78011	0.79974	95.0	0.79614	0.97175
15	20	91.2	0.74779	0.77109	95.0	0.76848	0.92859
20	20	90.4	0.75256	0.69630	96.2	0.76597	0.84085
20	30	88.2	0.75098	0.67300	93.2	0.76345	0.80121
30	20	89.2	0.74215	0.58970	94.0	0.75645	0.69848
30	30	89.6	0.73083	0.56661	94.2	0.74570	0.67123

want to maximize (2.2) with respect to p_i, $i = 1, 2, \dots, N$, where p_i w and θ satisfy constraints

$$\sum_{i=1}^{N} \{w(t_i) - w\}p_i = 0, \qquad \sum_{i=1}^{N} \psi(t_i, \theta)p_i = 0, \qquad \sum_{i=1}^{N} p_i = 1,$$

$$p_i \geq 0, \qquad i = 1, 2, \dots, N.$$

After profiling out (i.e., maximizing over) the p_i, we have the empirical likelihood $l(w, \theta)$ as

$$l(w, \theta) = -\sum_{i=1}^{N} \log\{1 + \lambda_1(w(t_i) - w) + \lambda_2\psi(t_i, \theta)\} - n \log w - N \log N, \tag{3.2}$$

where λ_1 and λ_2 are Lagrange multipliers and are determined by

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} \frac{w(t_i) - w}{1 + \lambda_1(w(t_i) - w) + \lambda_2\psi(t_i, \theta)} = 0, \\ &\frac{1}{N}\sum_{i=1}^{N} \frac{\psi(t_i, \theta)}{1 + \lambda_1(w(t_i) - w) + \lambda_2\psi(t_i, \theta)} = 0. \end{aligned} \tag{3.3}$$

Then we define the empirical likelihood ratio statistic as (2.9), and we can obtain confidence intervals for θ via the empirical likelihood ratio statistic. If we let

$$\psi(x, \theta) = 1(x \le \theta) - \frac{\gamma}{1-\gamma} 1(x > \theta),$$

then θ is the γ quantile of $F(x)$. Following Owen's (1988, 1990) arguments, we can get a confidence interval for θ. We do not give details here.

In the following, some simulations have been performed. We generated length-biased two-sample data by using the S language, and we considered estimation of a median. We assumed that the unbiased sample comes from the standard exponential distribution, and the other from the length-biased exponential distributions Gamma(2) and Gamma(3) which arise when $w(x) = x, x^2$,

