A MULTIPLE DIVERGENCE CRITERION FOR TESTING
BETWEEN SEPARATE HYPOTHESES

K.R. Sawyer

The University of Western Australia

Discussion Paper No. 81-16

September 1981

A MULTIPLE DIVERGENCE CRITERION FOR TESTING
BETWEEN SEPARATE HYPOTHESES

by

K.R. Sawyer

University of Western Australia

1. Introduction

Two families of hypotheses H_1 and H_2 are separate if an arbitrary simple hypothesis in H_1 cannot be obtained as the limit of simple hypotheses in H_2 . In discriminating between a set of M separate hypotheses, traditional statistical method has invoked binary comparisons. These binary comparisons include the Cox test (Cox (1961, 1962)), the information criterion (Sawyer (1981a)), and a number of artificial-nesting techniques (Davidson and Mackinnon (1979a, 1979b)).

Interactive and interpretative difficulties considerably undermine the viability of binary comparisons however. This has provided the motivation for multiple procedures which discriminate the models jointly, notably the multiple J-test (Davidson and Mackinnon (1979a)), the multiple P-test (Davidson and Mackinnon (1979b)), and the multiple Cox test (Sawyer (1981b)). None of these measures is based on a well established measure of hypothesis discrimination.

The statistic proposed in this paper, a variant of Jeffreys' divergence, is formulated for multiple comparisons of separate hypotheses. Like the information criterion, it is derived from an accepted measure of hypothesis discrimination. The multiple divergence criterion, as it will be termed, has a strong intuitive and interpretative appeal.

2. Divergence

Jeffreys (1948) and Kullback (1959) formalised the concept of divergence as a measure of the separation of hypotheses. Consider M hypotheses

$$H_i: \text{density } f_i(y, \alpha_i), \text{ likelihood } \ell_i(y, \alpha_i), \text{ loglikelihood } L_i(y, \alpha_i) \\ i=1, \dots, M. \quad (1)$$

Denote by

- (a) $\hat{\alpha}_i$, the maximum likelihood estimate of α_i ;
- (b) $L_{ik}(\alpha_i, \alpha_k) = L_i(\alpha_i) - L_k(\alpha_k)$, the log of the likelihood ratio;
- (c) $\text{plim}_{\hat{\alpha}_i}$, the probability limit under H_i evaluated at $\alpha_i = \hat{\alpha}_i$.

The non-directional divergence between two hypotheses H_i and H_k is given by

$$J(H_i: H_k) = J_{ik} = \int [f_i(y, \alpha_i) - f_k(y, \alpha_k)] \log \frac{f_i(y, \alpha_i)}{f_k(y, \alpha_k)} dy \quad (2)$$

and estimated asymptotically by

$$\hat{J}_{ik} = \text{plim}_{\hat{\alpha}_i} \frac{1}{T} L_{ik}(\hat{\alpha}_i, \hat{\alpha}_k) - \text{plim}_{\hat{\alpha}_k} \frac{1}{T} L_{ik}(\hat{\alpha}_i, \hat{\alpha}_k). \quad (3)$$

Consider the statistic J_{ik}^* , the non-directional divergence minus its asymptotic expectation, or estimate thereof, under H_i

$$J_{ik}^* = J_{ik} - \text{plim}_{\hat{\alpha}_i} J_{ik}^*. \quad (4)$$

Sawyer (1981a) showed that J_{ik}^* can be written

$$J_{ik}^* = -\eta_k' \left\{ (\hat{\alpha}_k - \alpha_{ik}) - \left(\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right)' (\hat{\alpha}_i - \alpha_i) \right\} \quad (5)$$

$$\text{where (a) } \alpha_{ik} = \text{plim}_{\hat{\alpha}_i} \hat{\alpha}_k \quad \text{and} \quad (6)$$

$$(b) \quad \eta_k = \frac{\partial}{\partial \alpha_k} \text{plim}_{\hat{\alpha}_i} \frac{1}{T} L_{ik}(\hat{\alpha}_i, \hat{\alpha}_k). \quad (7)$$

J_{ik}^* is then the scaled vector of deviations of $\hat{\alpha}_k - \alpha_{ik}$ from its asymptotic regression on $(\hat{\alpha}_i - \alpha_i)$.

