Empirical likelihood ratio tests with power one

Albert Vexler*, Li Zou

Department of Biostatistics, The State University of New York at Buffalo, NY 14206, USA

ABSTRACT

In the 1970s, Professor Robbins and his coauthors extended the Vile and Wald inequality in order to derive the fundamental theoretical results regarding likelihood ratio based sequential tests with power one. The law of the iterated logarithm confirms an optimal property of the power one tests. In parallel with Robbins's decision-making procedures, we propose and examine sequential empirical likelihood ratio (ELR) tests with power one. In this setting, we develop the nonparametric one- and two-sided ELR tests. It turns out that the proposed sequential ELR tests significantly outperform the classical nonparametric *t*-statistic-based counterparts in many scenarios based on different underlying data distributions.

MSC: 97K70, 62L05, 62G10, 62G20.

Keywords: Empirical likelihood, Law of the iterated logarithm, Power one, *t*-statistic, Vile and Wald inequality, Sequential tests.

1. Introduction

Robbins (1970) as well as Robbins and Siegmund (1970) proposed the parametric likelihood ratio type tests with power one. Towards this end, the classical inequality obtained by Ville (1939) and Wald (1947) was extended to cover cases when the alternative joint density functions have forms of integrated likelihood functions in the context related to Bayes factor type procedures (Vexler et al., 2016a). The extended Vile and Wald inequality was employed to develop probability inequalities based on sums of independent and identically distributed (i.i.d.)

*Corresponding author.
E-mail addresses: avexler@buffalo.edu (A.Vexler), lizou@buffalo.edu (L. Zou)

random variables, providing a scheme to construct very accurate sequential procedures. These sequential decision-making mechanisms employ threshold bounds, which increase as slowly as possible with respect to the involved test statistics. This optimal property is confirmed by the law of the iterated logarithm. Thus, any attempt to improve Robbins's sequential schemes could lead to tests with high error probabilities (Robbins, 1970).

In nonparametric settings, the modern statistical literature has introduced various results related to the *t*-statistic-based sequential tests with power one (Govindarajulu, 2004; Mukhopadhyay and De Silva, 2008; Sen, 1981). In this context, we assume that $X_1, X_2, \ldots$ are i.i.d. random variables. Consider the parameter $\theta = E(X_1)$. In order to test sequentially the one(two)-sided hypothesis $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$ $(H_1 : \theta \neq \theta_0)$, where θ_0 is known, one can define the stopping rules N_0 (N_1) by

$$N_k = \inf \left\{ n : n \geq m_k \text{ and } n^{1/2} W_{k,n} \geq \left(c_{m_k} \log(n) \right)^{1/2} \right\}, \tag{1.1}$$

where $W_{0,n} = T_n$, $W_{1,n} = |T_n|$, $T_n = \left(\bar{X}_n - \theta_0 \right) \left\{ (n-1)^{-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2 \right\}^{-1/2}$, $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$, $m_k \geq 3$ defines an initial sample size and $c_{m_k} > 0$ is a pre-specified constant, *k*=0, 1, respectively. The corresponding decision-making policies consist of the following algorithms 1) We stop sampling and reject H_0 in favor of H_1 when $N_k < \infty$; 2) We continue sampling indefinitely and do not reject H_0 if $N_k = \infty$, *k*=0, 1, respectively. In order to control the Type I error rates of the schemes mentioned above at a fixed significance level α, we set values of m_k and c_{m_k} at (1.1) to satisfy the equation $\alpha = (k+1) \left[1 - \Phi \left\{ \left(c_{m_k} \log(m_k) \right)^{1/2} \right\} \right] \left\{ 1 + c_{m_k}^{-1} + \log(m_k) \right\}$ related to the one-sided (*k*=0) or two-sided (*k*=1) hypothesis test, where $\Phi(y) = (2\pi)^{-1/2} \int_{-\infty}^{y} \exp(-x^2/2) dx$.

These rules are asymptotic with respect to $m \to \infty$ (Mukhopadhyay and De Silva, 2008; Sen, 1981).

In this paper, we derive the sequential empirical likelihood ratio (ELR) tests with power one. It is well-known that in retrospective settings, the empirical likelihood (EL) methodology can significantly improve *t*-statistic type procedures (Owen, 1990; Vexler et al., 2009). It turns out that the ELR test statistics can approximate corresponding optimal parametric likelihood ratios (Lazar and Mykland, 1998; Owen, 1990; Qin and Lawless, 1994; Vexler et al., 2014a). Note that in the retrospective context of the one-sided hypothesis testing, the conventional EL methodology should be modified with respect to the statement of the testing problem. In this aspect, the standard ELR tests may suffer from power loss; for details, see DiCiccio and Romano (1989). Towards this end, DiCiccio and Romano (1989) established the signed root adjusted retrospective ELR test statistic, which significantly reduces the one-sided coverage error. Alternatively El Barmi (1996) introduced the retrospective ELR method to test for or against a set of inequality constraints. We implement and refine the methodologies developed by DiCiccio and Romano (1989) and El Barmi (1996) to propose the one-sided ELR sequential tests.

We examine the proposed ELR tests and derive asymptotic expressions of their error probabilities. We prove that the proposed procedures have the Type I error probabilities that are asymptotically equivalent to those of the corresponding nonparametric *t*-statistic-based testing schemes (1.1), whereas the Type II error probabilities of the proposed tests are equal to zero. Note that commonly Robbins's *t*-test type algorithms focus on the parameter $\theta = E(X_1)$. In this paper, we extend the hypothesis of interest to include more general cases when the parameter of interest θ is a solution of the equation $E\{G(X_1, \theta)\} = 0$, where G is a specified function. The function G can be associated with the statement of problem regarding a generalized estimating equation (e.g., Qin and Lawless, 1994)

This paper is organized as follows. Section 2 briefly reviews the ELR techniques for the one- and two-sided hypothesis tests. In Section 3, we propose and examine sequential ELR tests with power one. Section 4 presents the Monte Carlo (MC) study to demonstrate that the proposed sequential ELR tests significantly outperform the nonparametric *t*-statistic-based counterparts in many scenarios related to various underlying data distributions. Section 5 provides some concluding remarks. Proofs of the theoretical results presented in this paper are outlined in the Appendix, and are shown in the Supplementary Materials (SM) in details.

2. One- and two-sided retrospective empirical likelihood ratio tests

In this section, we outline the EL methodology developed for the one- and two-sided hypothesis tests. Assume that the parameter θ satisfies $E\{G(X_1,\theta)\}=0$. Then, the ELR is

$$L(\theta)=\max_{0<p_1,\ldots,p_n<1}\left\{\prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i=1, \sum_{i=1}^{n} p_i G(X_i,\theta)=0\right\}. \tag{2.1}$$

In this framework, applying the Lagrange multipliers method, we can derive $L(\theta)=\prod_{i=1}^{n} p_i$ with $p_i=\{n+\lambda G(X_i,\theta)\}^{-1}$, i=1,…,n, where λ satisfies $\sum_{i=1}^{n} G(X_i,\theta)\{n+\lambda G(X_i,\theta)\}^{-1}=0$, see Owen (2001) and Vexler et al. (2014a), for details. In the simple case $G(X,\theta)=X-\theta$, we have the parameter $\theta=E(X_1)$. Let $\Pr_k$ be the probability measure corresponding to the hypothesis H_k, k=0, 1, respectively. For the two-sided hypothesis $H_0:\theta=\theta_0$ vs. $H_1:\theta\neq\theta_0$, the ELR test statistic is

$$J_{n,2}=-2\log\{L(\theta_0)\}. \tag{2.2}$$

The nonparametric Wilks theorem shows that $\Pr_0\left(J_{n,2}\geq u\right)\to\Pr\left(\chi_1^2\geq u\right)$ as $n\to\infty$.

Now, consider the one-sided hypothesis $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$. In this scenario the development of ELR test statistics is more complicated than the two-sided setting mentioned above. We outline the retrospective EL methods following the schemes: 1) The adjusted ELR test statistic developed by DiCiccio and Romano (1989) has the form

$$J_{n,0} = \left\{ \operatorname{sgn}\left(\hat{\theta}_n - \theta_0\right) J_{n,2}^{1/2} - n^{-1/2} \hat{\Lambda}_n \right\} \Big/ \left(1 + n^{-1} \hat{b}_n / 2\right), \tag{2.3}$$

where $\hat{\theta}_n$ satisfies $n^{-1} \sum_{i=1}^{n} G\left(X_i, \hat{\theta}_n\right) = 0$, sequences $\hat{\Lambda}_n$ and $\hat{b}_n$ are chosen such that $\Pr_0\left(J_{n,0} \ge u\right) = 1 - \Phi(u) + O\left(n^{-3/2}\right)$ as $n \to \infty$ (DiCiccio and Romano, 1989; Lazar and Mykland, 1998). In this context, we note that the conventional EL literature related to the signed root approach considers $\theta = h(E(X_1))$, where h is a smooth function; 2) The ELR approach shown in El Barmi (1996) employs the test statistic

$$J_{n,1} = -2 \log \left\{ L(\theta_0) \Big/ \sup_{\theta > \theta_0} L(\theta) \right\}, \tag{2.4}$$

where $L(\theta_0)$ is defined in (2.1). In this case, $\Pr_0\left(J_{n,1} \ge u\right) \to 0.5 I(u \le 0) + 0.5 \Pr\left(\chi_1^2 \ge u\right)$ as $n \to \infty$, where $I(\cdot)$ is the indicator function (see Theorem 3.2 shown in El Barmi, 1996).

3. Sequential ELR Procedures

In parallel with the sequential t-statistic-based nonparametric procedures shown in Section 1, we propose the following ELR based stopping rules

$$\tau_k = \inf \left\{ n : n \ge m_k , J_{n,k} \ge d_{n,k} \right\}, k{=}0, 1, 2, \tag{3.1}$$

where $J_{n,k}$, k=0, 1, 2, are defined in (2.1)-(2.3), $d_{n,0} = \left(c_{m_0} \log(n)\right)^{1/2}$, $d_{n,1} = c_{m_1} \log(n)$, $d_{n,2} = c_{m_2} \log(n)$, $c_{m_k} > 0$, k=0, 1, 2, are deterministic sequences that satisfy the equation $b_j \left[1 - \Phi\left\{\left(c_{m_j} \log\left(m_j\right)\right)^{1/2}\right\}\right]\left\{1 + c_{m_j}^{-1} + \log\left(m_j\right)\right\} = \alpha$ with $b_0 = b_1 = 1$, $b_2 = 2$, α denotes the pre-specified significance level, and $\Phi(y) = (2\pi)^{-1/2}\int_{-\infty}^{y}\exp(-x^2/2)dx$. The stopping rule τ_2 is related to the two-sided scenario, and τ_k (k=0, 1) correspond to the one-sided scenario. The new decision-making policies consist of the algorithms: 1) We stop sampling and reject H_0 in favor of H_1 when $\tau_k < \infty$; 2) We continue sampling indefinitely and do not reject H_0 if $\tau_k = \infty$, k=0, 1, 2.

In order to analyze the proposed sequential ELR tests defined in (3.1), we assume the following conditions:

(A1) $\partial G(u,\theta)/\partial\theta < 0$ (or $\partial G(u,\theta)/\partial\theta > 0$), for all u;

(A2) There exists functions $M_k(u)$, k=0, 1, 2, such that $0 < M_0(u) \le |\partial G(u,\theta)/\partial\theta| \le M_1(u)$, $|\partial^2 G(u,\theta)/\partial\theta^2| \le M_2(u)$ in a neighborhood of the true value θ that satisfies $E\{G(X_1,\theta)\} = 0$;

(A3) $E[\exp\{tG(X_1,\theta)\}] < \infty$, $E[\exp\{tM_k(X_1)\}] < \infty$, k=0, 1, 2, in some neighborhood of the point *t=0*.

