Hypotheses Testing from complex survey data using bootstrap weights: a unified approach

Jae-Kwang Kim
Iowa State University, Ames, Iowa, USA - jkim@iastate.edu

J.N.K. Rao
Carleton University, Ottawa, Canada - jrao@math.carleton.ca

Abstract

Standard statistical methods that do not take proper account of the complexity of survey design can lead to erroneous inferences when applied to survey data. In particular, the actual type I error rates of tests of hypotheses based on standard tests can be much bigger than the nominal level. Methods that take account of survey design features in testing hypotheses have been proposed, including Wald tests and quasi-score tests (Rao, Scott and Skinner 1998) that involve the estimated covariance matrices of parameter estimates. The bootstrap method of Rao and Wu (1988) is often applied at Statistics Canada to estimate the covariance matrices, using the data file containing columns of bootstrap weights. Standard statistical packages often permit the use of survey weighted test statistics and it is attractive to approximate their distributions under the null hypothesis by their bootstrap analogues computed from the bootstrap weights supplied in the data file. Beaumont and Bocci (2009) applied this bootstrap method to testing hypotheses on regression parameters under a linear regression model, using weighted F statistics. In this paper, we present a unified approach to the above method by constructing bootstrap approximations to weighted likelihood ratio statistics and weighted quasi-score statistics. We report the results of a simulation study on testing hypothesis under the generalized linear model setup.

Keywords: Likelihood ratio test; Pseudo maximum likelihood estimation; Resampling; Score test.

1. Introduction

Testing statistical hypothesis is one of the fundamental problem of statistics. In the parametric model approach, testing statistical hypothesis can be implemented using Wald test, likelihood ratio test, or score test. In each test, a test statistic is computed and then is compared with the $100\alpha\%$-quantile of the reference distribution which is the limiting distribution of the test statistic under the null hypothesis. The limiting distribution is often a chi-squared distribution due to the central limit theorem of the point estimators.

In survey sampling, however, the limiting distribution of the test statistic is not generally a chi-squared distribution. Rather, it can be expressed as a weighted sum of p independent random variables from $\chi^2(1)$ distribution and the weights depend on unknown parameters. To handle such problem, one may consider some correction of the test statistics to obtain a chi-square limiting distribution. Such approach usually involves computing the design effect (Rao and Scott, 1984) to the test statistics. Rao, Scott and Skinner (1998) uses this approach to obtain quasi-score test in survey data.

In this paper, we use a different approach of computing the limiting distribution using parametric bootstrap. Use of bootstrap to compute the limiting distribution of the test statistics under complex sampling has been discussed by Beaumont and Bocci (2009), but he mainly discussed the idea in the context of Wald test, but did not discuss any extension to likelihood ratio test nor to score test. In this proposed bootstrap approach, we present a unified approach of using the bootstrap method to obtain the limiting distribution of the test statistics under complex sampling. The sampling design is allowed to be informative. The proposed method can be applicable to any of the Wald test, likelihood ratio test, or score test. The proposed method is also directly applicable to the test of independence for categorical survey data which does not involve computing Rao-Scott correction.

In Section 2, basic setup is introduced and the bootstrap method for Wald test, discussed in Beaumont and Bocci (1984), is presented. In Section 3, the proposed bootstrap test using survey-weighted log-likelihood ratio is introduced. In Section 4, the proposed bootstrap test using survey-weighted quasi-score is presented. Results from simulation studies are presented in Section 6.

