Ann. Inst. Statist. Math.
Vol. 46, No. 1, 117–126 (1994)

SEMI-EMPIRICAL LIKELIHOOD RATIO CONFIDENCE INTERVALS FOR THE DIFFERENCE OF TWO SAMPLE MEANS

JING QIN

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

(Received February 18, 1993; revised June 2, 1993)

Abstract. We all know that we can use the likelihood ratio statistic to test hypotheses and construct confidence intervals in full parametric models. Recently, Owen (1988, *Biometrika*, **75**, 237–249; 1990, *Ann. Statist.*, **18**, 90–120) has introduced the empirical likelihood method in nonparametric models. In this paper, we combine these two likelihoods together and use the likelihood ratio to construct confidence intervals in a semiparametric problem, in which one model is parametric, and the other is nonparametric. A version of Wilks's theorem is developed.

Key words and phrases: Empirical likelihood, hypotheses tests, semi-empirical likelihood, Wilks's theorem.

1. Introduction

A problem arising in many different contexts is the comparison of two treatments or of one treatment with a control situation in which no treatment is applied. If the observations consist of the number of successes in a sequence of trials for each treatment, for example the number of cures of a certain disease, the problem becomes that of testing the equality of two binomial probabilities. In some cases, however, we don't know or perhaps only partially know the underlying distribution, but we still want to compare the two treatments.

Consider $N = n + m$ independent measurements in two samples. The first sample consists of n measurements $x_1, x_2, \dots, x_n$ recorded under one set of conditions, and the second sample consists of m measurements $y_1, y_2, \dots, y_m$ recorded under a different set of conditions. For instance, the x's might be blood pressure increases for n subjects who received drug A, while the y's are increases for m different subjects who received drug B. The problem is to compare the two population means, i.e. test $\mu_A = \mu_B$, or give a confidence interval for the difference of the two means $\Delta = \mu_A - \mu_B$. Suppose that based on our experience, we are quite sure of y's distributional form up to one parameter, say $G_\theta(y)$ (maybe the drug B has been used a long time), but for new drug A, it is hard to say x's distributional form, so we have no knowledge about x's distribution $F(x)$. How can we test

$H_0 : \mu_A = \mu_B$, or give a confidence interval for $\Delta = \mu_A - \mu_B$? In other words, we need to test the equality of population means based on one parametric model and one nonparametric model.

We all know that we can use the likelihood ratio statistic to test hypotheses and construct confidence intervals in full parametric models. Recently, Owen (1988, 1990) has introduced the empirical likelihood method in nonparametric models. In this paper, we will combine these two likelihoods together and develop a likelihood ratio test and confidence intervals for this semiparametric two sample problem. In Section 2 we briefly describe the empirical likelihood developed by Owen (1988, 1990). In Section 3, we give our main results. Section 4 gives some proofs. Section 5 presents some limited simulation results.

2. Empirical likelihood

The empirical likelihood method for constructing confidence regions was introduced by Owen (1988, 1990). It is a nonparametric method of inference. It 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.

Consider a random sample $x_1, x_2, \ldots, x_n$ of size n drawn from an unknown r-variate distribution F_0 having mean μ_0 and nonsingular covariance matrix Σ_0. Denote the j-th component as x_i^j $(j = 1, \ldots, r)$, so that $x_i = (x_i^1, \ldots, x_i^r)^\tau$. Let L be the empirical likelihood function for the mean. For a specific vector $\mu = (\mu^1, \ldots, \mu^r)^\tau$, $L(\mu)$ is defined to be the maximum value of $\prod p_i$ over all vectors $p = (p_1, \ldots, p_n)$ that satisfy the constraints

$$(2.1) \qquad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} x_i p_i = \mu, \qquad p_i \geq 0, \quad i = 1, \ldots, n.$$

