

# On the Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions

Yuichi Kitamura  
Department of Economics  
Yale University  
yuichi.kitamura@yale.edu

Andres Santos  
Department of Economics  
University of California - San Diego  
a2santos@ucsd.edu

Azeem M. Shaikh  
Department of Economics  
University of Chicago  
amshaikh@uchicago.edu

October 7, 2010

## Abstract

We show by example that empirical likelihood and other commonly used tests for moment restrictions are unable to control the (exponential) rate at which the probability of a Type I error tends to zero. It follows that the optimality of empirical likelihood asserted in Kitamura (2001) does not hold without additional assumptions. Under stronger assumptions than those in Kitamura (2001), we establish the following optimality result: (i) empirical likelihood controls the rate at which the probability of a Type I error tends to zero and (ii) among all procedures for which the probability of a Type I error tends to zero at least as fast, empirical likelihood maximizes the rate at which probability of a Type II error tends to zero for “most” alternatives. This result further implies that empirical likelihood maximizes the rate at which probability of a Type II error tends to zero for all alternatives among a class of tests that satisfy a weaker criterion for their Type I error probabilities.

**KEYWORDS:** Empirical Likelihood, Large Deviations, Hoeffding Optimality, Moment Restrictions

**ACKNOWLEDGMENTS:** The research of the first author has been supported by the National Science Foundation grants SES-0551271 and SES-0851759. The research of the third author has been supported by the National Science Foundation grant DMS-0820310.

# 1 Introduction

Let  $X_i, i = 1, \dots, n$ , be an i.i.d. sequence of random variables with distribution  $P \in \mathbf{P} \subseteq \mathbf{M}$  where  $\mathbf{M}$  denotes the set of probability distributions on  $\mathcal{X} \subseteq \mathbf{R}^d$  (with the Borel  $\sigma$ -algebra). Let  $\Theta \subseteq \mathbf{R}^r$  with  $m > r$  and  $g : \mathbf{R}^d \times \Theta \rightarrow \mathbf{R}^m$  be given, and define

$$\mathbf{P}_0 = \{P \in \mathbf{P} : E_P[g(X, \theta)] = 0 \text{ for some } \theta \in \Theta\} . \quad (1)$$

In this paper, we consider testing

$$H_0 : P \in \mathbf{P}_0 \quad \text{versus} \quad H_1 : P \in \mathbf{P}_1 , \quad (2)$$

where  $\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$ . Note that any non-randomized test of (2) can be identified with a partition  $(\Omega_{0,n}, \Omega_{1,n})$  of  $\mathbf{M}$ , wherein the test rejects the null hypothesis if the empirical distribution of the observations, denoted by  $\hat{P}_n$ , falls in  $\Omega_{1,n}$  and fails to reject it otherwise.

Following Owen (1988), Qin and Lawless (1994) propose a test of (2) based on empirical likelihood ratio. In order to describe their test, recall that the Kullback-Leibler divergence of  $P$  from  $Q$  is defined as

$$I(Q|P) = \begin{cases} \int \log\left(\frac{dQ}{dP}\right) dQ & \text{if } Q \ll P \\ \infty & \text{otherwise} \end{cases} ,$$

and let

$$\mathbf{M}_0(Q) = \{P \in \mathbf{M} : P \ll Q, Q \ll P, E_P[g(X, \theta)] = 0 \text{ for some } \theta \in \Theta\} . \quad (3)$$

Using this notation, their test can be described as rejecting the null hypothesis for large values of

$$\inf_{P \in \mathbf{M}_0(\hat{P}_n)} I(\hat{P}_n|P) . \quad (4)$$

It therefore corresponds, for some sequence of critical values  $\{\eta_n > 0 : n \geq 1\}$ , to the partition  $(\Lambda_0(\eta_n), \Lambda_1(\eta_n))$  of  $\mathbf{M}$ , where

$$\begin{aligned} \Lambda_0(\eta_n) &= \{Q \in \mathbf{M} : \inf_{P \in \mathbf{M}_0(Q)} I(Q|P) < \eta_n\} \\ \Lambda_1(\eta_n) &= \mathbf{M} \setminus \Lambda_0(\eta_n) . \end{aligned} \quad (5)$$

Here, the infimum over the empty set is understood to be infinity.

