MODIFIED LIKELIHOOD RATIO TEST FOR HOMOGENEITY IN A TWO-SAMPLE PROBLEM

Yuejiao Fu, Jiahua Chen and John D. Kalbfleisch

York University, University of British Columbia, University of Michigan

Supplementary Material

Lemma 1 *Under the conditions of Theorem 1,* $\log(\hat{\lambda}) = O_p(1)$ *and the modified MLEs,* $\hat{\theta}_1$ *and* $\hat{\theta}_2$*, both converge to* θ_0 *in probability.*

Lemma 2 *Under the same assumptions as in Theorem 2,* $\log(\hat{\lambda}) = O_p(1)$*,* $\hat{\theta}_j \to \theta_0$ *for* $j = 1, 2$*, and* $\hat{\xi} \to \xi_0$*, in probability.*

It can be seen that Lemma 1 is a special case of Lemma 2. Thus, we prove Lemma 2 only.

Proof of Lemma 2. The key step of the proof is to show that $\sup R_{1n} = \sup R_{1n}(\lambda, \theta_1, \theta_2, \xi) = O_p(1)$. We consider the following two cases: (1) $|\theta_2 - \theta_0| \le \epsilon$ and (2) $|\theta_2 - \theta_0| > \epsilon$.

For case (1), applying the classical asymptotic technique (Wolfowitz, 1949), we can easily get that for any $\epsilon > 0$

$$\sup_{|\theta_1 - \theta_0| > \epsilon} \left[\sum_{i=1}^{n_1} \log \left\{ \frac{f(x_{1i}; \theta_1, \xi)}{f(x_{1i}; \theta_0, \xi_0)} \right\} + \sum_{i=1}^{n_2} \log \left\{ \frac{f(x_{2i}; G, \xi)}{f(x_{2i}; \theta_0, \xi_0)} \right\} \right] \le -n\rho$$

for some $\rho > 0$. Hence, with a negative penalty $2C \log \lambda$, $\sup\{R_{1n} : |\theta_1 - \theta_0| > \epsilon\} \le O_p(1)$, bounded above by $O_p(1)$. At the same time, it is easy to see that $\sup\{R_{1n} : |\theta_1 - \theta_0| \le \epsilon, |\theta_2 - \theta_0| \le \epsilon\} = O_p(1)$. Hence, $\sup\{R_{1n} : |\theta_2 - \theta_0| \le \epsilon\} = O_p(1)$ for some small enough ϵ.

We now consider case (2). For any given $\epsilon > 0$, classical consistency results (Wald, 1949) for the MLEs over the restricted region $|\theta_2 - \theta_0| > \epsilon$ imply that the un-modified MLE of λ goes to 0 in probability. Hence, asymptotically, we need only consider the $\sup R_{1n}$ over the region of $|\theta_2 - \theta_0| > \epsilon$ and $\lambda \le \epsilon$.

Using the same inequality as before, we have

$$\begin{aligned} R_{1n} - 2C\log\lambda &= 2\sum_{i=1}^{n_1}\log\{f(x_{1i};\theta_1,\xi)/f(x_{1i};\theta_0,\xi_0)\} + 2\sum_{i=1}^{n_2}\log(1+\delta_i) \\ &\leq 2\sum_{i=1}^{n_1}\log\{f(x_{1i};\theta_1,\xi)/f(x_{1i};\theta_0,\xi_0)\} \\ &\quad +2\sum_{i=1}^{n_1}\delta_i - \sum_{i=1}^{n_2}\delta_i^2 + \frac{2}{3}\sum_{i=1}^{n_2}\delta_i^3, \end{aligned}$$

where $\delta_i = f(x_{2i};G,\xi)/f(x_{2i};G_0,\xi_0) - 1$. Due to the regularity conditions on $f(x;\theta,\xi)$, there is a quadratic expansion for $\sum_{i=1}^{n_1}\log\{f(x_{1i};\theta_1,\xi)/f(x_{1i};\theta_0,\xi_0)\}$ in $\theta_1-\theta_0$ and $\xi-\xi_0$.