2. Basic Setup

Suppose that a finite population U_N of size N is generated from a superpopulation model with density $f(y;\theta)$ for some $\theta \in \Theta \subset \mathcal{R}^p$. From the finite population U_N, a probability sample A of size n is selected with w_i being the sampling weight for unit i. We are interested in making inference about θ. The pseudo maximum likelihood estimator (PMLE) of θ is then obtained by

$$\hat{\theta} = \arg\max_{\theta \in \Theta} l_w(\theta),$$

where $l_w(\theta) = N^{-1}\sum_{i \in A} w_i \ln f(y_i;\theta)$.
Often, the solution $\hat{\theta}$ can be often obtained by solving the weighted score equation

$$\hat{S}_w(\theta) := \frac{\partial}{\partial \theta} l_w(\theta) = 0. \tag{1}$$

Under some regularity conditions (Fuller, 2009), we can establish

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \stackrel{\mathcal{L}}{\longrightarrow} N(0, nV(\hat{\theta})) \tag{2}$$

where

$$V(\hat{\theta}) = \mathcal{I}_\theta^{-1} V(\hat{S}_w) \mathcal{I}_\theta^{-1'}, \tag{3}$$

$\mathcal{I}_\theta = E\{I(\theta;Y)\}$, $I(\theta;y) = -\partial^2 \ln f(y;\theta)/\partial\theta\partial\theta'$, and $V(\hat{S}_w)$ is the variance of $\hat{S}_w(\theta)$ in (1).
Using the second order Taylor expansion, we can obtain

$$\begin{aligned} l_w(\theta) &= l_w(\hat{\theta}) + \hat{S}_w(\hat{\theta})'(\theta - \hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})'\hat{I}_w(\hat{\theta})(\theta - \hat{\theta}) + o_p(n^{-1}) \\ &= l_w(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})' E\{\hat{I}_w(\theta)\}(\theta - \hat{\theta}) + o_p(n^{-1}), \end{aligned}$$

where

$$\hat{I}_w(\theta) = N^{-1}\sum_{i \in A} w_i I(\theta; y_i) \tag{4}$$

and $I(\theta;y)$ is defined after (3). Thus, if we define

$$W_n(\theta) = -2n\left\{l_w(\theta) - l_w(\hat{\theta})\right\} \tag{5}$$

as a statistics for the likelihood ratio test, we can obtain

$$W_n(\theta) = n(\theta - \hat{\theta})' E\{\hat{I}_w(\theta)\}(\theta - \hat{\theta}) + o_p(1) \tag{6}$$

and, using (2),

$$n(\theta - \hat{\theta})' E\{\hat{I}_w(\theta)\}(\theta - \hat{\theta}) \stackrel{\mathcal{L}}{\longrightarrow} \sum_{i=1}^{p} c_i Z_i^2 \tag{7}$$

where $c_1, \cdots, c_p$ are the eigenvalues of $D = nV(\hat{\theta})E\{\hat{I}_w(\theta)\}$ and $Z_1, \cdots, Z_p$ are p independent random variables from the standard normal distribution. Therefore, we can establish a version of Wilks' theorem in survey sampling:

$$W_n(\theta) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{G} \equiv \sum_{i=1}^{p} c_i Z_i^2. \tag{8}$$

Unless the sampling design is the simple random sampling, the limiting distribution does not reduce to the standard chi-squared distribution. If $p = 1$ then we can use $c_1^{-1} W_n$ as the test statistic with $\chi^2(1)$ distribution as the limiting distribution under the null hypothesis.

3. Bootstrap calibration

We propose using a bootstrap method to approximating the limiting distribution in (7). Such a bootstrap calibration is very attractive because there is no need to derive the analytic form of the limiting distribution of the test statistic. To apply the bootstrap method, let w_i^* be the bootstrap weight of unit i and let $\hat{T}^* = \sum_{i \in A} w_i^* y_i$ be the bootstrap replicate of $\hat{T} = \sum_{i \in A} w_i y_i$, which is design unbiased for $T = \sum_{i \in U} y_i$. We assume that bootstrap weights are constructed such that the limiting distribution of $\sqrt{n} N^{-1}(\hat{T}^* - \hat{T})$ approximate the sampling distribution of $\sqrt{n} N^{-1}(\hat{T} - T)$. That is,

$$\lim_{n \to \infty} \sup_x \left| Pr\{\sqrt{n} N^{-1}(\hat{T}^* - \hat{T}) \leq x\} - Pr\{\sqrt{n} N^{-1}(\hat{T} - T) \leq x\} \right| = 0 \tag{9}$$