Kitamura (2001) claims the empirical likelihood ratio test is optimal for testing (2) in the large deviations sense of Hoeffding (1965) provided that the following conditions hold

$$P\{\sup_{\theta \in \Theta} \|g(X, \theta)\| = \infty\} = 0 \text{ for all } P \in \mathbf{P} \quad (6)$$

$$g(x, \theta) \text{ is continuous at every } \theta \in \Theta \text{ for each } x \in \mathbf{R}^d . \quad (7)$$

Specifically, part (a) of Theorem 2 in Kitamura (2001) asserts that for any  $\eta > 0$

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n \{ \hat{P}_n \in \Lambda_1(\eta) \} \leq -\eta, \quad (8)$$

whereas part (b) of the same theorem asserts that if another test  $(\Omega_{0,n}, \Omega_{1,n})$  satisfies

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n \{ \hat{P}_n \in \Omega_{1,n}^\delta \} \leq -\eta$$

for some  $\delta > 0$ , then

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n \{ \hat{P}_n \in \Omega_{0,n} \} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n \{ \hat{P}_n \in \Lambda_0(\eta) \}$$

for any  $P \in \mathbf{P}_1$ . Here, the notation  $A^\delta$  denotes the (open)  $\delta$ -“blowup” of a set  $A \subseteq \mathbf{M}$  with respect to Prokhorov-Lévy metric. More formally,  $A^\delta = \bigcup_{P \in A} B(P, \delta)$ , where  $B(P, \delta)$  denotes an open ball with center  $P$  and radius  $\delta$  with respect to the Prokhorov-Lévy metric.

In Section 2, we provide two examples that demonstrate that the empirical likelihood ratio test of Qin and Lawless (1994) in fact fails to control size as required in (8) without additional restrictions on  $\mathbf{P}$ . More specifically, we show that given the assumptions in Kitamura (2001), (8) is not satisfied for any  $\eta > 0$ . Interestingly, these examples also reveal that most commonly used tests of (2) suffer of the same deficiency. In Section 3, we use these examples to motivate restrictions on  $\mathbf{P}$  under which we establish a more limited version of Theorem 2 in Kitamura (2001). Proofs of all results are collected in a Supplemental Appendix.

**Remark 1.1.** Strictly speaking, Theorem 2 of Kitamura (2001) considers a test different from that of Qin and Lawless (1994) in that it omits the requirements  $P \ll Q$  and  $Q \ll P$  in (3). ■

## 2 Two Examples

We now provide two examples that illustrate the need for further restrictions on  $\mathbf{P}$  in order for (8) to hold. The family of distributions used in the first example is the same as the one used in Romano (1989) to show that the bootstrap does not behave well uniformly over certain classes of distributions and in Romano (2004) to show that the  $t$ -test has size one in finite samples over certain classes of distributions.

**Example 2.1.** Suppose  $d = 1$ ,  $m = 1$  and  $g(x, \theta) = x$  for all  $\theta \in \Theta$ . Let  $\mathbf{P}$  be any set of probability distributions satisfying (6) and (7) and containing

$$\mathbf{C}_0 = \{P_c : 0 < c < 1\},$$

where  $P_c$  is the measure with mass  $1 - c$  on  $c$  and mass  $c$  on  $-(1 - c)$ . Define the event

$$E_n = \{X_i = c \text{ for all } 1 \leq i \leq n\}$$

and notice that  $E_n$  implies that  $\mathbf{M}_0(\hat{P}_n) = \emptyset$ . Since the empirical likelihood ratio test rejects whenever  $\mathbf{M}_0(\hat{P}_n) = \emptyset$ , we obtain for any  $\eta > 0$  that

$$P_c^n\{\hat{P}_n \in \Lambda_1(\eta)\} \geq P_c^n\{E_n\} = (1 - c)^n . \quad (9)$$

Moreover, since  $\mathbf{C}_0 \subseteq \mathbf{P}_0$ , it follows from (9) that

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_1(\eta)\} \geq \sup_{c \in (0,1)} \log(1 - c) = 0 . \quad (10)$$

We conclude in particular that (8) cannot be satisfied for any  $\mathbf{P}$  that includes  $\mathbf{C}_0$ . ■

The above example suggests that if  $\mathbf{P}$  is “too rich” then the empirical likelihood ratio test cannot satisfy (8) for any value of  $\eta > 0$ . It is important to note that that this would remain true even if we were to allow the critical value  $\eta$  to depend on  $n$ . Furthermore, this shortcoming is not unique to empirical likelihood in that it is shared by many commonly used tests. In particular, Example 2.1 also applies to the  $t$ -test. To see this, let  $P_c$  and  $E_n$  be defined as in Example 2.1 and simply note that the sample mean,  $\bar{X}_n$ , equals  $c$  and the standard error,  $\hat{\sigma}_n$ , defined by

