

# Generalized Neyman–Pearson optimality of empirical likelihood for testing parameter hypotheses

Taisuke Otsu

Received: 1 June 2007 / Revised: 3 October 2007 / Published online: 24 April 2008  
© The Institute of Statistical Mathematics, Tokyo 2008

**Abstract** This paper studies the Generalized Neyman–Pearson (GNP) optimality of empirical likelihood-based tests for parameter hypotheses. The GNP optimality focuses on the large deviation errors of tests, i.e., the convergence rates of the type I and II error probabilities under fixed alternatives. We derive (i) the GNP optimality of the empirical likelihood criterion (ELC) test against all alternatives, and (ii) a necessary and a sufficient condition for the GNP optimality of the empirical likelihood ratio (ELR) test against each alternative.

**Keywords** Empirical likelihood · Generalized Neyman–Pearson optimality

## 1 Introduction

This paper studies the Generalized Neyman–Pearson (GNP) optimality of empirical likelihood-based tests for parameter hypotheses. Compared to the Neyman–Pearson optimality, the GNP optimality focuses on the large deviation errors of tests, i.e., the convergence rates of the type I and II error probabilities under fixed alternatives. Under some restriction on the convergence rate of the type I error probability, we maximize the convergence rates of the type II error probabilities under fixed alternatives. We derive (i) the GNP optimality of the empirical likelihood criterion (ELC) test against

---

T. Otsu would like to thank Yuichi Kitamura, Gustavo Soares, and two anonymous referees for helpful comments. Financial support from the National Science Foundation (SES-0720961) is gratefully acknowledged.

---

T. Otsu (✉)  
Cowles Foundation and Department of Economics, Yale University,  
P.O. Box 208281, New Haven, CT 06520-8281, USA  
e-mail: taisuke.otsu@yale.edu

all alternatives, and (ii) a necessary and a sufficient condition for the GNP optimality of the empirical likelihood ratio (ELR) test against each alternative.

Empirical likelihood proposed by Owen (1988) is nonparametric likelihood constructed from some moment restrictions. Although the empirical likelihood-based estimator and tests are asymptotically first-order equivalent to the generalized method of moments (GMM) estimator and tests, recent research found important differences between these methods. DiCiccio et al. (1991) and Newey and Smith (2004) showed desirable higher order properties of the empirical likelihood-based parameter hypothesis test and estimator, respectively. Kitamura (2001) considered power properties of the empirical likelihood-based overidentifying restriction test and showed that the empirical likelihood-based test has the GNP optimality against all fixed alternatives.

The motivation of this paper is as follows. Suppose that a statistician propose a moment restriction model to explain some data. The validity of the model can be checked by the empirical likelihood-based overidentifying restriction test whose GNP optimality is shown by Kitamura (2001). If the proposed moment restriction model is accepted, the next task for the statistician is to conduct some hypothesis testing for unknown parameters in the model. The main interest of the present paper is to investigate the GNP optimal properties of the empirical likelihood-based tests for the parameter hypothesis testing problem and to complement Kitamura's (2001) GNP optimality result for the overidentifying restriction testing problem.

To analyze the convergence rates of the type I and II error probabilities under fixed alternatives, large deviation theory plays a key role. See Dembo and Zeitouni (1998) for a review of large deviation theory. By utilizing these convergence rates, several efficiency or optimality criteria for tests are proposed, such as Bahadur and Chernoff efficiency. Serfling (1980, Chap. 10) provides a concise review on this topic. The notion of the GNP optimality is originally proposed by Hoeffding (1965) to compare the chi-square and likelihood ratio tests for multinomial distributions. Hoeffding's approach is extended to various setups by, e.g., Zeitouni and Gutman (1991) and Steinberg and Zeitouni (1992). Zeitouni et al. (1992) considered finite-state Markov process and derived a necessary and a sufficient condition for the GNP optimality of the likelihood ratio test. This paper extends the result of Zeitouni et al. (1992) to moment restriction models, where data can be continuously distributed and the distribution forms of the data are unspecified (i.e., semiparametric).

This paper is organized as follows. Section 2 introduces our basic setup. Section 3 presents the optimality results. Section 4 concludes. All proofs are in Appendix.

## 2 Setup

Let  $\{x_i\}_{i=1}^n$  be an iid sequence of  $d \times 1$  random vectors,  $\theta \in \Theta \subset \mathbb{R}^p$  be a  $p \times 1$  vector of unknown parameters,  $\theta_0 \in \Theta$  be the true value,  $g : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^q$  be an  $q \times 1$  vector of known functions, where  $q > p$  (overidentified). Consider the unconditional moment restriction model:

$$E[g(x, \theta_0)] = \int g(x, \theta_0) d\mu_0 = 0, \quad (1)$$

where  $\mu_0$  is the unknown true measure of  $x$ . Let  $\Theta_0$  be a subset of  $\Theta$ . Consider the composite parameter hypothesis testing problem:

$$H_0 : \theta_0 \in \Theta_0, \quad H_1 : \theta_0 \in \Theta \setminus \Theta_0. \quad (2)$$

This setup includes equality and inequality hypotheses for possibly nonlinear functions of the parameters. This paper studies optimal properties of empirical likelihood for testing  $H_0$ . From [Owen \(1988\)](#) and [Qin and Lawless \(1994\)](#), empirical likelihood at each  $\theta \in \Theta$  is defined as

$$L(\theta) = \sup_{\{p_i\}_{i=1}^n} \left\{ \prod_{i=1}^n p_i \mid p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g(x_i, \theta) = 0 \right\}. \quad (3)$$

Without the restriction  $\sum_{i=1}^n p_i g(x_i, \theta) = 0$ , the unconstrained empirical likelihood is

$$L^u = \sup_{\{p_i\}_{i=1}^n} \left\{ \prod_{i=1}^n p_i \mid p_i > 0, \sum_{i=1}^n p_i = 1 \right\} = n^{-n}.$$

For testing  $H_0$ , we consider the following two empirical likelihood-based test statistics.

