Strong model dependence in statistical analysis: goodness of fit is not enough for model choice

John COPAS* and Shinto EGUCHI**
* Department of Statistics, University of Warwick, Coventry, UK
** Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan

SUPPLEMENTARY APPENDIX

Appendix A: Proof of Inequality (17)

Lemma

Let M be a model and $G(x, M)$ be an omnibus goodness-of-fit test satisfying (11). Then, for sufficiently large sample size n and almost all x, there exists a (non-null) set of data vectors x^* for which $G(x^*, M) \leq g_\alpha$ and $|z(x^*, M)| > z_\alpha$.

Proof of Lemma

The proof follows the classic proof of the Neyman-Pearson Lemma. Consider the two sets

$$\{x \ : \ G(x, M) \leq g_\alpha\} \text{ and } \{x \ : \ |z(x, M)| \leq z_\alpha\} \ .$$

The assumption of an omnibus test means that, with positive probability, these sets are different. Both have the same asymptotic probability under model M, and so for sufficiently large n there must exist values of x^* (with positive probability) that belong to the first set but not to the second.

Proof of (17)

Inequality (17) is established if, for sufficiently large n and almost all data vectors x, we can find a model M for which

$$\sigma_M = \sigma^* \ , \ G(x, M) \leq g_\alpha \ , \ |z(x, M)| > z_\alpha \ . \qquad (32)$$

Let $\hat{\theta}$ be the MLE based on model M_0, and choose a vector γ for which

$$\frac{(\phi'^T \gamma)^2}{\gamma^T I \gamma} = \sigma^{*2} \ , \qquad (33)$$

where $\phi' = \partial\phi(\theta)/\partial\theta$ evaluated at $\hat{\theta}$. It is easy to show that γ satisfying (33) exists for any $\sigma^{*2} \leq \sigma^2 = \phi'^T I^{-1} \phi'$. Now consider the model N, with scalar parameter θ_N, that is defined by the linear function

$$F_N(\theta_N) = \hat{\theta} + \theta_N \gamma \ .$$

From (10), $D_N = \gamma$ and $\sigma_N = \sigma^*$. Model N gives an exact fit to the data in the sense that $F_N(0) = \hat{\theta}$, and so is empirically acceptable in the sense of this paper.

Using the *Lemma*, for sufficiently large sample size n and for almost all x, we can find another data vector x^* for which

$$G(x^*, N) \leq g_\alpha \ , \ |z(x^*, N)| > z_\alpha \ . \tag{34}$$

Let $\hat{\theta}^*$ be the MLE of θ for data x^* under model M_0, and $\hat{\theta}^*_N$ the corresponding MLE of θ_N under model N. Now consider another scalar parameter model, M, defined by

$$F_M(\theta_M) = \hat{\theta} + d + \theta_M \gamma \ ,$$

where $d = \hat{\theta}^* - F_N(\hat{\theta}^*_N)$. Model M is essentially the same as N but with the fitted values of θ shifted by the constant displacement d. Trivially, M also has $D_M = \gamma$ and $\sigma_M = \sigma^*$.

Geometrically, M and N can be thought of as two parallel straight lines in θ-space, with the point $\hat{\theta}$ on the N-line projecting onto the point $F_M(\hat{\theta}_M)$ on the M-line, and the point $\hat{\theta}^*$ on the M-line projecting on to the point $F_N(\hat{\theta}^*_N)$ on the N-line. Using standard first order likelihood approximations as in (12), we get

$$\hat{\theta}^*_N = \frac{\gamma^T I(\hat{\theta}^* - \hat{\theta})}{\gamma^T I \gamma} \ , \ \hat{\theta}_M = -\frac{\gamma^T I d}{\gamma^T I \gamma}$$

up to $O_p(n^{-1})$. The first of these equations gives

$$d = \hat{\theta}^* - (\hat{\theta} + \hat{\theta}^*_N \gamma) = (\hat{\theta}^* - \hat{\theta}) - \gamma \frac{\gamma^T I(\hat{\theta}^* - \hat{\theta})}{\gamma^T I \gamma} + O_p(n^{-1}) \ ,$$

and so, from the second, we get $\hat{\theta}_M = O_p(n^{-1})$. Hence

$$d = \hat{\theta}^* - F_N(\hat{\theta}^*_N) = -\{\hat{\theta} - F_M(\hat{\theta}_M)\} + O_p(n^{-1}) \ . \tag{35}$$

Geometrically, this means that the two projections between the N-line and the M-line are also (approximately) parallel. As $F_M(\hat{\theta}_M) = F_N(\hat{\theta}^*_N) + O_p(n^{-\frac{1}{2}})$, (35) and (11) give

$$G(x, M) = G(x^*, N) + O_p(n^{-\frac{1}{2}}) \ . \tag{36}$$

Also, from (35),

$$\phi(\hat{\theta}) - \phi\{F_M(\hat{\theta}_M)\} = -[\phi(\hat{\theta}^*) - \phi\{F_N(\hat{\theta}^*_N)\}] + O_p(n^{-1}) \ ,$$

and so, from (15),

$$z(x, M) = -z(x^*, N) + O_p(n^{-\frac{1}{2}}) \ . \tag{37}$$

Comparing (36) and (37) with (34) shows that, for sufficiently large n and almost all x, model M satisfies (32).

Appendix B: Proof of Equation (30)

To simplify the notation, define

$$b^2 = \hat{\theta}^T P \hat{\theta} \ , \ c = \xi^T P \xi \ , \ d = \xi^T P \hat{\theta} \ , \ v^2 = \hat{\theta}^T \hat{\theta} \ , \ r = \hat{\theta}^T \xi / v \ .$$

There is no loss of generality if we assume that $r \geq 0$.