in probability. Under simple random sampling or stratified random sampling, Bickel and Freedman (1984) proves (9). Extension to other complex designs to satisfy (9) is still an open problem in survey sampling. To discuss (9), we consider Poisson sampling. For each $i \in A$, obtain $m_i^* \sim Bin(\hat{N}_i, \pi_i)$ independently, where $\hat{N}_i = w_i$ and $\pi_i = w_i^{-1}$. The bootstrap weight is computed as $w_i^* = w_i m_i^*$.
Using the above bootstrap weights, we can compute

$$l_w^*(\theta) = N^{-1} \sum_{i \in A} w_i^* \ln f(y_i; \theta),$$

the pseudo log-likelihood function based on the bootstrap sample. Let $\hat{\theta}^*$ be the maximizer of $l_w^*(\theta)$. The following lemma shows that the bootstrap replicate $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ approximates the sampling distribution of $\sqrt{n}(\hat{\theta} - \theta)$ in large samples.

Lemma 1 *Under some regularity conditions, we have*

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{\mathcal{L}} N(0, \hat{V}(\hat{\theta})) \tag{10}$$

almost everywhere, as $n \to \infty$*, where*

$$\hat{V}(\hat{\theta}) = \{\hat{I}_w(\hat{\theta})\}^{-1} \hat{V}(\hat{S}_w) \{\hat{I}_w(\hat{\theta})\}^{-1'}.$$

Note that Lemma 1 is a bootstrap version of the CLT in (2) for the pseudo MLE of θ. Now, to establish a bootstrap version of the Wilks' theorem in (8), we compute

$$W_n^*(\hat{\theta}) = -2n \left\{ l_w^*(\hat{\theta}) - l_w^*(\hat{\theta}^*) \right\} \tag{11}$$

as the bootstrap version of $W_n(\theta)$ in (5). The following theorem shows that W^* in (15) follows the same asymptotic distribution of W_n in (13).

Theorem 1 *Under some regularity conditions, for sufficiently large* n*,*

$$W_n^* \xrightarrow{\mathcal{L}} \mathcal{G} \tag{12}$$

where $\mathcal{G}$ *is the same stable distribution in (8).*

4. Likelihood ratio test

Now, let $\Theta_0 (\subset \Theta)$ be the parameter space under the null hypothesis. Thus, the null hypothesis can be written as $H_0 : \theta \in \Theta_0$. For example, if $\theta = (\theta_1, \theta_2)$ and $H_0 : \theta_2 = \theta_2^{(0)}$ for some fixed constant $\theta_2^{(0)}$, we can write

$$\Theta_0 = \{\theta = (\theta_1, \theta_2) \in \Theta; \theta_2 = \theta_2^{(0)}\}.$$

Let $\hat{\theta}_1^{(0}$ be the PMLE of θ_1 under $H_0 : \theta_2 = \theta_2^{(0)}$. The profile PMLE $\hat{\theta}_1^{(0)}$ can be obtained by maximizing $l_w(\theta_1, \theta_2^{(0)})$ with respect to θ_1.
The (psuedo) likelihood ratio test statistic for testing $H_0 : \theta_2 = \theta_2^{(0)}$ is defined as

$$W_n = -2n \left\{ l_w(\hat{\theta}^{(0)}) - l_w(\hat{\theta}) \right\} \tag{13}$$

where $\hat{\theta}^{(0)} = (\hat{\theta}_1^{(0)}, \theta_2^{(0)})$. Under simple random sampling, W_n in (13) follows from $\chi^2(q)$ distribution with $q = p - p_0$ and $p_0 = \dim(\Theta_0)$.
The following theorem, proved by Lumley and Scott (2014), presents the limiting distribution of the likelihood ratio test statistics in (13).