(i) ELC test statistic:

$$\begin{aligned} \ell_C &= -2 \left\{ \sup_{\theta \in \Theta_0} \log L(\theta) - \log L^u \right\} \\ &= \inf_{\theta \in \Theta_0} \max_{\gamma \in \mathbb{R}^q} 2 \sum_{i=1}^n \log(1 + \gamma' g(x_i, \theta)), \end{aligned} \quad (4)$$

(ii) ELR test statistic:

$$\begin{aligned} \ell_R &= -2 \left\{ \sup_{\theta \in \Theta_0} \log L(\theta) - \sup_{\theta \in \Theta} \log L(\theta) \right\} \\ &= \inf_{\theta \in \Theta_0} \max_{\gamma \in \mathbb{R}^q} 2 \sum_{i=1}^n \log(1 + \gamma' g(x_i, \theta)) \\ &\quad - \inf_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}^q} 2 \sum_{i=1}^n \log(1 + \gamma' g(x_i, \theta)). \end{aligned} \quad (5)$$

Under  $H_0$  with certain regularity conditions,  $\ell_C$  and  $\ell_R$  converge in distribution to  $\chi^2(q)$  and  $\chi^2(p)$ , respectively ([Qin and Lawless 1994](#)). Note that the second term in (5) (i.e.,  $\ell_O = \ell_C - \ell_R$ ) is the empirical likelihood overidentifying restriction test statistic, which follows the  $\chi^2(q-p)$  limiting distribution. [Kitamura \(2001\)](#) derived the GNP optimality of  $\ell_O$  for testing the overidentifying restrictions.

We now introduce an information theoretic interpretation of empirical likelihood. Let  $\mathcal{M}$  be the space of probability measures on  $\mathbb{R}^d$ ,  $\mathcal{P}(\theta) = \{\mu \in \mathcal{M} : \int g(x, \theta) d\mu = 0\}$  be a set of measures which satisfy the moment restrictions (1) at  $\theta \in \Theta$ ,  $\mathcal{P} = \bigcup_{\theta \in \Theta} \mathcal{P}(\theta)$ , and  $\mathcal{P}_0 = \bigcup_{\theta \in \Theta_0} \mathcal{P}(\theta)$ . Based on this notation, the hypotheses  $H_0$  and  $H_1$  are written as

$$H_0 : \mu_0 \in \mathcal{P}_0, \quad H_1 : \mu_0 \in \mathcal{P} \setminus \mathcal{P}_0. \tag{6}$$

The relative entropy (or Kullback–Leibler divergence) of measures  $P$  and  $Q$  is defined as

$$I(P \parallel Q) = \int \log \left( \frac{dP}{dQ} \right) dP \quad \text{if } P \text{ is absolutely continuous with respect to } Q \\ = \infty \quad \text{otherwise.}$$

Let  $\mu_n$  be the empirical measure of  $\{x_i\}_{i=1}^n$ . It is known that

$$\frac{\ell_C}{2n} = \inf_{P \in \mathcal{P}_0} I(\mu_n \parallel P), \\ \frac{\ell_R}{2n} = \inf_{P \in \mathcal{P}_0} I(\mu_n \parallel P) - \inf_{P \in \mathcal{P}} I(\mu_n \parallel P), \tag{7}$$

i.e., the test statistics  $\ell_C$  and  $\ell_R$  are written as functions of  $\mu_n$ . Thus, the empirical likelihood-based tests can be defined by partitions of the space of measures  $\mathcal{M}$ .

(i) ELC test:

$$\text{accept } H_0 \quad \text{if } \mu_n \in \Lambda_{C1} = \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_0} I(\mu \parallel P) < \eta \right\}, \\ \text{reject} \quad \text{if } \mu_n \in \Lambda_{C2} = \mathcal{M} \setminus \Lambda_{C1}. \tag{8}$$

(ii) ELR test:

$$\text{accept } H_0 \quad \text{if } \mu_n \in \Lambda_{R1} = \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_0} I(\mu \parallel P) - \inf_{P \in \mathcal{P}} I(\mu \parallel P) < \eta' \right\}, \\ \text{reject} \quad \text{if } \mu_n \in \Lambda_{R2} = \mathcal{M} \setminus \Lambda_{R1}. \tag{9}$$

We consider a class of tests which are represented by partitions of  $\mathcal{M}$  for  $\mu_n$  and derive optimal properties of the partitions  $\Lambda_C = (\Lambda_{C1}, \Lambda_{C2})$  and  $\Lambda_R = (\Lambda_{R1}, \Lambda_{R2})$ . Let  $B(\mu, \delta)$  be an open ball of radius  $\delta \in (0, \infty)$  around  $\mu \in \mathcal{M}$ ,  $A^\delta = \cup_{\mu \in A} B(\mu, \delta)$  be a  $\delta$ -blowup of a set  $A$ , and  $P_0^n$  and  $P^n$  be the  $n$ -fold product measures of  $P_0$  and  $P$ , respectively. As a metric on  $\mathcal{M}$ , we use the Lévy metric, which is defined as

$$\rho(P_1, P_2) \equiv \inf \left\{ \epsilon > 0 : F_1(x - \epsilon \mathbf{e}) - \epsilon \leq F_2(x) \leq F_1(x + \epsilon \mathbf{e}) + \epsilon \text{ for all } x \in \mathbb{R}^d \right\},$$

for each  $P_1 \in \mathcal{M}$  and  $P_2 \in \mathcal{M}$ , where  $F_1$  and  $F_2$  are the distribution functions of  $P_1$  and  $P_2$ , respectively, and  $\mathbf{e} \equiv (1, \dots, 1)'$ . The Lévy metric is compatible with the weak topology on  $\mathcal{M}$  (see Dembo and Zeitouni (1998, Chap. D.2)). Our optimality criterion is called the Generalized Neyman–Pearson  $\delta$ -optimality (or simply the GNP optimality) by Zeitouni and Gutman (1991), which is a natural extension of the Neyman–Pearson optimality to analyze large deviation error properties of tests.

**Definition 1** (*Generalized Neyman–Pearson  $\delta$ -optimality*) A test defined by a partition  $\Lambda = (\Lambda_1, \Lambda_2)$  of  $\mathcal{M}$  is Generalized Neyman–Pearson  $\delta$ -optimal for a set  $\mathcal{A}$  if it satisfies

- (i)  $\sup_{P_0 \in \mathcal{P}_0} \limsup_{n \rightarrow \infty} n^{-1} \log P_0^n \{ \mu_n \in \Lambda_2 \} \leq -\eta;$
- (ii) for any test defined by some partition  $\Omega_n = (\Omega_{1,n}, \Omega_{2,n})$ , which satisfies

$$\sup_{P_0 \in \mathcal{P}_0} \limsup_{n \rightarrow \infty} n^{-1} \log P_0^n \{ \mu_n \in \Omega_{2,n}^\delta \} \leq -\eta \tag{10}$$

for some  $\delta \in (0, \infty)$ , we have

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_1 \} \leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Omega_{1,n} \} \tag{11}$$

for each  $P \in \mathcal{A}$ .