An explicit expression for $L(\mu)$ can be derived by a Lagrange multiplier argument. The maximum of $\prod_{i=1}^n p_i$ subject to (2.1) is attained when

$$(2.2) \qquad p_i = p_i(\mu) = n^{-1}\{1 + t^\tau(x_i - \mu)\}^{-1}$$

where $t = t(\mu)$ is an r-dimensional column vector given by

$$(2.3) \qquad \sum_{i=1}^{n} \{1 + t^\tau(x_i - \mu)\}^{-1}(x_i - \mu) = 0.$$

Since $\prod_{i=1}^n p_i$ attains its largest value over all vectors $p = (p_1, \ldots, p_n)$ satisfying $\sum_{i=1}^n p_i = 1$ when $p_i = n^{-1}$ $(i = 1, \ldots, n)$, it follows that the empirical likelihood function $L(\mu)$ is maximized at $\hat{\mu} = \bar{x} = n^{-1}\sum_{i=1}^n x_i$ and $L(\hat{\mu}) = n^{-n}$. The empirical likelihood ratio at the point μ is

$$(2.4) \qquad \frac{L(\mu)}{L(\hat{\mu})} = \prod_{i=1}^{n} \{1 + t^\tau(x_i - \mu)\}^{-1}$$

and minus twice the logarithm of this ratio is

$$W(\mu) = 2\sum_{i=1}^{n} \log\{1 + t^{\tau}(X_i - \mu)\}. \tag{2.5}$$

Under appropriate regularity conditions, Owen (1988, 1990) has proved that a version of Wilks's theorem holds, i.e. under $H_0 : \mu = \mu_0$, $W(\mu_0) \to \chi_r^2$.

There is an obvious extension of this to construct confidence interval for the difference Δ of two sample means. Let $x_1, x_2, \ldots, x_n$ be independently and identically distributed random variables with distribution function $F(x)$ and $y_1, y_2, \ldots, y_m$ be independently and identically distributed random variables with distribution function $G(y)$, where both $F(x)$ and $G(y)$ are unknown. We define $EL(\Delta)$ as the maximum value of $\prod_{i=1}^{n} p_i \prod_{j=1}^{m} q_j$ subject to constraints

$$p_i \geq 0, \qquad q_j \geq 0, \qquad \sum_{i=1}^{n} p_i = 1, \qquad \sum_{j=1}^{m} q_j = 1 \qquad \text{and}$$
$$\sum_{i=1}^{n} p_i x_i - \sum_{j=1}^{m} q_j y_j = \Delta.$$

The empirical likelihood ratio statistic for Δ is

$$EW(\Delta) = -2\log\left\{EL(\Delta)/\max_{\Delta} EL(\Delta)\right\}. \tag{2.6}$$

Easily we can show that under $H_0 : \Delta = \Delta_0$, the true difference, $EW(\Delta_0) \to \chi_{(1)}^2$. We do not give the details here.

3. Semi-empirical likelihood and main results

It is well known that likelihood based confidence intervals and tests perform well in parametric models. Owen's empirical likelihood ratio confidence interval can be used in nonparametric models. In this section, we consider a semi-empirical likelihood based confidence intervals for the difference of two means. Let $x_1, x_2, \ldots, x_n$; $y_1, y_2, \ldots, y_m$ be independent and suppose the x_i are identically distributed as unknown $F(x)$ with mean $\mu_1 = \int x dF(x)$ and the y_j are identically distributed as $G_\theta(y)$ with mean $\mu_2 = \int y dG_\theta(y) = \mu(\theta)$, where $G_\theta(y)$ is of known form depending on parameter θ. We assume that $G_\theta(y)$ has density function $g_\theta(y)$. The problem is to test $H_0 : \mu_1 = \mu_2 = \mu(\theta)$, or give a confidence interval for $\Delta_0 = \mu_1 - \mu(\theta)$. The semi-empirical likelihood function is

$$\prod_{i=1}^{n} dF(x_i) \prod_{j=1}^{m} g_\theta(y_j).$$

It has maximum value $n^{-n}\prod_{j=1}^m g_{\hat\theta}(y_j)$, where $\hat\theta$ is the MLE based on the second sample. Let