First we consider the bounds, over symmetric idempotent matrices P, of the value of d for a given value of c. To do this consider the Lagrangian

$$L(P) = \xi^T P \hat{\theta} - \frac{1}{2}\text{trace}\{\Lambda(P^2 - P)\} - \frac{1}{2}\mu\{\xi^T P \xi - c\} \ ,$$

where the Lagrange multipliers are the symmetric $k \times k$ matrix Λ and the scalar μ. We get

$$\frac{\partial}{\partial P} L(P) = Q - \Lambda P - P\Lambda + \Lambda = 0 \ , \tag{38}$$

where Q is the symmetric matrix

$$Q = \xi \hat{\theta}^T + \hat{\theta} \xi^T - \mu \xi \xi^T \ . \tag{39}$$

Pre-multiplying (38) by P, and noting that $P^2 = P$, gives $P\Lambda P = PQ$, from which we get the key equation

$$PQ = PQP \ ,$$

since $P\Lambda PP = PQP$. (We can also get the same equation using a Lagrangian in terms of A_M with the constraint $A_M A_M^T = I_{k_M}$).

Substituting (39) for Q in

$$\xi^T PQ\xi = \xi^T PQP\xi \ , \tag{40}$$

and identifying each term in the above notation, gives

$$rvc + d - \mu c = 2cd - \mu c^2 \ .$$

Similarly, replacing one or the other of the vectors ξ to the left and right of each side of (40) by $\hat{\theta}$ gives two further equations, and hence simultaneous equations from which we can eliminate b and μ to get

$$d^2 - 2rvcd = cr^2v^2 + c(1-c)v^2 \ .$$

This is a quadratic equation for d whose roots give the maximum and minimum

$$v[rc - \{c(1-c)(1-r^2)\}^{\frac{1}{2}}] \leq d \leq v[rc + \{c(1-c)(1-r^2)\}^{\frac{1}{2}}] \ . \tag{41}$$

To satisfy $|z| \leq z_\alpha$ we also need (29), which in this notation is

$$rv - \delta(1-c)^{\frac{1}{2}} \leq d \leq rv + \delta(1-c)^{\frac{1}{2}} \ . \tag{42}$$

The confidence limits are

$$CI_M = (d - \delta c^{\frac{1}{2}} \ , \ d + \delta c^{\frac{1}{2}}) \ .$$

For large n, we are interested in these limits when δ is small.

To prove (30), we first fix c and find the maximum of $d + \delta c^{\frac{1}{2}}$ for d satisfying both (41) and (42), and then maximize over c. This is $\max\{f_3(c)\}$ where

$$\begin{aligned} f_1(c) &= v[rc + \{c(1-c)(1-r^2)\}^{\frac{1}{2}}] \\ f_2(c) &= rv + \delta(1-c)^{\frac{1}{2}} \\ f_3(c) &= \min\{f_1(c), f_2(c)\} + \delta c^{\frac{1}{2}} \ . \end{aligned}$$

Elementary calculation shows that

$$\begin{array}{cc} f_1(c) < f_2(c) & \text{if } c < c^* \\ f_1(c) > f_2(c) & \text{if } c > c^* \ , \end{array}$$

where, to first order in δ,

$$c^* = r^2 + 2\delta \frac{r(1-r^2)^{\frac{1}{2}}}{v} \ ,$$

and so

$$f_3(c) = \left\{ \begin{array}{ll} f_1(c) + \delta c^{\frac{1}{2}} & \text{if } c < c^* \\ f_2(c) + \delta c^{\frac{1}{2}} & \text{if } c > c^* \end{array} \right\} . \ .$$

Note that $f_3(c)$ has a discontinuity in first derivative at $c = c^*$. The first derivative to the left is

$$f_3'(c^*-) = \frac{v}{2r} + O(\delta) \ ,$$

which is positive when δ is small. In fact $f_3(c)$ is an increasing function for $c < c^*$. The derivative to the right is

$$f_3'(c^*+) = \frac{\delta}{2r(1-r^2)^{\frac{1}{2}}}\{(1-r^2)^{\frac{1}{2}} - r\} = \left\{ \begin{array}{ll} > 0 & \text{if } r < 2^{-\frac{1}{2}} \\ < 0 & \text{if } r > 2^{-\frac{1}{2}} \end{array} \right\} \ .$$

It follows that the maximum of $f_3(c)$ is attained at $c = c^*$ if $r > 2^{-\frac{1}{2}}$ and at $c = c^{**}$ if $r < 2^{-\frac{1}{2}}$, where $c^{**} > c^*$ is the solution to

$$\frac{\partial}{\partial c}\{f_2(c) + \delta c^{\frac{1}{2}}\} = \frac{\delta}{2}\{c^{-\frac{1}{2}} - (1-c)^{-\frac{1}{2}}\} = 0 \ .$$

This gives $c^{**} = \frac{1}{2}$ which, for large n, exceeds $c^* = r^2 + O(\delta)$ when $r < 2^{-\frac{1}{2}}$. Thus we have equation (30),

$$\max_c\{f_3(c)\} = \left\{ \begin{array}{ll} f_3(\frac{1}{2}) = rv + 2^{\frac{1}{2}}\delta & \text{if } r \le 2^{-\frac{1}{2}} \\ f_3(x^*) = rv + \delta\{(1-r^2)^{\frac{1}{2}} + r) & \text{if } r > 2^{-\frac{1}{2}} \end{array} \right. \ .$$