This definition of the GNP optimality is slightly different from the one in Kitamura (2001), where the set  $\mathcal{A}$  is to be  $\mathcal{M}$ . To analyze the optimality of the ELR test, we adopt this definition. If the test  $\Lambda = (\Lambda_1, \Lambda_2)$  is consistent, the type I (i.e.,  $P_0^n \{ \mu_n \in \Lambda_2 \}$  for  $P_0 \in \mathcal{P}_0$ ) and type II (i.e.,  $P^n \{ \mu_n \in \Lambda_1 \}$  for  $P \in \mathcal{M} \setminus \mathcal{P}_0$ ) error probabilities converge to zero. As we will see below, these convergence rates (or large deviation errors) are exponentially small. Therefore, under the restriction on the exponential convergence rate of the type I error in (10), we minimize the exponential convergence rate of the type II error in (11). However, because of a rough nature of the large deviation theorem for  $\mu_n$  (see Theorem 1 below), we need a modification by the  $\delta$ -blowup, i.e., the rival test  $\Omega_n$  must satisfy the type I error restriction (10) for the  $\delta$ -blowup of  $\Omega_{2,n}$ . We can replace  $\Omega_{2,n}^\delta$  in (10) with  $\Omega_{2,n}$  if we consider a class of tests that satisfy

$$\lim_{\delta \rightarrow 0} \sup_{P_0 \in \mathcal{P}_0} \limsup_{n \rightarrow \infty} n^{-1} \log P_0^n \{ \mu_n \in \Omega_{2,n}^\delta \} = \sup_{P_0 \in \mathcal{P}_0} \limsup_{n \rightarrow \infty} n^{-1} \log P_0^n \{ \mu_n \in \Omega_{2,n} \}$$

(see, Kitamura 2001, pp. 1665–1666).

In order to analyze the convergence rates in Definition 1, we need to know large deviation properties of the empirical measure  $\mu_n$ . For our purpose, Sanov’s theorem (see, e.g., Deuschel and Stroock 1989, Theorem 3.2.17) plays a central role.

**Theorem 1** (Sanov) *Let  $\Sigma$  be the Polish space (i.e., complete separable metric space),  $\mathcal{M}(\Sigma)$  be the space of measures on  $\Sigma$  endowed with the Lévy metric and  $P \in \mathcal{M}(\Sigma)$ . Then*

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in G \} \leq - \inf_{\nu \in G} I(\nu \| P)$$

for each closed sets  $G \subset \mathcal{M}(\Sigma)$ , and

$$\liminf_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in H\} \geq - \inf_{v \in H} I(v \| P)$$

for each open sets  $H \subset \mathcal{M}(\Sigma)$ .

Sanov's theorem says that the large deviation properties of the empirical measure  $\mu_n$  is characterized by the relative entropy  $I$ . Intuitively, if the set  $G$  (or  $H$ ) is far from the true measure  $P$  in terms of the relative entropy, the exponential convergence rate of  $P^n \{\mu_n \in G\}$  (or  $P^n \{\mu_n \in H\}$ ) is fast. Note that the upper (resp. lower) bound of the convergence rate is available only for closed (resp. open) sets. This property is called a "rough nature" of the Sanov's theorem.

### 3 Main results

#### 3.1 Optimality of the empirical likelihood criterion test

We first present the GNP optimality of the ELC test in (8). We impose the following assumptions.

**Assumption 1** Assume that

- (i)  $\{x_j\}_{j=1}^n$  is iid,
- (ii)  $P \left\{ \sup_{\theta \in \Theta_0} \|g(x, \theta)\| = \infty \right\} = 0$  for each  $P \in \mathcal{P}_0$ ,
- (iii)  $g(x, \theta)$  is continuous for all  $x \in \mathbb{R}^d$  at each  $\theta \in \Theta_0$ .

Assumption 1 (i) is required to apply Sanov's theorem. Assumption 1 (ii) is a tightness condition, which guarantees that  $\sup_{\theta \in \Theta_0} \|g(x, \theta)\|$  is a random variable. Assumption 1 (iii) is required to obtain the first statement of Lemma 1 (i) (Appendix A.2), which ensures continuity of a functional of the relative entropy. The GNP optimality of the ELC test is obtained as follows.

**Theorem 2** (Optimality of the empirical likelihood criterion test)

*Under Assumption 1, the ELC test (8) is Generalized Neyman–Pearson  $\delta$ -optimal for  $\mathcal{M}$ .*

The proof is similar to that of Kitamura (2001, Theorem 2). We basically show that as far as the rival test  $\Omega_n$  satisfies the type I restriction (10), the acceptance region  $\Omega_{1,n}$  must satisfy  $\Lambda_{C1} \subseteq \Omega_{1,n}$  for sufficiently large  $n$ . Note that the GNP optimality of the ELC test holds for any alternative  $P \in \mathcal{M}$ . Thus, even if the moment restriction (1) does not hold (i.e.,  $P \in \mathcal{M} \setminus \mathcal{P}$ ), the ELC test maintains the optimal power property in (11).

#### 3.2 Optimality of the empirical likelihood ratio test

We next investigate the GNP optimality of the ELR test. To test the parameter hypothesis  $H_0$ , we often assume that the moment restriction (1) holds true and compare

the difference of the maximized values of empirical likelihood under  $H_0$  and  $H_1$ . The critical value  $\eta'$  for the ELR test in (9) is typically smaller than the critical value  $\eta$  for the ELC test in (8). From the GNP optimality of the ELC test (Theorem 2), our next question is: “when do the large deviation properties of the ELR test with the critical value  $\eta'$  become equivalent to those of the ELC test with the same critical value  $\eta'$ ?” Therefore, we set as  $\eta' = \eta$  in the following discussion. We add the following assumption.