$$R(F,\theta)=\frac{\prod_{i=1}^n dF(x_i)\prod_{j=1}^m g_\theta(y_j)}{n^{-n}\prod_{j=1}^m g_{\hat\theta}(y_j)},$$
$$C_{r,n}=\left\{\int xdF-\mu(\theta)\mid F\ll F_n, R(F,\theta)\ge r\right\},$$
$$\mathcal{R}(\Delta)=\sup_{F,\theta}\left\{R(F,\theta)\mid \int xdF-\mu(\theta)=\Delta, F\ll F_n\right\},$$

where $F\ll F_n$ denotes that F is absolutely continuous w.r.t F_n, i.e. the support of F is contained in the support of empirical distribution F_n. Then $\Delta\in C_{r,n}$ if and only if $\mathcal{R}(\Delta)\ge r$. We want to show that $-2\log\mathcal{R}(\Delta_0)\to\chi^2_{(1)}$. Without loss of generality, we assume that $\Delta_0=0$, and

$$\mathcal{R}(0)=\sup_{p_i,\theta}\prod_{i=1}^n(np_i)\prod_{j=1}^m g_\theta(y_j)\left[\prod_{j=1}^m g_{\hat\theta}(y_j)\right]^{-1}.$$

We first maximize the joint likelihood with restriction $\mu_1=\mu_2$, i.e.

$$\max_{p_1,\dots,p_n,\theta}\left\{\sum_{i=1}^n\log p_i+\sum_{j=1}^m\log g_\theta(y_j)\,\middle|\, p_i>0,\sum_i p_i=1,\sum_i p_ix_i=\mu(\theta)\right\}. \tag{3.1}$$

Let

$$H=\sum_{i=1}^n\log p_i+\sum_{j=1}^m\log g_\theta(y_j)+\gamma\left(1-\sum_i p_i\right)+n\lambda\left(\mu(\theta)-\sum_i p_ix_i\right),$$

then

$$\frac{\partial H}{\partial p_i}=p_i^{-1}-\gamma-n\lambda x_i=0,\Rightarrow p_i=\frac{1}{\gamma+n\lambda x_i},$$
$$0=\sum_i p_i\frac{\partial H}{\partial p_i}=n-\gamma-n\lambda\sum_i p_ix_i\Rightarrow\gamma=n-n\lambda\mu(\theta),\qquad\text{or}$$

$$p_i=\frac{1}{n}\frac{1}{1+\lambda(x_i-\mu(\theta))}.$$

Also,

$$\frac{\partial H}{\partial\theta}=\sum_j\frac{\partial\log g_\theta(y_j)}{\partial\theta}+n\lambda\mu'(\theta)=0,\qquad\text{i.e.}$$

$$\lambda(\theta)=-\frac{\sum_j\frac{\partial\log g_\theta(y_j)}{\partial\theta}}{n\mu'(\theta)}=-\frac{\frac{\partial l_2(\theta)}{\partial\theta}}{\mu'(\theta)} \tag{3.2}$$

where, $l_2 = (1/n)\sum_{j=1}^{m} \log g_\theta(y_j)$. By the side condition $\sum_{i=1}^{n} p_i(x_i - \mu(\theta)) = 0$, we have,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{x_i - \mu(\theta)}{1 + \lambda(\theta)(x_i - \mu(\theta))} = 0. \tag{3.3}$$

We will prove that there exists a root $\tilde{\theta}$ of this equation such that the root lies within an $O_p(n^{-1/2})$ neighborhood of the true value θ_0 when n is large enough.

In the following, we will make assumptions on the distribution $G_\theta(y)$ which coincide with the conditions of normality of the MLE in full parametric models, given in Lehmann (1983).

ASSUMPTIONS. (i) The parameter space Ω is an open interval.

(ii) The distributions $G_\theta(y)$ of y_j have common support, so that the set $A = \{y : g_\theta(y) > 0\}$ is independent of θ.

(iii) For every $y \in A$, the density $g_\theta(y)$ is differentiable three times with respect to θ.

(iv) The integral $\int g_\theta(y)dy$ can be twice differentiated under the integral sign.