Remarks: 1) It is clear that assumptions (A1)-(A2) are satisfied in the case when $G(u,\theta) = u - \theta$. Regarding assumption (A1), we refer the reader to, e.g., Vexler et al. (2014b) and Vexler et al. (2016b), for more details; 2) The assumption (A3) has a form that is similar to that required in theoretical evaluations regarding the nonparametric *t*-statistic based test with power one (Darling and Robbins, 1967); 3) The assumption (A2) can be substituted by the assumptions of Lemma 1 presented in Qin and Lawless (1994).

Thus, we have the following result.

Proposition 1. *Assume conditions (A1)-(A3) are satisfied. Then we have*

$$\lim_{m_j\to\infty}\left| \Pr_0(\tau_j < \infty) - b_j\left[1-\Phi\left\{(c_{m_j}\log(m_j))^{1/2}\right\}\right]\left[1+1/c_{m_j}+\log(m_j)\right]\right| = 0 \quad ,j=0,\ 1,\ 2,$$

where $b_0 = b_1 = 1$, $b_2 = 2$, $d_{n,j}$ *is defined in equation (3.1),* c_{m_j} *satisfies that* $b_j\left[1-\Phi\left\{\left(c_{m_j}\log\left(m_j\right)\right)^{1/2}\right\}\right]\left\{1+c_{m_j}^{-1}+\log\left(m_j\right)\right\}=\alpha$, α *denotes the pre-specified significance level, and* $\Phi(y)=(2\pi)^{-1/2}\int_{-\infty}^{y}\exp(-x^2/2)dx$.

Proposition 1 shows that the Type I error probabilities of the proposed ELR tests are asymptotically equal to those of their *t*-statistic-based counterparts defined in (1.1)-(1.2). Note that the conventional *t*-statistic based procedures focus on $G(u,\theta)=u-\theta$, whereas we consider more general situations.

The next proposition shows that the powers of the proposed ELR tests are equal to one.

Proposition 2. *Assume conditions (A1)-(A2) are satisfied and* $E\left|G\left(X_1,\theta\right)\right|^3 \leq \infty$. *Then*

$$\Pr_1\left\{\tau_k < \infty\right\} = 1, \qquad k=0,\ 1,\ 2.$$

4. Monte Carlo study

In this section we compare the performances of the nonparametric *t*-statistic-based tests defined in (1.1)-(1.2) and the new tests (3.1), evaluating the corresponding Type I error rates and expected sample sizes. Without loss of generality, we focus on the case of $G(u,\theta)=u-\theta$, where the parameter of interest is $\theta = E\left(X_1\right)$. This case is related to a common scenario that appears in many practical applications. Let the notations *t*-test, ELR, ELR1, and ELR2 denote the *t*-statistic-based test defined in (1.1), the proposed two-sided method defined in (3.1) based on the statistic (2.2), and the one-sided methods defined in (3.1) based on the statistics (2.3)-(2.4), respectively. It is known that, when observations are normally distributed, EL methods demonstrate good properties, whereas EL type procedures based on skewed data can lead to unstable results (e.g., Vexler et al., 2009). In this MC study, we selected Normal (*N*), Exponential (*Exp*), Chi-square

(*Chisq*), and Lognormal (*LN*) distribution functions to generate data points. At each baseline distribution, the MC experiments were replicated 10,000 times to generate underlying data points. We used *m*=50, 75 and 100 in the definitions of the considered test procedures (1.1), (1.2) and (3.1).

Table 1 shows the MC Type I error rates related to the two(one)-sided hypothesis. The expected significance level of the tests was chosen to be $\alpha = 0.05$. The MC Type I error rates of the *t*-statistic-based tests are very close to 0.05 when the observations are normally distributed. In the cases with skewed data distributions, the MC Type I error rates of the proposed ELR tests are almost uniformly closer to the expected level of 5% than those of the *t*-statistic-based tests. Consider the scenario with observations $X_i = 1 - \zeta_i$, where $\zeta_i \sim Chisq(1)$, *i*=1, 2,…. The MC Type I error rates of the proposed ELR tests are significantly closer to 0.05 than those of the *t*-statistic-based tests. For example, in the case with *m*=50 (two-sided), the Type I error rate of the *t*-statistic-based test is 0.105, whereas in this case the Type I error rate of the proposed ELR test is 0.067. We note that the MC results presented in this section are consistent with those of numerical comparisons between retrospective *t*-statistics and EL ratios shown in many relevant MC experiments published in the literature (e.g., Vexler et al., 2009; Zhou and Gao, 2000).

Table 2 shows the MC estimators of the expected sample sizes related to the performances of the considered sequential tests. In several scenarios, the MC expectations of sample sizes required by the new tests are a little larger than those related to the *t*-statistic-based tests. For example, when $X_i = 1.5 - \zeta_i$, where $\zeta_i \sim Exp(1)$, *i*=1, 2,…, with *m*=75 in the two-sided setting, the proposed ELR test has the MC expectation of the sample size that is equal to 80.5, which is larger than 77.2, the MC expectation of the sample size of the classical test. In this case, one can note that the relevant MC results shown in Table 1 demonstrate that the MC Type I error rate of the *t*-statistic-based test is 0.069 comparing with 0.052 of the new ELR test. Then, the smaller MC average sample sizes related to the classical tests do not show that the *t*-statistic based tests are somewhat better than the new tests.

We also note that, in this MC study, it turns out that the strategy used in the ELR2's construction provides a better Type I error rate controlling and smaller MC average sample sizes than that used in the ELR1's development.

Table 1. The MC Type I error rates of the classical tests and the new tests for two(one)-sided hypothesis: $H_0:\theta=0$ *vs* $H_1:\theta\neq 0$ ($H_1:\theta>0$) with $\theta=E(X_1)$. The expected Type I error rate is α =0.05.

Two-sided	Tests	*N*(0,2)	1-*Exp(1)*	1-*Chisq*(1)	1.13-*LN(0,0.5)*
m=50	*t*-test	0.053	0.077	0.105	0.067
	ELR	0.043	0.058	0.067	0.055
m=75	*t*-test	0.050	0.069	0.084	0.061
	ELR	0.045	0.052	0.058	0.050
m=100	*t*-test	0.049	0.066	0.079	0.058
	ELR	0.046	0.048	0.049	0.050
One-sided *m=50*	*t*-test	0.047	0.101	0.138	0.094
	ELR1	0.042	0.062	0.074	0.061
	ELR2	0.048	0.061	0.071	0.054
m=75	*t*-test	0.046	0.089	0.107	0.081
	ELR1	0.041	0.052	0.062	0.054
	ELR2	0.044	0.048	0.060	0.047
m=100	*t*-test	0.046	0.082	0.105	0.075
	ELR1	0.041	0.051	0.055	0.046
	ELR2	0.042	0.051	0.057	0.046

Table 2. The MC average sample sizes of the classical tests and the new tests for two(one)-sided hypothesis $H_0:\theta=0$ vs $H_1:\theta=0$ ($H_1:\theta>0$) with $\theta=E(X_1)$, where the alternative parameter values are θ=0.5 for $X_i=\zeta_i$, where $\zeta_i\sim N(0.5,2)$; θ=0.5 for $X_i=1.5-\zeta_i$, where

$\zeta_i \sim Exp(1)$; θ=0.8 for $X_i = 1.8 - \zeta_i$, where $\zeta_i \sim Chisq(1)$; θ=0.37 for $X_i = 1.5 - \zeta_i$, where $\zeta_i \sim LN(0, 0.5)$, i=1, 2,….

Expected Sample Sizes					
Two-sided	Tests	*N*(0.5,2)	1.5-*Exp*(1)	1.8-*Chisq*(1)	1.5-*LN*(0,0.5)
m=50	*t*-test	128.1	55.4	53.2	50.1
	ELR	131.3	62.2	59.4	50.6
m=75	*t*-test	132.4	77.2	76.1	75.1
	ELR	135.4	80.5	79.5	75.2
m=100	*t*-test	143.3	100.6	100.3	100.0
	ELR	145.3	102.3	101.8	100.0
One-sided *m=50*	*t*-test	109.3	53.7	52.3	51.0
	ELR1	114.7	59.8	57.7	53.6
	ELR2	109.8	57.8	56.3	52.8
m=75	*t*-test	117.2	76.3	75.8	75.2
	ELR1	120.8	79.3	78.6	76.2
	ELR2	119.7	78.0	77.5	75.7
m=100	*t*-test	130.3	100.5	100.2	100.1
	ELR1	132.7	101.6	101.3	100.3
	ELR2	131.2	101.1	100.8	100.2

5. Conclusion

We have developed the power one sequential procedures employing the EL methodology. We have evaluated the one- and two-sided ELR tests with power one, extending and improving the relevant nonparametric *t*-statistic-based procedures. To the best of our knowledge, perhaps, this paper presents a research that belongs to a first cohort of studies related to applications of the EL techniques in order to construct sequential statistical procedures. We aim that the present article will convince the readers of the usefulness of EL techniques in various statistical sequential aspects.

Acknowledgement

This research was supported by the National Institutes of Health (NIH) grant 1G13LM012241-01. We are grateful to the Editor, Professor Yimin Xiao, for his helpful comments that clearly improved this paper.

Appendix A. Proof Schemes of Propositions 1 and 2.

Proof of Proposition 1.

Without loss of generality, we consider the stopping rule τ_2 in the two-sided setting. In a similar manner to the analysis presented below, one can evaluate the procedures τ_0 and τ_1 defined in (3.1). Denote $lr_n(\theta_0) = \log\{L(\theta_0)\}$, $R_n(\theta) = \partial^3 lr_n(\theta)/\partial\theta^3$, $\partial\lambda(\theta)/\partial\theta = \lambda'(\theta)$, $\partial^2\lambda(\theta)/\partial\theta^2 = \lambda''(\theta)$, $\partial G(u,\theta)/\partial\theta = G'(u,\theta)$, and $\partial^2 G(u,\theta)/\partial\theta^2 = G''(u,\theta)$, where $L(\theta_0)$ is defined in (2.1). Applying Taylor's theorem to $lr_n(\theta_0)$ (see SM for details), we obtain the Type I error rate

$$\Pr_0(\tau_2 < \infty) = \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 - \frac{1}{3}\left(\theta_0 - \hat{\theta}_n\right)^3 R_n\left(\tilde{\theta}_n\right) > c_m \log(n) \; for\, some\, n \ge m\right\}, \quad (A.1)$$

where $\hat{\theta}_n$ is a root of $\sum_{i=1}^n G\left(X_i, \hat{\theta}_n\right) = 0$, $\tilde{\theta}_n = \theta_0 + \rho\left(\hat{\theta}_n - \theta_0\right)$, $\rho \in (0,1)$, and $\sigma_n^2 = n^{-1}\sum_{i=1}^n \left\{G\left(X_i, \hat{\theta}_n\right)\right\}^2 \Big/ \left\{n^{-1}\sum_{i=1}^n G'\left(X_i, \hat{\theta}_n\right)\right\}^2$.

In order to evaluate (A.1), we derive the next lemma.