**Assumption 2**  $g(x, \theta)$  is continuous for all  $x \in \mathbb{R}^d$  at each  $\theta \in \Theta$ .

This assumption is an extension of Assumption 1 (iii) and is required to obtain the second statement in Lemma 1 (i) (Appendix A.2). From  $\eta' = \eta$  and  $\inf_{P \in \mathcal{P}} I(\mu \| P) \geq 0$  for each  $\mu \in \mathcal{M}$ , we have  $\Lambda_{C1} \subseteq \Lambda_{R1}$  and  $\Lambda_{R2} \subseteq \Lambda_{C2}$ . From Theorem 2 and  $\Lambda_{R2} \subseteq \Lambda_{C2}$ , Definition 1 (i) is obviously satisfied for the ELR test  $\Lambda_R$ . On the other hand, from  $\Lambda_{C1} \subseteq \Lambda_{R1}$ , the type II error probability satisfies

$$P^n \{\mu_n \in \Lambda_{R1}\} = P^n \{\mu_n \in \Lambda_{C1}\} + P^n \{\mu_n \in \Lambda_{R1} \cap \Lambda_{C2}\}, \tag{12}$$

for each  $P \in \mathcal{M}$ . Therefore,  $\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{C1}\} \leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{R1}\}$  for each  $P \in \mathcal{M}$ , i.e., if the critical values are same, the ELC test has a faster convergence rate of the type II error probability than the ELR test. However, if the convergence rate of the component  $P^n \{\mu_n \in \Lambda_{R1} \cap \Lambda_{C2}\}$  in (12) is faster than that of  $P^n \{\mu_n \in \Lambda_{C1}\}$  for some  $P \in \mathcal{M}$ , i.e.,

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{R1} \cap \Lambda_{C2}\} \leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{C1}\}, \tag{13}$$

then we can still obtain  $\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{C1}\} = \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{R1}\}$  and the ELR test becomes GNP optimal at  $P$ . Thus, to obtain the GNP optimality of the ELR test, we need additional conditions for the alternatives. Let  $B_m = \{x \in \mathbb{R}^d : \sup_{\theta \in \Theta} \|g(x, \theta)\| \leq m, \|x\| \leq m\}$  for  $m \in \mathbb{N}$ , and

$$\begin{aligned} \tilde{\mathcal{M}} &= \{\mu \in \mathcal{M} : \mu\{B_m\} = 1 \text{ for some } m \in \mathbb{N}\}, \\ \Lambda_{R1,0} &= \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_0} I(\mu \| P) - \inf_{P \in \mathcal{P}} I(\mu \| P) \leq \eta \right\}, \\ \Lambda_{C1,0} &= \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_0} I(\mu \| P) \leq \eta \right\}, \quad \Lambda_{C2,0} = \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_0} I(\mu \| P) > \eta \right\}. \end{aligned}$$

Note that  $\lim_{m \rightarrow \infty} P\{B_m\} = 1$  for each  $P \in \mathcal{P}_0$  by Assumption 1 (ii). The GNP optimality of the ELR test is presented as follows.

**Theorem 3** (Optimality of the empirical likelihood ratio test)

(i) *Sufficient condition: Suppose that Assumptions 1 (i)–(ii) and 2 hold. If*

$$P \in \mathcal{M}^* = \left\{ P \in \mathcal{M} : P \left\{ \sup_{\theta \in \Theta} \|g(x, \theta)\| = \infty \right\} = 0, \right. \\ \left. \inf_{\mu \in \Lambda_{R1,0} \cap \Lambda_{C2}} I(\mu || P) \geq \inf_{\mu \in \Lambda_{C1} \cap \bar{\mathcal{M}}} I(\mu || P) \right\}, \quad (14)$$

*then the ELR test (9) is Generalized Neyman–Pearson  $\delta$ -optimal for  $P$ .*

(ii) *Necessary condition: Suppose that Assumptions 1 (i)–(ii) and 2 hold. If the ELR test (9) is Generalized Neyman–Pearson  $\delta$ -optimal for  $P$ , then*

$$P \in \mathcal{M}^{**} = \left\{ P \in \mathcal{M} : P \left\{ \sup_{\theta \in \Theta} \|g(x, \theta)\| = \infty \right\} = 0, \right. \\ \left. \inf_{\mu \in \Lambda_{R1} \cap \Lambda_{C2,0} \cap \bar{\mathcal{M}}} I(\mu || P) \geq \inf_{\mu \in \Lambda_{C1,0}} I(\mu || P) \right\}. \quad (15)$$

Obviously,  $\mathcal{M}^* \subseteq \mathcal{M}^{**}$ . The restriction  $P \{ \sup_{\theta \in \Theta} \|g(x, \theta)\| = \infty \} = 0$  in  $\mathcal{M}^*$  and  $\mathcal{M}^{**}$  implies that  $\lim_{m \rightarrow \infty} P \{B_m\} = 1$  for each  $P \in \mathcal{M}^*$  and  $\mathcal{M}^{**}$ . Theorem 3 (i) is a sufficient condition for the GNP optimality of the ELR test at  $P$ . By applying Sanov’s theorem (Theorem 1), the restriction  $\inf_{\mu \in \Lambda_{R1,0} \cap \Lambda_{C2}} I(\mu || P) \geq \inf_{\mu \in \Lambda_{C1} \cap \bar{\mathcal{M}}} I(\mu || P)$  in  $\mathcal{M}^*$  guarantees (13). Because of a “rough nature” of Sanov’s theorem, we need to use  $\Lambda_{R1,0} \cap \Lambda_{C2}$  and  $\Lambda_{C1} \cap \bar{\mathcal{M}}$  in the restriction instead of  $\Lambda_{R1} \cap \Lambda_{C2}$  and  $\Lambda_{C1}$ , respectively. Theorem 3 (ii) is a necessary condition for the GNP optimality of the ELR test at  $P$ . Again, because of the rough nature of Sanov’s theorem, the restriction  $\inf_{\mu \in \Lambda_{R1} \cap \Lambda_{C2,0} \cap \bar{\mathcal{M}}} I(\mu || P) \geq \inf_{\mu \in \Lambda_{C1,0}} I(\mu || P)$  in  $\mathcal{M}^{**}$  is slightly different from that in  $\mathcal{M}^*$ . The component  $\bar{\mathcal{M}}$  in  $\mathcal{M}^*$  and  $\mathcal{M}^{**}$  is required to apply a trimming argument of Groeneboom et al. (1979). If the support of  $x$  is finite or compact, we can drop  $\bar{\mathcal{M}}$  in the definitions of  $\mathcal{M}^*$  and  $\mathcal{M}^{**}$ . In finite state Markov models, Zeitouni et al. (1992) derived similar results for the (standard) likelihood ratio test. Our result can be considered as an extension to the ELR test, where the model is semiparametric.

### 4 Conclusion

This paper studies the Generalized Neyman–Pearson (GNP) optimality of the empirical likelihood criterion and ratio tests for testing parameter hypotheses. The GNP optimality is defined by extending the Neyman–Pearson optimality to the large deviation analog. We show that (i) the empirical likelihood criterion test is GNP optimal for all alternatives, and (ii) under additional conditions for the alternatives the empirical likelihood ratio test is also GNP optimal.