Theorem 2 *Under some regularity conditions, for sufficiently large n,*

$$W_n \xrightarrow{\mathcal{L}} \mathcal{G}_1 \equiv \sum_{i=1}^{q} c_i Z_i^2 \tag{14}$$

under $H_0 : \theta_2 = \theta_2^{(0)}$, where $c_1 \geq c_2 \geq \cdots \geq c_q > 0$ are the eigenvalues of $P = nV(\hat{\theta}_2)E\{I_{w22\cdot1}(\theta)\}$ and $Z_1, \cdots, Z_q$ are q independent random samples from the standard normal distribution, where

$$I_{w22\cdot1}(\theta) = I_{w22}(\theta) - I_{w21}(\theta)\{I_{w11}(\theta)\}^{-1} I_{w12}(\theta)$$

and $I_{wij}(\theta) = -E\{\partial^2 l_w(\theta)/(\partial\theta_i \partial\theta_j')\}$ for $i, j = 1, 2$.

Lumley and Scott (2014) proposed using the eigenvalues of the design effect matrix $\hat{P} = n\hat{V}(\hat{\theta}_2)I_{w22\cdot1}(\hat{\theta})$. The computation for $\hat{P}$ can be cumbersome. We consider an alternative method using a novel application of the bootstrap method in the next section.
Now, to discuss the hypothesis testing, the bootstrap replicate of the LR test statistics is computed by

$$W_n^* = -2n \left\{ l_w^*(\hat{\theta}_1^{*(0)}, \hat{\theta}_2) - l_w^*(\hat{\theta}^*) \right\}, \tag{15}$$

where $\hat{\theta}_1^{*(0)} = \arg\max_{\theta_1} l_w^*(\theta_1, \hat{\theta}_2)$ and $\hat{\theta}^* = \arg\max_{\theta} l_w^*(\theta)$. Thus, the bootstrap version of the profile MLE $\hat{\theta}_1(\theta_2^{(0)})$ is computed by maximizing the bootstrap profile likelihood at $\theta_2 = \hat{\theta}_2$.
The following theorem shows that W^* in (15) follows the same asymptotic distribution of W_n in (13).

Theorem 3 *Under some regularity conditions, for sufficiently large n,*

$$W_n^* \xrightarrow{\mathcal{L}} \mathcal{G}_1 \tag{16}$$

under $H_0 : \theta_2 = \theta_2^{(0)}$, where $\mathcal{G}_1$ is defined in Theorem 2.

By Theorem 3, we can use the bootstrap distribution of W_n^* in (15) to approximate the sampling distribution of the original test statistic under H_0. Thus, the p-value for testing $H_0 : \theta_2 = \theta_2^{(0)}$ using W_n can be obtained by computing the proportion of W_n^* greater than W_n.

5. Quasi-score test

We now consider a bootstrap method for the quasi-score test for $H_0 : \theta_2 = \theta_2^{(0)}$ under the setup of Section 4. The quasi-score test can be extended when only the first and the second moment conditions are assumed. We will first present the idea under the parametric model assumptions and then discuss quasi-score test.
Since $\hat{S}_w(\hat{\theta}) = 0$, we can use Taylor expansion of $S_w(\hat{\theta})$ with respect to θ to get

$$\hat{\theta} - \theta = [E\{I_w(\theta)\}]^{-1} \hat{S}_w(\theta) + o_p(n^{-1/2}). \tag{17}$$

Partitioning $\theta = (\theta_1, \theta_2)'$, $\hat{S}_w(\theta) = [\hat{S}_{w1}(\theta), \hat{S}_{w2}(\theta)]'$, and

$$\hat{I}_w(\theta) = \begin{bmatrix} \hat{I}_{w11}(\theta) & \hat{I}_{w12}(\theta) \\ \hat{I}_{w21}(\theta) & \hat{I}_{w22}(\theta) \end{bmatrix},$$

we can establish

$$\hat{\theta}_2 - \theta_2 = \left[E\{\hat{I}_{w2\cdot1}(\theta)\}\right]^{-1} S_{w2\cdot1}(\theta) + o_p(n^{-1/2}), \tag{18}$$

where

$$\hat{I}_{w2\cdot1}(\theta) = \hat{I}_{w22}(\theta) - \hat{I}_{w21}(\theta)\{\hat{I}_{w11}(\theta)\}^{-1}\hat{I}_{w12}(\theta)$$

and

$$\hat{S}_{w2\cdot1}(\theta) = \hat{S}_{w2}(\theta) - \hat{I}_{w21}(\theta)\{\hat{I}_{w11}(\theta)\}^{-1}S_{w1}(\theta).$$