Our aim is to expand terms related to the second sample as quadratic functions of $\theta_1-\theta_0$, $\xi-\xi_0$ and λ. (Because $\theta_2-\theta_0$ cannot be regarded as a small-o term, it is not part of the targeted quadratic function.) Toward this end, we write $\delta_i = (1-\lambda)(\theta_1-\theta_0)Y_{2i} + \lambda\theta_2 Y_{2i}(\theta_2,\xi_0) + (\xi-\xi_0)U_{2i} + e_i$ with

$$\begin{aligned} e_i &= (1-\lambda)(\theta_1-\theta_0)\{Y_{2i}(\theta_1,\xi) - Y_{2i}\} + \lambda\theta_2\{Y_{2i}(\theta_2,\xi) - Y_{2i}(\theta_2,\xi_0)\} \\ &\quad +(\xi-\xi_0)\{U_{2i}(\xi) - U_{2i}\}. \end{aligned}$$

We now establish the asymptotic orders of $\sum e_i$, $\sum e_i^2$ and $\sum |e_i|^3$. Notice that

$$Y_{2i}(\theta_1,\xi) - Y_{2i} = \{Y_{2i}(\theta_1,\xi) - Y_{2i}(\theta_0,\xi)\} + \{Y_{2i}(\theta_0,\xi) - Y_{2i}\}.$$

With some abuse of notation, we have

$$\begin{aligned} \sum\{Y_{2i}(\theta_1,\xi) - Y_{2i}\} &= (\theta_1-\theta_0)\sum Y'_\theta(\theta^*,\xi) + (\xi-\xi_0)\sum Y'_\xi(\theta_0,\xi^*) \\ &= (\theta_1-\theta_0)O_p(n_2^{1/2}) + (\xi-\xi_0)O_p(n_2^{1/2}), \end{aligned}$$

where the tightness condition (B5) is used in the last step. Hence, we have

$$\sum(\theta_1-\theta_0)\{Y_{2i}(\theta_1,\xi) - Y_{2i}\} = \{(\theta_1-\theta_0)^2 + (\xi-\xi_0)^2\}O_p(n_2^{1/2}).$$

In a similar way, we find

$$\sum\lambda\theta_2\{Y_{2i}(\theta_2,\xi) - Y_{2i}(\theta_2,\xi_0)\} = \lambda(\xi-\xi_0)O_p(n_2^{1/2}) = \{\lambda^2 + (\xi-\xi_0)^2\}O_p(n_2^{1/2})$$

and $\sum(\xi-\xi_0)\{U_{2i}(\xi) - U_{2i}\} = (\xi-\xi_0)^2 O_p(n_2^{1/2})$. Taking these results together, we obtain $\sum e_i = \{(\theta_1-\theta_0)^2 + \lambda^2 + (\xi-\xi_0)^2\}O_p(n_2^{1/2})$.

Next, we examine the order of $\sum e_i^2$. By the condition of uniform convergence in $Y_\theta'^2$ and $Y_\xi'^2$, we have

$$\begin{aligned}\sum(\theta_1-\theta_0)^2\{Y_{2i}(\theta_1,\xi)-Y_{2i}\}^2 &\leq (\theta_1-\theta_0)^2\{(\theta_1-\theta_0)^2+(\xi-\xi_0)^2\}O_p(n_2)\\ &= (\theta_1-\theta_0)^2 o(1)O_p(n_2).\end{aligned}$$

Here, $o(1)$ means a quantity that shrinks to 0 as $\theta_1-\theta_0\to 0$ and $\xi-\xi_0\to 0$. Along the same line, we have

$$\sum \lambda^2\theta_2^2\{Y_{2i}(\theta_2,\xi)-Y_{2i}(\theta_2,\xi_0)\}^2=\{(\theta_1-\theta_0)^2+(\xi-\xi_0)^2\}o(1)O_p(n_2)$$

and $\sum(\xi-\xi_0)^2\{U_{2i}(\xi)-U_{2i}\}=(\xi-\xi_0)^2 o(1)O_p(n_2)$. These order assessments lead to $\sum e_i^2=\{(\theta_1-\theta_0)^2+\lambda^2+(\xi-\xi_0)^2\}o(1)O_p(n_2)$, and similarly we also obtain $\sum|e_i|^3=\{(\theta_1-\theta_0)^2+\lambda^2+(\xi-\xi_0)^2\}o(1)O_p(n_2)$. Further, since we focus on small values of λ, we have $(1-\lambda)(\theta_2-\theta_0)=(\theta_2-\theta_0)(1+o(1))$.

