A Simple Distribution-Free Test for Nonnested Model Selection: Supplementary Materials

Kevin A. Clarke †
University of Rochester

Abstract

This Web appendix contains proofs of consistency and unbiasedness for the article "A Simple Distribution-Free Test for Nonnested Model Selection." The accompanying Stata code demonstrates how the Huth and Allee example was computed. The data must be obtained from the original authors.

January 4, 2007

†Kevin A. Clarke is Assistant Professor, Department of Political Science, University of Rochester, Rochester, NY 14627-0146. Email: kevin.clarke@rochester.edu. This work was supported by National Science Foundation Grant SES-0213771. I thank Paul K. Huth and Todd L. Allee for graciously sharing their data and code. Errors remain my own.

Properties of the Distribution-Free Test

Consistency

To prove consistency, we can make use of the following theorem.

Theorem .1 *(Lehmann 1951) Let T_n denote a sequence of test statistics for an α-level test of $H_0 : \theta \in \omega$ versus $H_1 : \theta \in \Omega - \omega$, such that the test based on T_n rejects H_0 if $T_n \geq c_n$ or $T_n \leq c'_n$. Suppose there exists a function $g(\theta)$ such that T_n converges in probability to $g(\theta)$ for every $\theta \in \Omega$. If, in addition,*

$$\begin{aligned} g(\theta) &= g_0 \quad \forall \;\; \theta \in \omega, \\ g(\theta) &\neq g_0 \quad \forall \;\; \theta \in \Omega - \omega, \end{aligned}$$

and

$$\begin{aligned} \lim_{n\to\infty} c_n &\leq g_0, \\ \lim_{n\to\infty} c'_n &\geq g_0, \end{aligned}$$

then T_n is a consistent sequence of tests for all alternatives in $H_1 : \theta \in \Omega - \omega$.

Let θ be the median, and let $g(\theta) = \Pr(D_i > \theta)$ where $D_i = \ln[f(Y_i|X_i;\beta_*) /g(Y_i|Z_i;\gamma_*)]$. Then $g(\theta) = (1/2)$ for $\theta \in \omega$, and $g(\theta) \neq (1/2)$ for $\theta \in \Omega - \omega$. Now, write the test statistic as $T_n = B/n$, where $B = \sum_{i=1}^{n} I_{(\theta,+\infty)}(d_i)$. We show that T_n converges in probability to $g(\theta)$ by showing convergence in quadratic mean. The expected value and variance of T_n are

$$\begin{aligned} E\left[\frac{B}{n}\right] &= \frac{1}{n}\sum_{i=1}^{n} E[I_{(\theta,+\infty)}(d_i)] = g(\theta), \\ V\left[\frac{B}{n}\right] &= \frac{1}{n^2}\sum_{i=1}^{n} V[I_{(\theta,+\infty)}(d_i)] = \frac{g(\theta)[1-g(\theta)]}{n}. \end{aligned}$$

The variance tends to zero as $n \to \infty$, so we have shown that $T_n \xrightarrow{p} g(\theta)$.

As consistency is a large-sample property, we can consider the large-sample approximation of the distribution-free test. Under the null hypothesis, the expected value of T_n (from above) is (1/2), and the variance is $(1/4n)$. For large n, the test therefore rejects the null when

$$\frac{\left|\frac{B}{n} - \frac{1}{2}\right|}{\sqrt{\frac{1}{4n}}} \geq z_{\frac{\alpha}{2}},$$

where $z_{\frac{\alpha}{2}}$ is the upper $100(\alpha/2)^{\text{th}}$ percentile of the standard normal. Rearranging, we see that the test rejects when

$$\left|\frac{B}{n}\right| \geq c_n = \frac{1}{2} + z_{\frac{\alpha}{2}}\sqrt{\frac{1}{4n}}.$$

As $n \to \infty$, $c_n \to (1/2) \leq g_0 = (1/2)$. As we have met the conditions of Lehmann's theorem, we can state that the test is consistent for all alternatives in $\mathrm{H}_1 : \theta \in \Omega - \omega$. (Similar proofs can be given for one-tailed tests.)

Unbiasedness

The distribution-free test is,

$$\begin{aligned} \mathrm{H}_0 : \mathrm{median}_0(D_i) &= \theta_0 \text{ versus} \\ \mathrm{H}_1 : \mathrm{median}_0(D_i) &< (>)\ \theta_0, \end{aligned}$$

where $D_i = \ln[f(Y_i|X_i;\beta_*)/g(Y_i|Z_i;\gamma_*)]$. Let $D_1,\ldots,D_n$ be i.i.d. $F(d-\theta)$, where θ is the median of the underlying distribution. We prove unbiasedness by noting that the distribution-free test reaches its natural significance level for every distribution in $F(d-\theta)$, and that its power function is monotonic. We prove the later point using the following theorem.

Theorem .2 *(Randles and Wolfe 1979) Suppose that for testing H_0 versus H_1 we reject H_0 for large (small) values of a test statistic $T(X_1,\ldots,X_n)$ that satisfies $T(x_1+k,\ldots,x_n+k) \geq (\leq)\ T(x_1,\ldots,x_n)$ for every $k \geq 0$ and $(x_1,\ldots,x_n)$. Then the test has a monotone power function in θ for the one-sample location problem; that is,*

$$\mathcal{P}_T(\theta,F) \leq \mathcal{P}_T(\theta',F) \text{ for } \theta \leq \theta',$$

and any continuous distribution with c.d.f. $F(\cdot)$.

The distribution-free test rejects for large (small) values of

$$B(D_i,\ldots,D_n) = \sum_{i=1}^{n} I_{(\theta_0,+\infty)}(d_i)$$

where I is the indicator function. When $k \geq 0$,

$$\begin{aligned} B(d_1+k,\ldots,d_n+k) &= \sum_{i=1}^{n} I_{(\theta_0,+\infty)}(d_i+k) \\ &= \sum_{i=1}^{n} I_{(\theta_0-k,+\infty)}(d_i) \\ &\geq B(d_1,\ldots,d_n). \end{aligned}$$

The test then has a monotone power function in θ, and therefore the distribution-free test is an unbiased of $H_0 : \text{median}_0(D_i) = \theta_0$ against $H_1 : \text{median}_0(D_i) > (<)\ \theta_0$.

References

Lehmann, E. L. 1951. "Consistency and Unbiasedness of Certain Nonparametric Tests." *Annals of Mathematical Statistics* 22:165–179.

Randles, Ronald H., and Douglas A. Wolfe. 1979. *Introduction to The Theory of Nonparametric Statistics*. New York: John Wiley and Sons.