The multiple divergence between H_i and all the other hypotheses can be defined as the vector of non-directional divergences between H_i and H_k , $k \neq i$, so that

$$J(H_i : H_k, k \neq i) = (J_{ik})_{k \neq i}. \quad (8)$$

As the vector of divergences from H_i to the other hypotheses, it is a proper measure of separation between them, representing an additive, non-negative discriminant function in $(M-1)$ Euclidean space. The multiple divergence is estimated asymptotically by

$$\hat{J} = (\hat{J}_{ik})_{k \neq i} = \left(\text{plim}_{\hat{\alpha}_i} \frac{1}{T} L_{ik}(\hat{\alpha}_i, \hat{\alpha}_k) - \text{plim}_{\hat{\alpha}_k} \frac{1}{T} L_{ik}(\hat{\alpha}_i, \hat{\alpha}_k) \right)_{k \neq i} \quad (9)$$

and will be used as a basis of the divergence criterion for multiple comparisons of separate hypotheses.

3. The Multiple Divergence Criterion

The multiple divergence criterion for testing an hypothesis H_i against $M-1$ alternatives H_k , $k \neq i$, is given by

$$S_i = J^* V^{-1} J^* \quad (10)$$

where

$$(a) \quad J^* = \hat{J} - \text{plim}_{\hat{\alpha}_i} \hat{J} = (J_{ik}^*)_{k \neq i} \quad (11)$$

$$(b) \quad V = (v_{kl})_{k, l \neq i} \text{ is a variance-covariance matrix defined by}$$

$$v_{kl} = \eta_k' \left\{ C_i(\hat{\alpha}_k, \hat{\alpha}_l) - \left[\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right]' V_i(\hat{\alpha}_i) \left[\frac{\partial \alpha_{il}}{\partial \alpha_i} \right] \right\} \eta_l \quad (12)$$

and

$$(c) \quad C_i, V_i \text{ represent the covariance and variance matrices under } H_i.$$

Theorem: Asymptotically, S_i is distributed as χ^2_{M-1} under H_i .

Proof: Sawyer (1981a) shows that under H_i , J_{ik}^* is asymptotically normally distributed with mean zero and variance

$$\eta_k' \left\{ v_i(\hat{\alpha}_k) - \left[\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right]' v_i(\hat{\alpha}_i) \left[\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right] \right\} \eta_i. \quad (13)$$

Furthermore, from (5) we note that

$$\begin{aligned} c_i(J_{ik}^*, J_{il}^*) &= \eta_k' c_i \left\{ (\hat{\alpha}_k - \alpha_{ik}) - \left[\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right]' (\hat{\alpha}_i - \alpha_i), (\hat{\alpha}_l - \alpha_{il}) - \left[\frac{\partial \alpha_{il}}{\partial \alpha_i} \right]' (\hat{\alpha}_i - \alpha_i) \right\} \eta_l \\ &= \eta_k' \left\{ c_i(\hat{\alpha}_k, \hat{\alpha}_l) - \left[\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right]' c_i(\hat{\alpha}_i, \hat{\alpha}_l) - c_i(\hat{\alpha}_k, \hat{\alpha}_i) \left[\frac{\partial \alpha_{il}}{\partial \alpha_i} \right] + \left[\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right]' v_i(\hat{\alpha}_i) \left[\frac{\partial \alpha_{il}}{\partial \alpha_i} \right] \right\} \eta_l. \end{aligned} \quad (14)$$

But asymptotically, Cox (1961, p. 112) shows that

$$c_i(\hat{\alpha}_i, \hat{\alpha}_l) = v_i(\hat{\alpha}_i) \left[\frac{\partial \alpha_{il}}{\partial \alpha_i} \right] \quad (15)$$

$$c_i(\hat{\alpha}_k, \hat{\alpha}_i) = \left[\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right] v_i(\hat{\alpha}_i). \quad (16)$$

