5 shows that there are only four mutually independent events. Hence any 5 of the 10 pairwise independent events are statistically dependent.

If the experiment consists of drawing n cards with replacement from a 52-card deck, then $N = 52^n = 2^{2n} \times 13^n$. There are only $3n$ mutually independent events, but there are $52^{n/2} + 1$ (for n even) pairwise independent events by Theorem 4 with $k = 52^{n/2}$. If one selects a random subset of 5 cards, then $N = \binom{52}{5} = (4 \times 7)^2 \times 3 \times 5 \times 13 \times 17$. Hence there are at least 33 pairwise independent events, but only 10 mutually independent events. Even the existence of 10 mutually independent events is surprising, since the selections are without replacement. The next example is related to this phenomenon.

If one takes a random sample of size n from a population of size r, then the selections are dependent. But $N = r \cdots (r - n + 1)$ implies that there are at least n mutually independent events, and in fact many more, since many of the factors of N must be composite. This can be understood by noting that with a relabeling of the outcomes, this experiment can be thought of as selecting n independent values, one from the set $\{1, 2, \ldots, r\}$, one from $\{1, 2, \ldots, r - 1\}$, and so forth.

We next consider Bernstein's example and a natural generalization. This well-known example goes as follows: Toss a fair coin two times. Let A_1 = "the first toss lands heads," A_2 = "the second toss lands heads," and A_3 = "an even number of tosses land heads." Here any two of A_1, A_2, A_3 are independent, but they are not mutually independent. The following generalizations extend this idea. Consider the experiment of tossing n fair coins. Let A_i = "the ith toss lands heads" for $i = 1, \ldots, n$. Let A_{n+1} = "the number of heads is even." It is well known and follows easily by mathematical induction that, for a fair coin, the probability of an even number of heads is one-half. It follows from this that any subset of n of the A_i ($i = 1, \ldots, n + 1$) are independent, whereas $A_1, \ldots, A_{n+1}$ are dependent. The same conclusion holds for the experiment of selecting n numbers at random with replacement from the set $\{0, 1, \ldots, p - 1\}$, for a prime p. Here A_i = "the ith number selected is zero," for $i = 1, \ldots, n$, and A_{n+1} = "the sum of the numbers is divisible by p." In this case $N = p^n$. Again the prime factorization of N is playing a role. For suppose that N were the product of n prime factors not all of which are the same. Then by the Corollary to Theorem 5 there would not exist $n + 1$ events any n of which are mutually independent. Wong (1972) gave a different example of A_i ($i = 1, \ldots, n + 1$) such that any subset of n are mutually independent, whereas $A_1, \ldots, A_{n+1}$ are dependent. It is interesting to note that in his example $N = 2^n$, the nth power of a prime.

The examples in this section illustrate that pairwise independent events that are not mutually independent occur in many common situations. We do not claim, however, that the events considered are of practical interest. Thus Feller's statement still stands. It would be interesting to find an example that contradicts it.

[*Received June 1986. Revised September 1986.*]

REFERENCES

Feller, W. (1950), *An Introduction to Probability Theory and Its Applications* (Vol. 1), New York: John Wiley.

LeVeque, W. (1977), *Fundamentals of Number Theory*, Reading, MA: Addison-Wesley.

Neuts, M. F. (1973), *Probability*, Boston: Allyn & Bacon.

Wong, C. K. (1972), "A Note on Mutually Independent Events," *The American Statistician*, 26, 27–38.

Another Look at Some Results on the Recursive Estimation in the General Linear Model

SIDDHARTHA CHIB, S. RAO JAMMALAMADAKA, and RAM C. TIWARI*

Written mainly for its pedagogical interest, this note deals with the computational formulas for the recursive updating of weighted least squares parameter estimates and the residual sum of squares in the general linear model under the assumption that the errors have a multivariate normal distribution. This approach simplifies considerably the derivations of Haslett (1985).

KEY WORDS: Multivariate normal; Weighted least squares; Maximum likelihood estimators; Score function.

*Siddhartha Chib is Assistant Professor, Department of Economics, University of Missouri, Columbia, MO 65211. S. Rao Jammalamadaka is Professor, Statistics and Applied Probability Program, University of California, Santa Barbara, CA 93106. Ram C. Tiwari is Assistant Professor, Department of Mathematics, University of North Carolina, Charlotte, NC 28223.