(v) The Fisher information $I(\theta) = E[\partial \log g_\theta/\partial\theta]^2$ satisfies $0 < I(\theta) < \infty$.

(vi) $|(\partial^3/\partial\theta^3) \log g_\theta(y)| < M(y)$, for all $y \in A$, $\theta_0 - c < \theta < \theta_0 + c$, with $E_{\theta_0}[M(y)] < \infty$.

THEOREM 3.1. *If $F(x)$ is a nondegenerate distribution function with $\int |x|^3 dF < \infty$, $\mu(\theta)$ is continuously differentiable at θ_0 with $\mu'(\theta_0) \neq 0$, g_θ satisfies the above assumptions* (i) *through* (vi), *and $n/m \to \gamma > 0$ as $n, m \to \infty$, then the* log *semi-empirical likelihood ratio statistic under the null hypothesis $\Delta_0 = 0$,*

$$\log \mathcal{R}(0) = \sum_{i=1}^{n} \log np_i(\tilde{\theta}) + \sum_{j=1}^{m} \log[g_{\tilde{\theta}}(y_j)/g_{\hat{\theta}}(y_j)]$$

satisfies $-2\log \mathcal{R}(0) \to \chi^2_{(1)}$ and $\lim_{n\to\infty} P(\Delta_0 \in C_{r,n}) = P(\chi^2_{(1)} \leq -2\log r)$.

4. Proofs

First we give a lemma.

LEMMA 4.1. *Under the conditions of Theorem* 3.1, *there exists a root $\tilde{\theta}$ of* (3.3), *such that $\tilde{\theta} - \theta_0 = O_p(n^{-1/2})$.*

PROOF. Let

$$\begin{aligned} h(\theta) &= \frac{1}{n}\sum_i \frac{x_i - \mu(\theta)}{1 + \lambda(\theta)(x_i - \mu(\theta))} \\ &= \bar{x} - \mu(\theta) - \lambda(\theta)\frac{1}{n}\sum_i \frac{(x_i - \mu(\theta))^2}{1 + \lambda(\theta)(x_i - \mu(\theta))}. \end{aligned}$$

First we prove that $h(\theta)=0$ has a root in an $O_p(n^{-q})$ neighborhood of θ_0, where $1/3<q<1/2$. In fact, note that

$$\frac{\partial l_2(\theta)}{\partial\theta}=\frac{\partial l_2(\theta_0)}{\partial\theta}+(\theta-\theta_0)\frac{\partial^2 l_2(\theta_0)}{\partial\theta^2}+\frac{1}{2}(\theta-\theta_0)^2\frac{\partial^3 l_2(\theta^*)}{\partial\theta^3}$$

where θ^* lies between θ_0 and θ. Since $\partial l_2(\theta_0)/\partial\theta=O_p(n^{-1/2})$, we have $\lambda(\theta)=O_p(n^{-q})$ when $\theta\in(\theta_0-n^{-q},\theta_0+n^{-q})$. By the assumption $E|x|^3<\infty$, we have $\max_{1\le i\le n}|x_i|<n^{1/3}$ for all but finitely many n, which implies $\lambda(\theta)(x_i-\mu(\theta))=o_p(1)$ in the interval $(\theta_0-n^{-q},\theta_0+n^{-q})$. Note that $h(\theta)$ is almost surely continuous in this interval for n large enough, and consider the signs of $n^q h(\theta_0+n^{-q})$ and $n^q h(\theta_0-n^{-q})$ for large n.

$$\begin{aligned}h(\theta_0+n^{-q})&=\bar{x}-\mu(\theta_0+n^{-q})\\&\quad-\lambda(\theta_0+n^{-q})\frac{1}{n}\sum_i\frac{(x_i-\mu(\theta_0+n^{-q}))^2}{1+\lambda(\theta_0+n^{-q})(x_i-\mu(\theta_0+n^{-q}))}.\end{aligned}$$