Lemma 1. *Assume conditions (A1)-(A3) are satisfied. Then*

$$\Pr_0\left\{n\left|\hat{\theta}_n - \theta_0\right| \ge a_n \; for \; some \; n \ge m\right\} = O\{1/\log\log(m)\},$$

where $a_n = \{2n \log\log(n)\}^{0.5}\sigma_G/\gamma$, $\sigma_G = \left[Var\{G(X_1, \theta_0)\}\right]^{0.5}$ *and* $\gamma = E\{M_0(X_1)\}$ *as* $m \to \infty$.

Appendix B of SM presents the proof of Lemma 1.

By virtue of Lemma 1, we will obtain upper and lower bounds for $\Pr_0(\tau_2 < \infty)$. To this end, we use (A.1) to show that

$$\Pr_0(\tau_2 < \infty)$$

$$= \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 - \frac{1}{3}\left(\theta_0 - \hat{\theta}_n\right)^3 R_n\left(\tilde{\theta}_n\right) > c_m \log(n) \textit{ for some } n \geq m, n\left|\hat{\theta}_n - \theta_0\right| \geq a_n \textit{ for some } n \geq m\right\}$$

$$+ \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 - \frac{1}{3}\left(\theta_0 - \hat{\theta}_n\right)^3 R_n\left(\tilde{\theta}_n\right) > c_m \log(n) \textit{ for some } n \geq m, n\left|\hat{\theta}_n - \theta_0\right| < a_n \textit{ for all } n \geq m\right\}$$

$$\leq \Pr_0\left\{n\left|\hat{\theta}_n - \theta_0\right| \geq a_n \textit{ for some } n \geq m\right\} + U_{0m},$$

where $\Pr_0\left\{n\left|\hat{\theta}_n - \theta_0\right| \geq a_n \textit{ for some } n \geq m\right\} = O\{1/\log\log(m)\}$, and

$$U_{0\mathrm{m}} = \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 + \frac{a_n^3}{n^3}\left|R_n\left(\tilde{\theta}_n\right)\right| > c_m \log(n) \textit{ for some } n \geq m\right\}.$$

Regarding the term $U_{0\mathrm{m}}$, we have the inequality

$$U_{0\mathrm{m}} = \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 + \frac{a_n^3}{n^3}\left|R_n\left(\tilde{\theta}_n\right)\right| > c_m \log(n) \textit{ for some } n \geq m, \left|R_n\left(\tilde{\theta}_n\right)\right| < n^{1+\delta} \textit{ for all } n \geq m\right\}$$

$$+ \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 + \frac{a_n^3}{n^3}\left|R_n\left(\tilde{\theta}_n\right)\right| > c_m \log(n) \textit{ for some } n \geq m, \left|R_n\left(\tilde{\theta}_n\right)\right| \geq n^{1+\delta} \textit{ for some } n \geq m\right\} \quad \text{(A.2)}$$

$$\leq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 + \frac{a_n^3}{n^{2-\delta}} > c_m \log(n) \textit{ for some } n \geq m\right\} + \Pr_0\left\{\left|R_n\left(\tilde{\theta}_n\right)\right| < n^{1+\delta} \textit{ for some } n \geq m\right\},$$

where δ is assumed to be in $(0, 0.5)$.

Now we consider the lower bound of $\Pr_0(\tau_2 < \infty)$. Noting that, by (A.1) we have

$$\Pr_0(\tau_2 < \infty)$$

$$\geq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 - \frac{1}{3}\left(\theta_0 - \hat{\theta}_n\right)^3 R_n\left(\tilde{\theta}_n\right) > c_m \log(n) \textit{ for some } n \geq m, n\left|\hat{\theta}_n - \theta_0\right| < a_n \textit{ for all } n \geq m\right\}$$

$$\geq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{a_n^3}{n^3}\left|R_n\left(\widetilde{\theta}_n\right)\right|>c_m\log(n)\ for\ some\ n\geq m, n\left|\hat{\theta}_n-\theta_0\right|<a_n\ for\ all\ n\geq m\right\}$$

$$\geq U_{1m}\text{ - }\Pr_0\left\{n\left|\hat{\theta}_n-\theta_0\right|\geq a_n\ for\ some\ n\geq m\right\},$$

where $\Pr_0\left\{n\left|\hat{\theta}-\theta_0\right|\geq a_n\ for\ some\ n\geq m\right\}=O\{1/\log\log(m)\}$, and

$$U_{1\mathrm{m}}=\Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{a_n^3}{n^3}\left|R_n\left(\widetilde{\theta}_n\right)\right|>c_m\log(n)\ for\ some\ n\geq m\right\}.$$

Regarding the term $U_{1\mathrm{m}}$, we have the following inequality

$$U_{1\mathrm{m}}\geq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{a_n^3}{n^3}\left|R_n\left(\widetilde{\theta}_n\right)\right|>c_m\log(n)\ for\ some\ n\geq m,\left|R_n\left(\widetilde{\theta}_n\right)\right|<n^{1+\delta}\ for\ all\ n\geq m\right\}$$

$$\geq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{a_n^3}{n^{2-\delta}}>c_m\log(n)\ for\ some\ n\geq m\right\}-\Pr_0\left\{\left|R_n\left(\widetilde{\theta}_n\right)\right|\geq n^{1+\delta}\ for\ some\ n\geq m\right\},\quad (A.3)$$

where the term $a_n^3 n^{-2+\delta}\to 0$ as $n\to\infty$. The following lemma analyzes the term $\Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2>c_m\log(n)\ for\ some\ n\geq m\right\}$ provided in (A.2) and (A.3).

Lemma 2. *Assume conditions (A1)-(A3) are satisfied. Then*

$$\lim_{m\to\infty}\left|\Pr_0\left\{n\sigma_n^{-2}\left(\theta_0-\hat{\theta}_n\right)^2>c_m\log(n)\ for\ some\ n\geq m\right\}-2\left[1-\Phi\left\{\left(c_m\log(m)\right)^{1/2}\right\}\right]\left\{1+c_m^{-1}+\log(m)\right\}\right|=0,$$

where c_m *satisfies* $2\left\{1-\Phi\left(\left(c_m\log(m)\right)^{1/2}\right)\right\}\left\{1+c_m^{-1}+\log(m)\right\}=\alpha$ *,* α *denotes the pre-specified significance level, and* $\Phi(y)=(2\pi)^{-1/2}\int_{-\infty}^{y}\exp(-x^2/2)dx$.

Appendix B of SM presents the proof of Lemma 2.

Thus, according to (A.2)-(A.3) and Lemma 2, we need to prove that $\Pr_0\left\{\left|R_n\left(\widetilde{\theta}_n\right)\right|>n^{1+\delta}\ for\ some\ n\geq m\right\}\to 0$ as $m\to\infty$ in order to analyze $\Pr_0\{\tau_2<\infty\}$. Since the remainder term $R_n\left(\widetilde{\theta}_n\right)$ is a function of the Lagrangian multiplier $\lambda(\theta)$ defined in (2.1), we begin with deriving theoretical results regarding $\lambda(\theta)$. These results might have an independent interest in evaluation of relevant EL problems. The following lemmas are proven in Appendix B of SM.

Lemma 3. *We have that* $\lambda \geq 0$ if and only if $\sum_{i=1}^{n} G(X_i, \theta) \geq 0$, and $\lambda < 0$ if and only if $\sum_{i=1}^{n} G(X_i, \theta) < 0$.

Lemma 4. *The Lagrange multiplier* λ *satisfies*

$$\lambda = \frac{\sum_{i=1}^{n} G(X_i, \theta)}{\sum_{i=1}^{n} \{G(X_i, \theta)\}^2 p_i}.$$

Lemma 5. *If* $\lambda \geq 0$, *we have* $0 \leq \lambda \leq n\sum_{i=1}^{n} G(X_i, \theta)\left[\sum_{i=1}^{n} \{G(X_i, \theta)\}^2 I\{G(X_i, \theta) < 0\}\right]^{-1}$; *If* $\lambda < 0$, *we have* $n\sum_{i=1}^{n} G(X_i, \theta)\left[\sum_{i=1}^{n} \{G(X_i, \theta)\}^2 I\{G(X_i, \theta) > 0\}\right]^{-1} \leq \lambda < 0$.

Lemmas 3-5 are applied to obtain the next result regarding remainder term $R_n(\tilde{\theta}_n)$ formulated in Lemma 6 below.

Lemma 6. *Assume conditions (A1)-(A3) are satisfied. Then*

$$\Pr_0\left\{\left|R_n(\tilde{\theta}_n)\right| > n^{1+\delta} \text{ for some } n \geq m\right\} = O\left\{(\log\log m)^{-1}\right\} \text{ as } m \to \infty,$$

where $\delta \in (0, 0.5)$.

By virtue of (A.2)-(A.3) and Lemma 6, we complete the proof of Proposition 1 with respect to the two-sided stopping rule τ_2.

Remark: In this paper we provide Lemma 6, employing Lemmas 1-5. This result can be confirmed by adapting Lemma 7 presented in Zhong and Ghosh (2016) that concludes with

$\left|\partial^3 lr_n(\theta)/\partial\theta^3\right| \le C_1 n$ almost surely, when θ is in a small neighborhood of $\hat{\theta}_n$, for relatively large *n*, where C_1 denotes a positive constant.

Proof of Proposition 2.

The proof of Proposition 2 is based on the following strategy. Assume that we consider the hypothesis $H_0:\theta=\theta_0$ *vs.* $H_1:\theta\ne\theta_0$, where θ_0 is known. Let $\theta_1\ne\theta_0$ satisfy $E\{G(X_1,\theta_1)\}=0$, under H_1. Under the alternative hypothesis, we show that, for large values of *n*, $-2lr_n(\theta_0)/n^{\varepsilon}>C$, whereas $c_m\log(n)/n^{\varepsilon}\to 0$, as $n\to\infty$, where $\varepsilon>0$ and C is a positive constant. This implies that the Type II error probability $\Pr_1\{-2lr_n(\theta_0)\le c_m\log(n)\, for\ all\ n\ge m\}=0$, since there exists *N* such that, for all $n>N$, the event $\{-2lr_n(\theta_0)\le c_m\log(n)\}$ is empty. Thus, the proposed two-sided ELR sequential procedure τ_2 defined at (3.1) has power one. The stopping rules τ_0 and τ_1 defined at (3.1) can be considered in a similar manner to that shown above (corresponding formal details are presented in the online SM.)

Appendix B. Supplementary Material for detailed proof.

Supplementary material related to this article can be found online.

References

Darling, D. and Robbins, H., 1967. Iterated logarithm inequalities. Proceedings of the National Academy of Sciences 57, 1188-1192.

DiCiccio, T.J. and Romano, J.P., 1989. On adjustments based on the signed root of the empirical likelihood ratio statistic. Biometrika 76, 447-456.

El Barmi, H., 1996. Empirical likelihood ratio test for or against a set of inequality constraints. Journal of Statistical Planning and Inference 55, 191-204.

Govindarajulu, Z., 2004. Sequential statistics. World Scientific Publishing Co Inc.

Lazar, N. and Mykland, P.A., 1998. An evaluation of the power and conditionality properties of empirical likelihood. Biometrika 85, 523-534.

Mukhopadhyay, N. and De Silva, B.M., 2008. Sequential methods and their applications. CRC press.

Owen, A., 1990. Empirical likelihood ratio confidence regions. The Annals of Statistics, 90-120.

Owen, A., 2001. Empirical likelihood. CRC press.

Qin, J. and Lawless, J., 1994. Empirical likelihood and general estimating equations. The Annals of Statistics, 300-325.

Robbins, H., 1970. Statistical methods related to the law of the iterated logarithm. The Annals of Mathematical Statistics 41, 1397-1409.