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{1 \leq i \leq n} (X_i - \bar{X}_n)^2 ,$$

equals zero whenever  $E_n$  occurs. Hence, for any critical value  $\eta > 0$ ,

$$P_c^n\{\sqrt{n}|\bar{X}_n| \geq \eta\hat{\sigma}_n\} \geq P_c^n\{E_n\} = (1 - c)^n . \quad (11)$$

By arguing as in (10), it follows that the  $t$ -test also fails to control size in the sense of Hoeffding (1965) whenever  $\mathbf{C}_0 \subseteq \mathbf{P}$ . Note that this failure persists even if we estimate the standard error without centering because (11) continues to hold provided  $\sqrt{n} \geq \eta$ . For this reason, Example 2.1 applies to the generalized method of moments (GMM)-based  $J$ -test proposed in Hansen (1982). A similar argument shows that Example 2.1 also applies to tests based on generalized empirical likelihood.

The simplicity of Example 2.1 is illustrative but potentially misleading. First, it is not enough to simply rule out discrete distributions. Indeed, one could modify Example 2.1 by “smoothing” out the mass on either side of zero and it would still apply. Second, it is also not sufficient to rule out distributions that have “too little” mass on one side of zero, as shown by Example 2.2 below.

**Example 2.2.** As in the previous example, assume  $d = 1$ ,  $m = 1$  and  $g(x, \theta) = x$  for all  $\theta \in \Theta$ . Let  $\mathbf{P}$  be any set of probability distributions satisfying (6) and (7) and containing

$$\mathbf{K}_0 = \{P_{K,c} = cD_{-1} + (1 - c)R_{K,c} : 0 < c < \frac{1}{2}, K \geq 2\} ,$$

where  $D_{-1}$  is the degenerate distribution at  $-1$ , and  $R_{K,c}$  is the distribution satisfying:

$$R_{K,c}\left\{X = \frac{-2c}{(1 - c)(K - 1)}\right\} = \frac{1}{2} , \quad R_{K,c}\left\{X = \frac{2Kc}{(1 - c)(K - 1)}\right\} = \frac{1}{2} .$$

As in Example 2.1, we show empirical likelihood fails to satisfy (8) by constructing an event under which it rejects too often. To this end, define

$$A_n = \{X_i \neq -1 \text{ for all } 1 \leq i \leq n\} .$$

Note that there is a unique measure  $F_{K,c}$  in  $\mathbf{M}_0(R_{K,c})$  given by:

$$F_{K,c} \left\{ X = \frac{-2c}{(1-c)(K-1)} \right\} = \frac{K}{1+K} , \quad F_{K,c} \left\{ X = \frac{2Kc}{(1-c)(K-1)} \right\} = \frac{1}{K} .$$

We can therefore conclude by direct calculation that there is a  $K_\eta$  sufficiently large for which

$$\inf_{P \in \mathbf{M}_0(R_{K_\eta,c})} I(R_{K_\eta,c}|P) = I(R_{K_\eta,c}|F_{K_\eta,c}) = \frac{1}{2} \log\left(\frac{1+K_\eta}{2K_\eta}\right) + \frac{1}{2} \log\left(\frac{1+K_\eta}{2}\right) > \eta .$$

Next, note that under  $R_{K_\eta,c}^n$  either  $\mathbf{M}_0(\hat{P}_n) = \emptyset$  or  $\mathbf{M}_0(\hat{P}_n) = \{F_{K_\eta,c}\}$ . It follows that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log R_{K_\eta,c}^n \{ \hat{P}_n \in \Lambda_1(\eta) \} &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log R_{K_\eta,c}^n \left\{ \inf_{P \in \mathbf{M}_0(\hat{P}_n)} I(\hat{P}_n|P) \geq \eta \right\} \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log R_{K_\eta,c}^n \left\{ I(\hat{P}_n|F_{K_\eta,c}) \geq \eta \cap \mathbf{M}_0(\hat{P}_n) \neq \emptyset \right\} \\ &= 0 , \end{aligned}$$