Note also that

$$\begin{aligned}\mu(\theta_0+n^{-q})&=\mu(\theta_0)+\mu'(\theta_0)n^{-q}+o(n^{-q}),\\ \lambda(\theta_0+n^{-q})&=-\left.\frac{\partial l_2(\theta_0+n^{-q})}{\partial\theta}\right/\mu'(\theta_0+n^{-q})\\&=-\left.\left[\frac{\partial l_2(\theta_0)}{\partial\theta}+\frac{\partial^2 l_2(\theta_0)}{\partial\theta^2}n^{-q}+o_p(n^{-q})\right]\right/\mu'(\theta_0+n^{-q})\\&=O_p(n^{-1/2})+[I_2/\mu'(\theta_0)]n^{-q}+o_p(n^{-q}),\\ \frac{1}{n}\sum_i\frac{(x_i-\mu(\theta_0+n^{-q}))^2}{1+\lambda(\theta_0+n^{-q})(x_i-\mu(\theta_0+n^{-q}))}&=S^2+o_p(1),\qquad\text{where}\end{aligned}$$

$$\frac{1}{n}\sum_i(x_i-\mu(\theta_0))^2=S^2,\qquad I_2=E\left(-\frac{\partial^2 l_2(\theta_0)}{\partial\theta^2}\right)>0,$$

so that

$$\begin{aligned}n^q h(\theta_0+n^{-q})&=n^q O_p(n^{-1/2})-\mu'(\theta_0)-I_2S^2/\mu'(\theta_0)+o_p(1)\\&=-(\mu'^2(\theta_0)+I_2S^2)(\mu'(\theta_0))^{-1}+o_p(1).\end{aligned}$$

Similarly

$$n^q h(\theta_0-n^{-q})=(\mu'^2(\theta_0)+I_2S^2)(\mu'(\theta_0))^{-1}+o_p(1),$$

i.e. $n^q h(\theta_0+n^{-q})$ and $n^q h(\theta_0-n^{-q})$ have opposite sign for large n. By the intermediate value theorem, there exists a root $\tilde\theta$ in $(\theta_0-n^{-q},\theta_0+n^{-q})$. Similar to the above argument, we have

$$0=h(\tilde\theta)=O_p(n^{-1/2})-(\mu'^2(\theta_0)+I_2S^2)(\mu'(\theta_0))^{-1}(\tilde\theta-\theta_0)+o_p(\tilde\theta-\theta_0),$$

i.e. $\tilde\theta-\theta_0=O_p(n^{-1/2})$. □

PROOF OF THEOREM 3.1. Note that $l_2(\theta) = (1/n)\sum_{j=1}^{m} \log g_\theta(y_j)$. Using a Taylor expansion, we have

$$l_2(\hat\theta) - l_2(\tilde\theta) = \frac{\partial l_2(\tilde\theta)}{\partial\theta}(\hat\theta - \tilde\theta) + 2^{-1}(\hat\theta - \tilde\theta)^2\frac{\partial^2 l_2(\tilde\theta)}{\partial\theta^2} + o_p(n^{-1}),$$
$$\mu(\hat\theta) - \mu(\tilde\theta) = \mu'(\tilde\theta)(\hat\theta - \tilde\theta) + o_p(n^{-1/2}).$$

From (3.2)

$$\lambda(\tilde\theta)(\mu(\tilde\theta) - \mu(\hat\theta)) = \frac{\partial l_2(\tilde\theta)}{\partial\theta}(\hat\theta - \tilde\theta) + o_p(n^{-1}). \tag{4.1}$$