Robbins, H. and Siegmund, D., 1970. Boundary crossing probabilities for the Wiener process and sample sums. The Annals of Mathematical Statistics 41, 1410-1429.

Sen, P.K., 1981. Sequential nonparametrics: invariance principles and statistical inference. Wiley New York.

Vexler, A., Hutson, A.D., and Chen, X., 2016a. Statistical Testing Strategies in the Health Sciences. CRC Press.

Vexler, A., Hutson, A.D., and Yu, J., 2014a. Empirical Likelihood Methods in Clinical Experiments. Methods and Applications of Statistics in Clinical Trials: Planning, Analysis, and Inferential Methods, Volume 2, 146-165.

Vexler, A., Liu, S., Kang, L., and Hutson, A.D., 2009. Modifications of the empirical likelihood interval estimation with improved coverage probabilities. Communications in Statistics-Simulation and Computation 38, 2171-2183.

Vexler, A., Tao, G., and Hutson, A.D., 2014b. Posterior expectation based on empirical likelihoods. Biometrika 101, 711-718.

Vexler, A., Zou, L., and Hutson, A.D., 2016b. Data-Driven Confidence Interval Estimation Incorporating Prior Information with an Adjustment for Skewed Data. The American Statistician 70, 243-249.

Ville, J., 1939. Etude critique de la notion de collectif. Gauthier-Villars Paris.

Wald, A., 1947. Sequential Analysis. Wiley, New York.

Zhong, X. and Ghosh, M., 2016. Higher-order properties of Bayesian empirical likelihood. Electronic Journal of Statistics 10, 3011-3044.

Zhou, X.-H. and Gao, S., 2000. One-sided confidence intervals for means of positively skewed distributions. The American Statistician 54, 100-104.

Supplementary Material to:

Empirical likelihood ratio tests with power one

Albert Vexler and Li Zou

Department of Biostatistics, University at Buffalo, State University of New York,

Buffalo, New York, USA.

Proofs

Proof of Proposition 1.

Without loss of generality, we consider the stopping rule τ_2 in the two-sided hypothesis setting. We obtain a third order expansion of the ELR test statistic $J_{n,2} = -2\log\{L(\theta_0)\}$, where $L(\theta_0) = \prod_{i=1}^n \{np_i(\theta_0)\}$, $p_i(\theta_0) = \{n + \lambda(\theta_0)G(X_i, \theta_0)\}^{-1}$, *i=1,…,n,* and $\lambda(\theta_0)$ satisfies $\sum_{i=1}^n G(X_i, \theta_0)\{n + \lambda G(X_i, \theta_0)\}^{-1} = 0$. Denote $lr_n(\theta_0) = \log\{L(\theta_0)\}$. Applying Taylor's theorem, we have

$$lr_n(\theta_0) = lr_n(\hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)\left.\frac{\partial lr_n(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_n} + \frac{1}{2}(\theta_0 - \hat{\theta}_n)^2\left.\frac{\partial^2 lr_n(\theta)}{\partial\theta^2}\right|_{\theta=\hat{\theta}_n} + \frac{1}{6}(\theta_0 - \hat{\theta}_n)^3\left.\frac{\partial^3 lr_n(\theta)}{\partial\theta^3}\right|_{\theta=\tilde{\theta}_n},$$

where $\tilde{\theta}_n = \theta_0 + \rho(\hat{\theta}_n - \theta_0)$, $\rho \in (0,1)$ and $\hat{\theta}_n$ satisfies $n^{-1}\sum_{i=1}^n G(X_i, \hat{\theta}_n) = 0$.

To simplify notations, denote $\partial\lambda(\theta)/\partial\theta = \lambda'(\theta)$, $\partial^2\lambda(\theta)/\partial\theta^2 = \lambda''(\theta)$, $\partial G(u,\theta)/\partial\theta = G'(u,\theta)$, and $\partial^2 G(u,\theta)/\partial\theta^2 = G''(u,\theta)$. By virtue of the definitions of $lr_n(\theta)$ and $\lambda(\theta)$, we can obtain

$$\frac{\partial lr_n(\theta)}{\partial\theta} = -\lambda(\theta)\sum_{i=1}^n G'(X_i,\theta)p_i,$$

$$\frac{\partial^2 lr_n(\theta)}{\partial\theta^2} = -\lambda'(\theta)\sum_{i=1}^n G'(X_i,\theta)p_i - \lambda(\theta)A(\theta),$$

$$\frac{\partial^3 lr_n(\theta)}{\partial\theta^3} = 2\{\lambda'(\theta)\}^2 \sum_{i=1}^{n} G'(X_i,\theta)G(X_i,\theta)p_i^2 + J(\theta),$$

$$\lambda'(\theta) = \frac{n\sum_{i=1}^{n} p_i^2 G'(X_i,\theta)}{\sum_{i=1}^{n} p_i^2 \{G(X_i,\theta)\}^2},$$

$$\lambda''(\theta) = \left[\sum_{i=1}^{n} p_i^2 \{G(X_i,\theta)\}^2\right]^{-1} \Big[\sum_{i=1}^{n} p_i^2 \{nG''(X_i,\theta) - 2\lambda'(\theta)G'(X_i,\theta)G(X_i,\theta)\}$$
$$- 2\sum_{i=1}^{n} p_i^3 \left\{nG'(X_i,\theta) - \lambda'(\theta)\{G(X_i,\theta)\}^2\right\}\{\lambda'(\theta)G(X_i,\theta) + \lambda(\theta)G'(X_i,\theta)\}\Big],$$

where $p_i = \{n + \lambda G(X_i,\theta)\}^{-1}$, $i=1,\dots,n$,

$A(\theta) = -\lambda(\theta)\sum_{i=1}^{n} \{G'(X_i,\theta)\}^2 p_i^2 + \sum_{i=1}^{n} G''(X_i,\theta)p_i - \lambda'(\theta)\sum_{i=1}^{n} G(X_i,\theta)G'(X_i,\theta)p_i^2$, and

$$J(\theta) = -2\lambda'(\theta)\sum_{i=1}^{n} G''(X_i,\theta)p_i + 2\lambda'(\theta)\lambda(\theta)\sum_{i=1}^{n} G'(X_i,\theta)p_i^2 - \lambda(\theta)A'(\theta) - \lambda''(\theta)\sum_{i=1}^{n} G'(X_i,\theta)p_i, \qquad \text{(S.1)}$$

see Vexler et al. (2014a) for details.

It is clear that the argument $\hat{\theta}_n$, which satisfies $n^{-1}\sum_{i=1}^{n} G(X_i,\hat{\theta}_n) = 0$, maximizes $lr_n(\theta)$. In this case, $p_i(\hat{\theta}_n) = n^{-1}$, $i=1,\dots,n$. Then we have

$$lr_n(\hat{\theta}_n) = \partial lr_n(\theta)/\partial\theta\big|_{\theta=\hat{\theta}_n} = \lambda(\hat{\theta}_n) = 0,$$

$$\lambda'(\hat{\theta}_n) = \frac{\sum_{i=1}^{n} G'(X_i,\hat{\theta}_n)}{n^{-1}\sum_{i=1}^{n} \{G(X_i,\hat{\theta}_n)\}^2}, \quad \left.\frac{\partial^2 lr_n(\theta)}{\partial\theta^2}\right|_{\theta=\hat{\theta}_n} = -\frac{n^{-1}\left\{\sum_{i=1}^{n} G'(X_i,\hat{\theta}_n)\right\}^2}{n^{-1}\sum_{i=1}^{n} \{G(X_i,\hat{\theta}_n)\}^2}.$$

Taking into account these results shown above and the Taylor expansion of the ELR statistic, we obtain $lr_n(\theta_0)$ in the form

$$lr_n(\theta_0) = -\frac{n}{2\sigma_n^2}(\theta_0 - \hat{\theta}_n)^2 + \frac{1}{6}(\theta_0 - \hat{\theta}_n)^3 \left.\frac{\partial^3 lr_n(\theta)}{\partial\theta^3}\right|_{\theta=\tilde{\theta}_n}, \qquad \text{(S.2)}$$

where $\sigma_n^2 = n^{-1}\sum_{i=1}^{n} \{G(X_i,\hat{\theta}_n)\}^2 \Big/ \left\{n^{-1}\sum_{i=1}^{n} G'(X_i,\hat{\theta}_n)\right\}^2$.

Let $\Pr_k$ be the probability measure corresponding to the hypothesis H_k, k=0, 1, respectively, and $R_n(\theta)=\partial^3 lr_n(\theta)/\partial\theta^3$. By (S.1), one can write $\Pr_0(\tau_2<\infty)$ in the form

$$\Pr_0(\tau_2<\infty)=\Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{1}{3}\left(\theta_0-\hat{\theta}_n\right)^3 R_n\left(\tilde{\theta}_n\right)>c_m\log(n)\ \textit{for some}\ n\geq m\right\}, \quad (S.3)$$

where $\tilde{\theta}_n=\theta_0+\rho\left(\hat{\theta}_n-\theta_0\right)$, $\rho\in(0,1)$.

In order to evaluate (S.3), we present the following lemma.

Lemma 1. *Assume conditions (A1)-(A3) are satisfied. Then*

$$\Pr_0\left\{n\left|\hat{\theta}_n-\theta_0\right|\geq a_n\ \textit{for some}\ n\geq m\right\}=O\{1/\log\log(m)\},$$

where $a_n=\{2n\log\log(n)\}^{0.5}\sigma_G/\gamma$, $\sigma_G=\left[\mathrm{var}\{G(X_1,\theta_0)\}\right]^{0.5}$ *and* $\gamma=E\{M_0(X_1)\}$ *as* $m\to\infty$.

Proof of Lemma 1.

By virtue of $n^{-1}\sum_{i=1}^n G\left(X_i,\hat{\theta}_n\right)=0$ and Taylor's theorem, we have

$$n^{-1}\sum_{i=1}^n G(X_i,\theta_0)+n^{-1}\sum_{i=1}^n G'(X_i,\theta_{1i})\left(\hat{\theta}_n-\theta_0\right)=0,$$

where $\theta_{1i}=\hat{\theta}_n+\omega_i\left(\theta_0-\hat{\theta}_n\right)$ and $\omega_i\in(0,1)$ for $i=1,\ldots,n$.