**Appendix: Mathematical appendix**

**Notation** Our notation closely follows that of Kitamura (2001). Let  $P^{(m)}$  be the conditional probability measure  $P^{(m)}\{C\} = P\{C|B_m\} = P\{C \cap B_m\} / P\{B_m\}$  for  $C \in \mathcal{B}^d$  and  $P \in \mathcal{M}$ ,  $\mathcal{M}(B_m)$  be the space of probability measures on  $B_m$ ,  $P^{(m)n}$  be the  $n$ -fold product measure of  $P^{(m)}$ , and  $\mathbf{1}\{B\}$ ,  $(B)^c$ , and  $\text{cl}(B)$  be the indicator function, complement, and closure of a set  $B$ , respectively. Define

$$\begin{aligned} \mathcal{P}_m(\theta) &= \left\{ \mu \in \mathcal{M} : \int g(x, \theta) \mathbf{1}\{B_m\} d\mu = 0 \right\}, \\ \mathcal{P}_{0,m} &= \cup_{\theta \in \Theta_0} \mathcal{P}_m(\theta), \quad \mathcal{P}_m = \cup_{\theta \in \Theta} \mathcal{P}_m(\theta), \\ \Lambda_{C1}^{(m)} &= \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_{0,m}} I(\mu \| P) < \eta \right\}, \\ \Lambda_{C2}^{(m)} &= \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_{0,m}} I(\mu \| P) \geq \eta \right\}, \\ \Lambda_{C2,\epsilon} &= \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_0} I(\mu \| P) \geq \eta - \epsilon \right\}, \\ \Lambda_{C2,\epsilon}^{(m)} &= \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_{0,m}} I(\mu \| P) \geq \eta - \epsilon \right\}, \\ \Lambda_{R1,\epsilon} &= \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_0} I(\mu \| P) - \inf_{P \in \mathcal{P}} I(\mu \| P) \leq \eta + \epsilon \right\}, \\ \Lambda_{R1,\epsilon}^{(m)} &= \left\{ \mu \in \mathcal{M} : \inf_{P \in \mathcal{P}_{0,m}} I(\mu \| P) - \inf_{P \in \mathcal{P}_m} I(\mu \| P) \leq \eta + \epsilon \right\}. \end{aligned}$$

**A.1 Proof of Theorem 2**

Since the proof is similar to that of Kitamura (2001, Theorem 2), it is omitted. A detailed proof is available from the author upon request.

**A.2 Proof of Theorem 3**

*A.2.1 Proof of (i)*

We set as  $\eta' = \eta$ . First, we show that  $\Lambda_{R2}$  satisfies Definition 1 (i). Since  $\inf_{P \in \mathcal{P}} I(\mu \| P) \geq 0$  for each  $\mu \in \mathcal{M}$ , we have  $\Lambda_{R2} \subseteq \Lambda_{C2}$ . Therefore,  $P_0^n \{\mu_n \in \Lambda_{R2}\} \leq P_0^n \{\mu_n \in \Lambda_{C2}\}$  for each  $P_0 \in \mathcal{P}_0$  and the result follows from Theorem 2.

Next, we show that  $\Lambda_{R1}$  satisfies Definition 1 (ii). From Theorem 2, it is sufficient to show that  $\Lambda_{R1}$  satisfies

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{R1}\} \leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{C1}\}, \quad (16)$$

for each  $P \in \mathcal{M}^*$ . Observe that  $\Lambda_{C1} \subseteq \Lambda_{R1}$ , i.e.,

$$P^n \{\mu_n \in \Lambda_{R1}\} = P^n \{\mu_n \in \Lambda_{C1}\} + P^n \{\mu_n \in \Lambda_{R1} \cap \Lambda_{C2}\}, \tag{17}$$

for each  $P \in \mathcal{M}$ . From (16) and (17), it is sufficient to show that

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{R1} \cap \Lambda_{C2}\} \leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{C1}\}, \tag{18}$$

for each  $P \in \mathcal{M}^*$ . Pick any  $P \in \mathcal{M}^*$ . Lemma 2 implies that for each  $\epsilon \in (0, \infty)$ , there exists  $m_\epsilon \in \mathbb{N}$  such that

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{\mu_n \in \Lambda_{R1} \cap \Lambda_{C2}\} \leq - \inf_{\mu \in \Lambda_{R1,0} \cap \Lambda_{C2}} I(\mu || P^{(m)}) + \epsilon, \tag{19}$$

for each  $m \geq m_\epsilon$ . Lemma 1 (iv) and (v) and the definition of  $\mathcal{M}^*$  in (14) imply that for each  $\epsilon \in (0, \infty)$ , there exists  $m'_\epsilon \in \mathbb{N}$  such that

$$\begin{aligned} - \inf_{\mu \in \Lambda_{R1,0} \cap \Lambda_{C2}} I(\mu || P^{(m)}) &\leq - \inf_{\mu \in \Lambda_{R1,0} \cap \Lambda_{C2}} I(\mu || P) + \epsilon \\ &\leq - \inf_{\mu \in \Lambda_{C1} \cap \tilde{\mathcal{M}}} I(\mu || P) + \epsilon \\ &\leq - \inf_{\mu \in \Lambda_{C1} \cap \tilde{\mathcal{M}}} I(\mu || P^{(m)}) + 2\epsilon \\ &\leq - \inf_{\mu \in \Lambda_{C1}} I(\mu || P^{(m)}) + 2\epsilon, \end{aligned} \tag{20}$$

for each  $m \geq m'_\epsilon$ . Lemma 1 (ii) and (iii) imply that

$$\begin{aligned} - \inf_{\mu \in \Lambda_{C1}} I(\mu || P^{(m)}) &= - \inf_{\mu \in \Lambda_{C1} \cap \mathcal{M}(B_m)} I(\mu || P^{(m)}) \\ &= - \inf_{\mu \in \Lambda_{C1}^{(m)} \cap \mathcal{M}(B_m)} I(\mu || P^{(m)}) \\ &\leq - \inf_{\mu \in \Lambda_{C1}^{(m)}} I(\mu || P^{(m)}), \end{aligned} \tag{21}$$

for each  $m \in \mathbb{N}$ . Since  $\Lambda_{C1}^{(m)}$  is an open set (from Lemma 1 (i)), Theorem 1 yields that

$$- \inf_{\mu \in \Lambda_{C1}^{(m)}} I(\mu || P^{(m)}) \leq \liminf_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \{\mu_n \in \Lambda_{C1}^{(m)}\}, \tag{22}$$

for each  $m \in \mathbb{N}$ . For each  $P \in \mathcal{M}^*$  and  $\epsilon \in (0, \infty)$ , there exists  $m''_\epsilon \in \mathbb{N}$  such that  $-\log P\{B_m\} \leq \epsilon$  for all  $m \geq m''_\epsilon$  (from  $P\{\sup_{\theta \in \Theta} \|g(x, \theta)\| = \infty\} = 0$ ). Thus, for