Expanding $\partial l_2(\tilde\theta)/\partial\theta$ at $\theta = \hat\theta$, and noting $\partial l_2(\hat\theta)/\partial\theta = 0$, we have

$$\frac{\partial l_2(\tilde\theta)}{\partial\theta} = \frac{\partial l_2(\hat\theta)}{\partial\theta} + \frac{\partial^2 l_2(\hat\theta)}{\partial\theta^2}(\tilde\theta - \hat\theta) + o_p(n^{-1/2}), \qquad \text{i.e.}$$

$$\tilde\theta - \hat\theta = \frac{\dfrac{\partial l_2(\tilde\theta)}{\partial\theta}}{\dfrac{\partial^2 l_2(\hat\theta)}{\partial^2\theta}} + o_p(n^{-1/2}) = -\lambda(\tilde\theta)\mu'(\tilde\theta)\left[\frac{\partial^2 l_2(\hat\theta)}{\partial\theta^2}\right]^{-1} + o_p(n^{-1/2}), \tag{4.2}$$

$$\begin{aligned}\mu(\tilde\theta) - \mu(\hat\theta) &= \mu'(\tilde\theta)(\tilde\theta - \hat\theta) + o_p(n^{-1/2}) \\ &= -\lambda(\tilde\theta)(\mu'(\tilde\theta))^2\left[\frac{\partial^2 l_2(\hat\theta)}{\partial\theta^2}\right]^{-1} + o_p(n^{-1/2}).\end{aligned} \tag{4.3}$$

From (3.3) we have

$$\begin{aligned}0 &= \frac{1}{n}\sum_i \frac{x_i - \mu(\tilde\theta)}{1 - \lambda(\tilde\theta)(\mu(\tilde\theta) - x_i)} \\ &= \frac{1}{n}\sum_i [1 - \lambda(\tilde\theta)(x_i - \mu(\tilde\theta))](x_i - \mu(\tilde\theta)) + o_p(n^{-1/2}), \qquad \text{i.e.}\end{aligned}$$

$$\bar x - \mu(\tilde\theta) = \lambda(\tilde\theta)\frac{1}{n}\sum_i (x_i - \mu(\tilde\theta))^2 + o_p(n^{-1/2}). \tag{4.4}$$

From (4.3), (4.4) we have

$$\bar x - \mu(\hat\theta) = \lambda(\tilde\theta)\left[\frac{1}{n}\sum_i (x_i - \mu(\tilde\theta))^2 - \frac{\mu'^2(\tilde\theta)}{\dfrac{\partial^2 l_2(\hat\theta)}{\partial\theta^2}}\right] + o_p(n^{-1/2}), \qquad \text{or}$$

$$\lambda(\tilde{\theta}) = (\bar{x} - \mu(\hat{\theta})) \left[\frac{1}{n} \sum_i (x_i - \mu(\tilde{\theta}))^2 - \frac{\mu'^2(\tilde{\theta})}{\dfrac{\partial^2 l_2(\hat{\theta})}{\partial \theta^2}} \right]^{-1} + o_p(n^{-1/2}). \tag{4.5}$$

So the empirical log likelihood ratio statistic is

$$\begin{aligned} \log \mathcal{R}(0) &= \sum_{i=1}^n \log np_i + \sum_{j=1}^m \log[g_{\tilde{\theta}}(y_j)/g_{\hat{\theta}}(y_j)] \\ &= -\sum_i \log[1 - \lambda(\tilde{\theta})(\mu(\tilde{\theta}) - x_i)] + n[l_2(\tilde{\theta}) - l_2(\hat{\theta})]. \end{aligned}$$

Since $\log(1+x) = x - (1/2)x^2 + o(x^2)$,

$$\begin{aligned} &-\sum_i \log[1 - \lambda(\tilde{\theta})(\mu(\tilde{\theta}) - x_i)] \\ &\quad = -n\lambda(\tilde{\theta})(\bar{x} - \mu(\tilde{\theta})) + \frac{1}{2}\lambda^2(\tilde{\theta}) \sum_i (x_i - \mu(\tilde{\theta}))^2 + o_p(1). \end{aligned}$$