1. INTRODUCTION

In a recent paper, Haslett (1985) derived computational formulas for the recursive updating of weighted least squares (WLS) parameter estimates and the residual sum of squares in the general linear model when more than one additional observation becomes available. These results generalize the updating expressions that were derived by Plackett (1950) for ordinary least squares estimators and by McGilchrist and Sandland (1979) for WLS estimators with only one additional observation.

In this article the results of Haslett (1985) are derived in a much simpler way by assuming that the errors have a multivariate normal distribution. We exploit the fact that under this assumption, WLS and maximum likelihood (ML) estimators are identical. This allows us to develop the recursions through a suitable Taylor series expansion of the score function (i.e., the first derivative of the log-likelihood). The method we propose is similar in some ways to the Newton–Raphson procedure of finding the roots of the likelihood equations and does not seem to have been exploited in the context of recursive updating. This approach circumvents the heavy algebra that has been necessary in all of the previous derivations.

2. PRELIMINARIES

Consider the linear model

$$Y_h = X_h \beta + e_h,$$

where $Y_h = (Y(1)', \ldots, Y(h)')'$, $X_h = (X(1)', \ldots, X(h)')'$, $e_h = (e(1)', \ldots, e(h)')'$, $Y(i)$ and $e(i)$ are $(n_i \times 1)$ random vectors, $X(i)$ is an $(n_i \times k)$ nonrandom matrix, and β is a $(k \times 1)$ unknown parameter vector. Let $N_h = \Sigma_{i=1}^{h} n_i$. Let $y(h+1)$ denote an incoming group of observations. Assume that $e_{h+1} \sim N(0, \Sigma_{h+1})$, where

$$\Sigma_{h+1} = \begin{bmatrix} \Sigma_j & C_h \\ - & \Sigma_{h+1,h+1} \end{bmatrix}$$

is a symmetric nonsingular matrix with $\Sigma_h = E\, e_h e_h'$, $C_h = E e_h e(h+1)'$, and $\Sigma_{h+1,h+1} = Ee(h+1)e(h+1)'$. Assume that Σ_{h+1} is completely known and that X is of full rank. Define

$$\mu_{2.1} = X(h+1)\beta + C_h' \Sigma_h^{-1}(y_h - X_h\beta),$$

$$\Sigma_{2.1} = \Sigma_{h+1,h+1} - C_h' \Sigma_h^{-1} C_h,$$

$$P = X(h+1) - C_h' \Sigma_h^{-1} X_h,$$

$$\Omega = \Sigma_{2.1} + P(X_h' \Sigma_h^{-1} X_h)^{-1} P'.$$

The pdf of Y_h is

$$f(y_h; \beta)\ \alpha\ \exp\{-\tfrac{1}{2}((y_h - X_h\beta)' \Sigma_h^{-1}(y_h - X_h\beta))\}, \tag{2.1}$$

where α is the proportionality symbol, and the conditional pdf of $Y(h+1)$ given y_h is

$$f(y(h+1)|y_h, \beta)$$
$$\alpha\ \exp\{-\tfrac{1}{2}((y(h+1) - \mu_{2.1})' \Sigma_{2.1}^{-1}(y(h+1) - \mu_{2.1}))\}. \tag{2.2}$$

Hence

$$\frac{\partial^2 \ln f(y_h; \beta)}{\partial\beta\, \partial\beta'} = -(X_h' \Sigma_h^{-1} X_h), \tag{2.3}$$

$$\frac{\partial \ln f(y(h+1)|y_h, \beta)}{\partial\beta} = P' \Sigma_{2.1}^{-1}(y(h+1) - \mu_{2.1}), \tag{2.4}$$

$$\frac{\partial^2 \ln f(y(h+1)|y_h, \beta)}{\partial\beta\partial\beta'} = -P' \Sigma_{2.1}^{-1} P. \tag{2.5}$$

3. UPDATING FORMULAS

This section contains the two main results, Theorems 3.3 and 3.4, which provide the recursive equations for the estimates of β and the residual sum of squares. In each case we exploit the assumption of multivariate normality to simplify the derivations.