all  $m \geq m'_\epsilon$  we have

$$\begin{aligned} & \liminf_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \{ \mu_n \in \Lambda_{C1}^{(m)} \} \\ &= \liminf_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \{ \mu_n \in \Lambda_{C1}^{(m)} \cap \mathcal{M}(B_m) \} \\ &= \liminf_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{C1}^{(m)} \cap \mathcal{M}(B_m) \} - \log P \{ B_m \} \\ &\leq \liminf_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{C1} \} + \epsilon, \end{aligned} \tag{23}$$

where the first equality follows from  $P^{(m)n} \{ \mu_n \notin \mathcal{M}(B_m) \} = 0$  for each  $m \in \mathbb{N}$  and  $n \in \mathbb{N}$  (by the definition of  $P^{(m)}$ ), the second equality follows from the definition of  $P^{(m)}$  and Assumption 1 (i), and the inequality follows from Lemma 1 (ii) and the set inclusion relationship. From (19)–(23), we have

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{R1} \cap \Lambda_{C2} \} \leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{C1} \} + 4\epsilon,$$

for each  $\epsilon \in (0, \infty)$  and  $P \in \mathcal{M}^*$ . Therefore, (18) is obtained.

### A.2.2 Proof of (ii)

Suppose that the ELR test (9) is Generalized Neyman–Pearson  $\delta$ -optimal at some  $P$ . Definition 1 (ii) implies that  $\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{R1} \} \leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{C1} \}$ . Since  $\Lambda_{C1} \subseteq \Lambda_{R1}$ , we must have

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{R1} \cap \Lambda_{C2} \} \leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{C1} \}. \tag{24}$$

Thus, it is sufficient to show that if  $P$  satisfies (24) then we have  $P \in \mathcal{M}^{**}$ .

First, consider the left hand side of (24). For each  $\epsilon \in (0, \infty)$ , there exists  $m_\epsilon \in \mathbb{N}$  such that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{R1} \cap \Lambda_{C2} \} \\ & \geq \liminf_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{R1} \cap \Lambda_{C2,0} \} \\ & \geq \liminf_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \{ \mu_n \in \Lambda_{R1}^{(m)} \cap \Lambda_{C2,0}^{(m)} \} - \epsilon, \end{aligned} \tag{25}$$

for each  $m \geq m_\epsilon$ , where the second inequality follows from the same argument as (23). Since  $\Lambda_{R1}^{(m)} \cap \Lambda_{C2,0}^{(m)}$  is open (from Lemma 1 (i)), Theorem 1 yields that

$$\begin{aligned} \liminf_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \{ \mu_n \in \Lambda_{R1}^{(m)} \cap \Lambda_{C2,0}^{(m)} \} & \geq - \inf_{\mu \in \Lambda_{R1}^{(m)} \cap \Lambda_{C2,0}^{(m)}} I(\mu || P^{(m)}) \\ & \geq - \inf_{\mu \in \Lambda_{R1}^{(m)} \cap \Lambda_{C2,0}^{(m)} \cap \mathcal{M}(B_m)} I(\mu || P^{(m)}) \end{aligned}$$

$$\begin{aligned}
 &= - \inf_{\mu \in \Lambda_{R1} \cap \Lambda_{C2,0} \cap \mathcal{M}(B_m)} I(\mu || P^{(m)}) \\
 &= - \inf_{\mu \in \Lambda_{R1} \cap \Lambda_{C2,0}} I(\mu || P^{(m)}), \tag{26}
 \end{aligned}$$

for each  $m \in \mathbb{N}$ , where the first and second equalities follow from Lemma 1 (ii) and (iii), respectively. Lemma 1 (iv) implies that for each  $\epsilon \in (0, \infty)$  there exists  $m'_\epsilon \in \mathbb{N}$  such that

$$\begin{aligned}
 - \inf_{\mu \in \Lambda_{R1} \cap \Lambda_{C2,0}} I(\mu || P^{(m)}) &\geq - \inf_{\mu \in \Lambda_{R1} \cap \Lambda_{C2,0} \cap \bar{\mathcal{M}}} I(\mu || P^{(m)}) \\
 &\geq - \inf_{\mu \in \Lambda_{R1} \cap \Lambda_{C2,0} \cap \bar{\mathcal{M}}} I(\mu || P) - \epsilon, \tag{27}
 \end{aligned}$$

for each  $m \geq m'_\epsilon$ .

Next, consider the right hand side of (24). Observe that

$$\begin{aligned}
 &\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{C1} \} \\
 &\leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{C1,0} \} \\
 &\leq \limsup_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \{ \mu_n \in \Lambda_{C1,0} \} \\
 &= \limsup_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \{ \mu_n \in \Lambda_{C1,0} \cap \mathcal{M}(B_m) \} \\
 &\leq \limsup_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \left\{ \mu_n \in \Lambda_{C1,0}^{(m)} \right\}, \tag{28}
 \end{aligned}$$

for each  $m \in \mathbb{N}$ , where the first inequality follows from the set inclusion relationship, the second inequality follows from the definition of  $P^{(m)}$  and  $P\{B_m\} \leq 1$  for all  $m \in \mathbb{N}$ , the equality follows from  $P^{(m)n}\{\mu_n \notin \mathcal{M}(B_m)\} = 0$  for each  $m \in \mathbb{N}$  and  $n \in \mathbb{N}$ , the last inequality follows from Lemma 1 (ii) and the set inclusion relationship. Since  $\Lambda_{C1,0}^{(m)}$  is closed (from Lemma 1 (i)), Theorem 1 yields that

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} n^{-1} \log P^{(m)n} \{ \mu_n \in \Lambda_{C1,0}^{(m)} \} &\leq - \inf_{\mu \in \Lambda_{C1,0}^{(m)}} I(\mu || P^{(m)}) \\
 &= - \inf_{\mu \in \Lambda_{C1,0}^{(m)} \cap \mathcal{M}(B_m)} I(\mu || P^{(m)}) \\
 &= - \inf_{\mu \in \Lambda_{C1,0} \cap \mathcal{M}(B_m)} I(\mu || P^{(m)}) \\
 &= - \inf_{\mu \in \Lambda_{C1,0}} I(\mu || P^{(m)}), \tag{29}
 \end{aligned}$$

for each  $m \in \mathbb{N}$ , where the equalities follow from Lemma 1 (ii) and (iii). Lemma 1 (v) implies that for each  $\epsilon \in (0, \infty)$  there exists  $m''_\epsilon \in \mathbb{N}$  such that

$$- \inf_{\mu \in \Lambda_{C1,0}} I(\mu || P^{(m)}) \leq - \inf_{\mu \in \Lambda_{C1,0}} I(\mu || P) + \epsilon, \tag{30}$$

for each  $m \geq m''_\epsilon$ .