Hence,

$$c_i(J_{ik}^*, J_{il}^*) = \eta_k' \left\{ c_i(\hat{\alpha}_k, \hat{\alpha}_l) - \left[\frac{\partial \alpha_{ik}}{\partial \alpha_i} \right]' v_i(\hat{\alpha}_i) \left[\frac{\partial \alpha_{il}}{\partial \alpha_i} \right] \right\} \eta_l \quad (17)$$

and J^* is asymptotically normally distributed $N_{M-1}(0, V)$ under H_i where V is as defined in (10).

Using a result of Rao (1973, p. 524) under H_i , S_i is asymptotically χ^2_{M-1} .

4. Application to Linear Gaussian Models

Consider M linear Gaussian models

$$H_i: y = X_i \beta_i + u_i \quad u_i \sim N(0, \sigma_i^2 I_T) \quad i=1, \dots, M \quad (18)$$

where X_i is non-stochastic or distributed independently of u_i and $\text{plim } \frac{1}{T} X_i' X_j$ exists and for $i \neq j$ is non-singular.

If $\hat{\beta}_i, \hat{\sigma}_i^2$ are the maximum likelihood estimates of β_i, σ_i^2 and $M_i = I - X_i (X_i' X_i)^{-1} X_i'$ is the residual sum of squares matrix of H_i , the loglikelihood of H_i can be written

$$L_i = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma_i^2 - \frac{1}{2\sigma_i^2} (y - X_i \beta_i)' (y - X_i \beta_i). \quad (19)$$

It is simple to deduce that $\beta_{ik}, \sigma_{ik}^2$ as defined by (6) are

$$\beta_{ik} = \text{plim} \left(\frac{X_k' X_k}{T} \right)^{-1} \left(\frac{X_k' X_i}{T} \right) \beta_i \quad (20)$$

$$\sigma_{ik}^2 = \sigma_i^2 + \text{plim } \beta_i' X_i' M_i X_i \beta_i / T \quad (21)$$

and are estimated by the maximum likelihood estimates.

Sawyer (1981a) showed that for linear Gaussian models

$$J_{ik}^* = \frac{1}{2} \log \left\{ \frac{\hat{\sigma}_k^2}{\hat{\sigma}_{ki}^2 \hat{\sigma}_{ik}^2} (\hat{\sigma}_{ik}^2 + \text{plim } \hat{\beta}_i' X_i' M_i M_i X_i \hat{\beta}_i / T) \right\}. \quad (22)$$

The following results determine the covariance matrix of J^* .

$$\text{plim } T C_i(\hat{\beta}_k, \hat{\beta}_l) = \sigma_i^2 \text{plim} \left(\frac{X_k' X_k}{T} \right)^{-1} \left(\frac{X_k' X_l}{T} \right) \left(\frac{X_l' X_l}{T} \right)^{-1} \quad (23)$$

$$\text{plim } T C_i(\hat{\beta}_k, \hat{\sigma}_l^2) = 2\sigma_i^2 \text{plim} \left(\frac{X_k' X_k}{T} \right) \left(\frac{X_k' M_l X_i \beta_i}{T} \right) \quad (24)$$

$$\text{plim } T C_i(\hat{\sigma}_k^2, \hat{\sigma}_l^2) = 4\sigma_i^2 \text{plim} \frac{\beta_i' X_i' M_l M_l X_i \beta_i}{T} + 2\sigma_i^4 \quad (25)$$

$$\frac{\partial \beta_{ik}}{\partial \beta_i} = \text{plim} \left(\frac{X_i' X_k}{T} \right) \left(\frac{X_k' X_k}{T} \right)^{-1} \quad (26)$$