Expanding $l_2(\hat{\theta})$ at $\tilde{\theta}$ and noting (4.1) and (4.2), we have

$$\begin{aligned} n[l_2(\tilde{\theta}) - l_2(\hat{\theta})] &= -n \left[\frac{\partial l_2(\tilde{\theta})}{\partial \theta}(\hat{\theta} - \tilde{\theta}) + \frac{1}{2}(\hat{\theta} - \tilde{\theta})^2 \frac{\partial^2 l_2(\tilde{\theta})}{\partial \theta^2} + o_p(n^{-1}) \right] \\ &= -n\lambda(\tilde{\theta})(\mu(\tilde{\theta}) - \mu(\hat{\theta})) - \frac{1}{2}\lambda^2(\tilde{\theta}) \frac{n\mu'^2(\tilde{\theta})}{\left(\dfrac{\partial^2 l_2(\hat{\theta})}{\partial \theta^2}\right)^2} \frac{\partial^2 l_2(\tilde{\theta})}{\partial \theta^2} + o_p(1). \end{aligned}$$

Hence by (4.3)–(4.5)

$$\begin{aligned} \log \mathcal{R}(0) = &- n\lambda(\tilde{\theta})(\bar{x} - \mu(\hat{\theta})) \\ &+ \frac{1}{2}\lambda^2(\tilde{\theta}) \left[\sum_i (x_i - \mu(\tilde{\theta}))^2 - n\mu'^2(\tilde{\theta}) \Bigg/ \frac{\partial^2 l_2(\hat{\theta})}{\partial \theta^2} \right] + o_p(1) \\ = &- \frac{1}{2} n(\bar{x} - \mu(\hat{\theta}))^2 \left[\frac{1}{n} \sum_i (x_i - \mu(\tilde{\theta}))^2 - \mu'^2(\tilde{\theta}) \Bigg/ \frac{\partial^2 l_2(\hat{\theta})}{\partial \theta^2} \right]^{-1} + o_p(1). \end{aligned}$$

Under $H_0 : \mu_1 = \mu_2 = \mu(\theta_0)$,

$$\sqrt{n}(\bar{x} - \mu(\hat{\theta})) = \sqrt{n}(\bar{x} - \mu(\theta_0)) + \sqrt{n}(\mu(\theta_0) - \mu(\hat{\theta})) \to N(0, \sigma_1^2 + \gamma\sigma_2^2)$$

where, $\sigma_1^2 = \mathrm{var}(x)$, and $\sigma_2^2 = \mu'^2(\theta_0)[-1/E(\partial^2 l_2(\theta_0)/\partial\theta^2)]$, hence

$$-2 \log \mathcal{R}(0) \to \chi^2_{(1)}.$$

□

COROLLARY 4.1. *Under the conditions of Theorem* 3.1, *let*

$$\sigma^2 = \sigma_1^2 + \gamma\sigma_2^2, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \mu(\hat{\theta}))^2 - \mu'^2(\hat{\theta}) \bigg/ \frac{\partial^2 l_2(\hat{\theta})}{\partial\theta^2}$$

and let τ *be any real constant. Then* $-2\log\mathcal{R}(\bar{x} - \mu(\hat{\theta}) + \tau\hat{\sigma}n^{-1/2}) \to \tau^2$ *in probability, and*

$$-2\log\mathcal{R}(\mu_1 - \mu(\theta_0) + \tau\sigma n^{-1/2}) \to \chi^2_{(1)}(\tau^2).$$

PROOF. By a minor modification of Theorem 3.1, we note that

$$\begin{aligned}
&-2\log\mathcal{R}(\bar{x} - \mu(\hat{\theta}) + \tau\hat{\sigma}n^{-1/2}) \\
&\qquad = n\{\bar{x} - \mu(\hat{\theta}) - [\bar{x} - \mu(\hat{\theta}) + \tau\hat{\sigma}n^{-1/2}]\}^2\hat{\sigma}^{-2} + o_p(1) \\
&\qquad = \tau^2 + o_p(1) \quad \text{and} \\
&-2\log\mathcal{R}(\mu_1 - \mu(\theta_0) + \tau\sigma n^{-1/2}) \\
&\qquad = n\{\bar{x} - \mu(\hat{\theta}) - [\mu_1 - \mu(\theta_0) + \tau\sigma n^{-1/2}]\}^2\hat{\sigma}^{-2} + o_p(1), \\
&\sqrt{n}\{\bar{x} - \mu_1 - (\mu(\hat{\theta}) - \mu(\theta_0)) - \tau\sigma n^{-1/2}\} \to N(-\tau\sigma, \sigma^2), \\
&\hat{\sigma}^2 \to \sigma^2 \quad \text{in prob.}
\end{aligned}$$