Finally, combining these results,  $P$  satisfies

$$\begin{aligned} - \inf_{\mu \in \Lambda_{R1} \cap \Lambda_{C2,0} \cap \bar{\mathcal{M}}} I(\mu||P) - 2\epsilon &\leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{R1} \cap \Lambda_{C2} \} \\ &\leq \limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{C1} \} \\ &\leq - \inf_{\mu \in \Lambda_{C1,0}} I(\mu||P) + 2\epsilon, \end{aligned}$$

for each  $\epsilon \in (0, \infty)$ , i.e.,  $P \in \mathcal{M}^{**}$ .

### A.3 Auxiliary lemmas

The following lemmas are derived under Assumptions 1 and 2.

- Lemma 1** (i) *Kitamura (2001, Lemma 1)*:  $\inf_{P \in \mathcal{P}_{0,m}} I(\mu||P)$  and  $\inf_{P \in \mathcal{P}_m} I(\mu||P)$  are continuous in  $\mu \in \mathcal{M}$ .
- (ii)  $A^{(m)} \cap \mathcal{M}(B_m) = A \cap \mathcal{M}(B_m)$  for each  $m \in \mathbb{N}$ , where  $(A^{(m)}, A) = (\Lambda_{C1}^{(m)}, \Lambda_{C1})$ ,  $(\Lambda_{C1,0}^{(m)}, \Lambda_{C1,0})$ ,  $(\Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)}, \Lambda_{R1,0} \cap \Lambda_{C2})$ ,  $(\Lambda_{R1}^{(m)} \cap \Lambda_{C2,0}^{(m)}, \Lambda_{R1} \cap \Lambda_{C2,0})$ , and  $(\Lambda_{R1,\epsilon}^{(m)} \cap \Lambda_{C2,\epsilon}^{(m)}, \Lambda_{R1,\epsilon} \cap \Lambda_{C2,\epsilon})$  for each  $\epsilon \in (0, \infty)$ .
- (iii)  $\inf_{\mu \in A \cap \mathcal{M}(B_m)} I(\mu||P^{(m)}) = \inf_{\mu \in A} I(\mu||P^{(m)})$  for each  $m \in \mathbb{N}$  and  $P \in \mathcal{M}$ , where  $A = \Lambda_{C1}$ ,  $\Lambda_{C1,0}^{(m)}$ ,  $\Lambda_{R1,0} \cap \Lambda_{C2}$ ,  $\Lambda_{R1} \cap \Lambda_{C2,0}$ , and  $\Lambda_{R1,\epsilon} \cap \Lambda_{C2,\epsilon}$  for each  $\epsilon \in (0, \infty)$ .
- (iv) *Groeneboom et al. (1979, Lemma 4.1)*: If  $\Lambda$  is a subset of  $\bar{\mathcal{M}} = \{ \mu \in \mathcal{M} : \mu\{B_m\} = 1 \text{ for some } m \in \mathbb{N} \}$ , then  $\lim_{m \rightarrow \infty} \inf_{\mu \in \Lambda} I(\mu||P^{(m)}) = \inf_{\mu \in \Lambda} I(\mu||P)$  for each  $P \in \mathcal{M}^*$ .
- (v) If  $\Lambda$  is a subset of  $\mathcal{M}$ , then  $\lim_{m \rightarrow \infty} \inf_{\mu \in \Lambda} I(\mu||P^{(m)}) \geq \inf_{\mu \in \Lambda} I(\mu||P)$  for each  $P \in \mathcal{M}^*$ .

*Proof* (ii) and (iii) are obtained from the proof of *Kitamura (2001, Lemma 3)*. (v) is obtained from the proof of *Groeneboom et al. (1979, Lemma 4.1)*.

**Lemma 2** For each  $\epsilon \in (0, \infty)$  and  $P \in \mathcal{M}^*$ , there exists  $m_\epsilon \in \mathbb{N}$  such that

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{R1} \cap \Lambda_{C2} \} \leq - \inf_{\mu \in \Lambda_{R1,0} \cap \Lambda_{C2}} I(\mu||P^{(m)}) + \epsilon,$$

for each  $m > m_\epsilon$ .

□

*Proof* A similar argument as the proof of *Kitamura (2001, Theorem 2)* yields that for each  $\epsilon \in (0, \infty)$  and  $P \in \mathcal{M}^*$ , there exists  $m_\epsilon \in \mathbb{N}$  such that

$$\limsup_{n \rightarrow \infty} n^{-1} \log P^n \{ \mu_n \in \Lambda_{R1} \cap \Lambda_{C2} \} \leq - \inf_{\mu \in \Lambda_{R1,\epsilon} \cap \Lambda_{C2,\epsilon}} I(\mu||P^{(m)}) + \epsilon/2,$$

for all  $m > m_\epsilon$ . From Lemma 1 (ii) and (iii),

$$\begin{aligned} \inf_{\mu \in \Lambda_{R1,0} \cap \Lambda_{C2}} I(\mu \| P^{(m)}) &= \inf_{\mu \in \Lambda_{R1,0} \cap \Lambda_{C2} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)}) \\ &= \inf_{\mu \in \Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)}), \\ \inf_{\mu \in \Lambda_{R1,\epsilon} \cap \Lambda_{C2,\epsilon}} I(\mu \| P^{(m)}) &= \inf_{\mu \in \Lambda_{R1,\epsilon} \cap \Lambda_{C2,\epsilon} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)}) \\ &= \inf_{\mu \in \Lambda_{R1,\epsilon}^{(m)} \cap \Lambda_{C2,\epsilon}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)}), \end{aligned}$$

for each  $m \in \mathbb{N}$ ,  $P \in \mathcal{M}^*$ , and  $\epsilon \in (0, \infty)$ . Therefore, it is sufficient to show that

$$\lim_{\epsilon \rightarrow 0} \inf_{\mu \in \Lambda_{R1,\epsilon}^{(m)} \cap \Lambda_{C2,\epsilon}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)}) = \inf_{\mu \in \Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)})$$

for each  $m \in \mathbb{N}$  and  $P \in \mathcal{M}^*$ .