Thus, the quasi-score test for $H_0 : \theta_2 = \theta_2^{(0)}$ can be constructed by computing

$$X_s^2 = \hat{S}_{w2\cdot1}(\hat{\theta}^{(0)})'\{\hat{I}_{w2\cdot1}(\hat{\theta}^{(0)})\}^{-1}\hat{S}_{w2\cdot1}(\hat{\theta}^{(0)}),$$

where $\hat{\theta}^{(0)}$ is the MLE of θ under H_0. Since $\hat{\theta}^{(0)}$ satisfies $S_{w1}(\hat{\theta}^{(0)}) = 0$, we can write

$$X_s^2 = \hat{S}_{w2}(\hat{\theta}^{(0)})'\{\hat{I}_{w2\cdot1}(\hat{\theta}^{(0)})\}^{-1}\hat{S}_{w2}(\hat{\theta}^{(0)}). \tag{19}$$

Under SRS, X_s^2 will follow $\chi^2(q)$ distribution. Under complex sampling, it will have the same asymptotic distribution as W_n in (13).
The bootstrap calibration method applied to X_s^2 can be used to construct a bootstrap test. That is, we can compute

$$X_s^{2*} = \hat{S}^*_{w2}(\hat{\theta}^{*(0)})'\{\hat{I}^*_{w2\cdot1}(\hat{\theta}^{*(0)})\}^{-1}\hat{S}^*_{w2}(\hat{\theta}^{*(0)})$$

to approximate the limiting distribution of X_s^2 in (19). The bootstrap distribution can be used to control the size of the test based on X_s^2 in (19).

6. Simulation Study

To test our theory, we perform a limited simulation study. The finite population of $N = 10,000$ with measurement (x, y) is generated from the following way:

1. The finite population is partitioned into 5 groups of size $N_g = 2,000$, $g = 1, 2, 3, 4, 5$.
2. In group g, generate $x_{gi} \sim N(-1 + 0.5g, 1)$ and compute p_{gi} with

$$\text{logit}(p_{gi}) = \beta_1 + \beta_2 x_{gi},$$

   where $(\beta_1, \beta_2) = (-1, 0.5)$. Generate $y_{gi} \sim Bernoulli(p_{gi})$.
3. Construct 10 strata by a cross-classification of groups and y-values. For example, stratum one consists of elements with $y = 1$ in group 1 and stratum two consists of elements with $y = 0$ in group 2.

From the finite population, we perform stratified random sampling with size $n_h = 20$ for each $h = 1, \cdots, H$. The sampling weights are computed using $w_{hi} = n_h^{-1} N_h$. The sampling design is a special case of case-control design and the sampling design is informative in the sense that we cannot ignore the sampling weights into estimation of βs.
We are interested in testing $H_0 : \beta_2 = 0.5$. We considered three setups for the simulation: $\beta_2 = 0.5, 0.4, 0.3$. Four test procedures are considered:

1. Naive LR method: Use W_n in (13) with reference distribution $\chi^2(1)$.
2. Naive score method: Use X_s^2 in (19) (with normalized weights) with reference distribution $\chi^2(1)$.
3. Bootstrap LR method: Use W_n in (13) with the boostrap-based reference distribution.
4. Bootstrap score method: Use X_s^2 in (19) with bootstrap reference distribution.