□

5. Simulation results

In this section, we give some limited simulation results. We compared three methods of obtaining confidence intervals for the difference of two sample means. The first one is based on the empirical likelihood ratio statistic (ELR) in (2.6) without a distribution form assumption on $F(x)$ and $G(y)$. The second one is based on the semi-empirical likelihood ratio statistic (SLR) with parametric assumption only on $G(y)$. The third method is based on the parametric likelihood ratio statistic (PLR) with parametric assumption on both $F(x)$ and $G(y)$. We generated data by using the S language. From each sample, 90% and 95% empirical, semi-empirical and parametric likelihood ratio confidence intervals were computed. In Tables 1 and 2, we reported the estimated true coverage, mean length and mean value of midpoint of those three likelihood confidence intervals. Each value in those tables was the average of 1000 simulations. We considered the parametric models with distribution $F(x)$ from $\nu\chi^2_{(1)}$ and $G(y)$ from $\log N(\mu, 1)$ and $F(x)$ from $\exp(\theta_1)$ and $G(y)$ from $\exp(\theta_2)$ in Tables 1 and 2 respectively. From those tables we can see that the performance of the semi-empirical likelihood ratio statistic lies between empirical and parametric likelihood ratio statistics. All empirical coverage levels are close to the nominal levels when the sample size is moderately large.

Table 1. $x \sim \nu\chi^2_{(1)}$, $\nu = 1$; $y \sim \log N(\mu, 1)$, $\mu = 0$, $\Delta_0 = -0.64872$.

		90% CI			95% CI		
		Cov.	Av.length	Av.midpt.	Cov.	Av.length	Av.midpt.
	ELR	80.0	2.30610	−0.72633	87.2	2.76668	−0.75193
$n = m = 10$	SLR	86.9	2.41986	−0.84189	92.6	2.94609	−0.89449
	PLR	88.8	2.83978	−0.64783	94.7	3.62803	−0.56925
	ELR	82.7	1.77102	−0.69261	89.1	2.14627	−0.71981
$n = m = 20$	SLR	88.3	1.66106	−0.70071	93.7	2.01150	−0.71345
	PLR	89.9	1.79548	−0.64089	94.7	2.21963	−0.61735
	ELR	87.1	1.31856	−0.70201	92.3	1.60124	−0.72544
$n = m = 40$	SLR	89.6	1.15677	−0.66525	94.0	1.39479	−0.66685
	PLR	90.1	1.19361	−0.64866	94.6	1.44971	−0.64127

Table 2. $x \sim \exp(\theta_1)$, $\theta_1 = 1$; $y \sim \exp(\theta_2)$, $\theta_2 = 2$, $\Delta_0 = 0.5$.

		90% CI			95% CI		
		Cov.	Av.length	Av.midpt.	Cov.	Av.length	Av.midpt.
	ELR	83.5	1.06239	0.52632	89.0	1.26948	0.54319
$n = m = 10$	SLR	84.4	1.12938	0.49744	90.0	1.37321	0.49589
	PLR	89.9	1.34585	0.59699	95.0	1.69466	0.64915
	ELR	86.8	0.79741	0.53490	92.5	0.95957	0.54985
$n = m = 20$	SLR	87.5	0.81504	0.52805	93.1	0.98549	0.53892
	PLR	90.7	0.88261	0.55829	95.7	1.08180	0.58457
	ELR	87.7	0.57942	0.52450	92.8	0.69752	0.53450
$n = m = 40$	SLR	87.7	0.58419	0.52303	92.9	0.70385	0.53228
	PLR	88.4	0.60522	0.53149	94.0	0.73150	0.54466

Acknowledgements

The author wishes to thank Professors J. F. Lawless, A. B. Owen and the referee for many useful suggestions.

References

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

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.