Hence

$$\begin{aligned}&R_{1n}-2C\log\lambda\\ &\quad\leq 2(\theta_1-\theta_0)\sum_{i=1}^{n_1}Y_{1i}+2(\xi-\xi_0)\sum_{i=1}^{n_1}U_{1i}\\ &\quad+2(\theta_1-\theta_0)\sum_{i=1}^{n_2}Y_{2i}+2\lambda\theta_2\sum_{i=1}^{n_2}Y_{2i}(\theta_2,\xi_0)+2(\xi-\xi_0)\sum_{i=1}^{n_2}U_{2i}\\ &\quad-\left[\sum_{i=1}^{n_1}\{(\theta_1-\theta_0)Y_{1i}+(\xi-\xi_0)U_{1i}\}^2\right.\\ &\quad+\left.\sum_{i=1}^{n_2}\{(\theta_1-\theta_0)Y_{2i}+\lambda\theta_2 Y_{2i}(\theta_2,\xi_0)+(\xi-\xi_0)U_{2i}\}^2\right]\\ &\quad+\{(\theta_1-\theta_0)^2+\lambda^2+(\xi-\xi_0)^2\}o(1)O_p(n).\end{aligned}$$

After division by n, the quadratic term in the above expression converges to

$$\begin{pmatrix}\theta_1-\theta_0\\ \xi-\xi_0\\ \lambda\theta_2\end{pmatrix}^\tau\begin{pmatrix}\sigma_Y^2 & \sigma_{YU} & \rho\sigma_{Y(\theta_2)Y}\\ \sigma_{YU} & \sigma_U^2 & \rho\sigma_{Y(\theta_2)U}\\ \rho\sigma_{Y(\theta_2)Y} & \rho\sigma_{Y(\theta_2)U} & \rho\sigma_{Y(\theta_2)}^2\end{pmatrix}\begin{pmatrix}\theta_1-\theta_0\\ \xi-\xi_0\\ \lambda\theta_2\end{pmatrix},$$

where $\sigma_{Y(\theta_2)Y}=\mathrm{Cov}(Y_{2i}(\theta_2,\xi_0),Y_{2i})$ and $\sigma_{Y(\theta_2)U}=\mathrm{Cov}(Y_{2i}(\theta_2,\xi_0),U_{2i})$. The symmetric matrix can be further written as

$$(1-\rho)\begin{pmatrix} \sigma_Y^2 & \sigma_{YU} & 0 \\ \sigma_{YU} & \sigma_U^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \rho \begin{pmatrix} \sigma_Y^2 & \sigma_{YU} & \sigma_{Y(\theta_2)Y} \\ \sigma_{YU} & \sigma_U^2 & \sigma_{Y(\theta_2)U} \\ \sigma_{Y(\theta_2)Y} & \sigma_{Y(\theta_2)U} & \sigma^2_{Y(\theta_2)} \end{pmatrix}.$$

The identifiability condition, $\sigma_{YU}^2 < \sigma_Y^2\sigma_U^2$, implies that it is positive definite, regardless of the value of θ_2.

Thus, due to tightness of $\sum Y_{2i}(\theta)$, we have

$$\sup_{|\theta_2-\theta_0|>\epsilon} R_{1n} \leq$$

$$\frac{1}{n}\begin{pmatrix} \sum Y_{1i} + \sum Y_{2i} \\ \sum U_{1i} + \sum U_{2i} \\ \sum Y_{2i}(\theta_2, \xi_0) \end{pmatrix}^\tau \begin{pmatrix} \sigma_Y^2 & \sigma_{YU} & \rho\sigma_{Y(\theta_2)Y} \\ \sigma_{YU} & \sigma_U^2 & \rho\sigma_{Y(\theta_2)U} \\ \rho\sigma_{Y(\theta_2)Y} & \rho\sigma_{Y(\theta_2)U} & \rho\sigma^2_{Y(\theta_2)} \end{pmatrix} \begin{pmatrix} \sum Y_{1i} + \sum Y_{2i} \\ \sum U_{1i} + \sum U_{2i} \\ \sum Y_{2i}(\theta_2, \xi_0) \end{pmatrix} + o_p(1)$$

$$= O_p(1).$$

It follows that $\sup R_{1n} = O_p(1)$. Let $\hat\lambda$ be the maximizer of $R_{1n}(\lambda, \theta_1, \theta_2, \xi)$, it follows that $\log(\hat\lambda) = O_p(1)$. Thus, for any given small positive number $\epsilon > 0$, we can find some $\delta > 0$ such that $P(\hat\lambda > \delta) > 1 - \epsilon$. For asymptotic considerations, this result allows us to discuss the problem further under the constraint $\lambda > \delta$ for some $\delta > 0$. With this restriction, the parameter space for G is compact, and the penalty term $\log(\lambda)$ has negligible influence in the modified likelihood. The consistency of $\hat G$ for G is the consequence of the classical result of Wald (1949). With $\hat\lambda > \delta > 0$ in probability, we must have $\hat\theta_j \to \theta_0$ for $j = 1, 2$. This completes the proof.

References

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595-601.

Wolfowitz, J. (1949). On Wald's proof of the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 601-602.