We used $\alpha = 0.05$ significant level for each test. In the bootstrap, we used $M = 1,000$ bootstrap samples. Table 1 presents the results of the simulation study based on $B = 1,000$ Monte Carlo samples. The simulation results shows good performance of the proposed bootstrap estimators. (The testing power at $\beta_2 = 0.5$ is close to the significance level $\alpha = 0.05$.) Bootstrap score test is slightly better than bootstrap LR test in terms of statistical power.

Table 1: Power of the test procedures

Method	β_2		
	0.5	0.4	0.3
Naive LR	0.001	0.021	0.083
Naive Score	0.002	0.022	0.098
Bootstrap LR	0.056	0.171	0.494
Bootstrap Score	0.055	0.182	0.514

References

Beaumont, J.-F. and Bocci, C. (2009). A practical bootstrap method for testing hypothesis from survey data. *Survey Methodology*, **35**, 25–35.

Bickel, P.J. and Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling, *The Annals of Statistics* **12**, 470–482.

Fuller, W.A. (2009). *Sampling Statistics*, Wiley.

Lumley, T. and Scott, A. (2014). Tests for regression models fitted to survey data, *Australian and New Zealand Journal of Statistics* **56**, 1–14.

Rao, J. N. K. and Scott, A. J. (1984). On chi-squared tests for multistage contingency tables with cell proportions estimated from survey data. *Annals of Statistics*, **12**, 46–60.

Rao, J. N. K., Scott, A. J., and Skinner, C. J. (1998). Quasi-score tests with survey data. *Statistica Sinica*, **8**, 1059–1070.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, **83**, 231–241.

Appendix

A. Proof of Lemma 1

Let $\hat{S}^*_w(\theta) = \partial l^*_w(\theta)/\partial\theta$ be the score function derived from $l^*_w(\theta) = N^{-1}\sum_{i\in A} w^*_i \ln f(y_i;\theta)$. We first show that

$$\sqrt{n}\left\{\hat{S}^*_w(\theta) - \hat{S}_w(\theta)\right\} \xrightarrow{\mathcal{L}} N[0, n\hat{V}\{\hat{S}_w(\theta)\}], \quad (A.1)$$

almost everywhere, which in turn implies

$$\sqrt{n}\hat{S}^*_w(\hat{\theta}) \xrightarrow{\mathcal{L}} N[0, n\hat{V}\{\hat{S}_w\}], \quad (A.2)$$

because $\hat{S}_w(\hat{\theta}) = 0$.

To show (A.1), note that we can write

$$\hat{S}^*_w(\theta) - \hat{S}_w(\theta) = \frac{1}{N}\sum_{i\in A}\hat{N}_i\left(\frac{m^*_i}{\hat{N}_i\pi_i} - 1\right)S(\theta; y_i)$$

where $m^*_i \sim Bin(\hat{N}_i, \pi_i)$, we can apply Lindeberg central limit theorem to show that, conditional on the sample,

$$\sqrt{n}\left\{\hat{S}^*_w(\theta) - \hat{S}_w(\theta)\right\} \xrightarrow{\mathcal{L}} N\left[0, nV_*\{\hat{S}^*_w(\theta)\}\right], \quad (A.3)$$

where

$$V_*\{\hat{S}_w^*(\theta)\} = \frac{1}{N^2}\sum_{i\in A}\hat{N}_i\frac{(1-\pi_i)}{\pi_i}S(\theta;y_i)^{\otimes 2}$$

is the variance of $\hat{S}_w^*(\theta)$ over the bootstrap distribution and $B^{\otimes 2}=BB'$. Since $\hat{N}_i=w_i=\pi_i^{-1}$,

$$V_*\{\hat{S}_w^*(\theta)\}=\frac{1}{N^2}\sum_{i\in A}\frac{(1-\pi_i)}{\pi_i^2}S(\theta;y_i)^{\otimes 2}=\hat{V}\{\hat{S}_w(\theta)\}.$$

Therefore, (A.3) implies (A.1).