Now pick any  $P \in \mathcal{M}^*$  and  $m \in \mathbb{N}$ . Since the sequence of the sets  $A_\epsilon^{(m)} = \{\Lambda_{R1,\epsilon}^{(m)} \cap \Lambda_{C2,\epsilon}^{(m)} \cap \mathcal{M}(B_m)\}_{\epsilon \in (0,\infty)}$  is non-increasing as  $\epsilon \downarrow 0$ , the sequence  $\{\inf_{\mu \in A_\epsilon^{(m)}} I(\mu \| P^{(m)})\}_{\epsilon \in (0,\infty)}$  is non-decreasing as  $\epsilon \downarrow 0$ . Also, this sequence  $\{\inf_{\mu \in A_\epsilon^{(m)}} I(\mu \| P^{(m)})\}_{\epsilon \in (0,\infty)}$  is bounded above by  $\inf_{\mu \in \Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)})$ . Therefore, the sequence  $\{\inf_{\mu \in A_\epsilon^{(m)}} I(\mu \| P^{(m)})\}_{\epsilon \in (0,\infty)}$  converges as  $\epsilon \downarrow 0$  to some limit  $\bar{I} \leq \inf_{\mu \in \Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)})$ , and we can take a decreasing sequence of positive numbers  $\{\epsilon_l\}_{l \in \mathbb{N}}$  such that  $\{\inf_{\mu \in A_{\epsilon_l}^{(m)}} I(\mu \| P^{(m)})\}_{l \in \mathbb{N}}$  increases to  $\bar{I}$  as  $l \rightarrow \infty$ .

Lemma 1 (i) implies that  $\Lambda_{R1,\epsilon_l}^{(m)}$  and  $\Lambda_{C2,\epsilon_l}^{(m)}$  are closed sets for each  $l \in \mathbb{N}$ . Since  $\mathcal{M}(B_m)$  is compact (see, e.g., Dembo and Zeitouni 1998, Appendix D.2),  $A_{\epsilon_l}^{(m)}$  is also compact for each  $l \in \mathbb{N}$ . Thus, the lower semicontinuity of  $I(\mu \| P^{(m)})$  for  $\mu \in A_{\epsilon_l}^{(m)}$  implies that the infimum  $\inf_{\mu \in A_{\epsilon_l}^{(m)}} I(\mu \| P^{(m)})$  is attained on the compact set  $A_{\epsilon_l}^{(m)}$ , i.e., there exists  $\mu_l \in A_{\epsilon_l}^{(m)}$  such that  $I(\mu_l \| P^{(m)}) = \inf_{\mu \in A_{\epsilon_l}^{(m)}} I(\mu \| P^{(m)})$  for each  $l \in \mathbb{N}$ .

Now consider the sequence of the sets  $\{\text{cl}(\{\mu_l\}_{l \geq l'})\}_{l' \in \mathbb{N}}$ . Since  $\text{cl}(\{\mu_l\}_{l \geq l'}) \subseteq A_{\epsilon_{l'}}^{(m)}$  for each  $l' \in \mathbb{N}$ ,  $\text{cl}(\{\mu_l\}_{l \geq l'})$  is compact for each  $l' \in \mathbb{N}$ . Thus, since  $\{\text{cl}(\{\mu_l\}_{l \geq l'})\}_{l' \in \mathbb{N}}$  is a non-increasing sequence of non-empty compact sets, the Heine–Borel theorem implies that  $\bigcap_{l'=1}^\infty \text{cl}(\{\mu_l\}_{l \geq l'})$  is non-empty. Pick any  $\bar{\mu} \in \bigcap_{l'=1}^\infty \text{cl}(\{\mu_l\}_{l \geq l'})$ . Since  $I(\mu_l \| P^{(m)}) \leq \bar{I}$  for each  $l \in \mathbb{N}$ , we have  $\{\mu_l\}_{l \geq l'} \subseteq \{\mu \in \mathcal{M} : I(\mu \| P^{(m)}) \leq \bar{I}\}$ . Since  $\{\mu \in \mathcal{M} : I(\mu \| P^{(m)}) \leq \bar{I}\}$  is closed (from the lower semicontinuity of  $I(\mu \| P^{(m)})$  for  $\mu \in \mathcal{M}$ ), we have  $\text{cl}(\{\mu_l\}_{l \geq l'}) \subseteq \{\mu \in \mathcal{M} : I(\mu \| P^{(m)}) \leq \bar{I}\}$  for each  $l' \in \mathbb{N}$ , which implies  $I(\bar{\mu} \| P^{(m)}) \leq \bar{I}$ . On the other hand, since  $\bar{\mu} \in A_{\epsilon_l}^{(m)}$  for each  $l \in \mathbb{N}$ , we have  $\bar{\mu} \in \Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)} \cap \mathcal{M}(B_m)$ , which implies  $\bar{I} \leq \inf_{\mu \in \Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)})$ .

$I(\mu \| P^{(m)}) \leq I(\bar{\mu} \| P^{(m)})$ . Combining these results,

$$I(\bar{\mu} \| P^{(m)}) \leq \bar{I} \leq \inf_{\mu \in \Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)}) \leq I(\bar{\mu} \| P^{(m)}),$$

i.e.,  $\{\inf_{\mu \in A_{\varepsilon_l}^{(m)}} I(\mu \| P^{(m)})\}_{l \in \mathbb{N}}$  increases to the limit  $\bar{I} = \inf_{\mu \in \Lambda_{R1,0}^{(m)} \cap \Lambda_{C2}^{(m)} \cap \mathcal{M}(B_m)} I(\mu \| P^{(m)})$  as  $l \rightarrow \infty$ . Therefore, the conclusion is obtained.  $\square$

## References

- Dembo, A., Zeitouni, O. (1998). *Large deviations techniques and applications*, (2nd ed.). Heidelberg: Springer.
- Deuschel, J. D., Stroock, D. W. (1989). *Large Deviations*. New York: Academic Press.
- DiCiccio, T. J., Hall, P., Romano, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Annals of Statistics*, 19, 1053–1061.
- Groeneboom, P., Oosterhoff, J., Ruymgaart, F. H. (1979). Large deviation theorems for empirical probability measures. *Annals of Probability*, 7, 553–586.
- Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Annals of Mathematical Statistics*, 36, 369–401.
- Kitamura, Y. (2001). Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, 69, 1661–1672.
- Newey, W. K., Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72, 219–255.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75, 237–249.
- Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, 22, 300–325.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.
- Steinberg, Y., Zeitouni, O. (1992). On tests for normality. *IEEE Transactions on Information Theory*, 38, 1779–1787.
- Zeitouni, O., Gutman, M. (1991). On universal hypothesis testing via large deviations. *IEEE Transactions on Information Theory*, 37, 285–290.
- Zeitouni, O., Ziv, J., Merhav, N. (1992). When is the generalized likelihood ratio test optimal? *IEEE Transactions on Information Theory*, 38, 1597–1602.