Proposition 3.1.

$$X_{h+1}' \Sigma_{h+1}^{-1} X_{h+1} = X_h' \Sigma_h^{-1} X_h + P' \Sigma_{2.1}^{-1} P. \tag{3.1}$$

Proof. Since the natural logarithm of the pdf of Y_{h+1} can be written as

$$\ln f(y_{h+1}; \beta) = \ln f(y_h; \beta) + \ln f(y(h+1)|y_h, \beta),$$

we have that the Hessians are connected by

$$\frac{\partial^2 \ln f(y_{h+1}; \beta)}{\partial\beta\, \partial\beta'} = \frac{\partial^2 \ln f(y_h; \beta)}{\partial\beta\, \partial\beta'} + \frac{\partial^2 \ln f(y(h+1)|y_h, \beta)}{\partial\beta\, \partial\beta'}.$$

Now the result follows using (2.3) and (2.5).

Using Proposition 3.1 alone, we obtain the following result that updates the variance–covariance matrix of the estimates of β. It should be noticed that the standard method for deriving such a result involves actually inverting the relevant matrices.

Proposition 3.2.

$$(X_{h+1}' \Sigma_{h+1}^{-1} X_{h+1})^{-1} = (X_h' \Sigma_h^{-1} X_h)^{-1}$$
$$- (X_h' \Sigma_h^{-1} X_h)^{-1} P' \Sigma_{2.1}^{-1} P (X_{h+1}' \Sigma_{h+1}^{-1} X_{h+1})^{-1}. \tag{3.2}$$

Proof. The proof follows by premultiplying and postmultiplying both sides of Equation (3.1) by $(X_h' \Sigma_h^{-1} X_h)^{-1}$ and $(X_{h+1}' \Sigma_{h+1}^{-1} X_{h+1})^{-1}$ and rearranging terms.

Remark 3.3. Sometimes the preceding recursion is stated in an alternative form, namely

$$(X_{h+1}' \Sigma_{h+1}^{-1} X_{h+1})^{-1} = (X_h' \Sigma_h^{-1} X_h)^{-1}$$
$$- (X_h' \Sigma_h^{-1} X_h)^{-1} P' \Omega^{-1} P (X_h' \Sigma_h^{-1} X_h)^{-1}. \tag{3.3}$$

[See Haslett (1985), p. 185.]

That (3.2) and (3.3) are identical is easy to establish. First note that

$$\Omega^{-1} = \Sigma_{2.1}^{-1} - \Sigma_{2.1}^{-1} P (X_{h+1}' \Sigma_{h+1}^{-1} X_{h+1})^{-1} P' \Sigma_{2.1}^{-1}, \tag{3.4}$$

which is proved by multiplying (3.4) by Ω and using Proposition 3.2. Now we only need to show that

$$P' \Sigma_{2.1}^{-1} P (X_{h+1}' \Sigma_{h+1}^{-1} X_{h+1})^{-1}$$
$$= P' \Omega^{-1} P (X_h' \Sigma_h^{-1} X_h)^{-1}. \tag{3.5}$$

Substituting for Ω^{-1} from (3.4) and using (3.2) yields the equality. The next result gives the updating expression for the WLS estimates of β, which, under normality, are also the ML estimates. This allows us to develop the recursions through a Taylor series of the score function based on $f(y_{h+l}; \beta)$, and evaluated at $\hat{\beta}_{h+1}$, around $\hat{\beta}_h$. By this approach much of the algebra that characterized earlier derivations is eliminated.

Theorem 3.4. Let $\hat{\beta}_h = (X_h' \Sigma_h^{-1} X_h)^{-1} X_h' \Sigma_h^{-1} y_h$ be the ML estimator of β corresponding to y_h based on groups 1 to h. Then the recursive estimator for $\hat{\beta}_h$ is given by

$$\hat{\beta}_{h+1} = \hat{\beta}_h + (X_h'\Sigma_h^{-1}X_h)^{-1}P'\Omega^{-1}[y(h+1) - \hat{\mu}_{2.1}],$$

where

$$\hat{\mu}_{2.1} = X(h+1)\hat{\beta}_h + C_h'\Sigma_h^{-1}(y_h - X_h\hat{\beta}_h).$$

Proof. Expanding $\partial/\partial\beta \ln f(y_{h+1}; \hat{\beta}_{h+1})$ around $\hat{\beta}_h$ in a Taylor series and noticing that the pdf is quadratic in β yields

$$\frac{\partial}{\partial\beta}\ln f(y_{h+1}; \hat{\beta}_{h+1}) = 0 = \frac{\partial}{\partial\beta}\ln f(y_{h+1}; \hat{\beta}_h) + \frac{\partial^2}{\partial\beta\,\partial\beta'}\ln f(y_{h+1}; \hat{\beta}_h)(\hat{\beta}_{h+1} - \hat{\beta}_h).$$