Now, using that $\hat{S}_w^*(\theta)$ converges uniformly in probability to $\hat{S}_w(\theta)$ and $\hat{\theta}$ is the unique solution to $\hat{S}_w(\theta)=0$ almost everywhere, the solution $\hat{\theta}^*$ to $\hat{S}_w^*(\theta)=0$ converges in probability to $\hat{\theta}$ almost everywhere. Thus, using the second order Taylor expansion, we can obtain

$$\begin{aligned}0=\hat{S}_w^*(\hat{\theta}^*) &= \hat{S}_w^*(\hat{\theta})-\left\{\hat{I}_w^*(\hat{\theta})\right\}(\hat{\theta}^*-\hat{\theta})\\ &= \hat{S}_w^*(\hat{\theta})-\left\{\hat{I}_w(\hat{\theta})+o_p(1)\right\}(\hat{\theta}^*-\hat{\theta})\end{aligned}$$

where $\hat{I}_w^*(\theta)=-\partial\hat{S}_w^*(\theta)/\partial\theta$ and $\hat{I}_w(\theta)$ is defined in (4). Therefore, we have

$$\hat{\theta}^*-\hat{\theta}=\left\{\hat{I}_w(\hat{\theta})\right\}^{-1}\hat{S}_w^*(\hat{\theta})+o_p(n^{-1/2}). \quad (A.4)$$

Combining (A.2) with (A.4), we can establish (10).

B. Proof of Theorem 1

Using (10) and by the second order Taylor expansion, we have

$$l_w^*(\hat{\theta})=l_w^*(\hat{\theta}^*)+\hat{S}_w^*(\hat{\theta}^*)'(\hat{\theta}-\hat{\theta}^*)-(\hat{\theta}^*-\hat{\theta})'\hat{I}_w^*(\hat{\theta})(\hat{\theta}^*-\hat{\theta})+o_p(n^{-1}). \quad (B.1)$$

Since $\hat{S}_w^*(\hat{\theta}^*)=0$ and $\hat{I}_w^*(\hat{\theta})=\hat{I}_w(\hat{\theta})+o_p(1)$, we have

$$\begin{aligned}-2n\left\{l_w^*(\hat{\theta})-l_w^*(\hat{\theta}^*)\right\} &= n(\hat{\theta}^*-\hat{\theta})'\hat{I}_w(\hat{\theta})(\hat{\theta}^*-\hat{\theta})+o_p(1)\\ &= nS_w^*(\hat{\theta})'\{\hat{I}_w(\hat{\theta})\}^{-1}S_w^*(\hat{\theta})+o_p(1), \quad (B.2)\end{aligned}$$

where the second equality follows from (A.4). Now, by (A.2), we can show that W_n^* converges in distribution to the weighted sum of p independent chi-squared distribution where the weights are the eigenvalues of $n\{\hat{I}_w(\hat{\theta})\}^{-1}\hat{V}\{\hat{S}_w\}$ which converges in probability to $n\mathcal{I}_\theta^{-1}V\{\hat{S}_w(\theta)\}$. Therefore, (12) is established.

C. Proof of Theorem 3

Since $\hat{\theta}^{*(0)}=(\hat{\theta}_1^{*(0)},\hat{\theta}_2)$ where $\hat{\theta}_1^{*(0)}$ is the maximizer of $l_w^*(\theta_1,\hat{\theta}_2)$, we can obtain, similarly to (B.1),

$$l_w^*(\hat{\theta}) = l_w^*(\hat{\theta}^*)+\hat{S}_{w1}^*(\hat{\theta}^{*(0)})'(\hat{\theta}_1-\hat{\theta}_1^{*(0)})-(\hat{\theta}_1^{*(0)}-\hat{\theta}_1)'\hat{I}_{w11}^*(\hat{\theta})(\hat{\theta}_1^{*(0)}-\hat{\theta}_1)+o_p(n^{-1}).$$