Denote the event $B_n=\left\{n^{-1}\left|\sum_{i=1}^n\left(M_0(X_i)-\gamma\right)\right|\geq\sigma_m(2\log\log n)^{0.5}n^{-0.5}\right\}$, where $\sigma_m=\left[\mathrm{var}\{M_0(X_1)\}\right]^{0.5}$. One can use the result of Darling and Robbins (1967) to prove that $\Pr_0\{B_n\ \textit{for some}\ n\geq m\}=O\{1/\log\log(m)\}$ as $m\to\infty$. Then

$$\Pr_0\left\{n\left|\hat{\theta}_n-\theta_0\right|\geq a_n\ \textit{for some}\ n\geq m\right\}=\Pr_0\left\{\left|\sum_{i=1}^n G(X_i,\theta_0)\right|\geq a_n\left|n^{-1}\sum_{i=1}^n G'(X_i,\theta_{1i})\right|\ \textit{for some}\ n\geq m\right\}$$

$$\leq\Pr_0\left\{\left|\sum_{i=1}^n G(X_i,\theta_0)\right|\geq a_n n^{-1}\sum_{i=1}^n M_0(X_i)\ \textit{for some}\ n\geq m\right\}$$

$$= \Pr_0 \left\{ \left| \sum_{i=1}^{n} G(X_i, \theta_0) \right| \geq a_n n^{-1} \sum_{i=1}^{n} M_0(X_i) \ \textit{for some} \ n \geq m, \ B_n \ \textit{for some} \ n \geq m \right\}$$

$$+ \Pr_0 \left\{ \left| \sum_{i=1}^{n} G(X_i, \theta_0) \right| \geq a_n n^{-1} \sum_{i=1}^{n} M_0(X_i) \ \textit{for some} \ n \geq m, \ B_n^C \ \textit{for all} \ n \geq m \right\}$$

$$\leq \Pr_0 \{ B_n \ \textit{for some} \ n \geq m \} + \Pr_0 \left\{ \left| \sum_{i=1}^{n} G(X_i, \theta_0) \right| \geq a_n \left(\gamma - \frac{\sigma_m (2 \log \log n)^{0.5}}{n^{0.5}} \right) \textit{for some} \ n \geq m \right\}$$

$$= O\{1 / \log \log(m)\} + \Pr_0 \left\{ \left| \sum_{i=1}^{n} G(X_i, \theta_0) \right| \geq \sigma_G (2n \log \log n)^{0.5} \ \textit{for some} \ n \geq m \right\} \ \textit{as} \ m \to \infty .$$

Note that the result of Darling and Robbins (1967) implies that $\Pr_0 \left\{ \left| \sum_{i=1}^{n} G(X_i, \theta_0) \right| \geq \sigma_G (2n \log \log n)^{0.5} \ \textit{for some} \ n \geq m \right\} = O\{1 / \log \log(m)\}$ as $m \to \infty$. Thus, we have $\Pr_0 \left\{ n \left| \hat{\theta}_n - \theta_0 \right| \geq a_n \ \textit{for some} \ n \geq m \right\} = O\{1/\log \log(m)\}$ as $m \to \infty$.

The proof of Lemma 1 is complete.

By virtue of Lemma 1, we will obtain upper and lower bounds for $\Pr_0(\tau_2 < \infty)$. To this end, we use (A.1) to show that $\Pr_0(\tau_2 < \infty)$

$$= \Pr_0 \left\{ \frac{n}{\sigma_n^2} (\theta_0 - \hat{\theta}_n)^2 - \frac{1}{3} (\theta_0 - \hat{\theta}_n)^3 R_n(\tilde{\theta}_n) > c_m \log(n) \ \textit{for some} \ n \geq m, n \left| \hat{\theta}_n - \theta_0 \right| \geq a_n \ \textit{for some} \ n \geq m \right\}$$

$$+ \Pr_0 \left\{ \frac{n}{\sigma_n^2} (\theta_0 - \hat{\theta}_n)^2 - \frac{1}{3} (\theta_0 - \hat{\theta}_n)^3 R_n(\tilde{\theta}_n) > c_m \log(n) \ \textit{for some} \ n \geq m, n \left| \hat{\theta}_n - \theta_0 \right| < a_n \ \textit{for all} \ n \geq m \right\}$$

$$\leq \Pr_0 \left\{ n \left| \hat{\theta}_n - \theta_0 \right| \geq a_n \ \textit{for some} \ n \geq m \right\} + U_{0m},$$

where $\Pr_0 \left\{ n \left| \hat{\theta}_n - \theta_0 \right| \geq a_n \ \textit{for some} \ n \geq m \right\} = O\{1 / \log \log(m)\}$, and

$$U_{0\mathrm{m}} = \Pr_0 \left\{ \frac{n}{\sigma_n^2} (\theta_0 - \hat{\theta}_n)^2 + \frac{a_n^3}{n^3} \left| R_n(\tilde{\theta}_n) \right| > c_m \log(n) \ \textit{for some} \ n \geq m \right\}.$$

Regarding the term U_{0m}, we have the inequality

$$U_{0m} = \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2+\frac{a_n^3}{n^3}\left|R_n\left(\tilde{\theta}_n\right)\right|>c_m\log(n)\ for\ some\ n\geq m,\left|R_n\left(\tilde{\theta}_n\right)\right|<n^{1+\delta}\ for\ all\ n\geq m\right\}$$

$$+\Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2+\frac{a_n^3}{n^3}\left|R_n\left(\tilde{\theta}_n\right)\right|>c_m\log(n)\ for\ some\ n\geq m,\left|R_n\left(\tilde{\theta}_n\right)\right|\geq n^{1+\delta}\ for\ some\ n\geq m\right\} \quad \text{(S.4)}$$

$$\leq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2+\frac{a_n^3}{n^{2-\delta}}>c_m\log(n)\ for\ some\ n\geq m\right\}+\Pr_0\left\{\left|R_n\left(\tilde{\theta}_n\right)\right|<n^{1+\delta}\ for\ some\ n\geq m\right\},$$

where δ is assumed to be in $(0, 0.5)$ and the term $a_n^3 n^{-2+\delta}\to 0$ as $n\to\infty$.

Now, we consider the lower bound for $\Pr_0(\tau_2<\infty)$. Noting that, by (S.3) we have

$\Pr_0(\tau_2<\infty)$

$$\geq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{1}{3}\left(\theta_0-\hat{\theta}_n\right)^3R_n\left(\tilde{\theta}_n\right)>c_m\log(n)\ for\ some\ n\geq m, n\left|\hat{\theta}_n-\theta_0\right|<a_n\ for\ all\ n\geq m\right\}$$

$$\geq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{a_n^3}{n^3}\left|R_n\left(\tilde{\theta}_n\right)\right|>c_m\log(n)\ for\ some\ n\geq m, n\left|\hat{\theta}_n-\theta_0\right|<a_n\ for\ all\ n\geq m\right\}$$

$$\geq U_{1m} - \Pr_0\left\{n\left|\hat{\theta}_n-\theta_0\right|\geq a_n\ for\ some\ n\geq m\right\},$$

where $\Pr_0\left\{n\left|\hat{\theta}_n-\theta_0\right|\geq a_n\ for\ some\ n\geq m\right\}=O\{1/\log\log(m)\}$, and

$$U_{1m} = \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{a_n^3}{n^3}\left|R_n\left(\tilde{\theta}_n\right)\right|>c_m\log(n)\ for\ some\ n\geq m\right\}.$$

Regarding the term U_{1m}, we have the following inequality

$$U_{1m} \geq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2-\frac{a_n^3}{n^3}\left|R_n\left(\tilde{\theta}_n\right)\right|>c_m\log(n)\ for\ some\ n\geq m,\left|R_n\left(\tilde{\theta}_n\right)\right|<n^{1+\delta}\ for\ all\ n\geq m\right\}$$

$$\geq \Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}_n\right)^2+\frac{a_n^3}{n^{2-\delta}}>c_m\log(n)\ for\ some\ n\geq m\right\}-\Pr_0\left\{\left|R_n\left(\tilde{\theta}_n\right)\right|\geq n^{1+\delta}\ for\ some\ n\geq m\right\}, \quad \text{(S.5)}$$

where the term $a_n^3 n^{-2+\delta} \to 0$ as $n \to \infty$.

The following lemma analyzes the term $\Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0 - \hat{\theta}_n\right)^2 > c_m \log(n) \text{ for some } n \geq m\right\}$ provided in in (S.4) and (S.5).

Lemma 2. *Assume conditions (A1)-(A3) are satisfied. Then*

$$\lim_{m\to\infty}\left|\Pr_0\left\{n\sigma_n^{-2}\left(\theta_0 - \hat{\theta}_n\right)^2 > c_m \log(n) \text{ for some } n \geq m\right\} - 2\left[1 - \Phi\left\{(c_m \log(m))^{1/2}\right\}\right]\left\{1 + c_m^{-1} + \log(m)\right\}\right| = 0,$$

where c_m *satisfies* $2\left[1 - \Phi\left\{(c_m \log(m))^{1/2}\right\}\right]\left\{1 + c_m^{-1} + \log(m)\right\} = \alpha$ *,* α *denotes the pre-specified significance level, and* $\Phi(y) = (2\pi)^{-1/2}\int_{-\infty}^{y}\exp(-x^2/2)dx$.

Proof of Lemma 3.

Applying Taylor's theorem to $n^{-1}\sum_{i=1}^{n} G\left(X_i, \hat{\theta}_n\right) = 0$ with $\hat{\theta}_n$ around θ_0, we obtain

$$\frac{n^{1/2}}{\sigma_n}\left(\hat{\theta}_n - \theta_0\right) = -\frac{n^{-1/2}\sum_{i=1}^{n} G(X_i, \theta_0)}{n^{-1}\sum_{i=1}^{n} G'(X_i, \theta_0)\sigma_n} - \frac{n^{-1/2}\sum_{i=1}^{n} G''(X_i, \theta_{0i})\left(\hat{\theta}_n - \theta_0\right)^2}{2n^{-1}\sum_{i=1}^{n} G'(X_i, \theta_0)\sigma_n},$$

where $\theta_{0i} = \theta_0 + \rho_{0i}\left(\hat{\theta}_n - \theta_0\right)$, $\rho_{0i} \in (0,1)$. This result will be used to evaluate $n^{1/2}\left(\hat{\theta}_n - \theta_0\right)\sigma_n^{-1}$.

Then by virtue of Lemma 1, it follows that

$$\Pr_0\left[\left|\frac{n^{1/2}}{\sigma_n}\left(\hat{\theta}_n - \theta_0\right)\right| > \{c_m \log(n)\}^{1/2} \text{ for some } n \geq m\right]$$

$$\leq \Pr_0\left[\left|\frac{n^{-1/2}\sum_{i=1}^{n} G(X_i, \theta_0)}{n^{-1}\sum_{i=1}^{n} G'(X_i, \theta_0)\sigma_n}\right| + \left|\frac{n^{-1/2}\sum_{i=1}^{n} M_2(X_i)\left(\hat{\theta}_n - \theta_0\right)^2}{2n^{-1}\sum_{i=1}^{n} M_0(X_i, \theta_0)\sigma_n}\right| \geq \{c_m \log(n)\}^{1/2} \text{ for some } n \geq m, \right.$$

$$\left. n\left|\hat{\theta}_n - \theta_0\right| \geq a_n \text{ for some } n \geq m\right]$$

$$+ \Pr_0\left[\left|\frac{n^{-1/2}\sum_{i=1}^{n} G(X_i, \theta_0)}{n^{-1}\sum_{i=1}^{n} G'(X_i, \theta_0)\sigma_n}\right| + \left|\frac{n^{-1/2}\sum_{i=1}^{n} M_2(X_i)\left(\hat{\theta}_n - \theta_0\right)^2}{2n^{-1}\sum_{i=1}^{n} M_0(X_i)\sigma_n}\right| \geq \{c_m \log(n)\}^{1/2} \text{ for some } n \geq m, \right.$$

$$n\left|\hat{\theta}_n - \theta_0\right| < a_n \ for \ all \ n \geq m \Big]$$

$$\leq \Pr_0\left[\left|\frac{n^{-1/2}\sum_{i=1}^n G(X_i,\theta_0)}{n^{-1}\sum_{i=1}^n G'(X_i,\theta_0)\sigma_n}\right| + \left|\frac{n^{-1/2}\sum_{i=1}^n M_2(X_i)n^{-2}a_n^2}{2n^{-1}\sum_{i=1}^n M_0(X_i)\sigma_n}\right| \geq \{c_m \log(n)\}^{1/2} \ \ for\ some\ n \geq m\right]$$

$$+ O\{1/\log\log(m)\} \text{ as } m \to \infty .$$

According to Darling and Robbins (1967), we have $\Pr_0\{B_{n,k}\ for\ some\ n \geq m\} = O\{1/\log\log(m)\}$ as $m \to \infty$, where the event $B_{n,k} = \left\{n^{-1}\left|\sum_{i=1}^n (M_k(X_i) - \gamma_k)\right| \geq \sigma_{m,k}(2\log\log n)^{0.5} n^{-0.5}\right\}$, $\gamma_k = E\{M_k(X_1)\}$ and $\sigma_{m,k} = [\mathrm{var}\{M_k(X_1)\}]^{0.5}$, k=0, 2. Then it is clear that as $m \to \infty$

$$\Pr_0\left\{\left|\frac{n^{1/2}}{\sigma_n}(\theta_0 - \hat{\theta})\right| \geq \{c_m \log(n)\}^{1/2} \ \ for\ some\ n \geq m\right\}$$

$$\leq \Pr_0\left\{\left|\frac{n^{-1/2}\sum_{i=1}^n G(X_i,\theta_0)}{n^{-1}\sum_{i=1}^n G'(X_i,\theta_0)\sigma_n}\right| + C_1 \frac{\log\log(n)}{n^{1/2}} \geq \{c_m \log(n)\}^{1/2} \ for\ some\ n \ \geq m\right\} + O\left[\{\log\log(m)\}^{-1}\right],$$

where C_1 is a positive constant.