where the final inequality follows from observing that  $I(\hat{P}_n|F_{K_\eta,c})$  converges in probability to  $I(R_{K_\eta,c}|F_{K_\eta,c})$  and  $\mathbf{M}_0(\hat{P}_n) \neq \emptyset$  with probability approaching one under  $R_{K_\eta,c}^n$ . We conclude

$$\begin{aligned} &\sup_{P \in \mathbf{K}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n \{ \hat{P}_n \in \Lambda_1(\eta) \} \\ &\geq \sup_{c \in (0, \frac{1}{2})} \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{K_\eta,c}^n \{ \hat{P}_n \in \Lambda_1(\eta) | A_n \} + \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{K_\eta,c}^n \{ A_n \} \right\} \\ &= \sup_{c \in (0, \frac{1}{2})} \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} \log R_{K_\eta,c}^n \{ \hat{P}_n \in \Lambda_1(\eta) \} + \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{K_\eta,c}^n \{ A_n \} \right\} \\ &= \sup_{c \in (0, \frac{1}{2})} \log(1-c) \\ &= 0 . \end{aligned}$$

Therefore, (8) cannot be satisfied for any  $\mathbf{P}$  that includes  $\mathbf{K}_0$ . ■

Both Examples 2.1 and 2.2 rely on a sequence of distributions for which the rate at which the probability of a Type I error tends to zero itself tends to zero. In Example 2.1 this sequence is denoted by  $P_c$  and in Example 2.2 this sequence is denoted by  $P_{K_\eta,c}$ . While they appear different in nature, these sequences are in fact linked together by the observation that

$$\lim_{c \rightarrow 0} P_c = \lim_{c \rightarrow 0} P_{K_\eta,c} = D_0 , \tag{12}$$

where  $D_0$  is the degenerate distribution at 0 and the limits should be interpreted in the weak topology. In the special case where  $\mathcal{X} = [-1, 1]$ ,  $d = 1$ ,  $m = 1$ , and  $g(x, \theta) = x$  for all  $\theta \in \Theta$ , it will follow from our general results in the next section that  $D_0$  is a limit point of any sequence of distributions for which the (exponential) rate at which the probability of a Type I error tends to zero itself tends to zero. Therefore, by requiring that  $D_0$  not be a limit point of  $\mathbf{P}$ , we will show that the test of Qin and Lawless (1994) in this setting satisfies (8) at least for all sufficiently small  $\eta > 0$ .

**Remark 2.1.** It is worth noting the implications of Example 2.1 for the ability of the empirical likelihood ratio test to control size in finite samples when one only imposes requirements (6) and (7) on  $\mathbf{P}$ . To this end, recall the setup of Example 2.1 and note that (9) implies

$$\sup_{P \in \mathbf{P}_0} P\{\hat{P}_n \in \Lambda_1(\eta)\} \geq \sup_{c \in (0,1)} P_c\{\hat{P}_n \in \Lambda_1(\eta)\} = 1 .$$

Indeed, when one imposes only (6) and (7) on  $\mathbf{P}$ , it follows from Bahadur and Savage (1956) that there is no “reasonable” test in this setting in the sense that no test can have power against any alternative greater than size. On the other hand, Example 2.1 applies even if we impose the additional requirement that  $\mathcal{X} = [-1, 1]$ . With this additional requirement, the results of Bahadur and Savage (1956) no longer apply and reasonable tests exist. See, for example, Romano and Wolf (2000). ■

**Remark 2.2.** In parametric models where  $\mathbf{P}$  consists of distributions with common finite support, the likelihood ratio test for  $H_0 : E_P[X] = 0$  versus  $H_1 : E_P[X] \neq 0$  rejects for large values of

$$\inf_{P \in \mathbf{P}_0} I(\hat{P}_n|P) , \tag{13}$$

which equals (4) with probability approaching one under any  $P \in \mathbf{P}$  as  $n$  tends to infinity; see Newey and Smith (2004). In more general settings, however, it is important that the empirical likelihood ratio test takes the infimum over  $\mathbf{M}_0(\hat{P}_n)$  rather than over  $\mathbf{P}_0$ . For example, if  $\mathbf{P}$  is only required to satisfy (6) and (7), it is straightforward to see (13) equals zero while the test statistic of Qin and Lawless (1994) converges to a chi-squared random variable. ■

### 3 The Main Result

Our main result will require the following assumptions:

**Assumption 3.1.**  $\mathcal{X}$  and  $\Theta$  are compact subsets of  $\mathbf{R}^d$  and  $\mathbf{R}^r$ , respectively.