$$\frac{\partial \beta_{ik}}{\partial \sigma_i^2} = 0 \quad (27)$$

$$\frac{\partial \sigma_{ik}^2}{\partial \beta_i} = 2 \text{plim} \frac{X_i' M_k X_i \beta_i}{T} \quad (28)$$

$$\frac{\partial \sigma_{ik}^2}{\partial \sigma_i^2} = 1. \quad (29)$$

η_k , as defined by (7) is given by

$$\eta_k = \frac{1}{\sigma_{ki}^2} \left\{ \begin{array}{l} - \text{plim } X_k' M_i X_k \beta_k / T \\ \frac{1}{2\sigma_k^2} \text{plim } \beta_k' X_k' M_i X_k \beta_k / T \end{array} \right\}. \quad (30)$$

Using (12) and (23)-(30), we see that the elements of the covariance matrix of J^* are

$$v_{kl} = \text{plim } \frac{1}{T} \frac{\sigma_i^2}{\sigma_{ki}^2 \sigma_{li}^2} \left\{ \begin{array}{l} \beta_k' X_k' M_i (I - M_k) M_i (I - M_l) M_i X_l \beta_l / T \\ - \frac{(\beta_k' X_k' M_i X_k \beta_k / T)}{\sigma_k^2} \beta_i' X_i' M_i M_i (I - M_l) M_i X_l \beta_l / T \\ - \frac{(\beta_l' X_l' M_i X_l \beta_l / T)}{\sigma_l^2} \beta_i' X_i' M_i M_i (I - M_k) M_i X_k \beta_k / T \\ + \frac{(\beta_k' X_k' M_i X_k \beta_k / T) (\beta_l' X_l' M_i X_l \beta_l / T)}{\sigma_k^2 \sigma_l^2} \beta_i' X_i' M_i M_i M_i X_i \beta_i / T \end{array} \right\}. \quad (31)$$

v_{kl} is estimated using the maximum likelihood estimates $\hat{\beta}_i, \hat{\beta}_k, \hat{\beta}_l, \hat{\sigma}_i^2, \hat{\sigma}_k^2, \hat{\sigma}_l^2$.

The divergence vector, given by (22), normalised by its covariance matrix as defined in (31) is then tested against a χ^2 random variable with $M-1$ degrees of freedom, a significant value indicating rejection of H_i . The procedure is repeated for each hypothesis as maintained.

5. Discussion

The multiple divergence criterion proposed above is intuitively appealing since it is based on a well-defined distance measure between an hypothesis and several alternatives. The statistic can be applied to likelihoods which are separate in the sense of distinct error distributions such as the lognormal, exponential and Gaussian distributions. It can also be used to test for functional form, for example across a power family of transformations. Future research should be directed to these questions, as well as the obvious comparisons with the other multiple procedures cited above.

References

- Cox, D.R. (1961). "Tests of Separate Families of Hypotheses", in
 Proceedings of the Fourth Berkeley Symposium on Mathematical
 Statistics and Probability, Vol. 1. University of California Press,
 Berkeley. 105-123.
- _____ (1962). "Further Results on Tests of Separate Families of
 Hypotheses." *Journal of the Royal Statistical Society, Series B*,
 24, 406-424.
- Davidson, R. and J.G. Mackinnon (1979a). "Several Tests for Model Specification
 in the Presence of Alternative Hypotheses." Discussion Paper No. 336,
 Department of Economics, Queen's University.
- _____ (1979b). "Relations Among Some Non-Nested
 Hypothesis Tests." Discussion Paper No. 369, Department of Economics,
 Queen's University.
- Jeffreys, H. (1948). *Theory of Probability*. 2nd edition. Oxford
 University Press, London.
- Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*.
 2nd edition. Wiley, New York.
- Sawyer, K.R. (1981a). "Testing Separate Families of Hypotheses: An
 Information Criterion." Discussion Paper No. 27, Centre for
 Econometrics and Decision Sciences, University of Florida.
- _____ (1981b). "Multiple Model Selection." Discussion Paper No. 32,
 Centre for Econometrics and Decision Sciences, University of
 Florida.