In a similar manner one can obtain the lower bound that as $m \to \infty$

$$\Pr_0\left\{\left|\frac{n^{1/2}}{\sigma_n}(\theta_0 - \hat{\theta})\right| \geq \{c_m \log(n)\}^{1/2} \ \ for\ some\ n \geq m\right\}$$

$$\geq \Pr_0\left\{\left|\frac{n^{-1/2}\sum_{i=1}^n G(X_i,\theta_0)}{n^{-1}\sum_{i=1}^n G'(X_i,\theta_0)\sigma_n}\right| - C_2 \frac{\log\log(n)}{n^{1/2}} \geq \{c_m \log(n)\}^{1/2} \ for\ some\ n \geq m\right\} - O\left[\{\log\log(m)\}^{-1}\right],$$

where C_2 is a positive constant.

Thus, the above analyses imply that

$$\lim_{m\to\infty}\left|\Pr_0\left\{\frac{n}{\sigma_n^2}\left(\theta_0-\hat{\theta}\right)^2 \geq c\log(n)\ \textit{for some}\ n\geq m\right\}-\Pr_0\left\{\left|\frac{n^{-1/2}\sum_{i=1}^{n}G(X_i,\theta_0)}{n^{-1}\sum_{i=1}^{n}G'(X_i,\theta_0)\sigma_n}\right|\geq c_m\log(n)\ \textit{for some}\ n\geq m\right\}\right|=0$$

Since $G(X_i,\theta_0)$, *i=1,...,n,* are i.i.d. random variables with $E\{G(X_1,\theta_0)\}=0$, we directly use the result presented in Mukhopadhyay and De Silva (2008, pp. 81-82) and Sen (1981, p. 238) to conclude that

$$\lim_{m\to\infty}\left|\Pr_0\left\{\left|\frac{n^{-1/2}\sum_{i=1}^{n}G(X_i,\theta_0)}{n^{-1}\sum_{i=1}^{n}G'(X_i,\theta_0)\sigma_n}\right|> c_m\log(n)\ \textit{for some}\ n\geq m\right\}-2\left[1-\Phi\left\{(c_m\log(m))^{1/2}\right\}\right]\left\{1+c_m^{-1}+\log(m)\right\}\right|=0.$$

This completes the proof of Lemma 2.

Lemma 2 evaluates the Type I error probability formula related to the proposed stopping rule τ_2. According to equations (S.3)-(S.4) and Lemma 2, next we complete the proof of Proposition 1 by proving that $\Pr_0\left\{\left|R_n\left(\tilde{\theta}_n\right)\right|>n^{1+\delta}\ \textit{for some}\ n\geq m\right\}\to 0$ as $m\to\infty$. Since the remainder term $R_n\left(\tilde{\theta}_n\right)$ is a function of the Lagrangian multiplier $\lambda(\theta)$ defined in (2.1), we begin with deriving theoretical results regarding $\lambda(\theta)$. These results might have an independent interest in evaluation of relevant EL problems.

Lemma 3. *We have that* $\lambda\geq 0$ if and only if $\sum_{i=1}^{n}G(X_i,\theta)\geq 0$, and $\lambda<0$ if and only if $\sum_{i=1}^{n}G(X_i,\theta)<0$.

Proof of Lemma 3.

The forms, $p_i=\{n+\lambda G(X_i,\theta)\}^{-1}$, $i=1,..,n$, and $n^{-n}\geq\prod_{i=1}^{n}p_i$ imply that

$$0\leq -lr(\theta)=\log\left(n^{-n}/\prod_{i=1}^{n}p_i\right)=\sum_{i=1}^{n}\log\{1+\lambda G(X_i,\theta)/n\}.$$

Using the inequality $\log(1+s)\leq s$ for $s>-1$, we obtain

$$-lr(\theta) = \sum_{i=1}^{n} \log\{1 + \lambda G(X_i, \theta)/n\}$$

$$\leq \sum_{i=1}^{n} \lambda G(X_i, \theta)/n = \lambda \sum_{i=1}^{n} G(X_i, \theta)/n .$$

This completes the proof of Lemma 3.

Lemma 4. *The Lagrange multiplier* λ *satisfies*

$$\lambda = \frac{\sum_{i=1}^{n} G(X_i, \theta)}{\sum_{i=1}^{n} \{G(X_i, \theta)\}^2 p_i} .$$

Proof of Lemma 4,

The constraint $\sum_{i=1}^{n} G(X_i, \theta) p_i = 0$ with $p_i = \{n + \lambda G(X_i, \theta)\}^{-1}$, $i = 1,..,n$, in (4) implies that

$$\sum_{i=1}^{n} G(X_i, \theta) = \sum_{i=1}^{n} \{G(X_i, \theta)(1 - p_i)\} + \sum_{i=1}^{n} \{G(X_i, \theta) p_i\}$$

$$= \sum_{i=1}^{n} G(X_i, \theta)(1 - p_i) = \sum_{i=1}^{n} G(X_i, \theta) \left\{ \frac{n + \lambda G(X_i, \theta) - 1}{n + \lambda G(X_i, \theta)} \right\}$$

$$= \lambda \sum_{i=1}^{n} \{G(X_i, \theta)\}^2 p_i .$$

This completes the proof of Lemma 4.

Lemma 5. *If* $\lambda \geq 0$, *we have* $0 \leq \lambda \leq n \sum_{i=1}^{n} G(X_i, \theta) \left[\sum_{i=1}^{n} \{G(X_i, \theta)\}^2 I\{G(X_i, \theta) < 0\} \right]^{-1}$ *; If* $\lambda < 0$, *we have* $n \sum_{i=1}^{n} G(X_i, \theta) \left[\sum_{i=1}^{n} \{G(X_i, \theta)\}^2 I\{G(X_i, \theta) > 0\} \right]^{-1} \leq \lambda < 0$.

Proof of Lemma 5,

Having $\lambda \geq 0$, we obtain

$$\sum_{i=1}^{n} \{G(X_i, \theta)\}^2 p_i \geq \sum_{i=1}^{n} \left[G(X_i, \theta)^2 p_i I\{G(X_i, \theta) < 0\} \right]$$

$$= \sum_{i=1}^{n}\left[\{G(X_i,\theta)\}^2 \frac{I\{G(X_i,\theta)<0\}}{n+\lambda G(X_i,\theta)} \right] \geq \sum_{i=1}^{n}\left[\{G(X_i,\theta)\}^2 \frac{1}{n} I\{G(X_i,\theta)<0\} \right],$$

where $p_i = \{n+\lambda G(X_i,\theta)\}^{-1}$, $i=1,...,n$.

Applying the results of Lemmas 3 and 4 to the above inequality yields

$$0 \leq \lambda < n\sum_{i=1}^{n} G(X_i,\theta)\left[\sum_{i=1}^{n} \{G(X_i,\theta)\}^2 I(G(X_i,\theta)<0) \right]^{-1}.$$

It follows similarly that when $\lambda < 0$, we have

$$n\sum_{i=1}^{n} G(X_i,\theta)\left[\sum_{i=1}^{n} \{G(X_i,\theta)\}^2 I\{G(X_i,\theta)>0\} \right]^{-1} \leq \lambda < 0.$$

This completes the proof of Lemma 5.

Lemmas 3-5 are applied to obtain the next result regarding the remainder term $R_n(\tilde{\theta}_n)$ presented in (S.3).

Lemma 6. *Assume conditions (A1)-(A3) are satisfied. Then*

$$\Pr_0\left\{ \left| R_n(\tilde{\theta}_n) \right| > n^{1+\delta} \text{ for some } n \geq m \right\} = O\left\{ (\log\log m)^{-1} \right\} \text{ as } m \to \infty,$$

where $\delta \in (0, 0.5)$.

Proof of Lemma 6,

Note that

$$\Pr_0\left\{ \left| R_n(\tilde{\theta}_n) \right| > n^{1+\delta} \text{ for some } n \geq m \right\} = \Pr_0\left\{ \bigcup_{n\geq m} \left[\left| R_n(\tilde{\theta}_n) \right| > n^{1+\delta} \right] \right\}$$

$$= \Pr_0\left\{ \bigcup_{n\geq m} \left[\left\{ \left| R_n(\tilde{\theta}_n) \right| > n^{1+\delta} \right\} \bigcap \left\{ \left(\lambda(\tilde{\theta}_n) \geq 0 \right) \bigcup \left(\lambda(\tilde{\theta}_n) < 0 \right) \right\} \right] \right\}$$

$$\leq \Pr_0\{Z_1\} + \Pr_0\{Z_2\}, \tag{S.6}$$

where $Z_1 = \bigcup_{n\geq m} \left[\left\{ \left| R_n(\tilde{\theta}_n) \right| > n^{1+\delta} \right\} \bigcap \left\{ \lambda(\tilde{\theta}_n) \geq 0 \right\} \right]$ and $Z_2 = \bigcup_{n\geq m} \left[\left\{ \left| R_n(\tilde{\theta}_n) \right| > n^{1+\delta} \right\} \bigcap \left\{ \lambda(\tilde{\theta}_n) < 0 \right\} \right]$.

Next we will show that $Pr_0\{Z_k\} = O\left\{(log\,log\,m)^{-1}\right\}, k = 1,2,$ as $m \to \infty$.

Define event $V_n = \left\{\left|\hat{\theta}_n - \theta_0\right| \le n^{-1} a_n\right\}$, where $a_n = \{2n \log\log(n)\}^{0.5} \sigma_G / \gamma$ is from Lemma 1. Then, we apply Lemma 1 to obtain the following inequality

$$\Pr_0\{Z_1\} = \Pr_0\{Z_1, V_n \textit{ for all } n \ge m\} + \Pr_0\{Z_1, V_n^c \textit{ for some } n \ge m\}$$

$$\le \Pr_0\{Z_1, V_n \textit{ for all } n \ge m\} + \Pr_0\{V_n^c \textit{ for some } n \ge m\}$$

$$\le \Pr_0\{Z_1, V_n \textit{ for all } n \ge m\} + O\left\{(\log\log m)^{-1}\right\} \text{ as } m \to \infty, \tag{S.7}$$

where $V_n^c = \left\{\left|\hat{\theta}_n - \theta_0\right| > n^{-1} a_n\right\}$.