**Assumption 3.2.**  $g : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^m$  is continuous in both of its arguments.

The compactness of  $\mathcal{X}$ , imposed in Assumption 3.1, ensures that  $\mathbf{M}$  is itself compact in the weak topology, a crucial point in the proof of our main result. As implied by Examples 2.1 and 2.2,

however, additional requirements must be imposed on the set of probability measures  $\mathbf{P}$  in order for empirical likelihood to be able to control size as in (8). To this end, let

$$\Theta_0(P) = \{\theta \in \Theta : E_P[g(X, \theta)] = 0\} \quad (14)$$

$$\mathbf{M}_0 = \{P \in \mathbf{M} : E_P[g(X, \theta)] = 0 \text{ for some } \theta \in \Theta\} . \quad (15)$$

Additionally, for  $s(P, \theta)$  the dimension of the convex hull of the support of  $g(X, \theta)$  under  $P$ , define

$$\mathbf{D}_0 = \{P \in \mathbf{M} : \exists \theta \in \Theta_0(P) \text{ with } s(P, \theta) < m\} , \quad (16)$$

which we note is equal to  $\{D_0\}$  when  $d = 1$  and  $g(x, \theta) = x$  for all  $\theta \in \Theta$ .

The set  $\mathbf{P}$  is then restricted by requiring that it satisfy the following additional assumptions:

**Assumption 3.3.**  $\mathbf{P} \subseteq \mathbf{M}$  is closed in the weak topology.

**Assumption 3.4.** For each  $P \in \mathbf{P}$ ,  $\Theta_0(P)$  is either empty or a singleton denoted  $\theta_0(P)$ .

**Assumption 3.5.** For some  $\epsilon > 0$ ,  $\mathbf{P} \cap \mathbf{D}_0^\epsilon = \emptyset$ .

Given the compactness of  $\mathbf{M}$ , Assumption 3.3 implies  $\mathbf{P}$  is compact as well. It is left as a high level assumption, but we note closed sets in the weak topology are easily constructed by imposing moment restrictions on bounded continuous functions. Assumption 3.4 is employed to show that  $\theta_0(P)$  is continuous in  $P$  under the weak topology. Since we are typically interested in cases where  $m > r$ , we feel that Assumption 3.4 is not particularly restrictive. Assumption 3.5 is made precisely to avoid Examples 2.1 and 2.2 as it ensures  $\mathbf{D}_0$  contains no limit points of  $\mathbf{P}$ . We note that, provided Assumption 3.4 is met, we may let  $\mathbf{P} = \mathbf{M} \setminus \mathbf{D}_0^\epsilon$ .

We are now in a position to state our main result:

**Theorem 3.1.** Let  $X_i, i = 1, \dots, n$  be an i.i.d. sequence of random variables with distribution  $P \in \mathbf{P}$ . Let  $(\Lambda_0(\eta), \Lambda_1(\eta))$  be defined as in (5). Suppose Assumptions 3.1 - 3.5 hold. Then,

(a) There exists  $\bar{\eta} > 0$  such that for all  $0 < \eta \leq \bar{\eta}$  we have that

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n \{\hat{P}_n \in \Lambda_1(\eta)\} \leq -\eta .$$

(b) If a test  $(\Omega_{0,n}, \Omega_{1,n})$  satisfies

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n \{\hat{P}_n \in \Omega_{1,n}^\delta\} \leq -\eta \quad (17)$$

for some  $\delta > 0$ , then

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log Q^n \{\hat{P}_n \in \Omega_{0,n}\} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log Q^n \{\hat{P}_n \in \Lambda_0(\eta)\}$$

for any  $Q \in \mathbf{P}$  satisfying

$$\inf_{P \in \overline{\mathbf{M}_0 \setminus \mathbf{P}_0}} I(Q|P) \geq \inf_{P \in \mathbf{P}_0 \setminus \overline{\mathbf{M}_0 \setminus \mathbf{P}_0}} I(Q|P) . \quad (18)$$