where $\hat{S}_{w1}^*(\theta)=\partial l_w^*(\theta)/\partial\theta_1$ and $\hat{I}_{w11}^*(\theta)=\partial^2 l_w^*(\theta)/(\partial\theta_1\partial\theta_1')$. By definition of $\hat{\theta}^{*(0)}$, we have $\hat{S}_{w1}^*(\hat{\theta}^{*(0)})=0$. Thus, using $\hat{I}_{w11}^*(\hat{\theta})=\hat{I}_{w11}(\hat{\theta})+o_p(1)$, we have

$$\begin{aligned}-2n\left\{l_w^*(\hat{\theta})-l_w^*(\hat{\theta}^{*(0)})\right\} &= n(\hat{\theta}_1^{*(0)}-\hat{\theta}_1)'\hat{I}_{w11}(\hat{\theta})(\hat{\theta}_1^*-\hat{\theta}_1)+o_p(1)\\ &= nS_{w1}^*(\hat{\theta})'\{\hat{I}_{w11}(\hat{\theta})\}^{-1}S_{w1}^*(\hat{\theta})+o_p(1), \quad (C.1)\end{aligned}$$

where the last equality follows from

$$\hat{\theta}_1^{*(0)}-\hat{\theta}_1=\left\{\hat{I}_{w11}(\hat{\theta})\right\}^{-1}\hat{S}_{w1}^*(\hat{\theta})+o_p(n^{-1/2}).$$

Thus, combining (B.2) with (C.1), we have

$$\begin{aligned} W_n^* &= -2n\left\{l_w^*(\hat{\theta}^{*(0)}) - l_w^*(\hat{\theta}^*)\right\} \\ &= -2n\left\{l_w^*(\hat{\theta}) - l_w^*(\hat{\theta}^*)\right\} + 2n\left\{l_w^*(\hat{\theta}) - l_w^*(\hat{\theta}^{*(0)})\right\} \\ &= n\begin{pmatrix} S_{w1}^*(\hat{\theta}) \\ S_{w2}^*(\hat{\theta}) \end{pmatrix}' \begin{bmatrix} \hat{I}_{w11}(\hat{\theta}) & \hat{I}_{w12}(\hat{\theta}) \\ \hat{I}_{w21}(\hat{\theta}) & \hat{I}_{w22}(\hat{\theta}) \end{bmatrix}^{-1} \begin{pmatrix} S_{w1}^*(\hat{\theta}) \\ S_{w2}^*(\hat{\theta}) \end{pmatrix} \\ &\quad -nS_{w1}^*(\hat{\theta})'\{\hat{I}_{w11}(\hat{\theta})\}^{-1}S_{w1}^*(\hat{\theta}) + o_p(1) \\ &= n\{S_{w2}^*(\hat{\theta}) - \hat{B}_{21}S_{w1}^*(\hat{\theta})\}'\{\hat{I}_{w22\cdot1}(\hat{\theta})\}^{-1}\{S_{w2}^*(\hat{\theta}) - \hat{B}_{21}S_{w1}^*(\hat{\theta})\} \end{aligned}$$

where

$$\hat{B}_{21} = \hat{I}_{w21}(\hat{\theta})\{\hat{I}_{w11}(\hat{\theta})\}^{-1}$$

and

$$\hat{I}_{w22\cdot1}(\hat{\theta}) = \hat{I}_{w22}(\hat{\theta}) - \hat{I}_{w21}(\hat{\theta})\{\hat{I}_{w11}(\hat{\theta})\}^{-1}\hat{I}_{w12}(\hat{\theta}).$$

Therefore, by (A.2), using the same argument for proving Theorem 1, we can show that W_n^* converges in distribution to the weighted sum of q independent chi-squared distribution where the weights are the eigenvalues of $n\{\hat{I}_{w22\cdot1}(\hat{\theta})\}^{-1}\hat{V}\{\hat{S}_{w2|1}\}$ which converges in probability to $n\mathcal{I}_{22\cdot1}^{-1}V\{\hat{S}_{w2\cdot1}(\theta)\}$.