Rewrite

$$R_n(\tilde{\theta}_n) = Q(\tilde{\theta}_n) + J(\tilde{\theta}_n),$$

where $Q(\tilde{\theta}_n) = 2\{\lambda'(\tilde{\theta}_n)\}^2 \sum_{i=1}^{n} G(X_i, \tilde{\theta}_n) G'(X_i, \tilde{\theta}_n) p_i^2(\tilde{\theta}_n)$ and $J(\tilde{\theta}_n)$ is defined in (S.1). Denote $\delta_1 \in (0, \delta)$ and event $W_1 = \left\{\left|Q(\tilde{\theta}_n)\right| < n^{1+\delta_1}\right\}$. Then

$$\Pr_0\{Z_1, V_n \textit{ for all } n \ge m\}$$

$$\le \Pr_0\left\{\left|Q(\tilde{\theta}_n)\right| + \left|J(\tilde{\theta}_n)\right| > n^{1+\delta} \textit{ and } \lambda(\tilde{\theta}_n) \ge 0 \textit{ for some } n \ge m, V_n \textit{ for all } n \ge m, W_1 \textit{ for all } n \ge m\right\}$$

$$+ \Pr_0\left\{\left|Q(\tilde{\theta}_n)\right| + \left|J(\tilde{\theta}_n)\right| > n^{1+\delta} \textit{ and } \lambda(\tilde{\theta}_n) \ge 0 \textit{ for some } n \ge m, V_n \textit{ for all } n \ge m, W_1^c \textit{ for some } n \ge m\right\}$$

$$\le \Pr_0\left\{\left|Q(\tilde{\theta}_n)\right| > n^{1+\delta_1} \textit{ and } \lambda(\tilde{\theta}_n) \ge 0 \textit{ for some } n \ge m, V_n \textit{ for all } n \ge m\right\}$$

$$+ \Pr_0\left\{\left|J(\tilde{\theta}_n)\right| > n^{1+\delta} - n^{1-\delta_1} \textit{ and } \lambda(\tilde{\theta}_n) \ge 0 \textit{ for some } n \ge m, V_n \textit{ for all } n \ge m\right\}.$$

To show that

$$\Pr_0\left\{\left|Q(\tilde{\theta}_n)\right| > n^{1+\delta_1} \textit{ and } \lambda(\tilde{\theta}_n) \ge 0 \textit{ for some } n \ge m, V_n \textit{ for all } n \ge m\right\} \to 0 \text{ as } m \to \infty,$$

it suffices to prove that when $\left|\hat{\theta}_n - \theta_0\right| \le n^{-1} a_n$, we have $p_i\left(\widetilde{\theta}_n\right) = O\left(n^{-1}\right)$ for all *i=1,…,n,* and $\lambda\left(\widetilde{\theta}_n\right) = O\left[\{2n \log\log(n)\}^{1/2}\right]$ for all $n \ge m$ as $m \to \infty$.

Since by assumption (A1) $G'(X_i, \theta) < 0$ for all *i=1,…,n,* we obtain the following results, for $\left|\hat{\theta}_n - \theta_0\right| \le n^{-1} a_n$,

$$\left|G\left(X_i, \widetilde{\theta}_n\right)\right| \le \max\left\{\left|G\left(X_i, \theta_0 - n^{-1} a_n\right)\right|, \left|G\left(X_i, \theta_0 + n^{-1} a_n\right)\right|\right\},$$

$$I\left\{G\left(X_i, \widetilde{\theta}_n\right) < 0\right\} \ge I\left\{G\left(X_i, \theta_0 - n^{-1} a_n\right) < 0\right\}, \text{ and}$$

$$\sum_{i=1}^n G\left(X_i, \widetilde{\theta}_n\right) \le \sum_{i=1}^n G\left(X_i, \theta_0 - n^{-1} a_n\right), \tag{S.8}$$

where $\widetilde{\theta}_n = \theta_0 + \rho\left(\hat{\theta}_n - \theta_0\right)$, $\rho \in (0,1)$.

These results and Lemma 4 imply that

$$0 \le \lambda\left(\widetilde{\theta}_n\right) \le n \sum_{i=1}^n G\left(X_i, \widetilde{\theta}_n\right) \left[\sum_{i=1}^n \left\{G\left(X_i, \widetilde{\theta}_n\right)\right\}^2 I\left\{G\left(X_i, \widetilde{\theta}_n\right) < 0\right\}\right]^{-1}$$

$$\le n \sum_{i=1}^n G\left(X_i, \theta_0 - n^{-1} a_n\right) \left[\sum_{i=1}^n \left\{G\left(X_i, \theta_0 + n^{-1} a_n\right)\right\}^2 I\left\{G\left(X_i, \widetilde{\theta}_n\right) < 0\right\}\right]^{-1}$$

$$\le n \sum_{i=1}^n G\left(X_i, \theta_0 - n^{-1} a_n\right) \left[\sum_{i=1}^n \left\{G\left(X_i, \theta_0 + n^{-1} a_n\right)\right\}^2 I\left\{G\left(X_i, \theta_0 - n^{-1} a_n\right) < 0\right\}\right]^{-1}.$$

In a similar manner to the analysis shown above, we have

$$0 \ge \lambda\left(\widetilde{\theta}_n\right) \ge n \sum_{i=1}^n G\left(X_i, \theta_0 + n^{-1} a_n\right) \left[\sum_{i=1}^n \left\{G\left(X_i, \theta_0 - n^{-1} a_n\right)\right\}^2 I\left\{G\left(X_i, \theta_0 + n^{-1} a_n\right) > 0\right\}\right]^{-1}.$$

Then, it is clear that

$$\lambda\left(\widetilde{\theta}_n\right) = O\left[\{2n \log\log(n)\}^{1/2}\right], \text{ for all } n \ge m \text{ as } m \to \infty. \tag{S.9}$$

Thus, by virtue of (S.8) and (S.9), we conclude that, for $\left|\hat{\theta}_n - \theta_0\right| \le n^{-1} a_n$,

$p_i\left(\tilde{\theta}_n\right) = O\left(n^{-1}\right)$ for all $i=1,\ldots,n$ and $n \ge m$, where the definition $p_i = \left\{n + \lambda G\left(X_i, \theta\right)\right\}^{-1}$ is taken into account.

This leads to the conclusion

$$\left|Q\left(\tilde{\theta}_n\right)\right| = \left|2\left\{\lambda'\left(\tilde{\theta}_n\right)\right\}^2 \sum_{i=1}^{n} G\left(X_i, \tilde{\theta}_n\right) G'\left(X_i, \tilde{\theta}_n\right) p_i^2\left(\tilde{\theta}_n\right)\right| = O(\log\log n) \text{ for } n \ge m \text{ as } m \to \infty .$$

Since $(\log\log n)/n^{1+\delta_1} \to 0$ as $m \to \infty$, it is clear that

$$\Pr_0\left\{\left|Q\left(\tilde{\theta}_n\right)\right| > n^{1+\delta_1} \text{ and } \lambda\left(\tilde{\theta}_n\right) \ge 0 \text{ for some } n \ge m, V_n \text{ for all } n \ge m\right\} = 0 \text{ as } m \to \infty .$$

Similarly to the analysis of $Q\left(\tilde{\theta}_n\right)$, one can prove that

$$\Pr_0\left\{\left|J\left(\tilde{\theta}_n\right)\right| > n^{1+\delta} - n^{1-\delta_1} \text{ and } \lambda\left(\tilde{\theta}_n\right) \ge 0 \text{ for some } n \ge m, V_n \text{ for all } n \ge m\right\} = 0 \text{ as } m \to \infty .$$

The above results lead to

$$\Pr_0\left\{Z_1,\ V_n \text{ for all } n \ge m\right\} = 0 \text{ as } m \to \infty . \qquad \text{(S.10)}$$

By virtue of (S.7) and (S.10), we conclude that

$$\Pr_0\left\{Z_1\right\} = O\left\{(\log\log m)^{-1}\right\} \text{ as } m \to \infty .$$

In a similar manner to the considerations shown above, one can show that

$$\Pr_0\left\{Z_2\right\} = O\left\{(\log\log m)^{-1}\right\} \text{ as } m \to \infty .$$

The proof of Lemma 6 is complete.

By virtue of Lemmas 2 and 6, inequality (S.5) has the form

$$\Pr_0\left\{N_1 < \infty\right\} - O\left\{(\log\log m)^{-1}\right\} \le \Pr_0\left\{\tau_2 < \infty\right\} \le \Pr_0\left\{N_1 < \infty\right\} + O\left\{(\log\log m)^{-1}\right\} \text{ as } m \to \infty .$$

This result completes the proof of Proposition 1 regarding the stopping rule τ_2. Proofs related to the one-sided stopping rules τ_0 and τ_1 defined in (3.1) can be directly derived following the algorithm shown above, and then we omitted these proofs.

Remark: In this paper we provide Lemma 6, employing Lemmas 1-5. This result can be confirmed by adapting Lemma 7 presented in Zhong and Ghosh (2016) that concludes with $\left|\partial^3 lr_n(\theta)/\partial\theta^3\right| \le C_1 n$ almost surely, when θ is in a neighborhood of $\hat{\theta}_n$, for relatively large *n*, where C_1 denotes a positive constant.

Proof of Proposition 2.

Assume that we consider the hypothesis $H_0 : \theta = \theta_0$ *vs.* $H_1 : \theta \ne \theta_0$, where θ_0 is known. Let $\theta_1 \ne \theta_0$ satisfy $E\{G(X_1, \theta_1)\} = 0$, under H_1. To prove the property $\Pr_1\{-2lr_n(\theta_0) \le c_m \log(n) \textit{ for all } n \ge m\} = 0$, we show that, for large values of *n*, $-2lr_n(\theta_0)/n^{\varepsilon} > C$, whereas $c_m \log(n)/n^{\varepsilon} \to 0$ as $n \to \infty$, where $\varepsilon > 0$ and C is a positive constant. Consider the following proof scheme: 1) Taking into account Lemma 1 of Qin and Lawless (1994) and Lemma A1 of Vexler et al. (2014b), we will obtain that: *i*) in the case of $\theta_0 > \theta_1$, for relatively large values of *n*, we have $\theta_0 > \theta_1 + n^{-1/3}$, $\left|\hat{\theta}_n - \theta_1\right| \le n^{-1/3}$ a.s., and then $\theta_0 > \theta_1 + n^{-1/3} > \hat{\theta}_n$ a.s., which implies $-2lr_n(\theta_0) \ge -2lr_n(\theta_1 + n^{-1/3})$, as $n \to \infty$; *ii*) in the case of $\theta_0 < \theta_1$, for relatively large values of *n*, we have $\theta_0 < \theta_1 - n^{-1/3}$, $\left|\hat{\theta}_n - \theta_1\right| \le n^{-1/3}$ a.s., and then $\theta_0 < \theta_1 - n^{-1/3} < \hat{\theta}_n$ a.s., which implies $-2lr_n(\theta_0) \ge -2lr_n(\theta_1 - n^{-1/3})$, where *n* is such that $\theta_0 \notin (\theta_1 - n^{-1/3}, \theta_1 + n^{-1/3})$; 2) We will demonstrate that, when *n* is relatively large, we have $-2lr_n(\theta_1 \pm n^{-1/3}) = O(n^{1/3}) \to +\infty$, as $n \to \infty$, under H_1; 3) By (*i*) and (*ii*), it follows that there exists N such that, for all $n \ge N$, the event $\{-2lr_n(\theta_0) \le c_m \log(n)\}$ is empty. Thus, we will conclude that $\Pr_1\{-2lr_n(\theta_0) \le c_m \log(n) \textit{ for all } n \ge m\} = 0$.