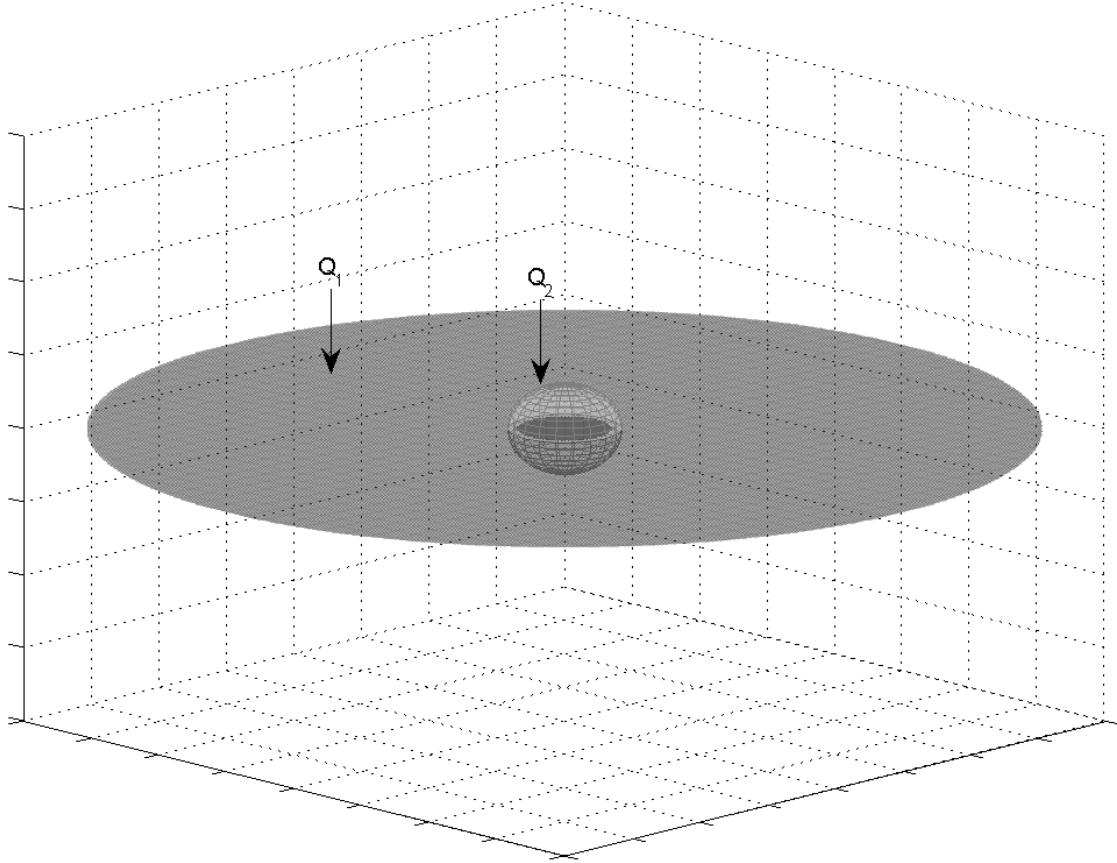


Figure 1: The larger and smaller (two-dimensional) ellipses represents  $\mathbf{M}_0$  and  $\mathbf{M}_0 \setminus \mathbf{P}_0$ , respectively, while the mesh ball represents  $\overline{\mathbf{M}_0 \setminus \mathbf{P}_0}^{\sqrt{\frac{\eta}{2}}}$ . Note  $Q_1$  is closer to  $\mathbf{P}_0 \setminus \overline{\mathbf{M}_0 \setminus \mathbf{P}_0}^{\sqrt{\frac{\eta}{2}}}$  than to  $\overline{\mathbf{M}_0 \setminus \mathbf{P}_0}^{\sqrt{\frac{\eta}{2}}}$  while the opposite is true for  $Q_2$ . Part (b) of Theorem 3.1 applies to  $Q_1$ , but not  $Q_2$ .

PROOF: See Supplemental Appendix. ■

Theorem 3.1 establishes the desired optimality property of empirical likelihood to test (2). First, the probability of a Type I error when using empirical likelihood tends to zero at a(n exponential) rate that is bounded away from zero on  $\mathbf{P}_0$ . Second, for any distribution  $Q$  satisfying (18), the probability of a Type II error when using empirical likelihood vanishes at a(n exponential) rate at least as fast as that of any non-randomized test satisfying the requirement (17). We emphasize the non-local nature of the optimality property, in that it holds for all distributions  $Q$  satisfying (18). Heuristically, condition (18) demands that  $Q$  be “closer” to the subset of  $\mathbf{M}_0$  over which we demand control of the rate at which the probability of a Type I error tends to zero than to the subset of  $\mathbf{M}_0$  over which we do not.