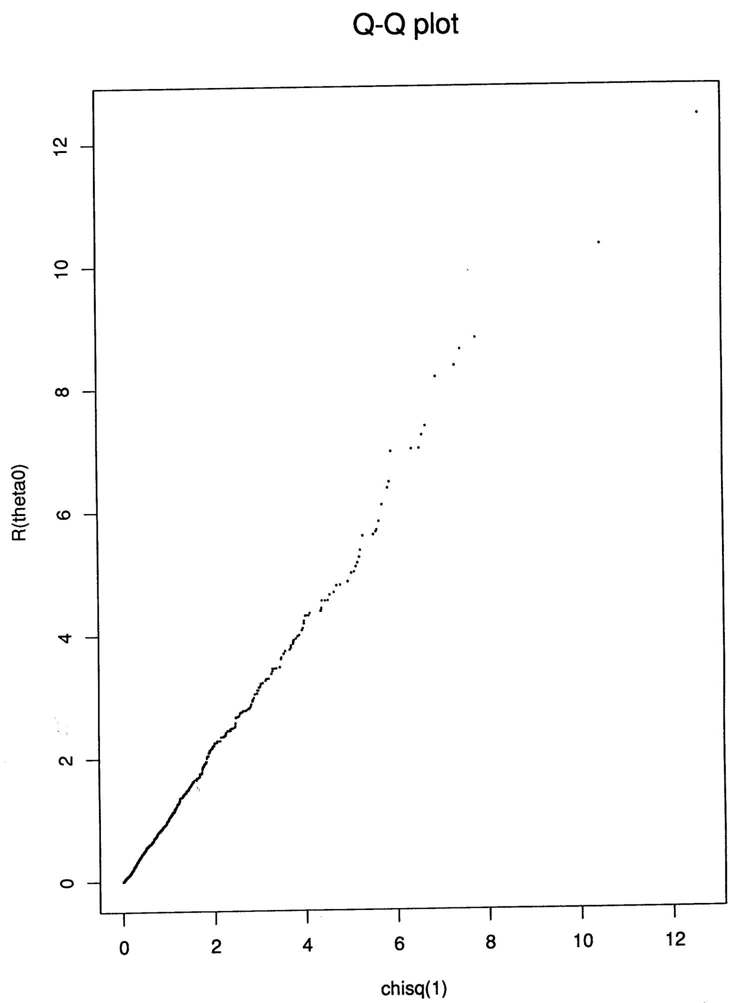

FIG. 1.

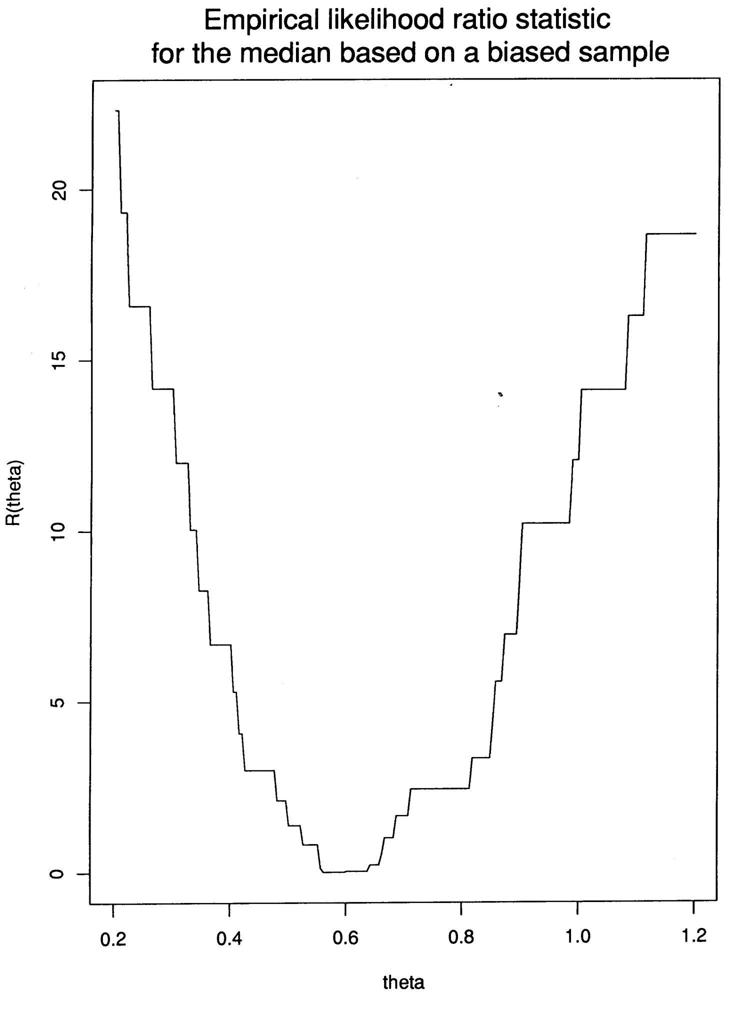

FIG. 2.

respectively. From each sample, the 90 and 95% empirical likelihood confidence intervals were computed.

In Table 1, we report the estimated true coverage, mean value of midpoint and mean length of the empirical likelihood confidence intervals for the median. Each value in the table is the average of 1000 simulations. The true coverage level is close to the nominal level. For the less biased distribution Gamma(2) [compared with Gamma(3)], the average lengths are shorter. Figure 1 is the Q–Q plot for the 1000 replications of $R(\theta_0)$ versus standard $\chi^2_{(1)}$. The approximation appears satisfactory.