Attending to step (2) above, we note that, by (2.1), we have $-2lr_n(\theta_1 + n^{-1/3}) = 2\sum_{i=1}^{n}\log(1+\xi_i)$, where $\xi_i = n^{-1}\lambda(\theta_1 + n^{-1/3})G(X_i, \theta_1 + n^{-1/3})$. In order to evaluate $-2lr_n(\theta_1 + n^{-1/3})$, we derive an upper bound and an approximate order for $\lambda(\theta_1 + n^{-1/3})$. To this end, we apply Lemma 5 to obtain that

$$\left|\lambda(\theta_1 + n^{-1/3})\right| \le \max\left[n\left|\sum_{i=1}^{n} G(X_i, \theta_1 + n^{-1/3})\right|(W_{1n})^{-1},\ n\left|\sum_{i=1}^{n} G(X_i, \theta_1 + n^{-1/3})\right|(W_{2n})^{-1}\right],$$

where $W_{1n} = \sum_{i=1}^{n}\{G(X_i, \theta_1 + n^{-1/3})\}^2 I\{G(X_i, \theta_1 + n^{-1/3}) < 0\}$,

and $W_{2n} = \sum_{i=1}^{n}\{G(X_i, \theta_1 + n^{-1/3})\}^2 I\{G(X_i, \theta_1 + n^{-1/3}) > 0\}$.

By virtue of condition (A2), one can use the Taylor theorem to show that under H_1, we have

$$n^{-1}\left|\sum_{i=1}^{n} G(X_i, \theta_1 + n^{-1/3})\right| = O(n^{-1/3}),\ n^{-1}W_{1n} = O(1) \text{ and } n^{-1}W_{2n} = O(1).$$

Taking into account the results above, we obtain that $\lambda(\theta_1 + n^{-1/3}) = O(n^{2/3})$ as $n \to \infty$.

Using the Borel-Cantelli lemma, for $|\theta - \theta_1| \le n^{-1/3}$, we have $\max_{1\le i\le n}\left|G(X_i, \theta_1 + n^{-1/3})\right| = o(n^{1/3})$, since $E\left|G(X_1, \theta)\right|^3 < \infty$. Thus $\max_{1\le i\le n}|\xi_i| = o(1)$, where $\xi_i = n^{-1}\lambda(\theta_1 + n^{-1/3})G(X_i, \theta_1 + n^{-1/3})$.

By definition (2.1) of $-2lr_n(\theta_1 + n^{-1/3})$, we have $n^{-1}\sum_{i=1}^{n} G(X_i, \theta_1 + n^{-1/3})\{1+\xi_i\}^{-1} = 0$. Expanding $n^{-1}\sum_{i=1}^{n} G(X_i, \theta_1 + n^{-1/3})\{1+\xi_i\}^{-1} = 0$ with respect to ξ_i around 0, we obtain

$$n^{-1}\sum_{i=1}^{n} G(X_i, \theta_1 + n^{-1/3})\left\{1 - \xi_i + (1+\omega_{1i}^*)^{-2}\xi_i^2\right\} = 0, \tag{S.11}$$

where $\omega_{1i}^* = \rho_{1i}\xi_i$ with $\rho_{1i} \in (0,1)$. Note that $\lambda(\theta_1 + n^{-1/3}) = O_p(n^{2/3})$ and $\omega_{1i}^* = o(1)$, for all i=1,…,n, since $\max_{1\le i\le n}|\xi_i| = o(1)$. We rewrite (S.11) in the form

$$n^{-1}\sum_{i=1}^{n} G\left(X_i,\theta_1+n^{-1/3}\right) - \lambda n^{-2}\sum_{i=1}^{n}\left\{G\left(X_i,\theta_1+n^{-1/3}\right)\right\}^2 + O\left(n^{-2/3}\right)n^{-1}\sum_{i=1}^{n}\{1+o(1)\}^{-2}\left\{G\left(X_i,\theta_1+n^{-1/3}\right)\right\}^3$$
$$=0.$$

Then

$$n^{-1}\sum_{i=1}^{n} G\left(X_i,\theta_1+n^{-1/3}\right) - \lambda n^{-2}\sum_{i=1}^{n}\left\{G\left(X_i,\theta_1+n^{-1/3}\right)\right\}^2 + O\left(n^{-2/3}\right)=0. \qquad \text{(S.12)}$$

This leads to

$$\lambda\left(\theta_1+n^{-1/3}\right)=\frac{\sum_{i=1}^{n} G\left(X_i,\theta_1+n^{-1/3}\right)}{n^{-1}\sum_{i=1}^{n}\left\{G\left(X_i,\theta_1+n^{-1/3}\right)\right\}^2}+O\left(n^{1/3}\right), \qquad \text{as } n\to\infty. \qquad \text{(S.13)}$$

Applying the Taylor series expansion to $-2lr_n\left(\theta_1+n^{-1/3}\right)$ with respect to ξ_i around 0, we have

$$-2lr_n\left(\theta_1+n^{-1/3}\right)=2\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\xi_i^2+\frac{2}{3}\sum_{i=1}^{n}\left(1+\omega_{2i}^*\right)^{-3}\xi_i^3,$$

where $\omega_{2i}^*=\rho_{2i}\xi_i$ with $\rho_{2i}\in(0,1)$. Note that, $\lambda\left(\theta_1+n^{-1/3}\right)=O_p\left(n^{2/3}\right)$ and $\omega_{2i}^*=o(1)$, for all i=1,…,n, since $\max_{1\le i\le n}\left|\xi_i\right|=o(1)$. Then, by virtue of (S.13), we have

$$-2lr_n\left(\theta_1+n^{-1/3}\right)=2\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\xi_i^2+n^{-2}\left\{\lambda\left(\theta_1+n^{-1/3}\right)\right\}^3\left[n^{-1}\sum_{i=1}^{n}\{1+o(1)\}^{-3}\left\{G\left(X_i,\theta_1+n^{-1/3}\right)\right\}^3\right]$$
$$=2\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\xi_i^2+n^{-2}O\left(n^2\right)O(1)$$
$$=\frac{n\left\{n^{-1}\sum_{i=1}^{n} G\left(X_i,\theta_1+n^{-1/3}\right)\right\}^2}{n^{-1}\sum_{i=1}^{n}\left\{G\left(X_i,\theta_1+n^{-1/3}\right)\right\}^2}+o\left(n^{1/3}\right).$$

Taking into account assumption (A2) and the first order Taylor expansion applied to $G\left(X_i,\theta_1+n^{-1/3}\right)$, one can easily show that $n^{-1}\sum_{i=1}^{n} G\left(X_i,\theta_1+n^{-1/3}\right)=O\left(n^{-1/3}\right)$ and $n^{-1}\sum_{i=1}^{n}\left\{G\left(X_i,\theta_1+n^{-1/3}\right)\right\}^2=O(1)$, under H_1: $E\{G(X_1,\theta_1)\}=0$. Thus, the main term

$n\left\{n^{-1}\sum_{i=1}^{n} G\left(X_i, \theta_1 + n^{-1/3}\right)\right\}^2 \left[n^{-1}\sum_{i=1}^{n}\left\{G\left(X_i, \theta_1 + n^{-1/3}\right)\right\}^2\right]^{-1}$ of $-2lr_n\left(\theta_1 + n^{-1/3}\right)$ converges to $+\infty$ in an order that is more than $n^{1/3}$, as $n \to \infty$.

In a similar manner to the analysis related to $-2lr_n\left(\theta_1 + n^{-1/3}\right)$, one can show that $-2lr_n\left(\theta_1 - n^{-1/3}\right) = O\left(n^{1/3}\right) \to +\infty$, as $n \to \infty$.

In the context of step (1) of the proposed proof scheme, we use the following lemmas provided in Vexler et al. (2014b) and Qin and Lawless (1994), respectively.

Lemma 7. *Assume condition (A1) is satisfied. Define* $\hat{\theta}_n$ *to be a root of the equation* $n^{-1}\sum_{i=1}^{n} G\left(X_i, \hat{\theta}_n\right) = 0$ *. Then the argument* $\hat{\theta}_n$ *is a global minimum of the function* $-2lr_n(\theta) = 2\sum_{i=1}^{n}\log\left\{1 + n^{-1}\lambda(\theta)G\left(X_i, \theta\right)\right\}$ defined in (2.1), that decreases and increases monotonically for $\theta < \hat{\theta}_n$ and $\theta > \hat{\theta}_n$, respectively.

Lemma 8. *Assume conditions (A1)-(A2) are satisfied. Then, with probability 1,* $\hat{\theta}_n \in \left(\theta_1 - n^{-1/3}, \theta_1 + n^{-1/3}\right)$*, as* $n \to \infty$.

The results (S.14)-(S.15) and Lemmas 7-8 imply that in the case of $\theta_0 > \theta_1$, for relatively large values of *n*, we have $\theta_0 > \theta_1 + n^{-1/3}$, $\left|\hat{\theta}_n - \theta_1\right| \le n^{-1/3}$ a.s., and then $\theta_0 > \theta_1 + n^{-1/3} > \hat{\theta}_n$ a.s., which leads to $-2lr_n(\theta_0) \ge -2lr_n\left(\theta_1 + n^{-1/3}\right) = O\left(n^{1/3}\right) \to +\infty$, as $n \to \infty$; in the case of $\theta_0 < \theta_1$, for relatively large values of *n*, we have $\theta_0 < \theta_1 - n^{-1/3}$, $\left|\hat{\theta}_n - \theta_1\right| \le n^{-1/3}$ a.s., and then $\theta_0 < \theta_1 - n^{-1/3} < \hat{\theta}_n$ a.s., which leads to $-2lr_n(\theta_0) \ge -2lr_n\left(\theta_1 - n^{-1/3}\right) = O\left(n^{1/3}\right) \to +\infty$, where *n* is such that $\theta_0 \notin \left(\theta_1 - n^{-1/3}, \theta_1 + n^{-1/3}\right)$. Then we conclude that $\Pr_1\left\{-2lr_n(\theta_0) \le c_m \log(n)\ \textit{for all}\ n \ge m\right\} = 0$. Thus, the proposed two-sided ELR sequential procedure τ_2 defined at (3.1) has power one. Proofs related to the stopping rules τ_0 and τ_1 defined at (3.1) can be performed in a similar manner to that shown above.

REFERENCES

Darling, D. and Robbins, H., 1967. Iterated logarithm inequalities. Proceedings of the National Academy of Sciences 57, 1188-1192.

Mukhopadhyay, N. and De Silva, B.M., 2008. Sequential methods and their applications. CRC press.

Qin, J. and Lawless, J., 1994. Empirical likelihood and general estimating equations. The Annals of Statistics, 300-325.

Sen, P.K., 1981. Sequential nonparametrics: invariance principles and statistical inference. Wiley New York.

Vexler, A., Hutson, A.D., and Yu, J., 2014a. Empirical Likelihood Methods in Clinical Experiments. Methods and Applications of Statistics in Clinical Trials: Planning, Analysis, and Inferential Methods, Volume 2, 146-165.

Vexler, A., Tao, G., and Hutson, A.D., 2014b. Posterior expectation based on empirical likelihoods. Biometrika 101, 711-718.

Zhong, X. and Ghosh, M., 2016. Higher-order properties of Bayesian empirical likelihood. Electronic Journal of Statistics 10, 3011-3044.