As illustrated by Examples 2.1 and 2.2, commonly used tests for (2) also fail to control the (exponential) rate at which the probability of a Type I error tends to zero if we allow for distributions in neighborhoods of  $\mathbf{D}_0$ . Rather than examining the performance of tests on restricted sets  $\mathbf{P}$  that



exclude  $\mathbf{D}_0$ , an alternative way of comparing among such procedures is to require

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{P}_n \in (\Omega_{1,n} \setminus \mathbf{D}_0^\epsilon)^\delta\} \leq -\eta \quad (19)$$

instead of (17). Requirement (19) should not be interpreted as “size” control, but rather as a benchmark for tests that have difficulty controlling the rate at which the Type I error tends to zero in neighborhoods of  $\mathbf{D}_0$ . Given the weaker criterion (19), it is clear that any optimal test must satisfy  $\mathbf{D}_0^\epsilon \subseteq \Omega_{1,n}$ . For this reason, consider

$$\begin{aligned} \tilde{\Lambda}_0(\eta) &= \Lambda_0(\eta) \setminus \mathbf{D}_0^\epsilon \\ \tilde{\Lambda}_1(\eta) &= \Lambda_1(\eta) \cup \mathbf{D}_0^\epsilon, \end{aligned} \quad (20)$$

where the dependence on  $\epsilon$  is omitted in the notation. Note that the tests  $(\Lambda_0(\eta), \Lambda_1(\eta))$  and  $(\tilde{\Lambda}_0(\eta), \tilde{\Lambda}_1(\eta))$  differ only on the event  $\hat{P}_n \in \mathbf{D}_0^\epsilon$ . We can use Theorem 3.1 to show that the optimal test in this framework is given by the modified empirical likelihood test defined by (20).

**Corollary 3.1.** *Let  $X_i, i = 1, \dots, n$  be an i.i.d. sequence of random variables with distribution  $P \in \mathbf{P}$ . Let  $(\tilde{\Lambda}_0(\eta), \tilde{\Lambda}_1(\eta))$  be defined as in (20). Under Assumptions 3.1, 3.2, 3.3 and 3.4:*

(a) *There exists  $\bar{\eta}(\epsilon) > 0$  such that for all  $0 < \eta \leq \bar{\eta}(\epsilon)$  we have that*

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \tilde{\Lambda}_1(\eta) \setminus \mathbf{D}_0^\epsilon\} \leq -\eta.$$

(b) *If a test  $(\Omega_{0,n}, \Omega_{1,n})$  satisfies*

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in (\Omega_{1,n} \setminus \mathbf{D}_0^\epsilon)^\delta\} \leq -\eta \quad (21)$$

*for some  $\delta > 0$ , then for every probability measure  $Q \in \mathbf{P}$ ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log Q^n\{\hat{P}_n \in \Omega_{0,n}\} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log Q^n\{\hat{P}_n \in \tilde{\Lambda}_0(\eta)\}. \quad (22)$$

PROOF: See Supplemental Appendix. ■

We reiterate that (19) differs from (17) only in how the former treats distributions that are close to the set  $\mathbf{D}_0$ . Remarkably, as a result of this rather simple modification, it is possible to remove almost all of the requirements on  $\mathbf{P}$ . Moreover, in contrast to Theorem 3.1, (22) holds without qualifications on  $Q$ . This result may therefore provide some guidance when choosing among tests that have difficulty controlling the rate at which the Type I error tends to zero in neighborhoods of  $\mathbf{D}_0$ , such as tests based on (generalized) empirical likelihood or the GMM-based  $J$ -test.

## References

- BAHADUR, R. R. and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, **27** 1115–1122.
- HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50** 891–916.
- 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.
- NEWNEY, W. K. and SMITH, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, **72** 219–255.
- OWEN, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75** 237–249.
- QIN, J. and LAWLESS, J. (1994). Empirical likelihood and generalized estimating equations. *Annals of Statistics*, **22** 300–325.
- ROMANO, J. P. (1989). Do bootstrap confidence intervals behave well uniformly in  $P$ ? *Canadian Journal of Statistics*, **17** 75–80.
- ROMANO, J. P. (2004). On nonparametric testing, the uniform behavior of the  $t$ -test, and related problems. *Scandinavian Journal of Statistics*, **31** 567–584.
- ROMANO, J. P. and WOLF, M. (2000). Finite sample nonparametric inference and large sample efficiency. *Annals of Statistics*, **28** 756–778.