Figure 2 shows minus twice the log empirical likelihood ratio statistic $R(\theta)$ versus median θ based on one sample [$m = n = 30$, F is exp(1) and G is Gamma(2)].

4. Discussion and extensions. Without essential difficulty our method in this paper can be generalized to the case of several independent samples, each subject to a different form of selection bias as in Vardi (1985) and Gill, Vardi and Wellner (1988).

Observing the result in Lemma 3, we know that when $\theta = \theta_0$, that is, we know that $E_F(x) = \theta_0$, the asymptotic variance of the estimator $\tilde{w}$ is less than $w_0^2(w_0 - \delta)/\{k_0(1 - k_0)\delta\}$, which is the asymptotic variance of the estimator $\hat{w}$, when θ is unknown. Qin and Lawless (1991) have found that the distribution function estimate based on the empirical likelihood method with auxiliary information is asymptotically equivalent to Haberman's (1984) estimate which minimizes the Kullback–Leibler distance measure from an empirical distribution subject to linear constraints. A generalization of this result is to consider semiparametric models based on biased samples, where the number of estimating equations is greater than the number of parameters [see Qin and Lawless (1991) in the unbiased sample case]. Also, in some situations, $w(t) = w(t, \theta)$ can depend on some additional parameters; see Vardi (1985).

Density estimation is another interesting problem. Recently, Jones (1991) proposed a likelihood-type kernel estimate

$$\hat{f}(x) = \sum_{i=1}^{N} \hat{p}_i K_h(x - t_i),$$

where $K_h(x) = h^{-1}K(h^{-1}x)$ and $\hat{p}_i = N^{-1}\{kw(t_i)/\hat{w} + (1 - k)\}^{-1}$, where $\hat{w}$ satisfies (2.5). This estimator has various advantages over some alternative estimators. It has better asymptotic mean integrated squared error properties and it is more readily extendible to related problems such as density derivative estimation. One point worth mentioning is that when w is known, the p_i can be obtained by maximizing (2.2) with respect to p_i only, subject to (2.3). An empirical likelihood confidence band can be considered following the work of Hall and Owen (1989), beginning by maximizing (2.2) subject to constraints

$$\sum_{i=1}^{N} p_i\{w(t_i) - w\} = 0, \qquad \sum_{i=1}^{N} p_i\{K_h(x - t_i) - f(x)\} = 0, \qquad \sum_{i=1}^{N} p_i = 1,$$
$$p_i \geq 0, \qquad i = 1, 2, \dots, N.$$

Finally, higher-order expansions need to be developed to compare the empirical likelihood confidence interval and the usual confidence interval and to assess the difference between parametric likelihood and empirical likelihood in the biased sample problems. We will consider these topics in future communications.

Acknowledgments. In my thesis proposal at University of Waterloo, I mentioned the possibility of using empirical likelihood in the biased sample problem. Recent conversation with Professor Art Owen has encouraged me to attack this. Many thanks to Professors J. F. Lawless, A. B. Owen and M. E. Thompson for useful suggestions. The detailed comments made by the referees improved the presentation of this paper.

REFERENCES

GILL, R. D., VARDI, Y. and WELLNER, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16** 1069–1112.

HABERMAN, S. J. (1984). Adjustment by minimum discriminant information. *Ann. Statist.* **12** 971–988.

HALL, P. and OWEN, A. B. (1989). Empirical likelihood confidence bands in curve estimation. Technical Report 324, Stanford Univ.

JONES, M. C. (1991). Kernel density estimation for length biased data. *Biometrika* **78** 511–519.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

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

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

QIN, J. (1991). Likelihood and empirical likelihood ratio confidence intervals in two sample semiparametric models. Technical Report 91-6, Univ. Waterloo.

QIN, J. and LAWLESS, J. F. (1991). Empirical likelihood and general estimating equations. Technical Report 91-10, Univ. Waterloo.

VARDI, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10** 616–620.

VARDI, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–203.

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9** 60–62.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO
CANADA N2L-3G1