Now since the first term on the right side is $(\partial/\partial\beta)\ln f(y(h+1)|y_h, \hat{\beta}_h)$, which equals $P'\Sigma_{2.1}^{-1}(y(h+1) - \hat{\mu}_{2.1})$ because of (2.4), and the second derivative term equals $(X_{h+1}'\Sigma_{h+1}^{-1}X_{h+1})$, we get

$$\hat{\beta}_{h+1} = \hat{\beta}_h + (X_{h+1}'\Sigma_{h+1}^{-1}X_{h+1})^{-1}P' \times \Sigma_{2.1}^{-1}[y(h+1) - \hat{\mu}_{2.1}]. \quad (3.6)$$

The result follows on applying (3.5).

As a consequence of the recursion formulas we can derive the updating expression for the residual sum of squares

$$\text{SSE}_h(\hat{\beta}_h) = (y_h - X_h\hat{\beta}_h)'\Sigma_h^{-1}(y_h - X_h\hat{\beta}_h),$$

where the subscript h indicates the number of groupings involved.

Theorem 3.5. If

$$\text{SSE}_h(\hat{\beta}_h) = (y_h - X_h\hat{\beta}_h)'\Sigma_h^{-1}(y_h - X_h\hat{\beta}_h)$$

is the residual sum of squares, then

$$\text{SSE}_{h+1}(\hat{\beta}_{h+1}) = \text{SSE}_h(\hat{\beta}_h) + [y(h+1) - \hat{\mu}_{2.1}]'\Omega^{-1}[y(h+1) - \hat{\mu}_{2.1}].$$

Proof. Expanding $\text{SSE}_{h+1}(\hat{\beta}_{h+1})$ in a Taylor series around $\hat{\beta}_h$ and collecting the second and third terms of the expansion, we get

$$\text{SSE}_{h+1}(\hat{\beta}_{h+1}) = \text{SSE}_{h+1}(\hat{\beta}_h) - (\hat{\beta}_{h+1} - \hat{\beta}_h)'(X_{h+1}'\Sigma_{h+1}^{-1}X_{h+1})(\hat{\beta}_{h+1} - \hat{\beta}_h). \quad (3.7)$$

Equating the exponents in

$$f(y_{h+1}; \beta) = f(y_h; \beta) \cdot f(y(h+1)|y_h, \beta),$$

we have that for all β, including $\hat{\beta}_h$,

$$(y_{h+1} - X_{h+1}\beta)'\Sigma_{h+1}^{-1}(y_{h+1} - X_{h+1}\beta) = (y_h - X_h\beta)'\Sigma_h^{-1}(y_h - X_h\beta) + [y(h+1) - \mu_{2.1}]'\Sigma_{2.1}^{-1}[y(h+1) - \mu_{2.1}]. \quad (3.8)$$

Evaluating (3.8) at $\hat{\beta}_h$ and using (3.6) and (3.7) gives

$$\text{SSE}_{h+1}(\hat{\beta}_{h+1}) = \text{SSE}_h(\hat{\beta}_h) + (y(h+1) - \hat{\mu}_{2.1})' \times [\Sigma_{2.1}^{-1} - \Sigma_{2.1}^{-1}P(X_{h+1}'\Sigma_{h+1}^{-1}X_{h+1})^{-1}P'\Sigma_{2.1}^{-1}] \times (y(h+1) - \hat{\mu}_{2.1}).$$

From (3.4) the matrix in square brackets is Ω^{-1} and hence the result is proved.

[*Received June 1986.*]

REFERENCES

Haslett, S. (1985), "Recursive Estimation of the General Linear Model With Dependent Errors and Multiple Additional Observations," *Australian Journal of Statistics*, 27, No. 2, 183–188.

McGilchrist, C. A., and Sandland, R. L. (1979), "Recursive Estimation of the General Linear Model With Dependent Errors," *Journal of the Royal Statistical Society*, Ser. B, 41, 65–68.

Plackett, R. L. (1950), "Some Theorems in Least Squares," *Biometrika*, 37, 149–157.

Accent on Teaching Materials

HARRY O. POSTEN, Section Editor

In this section *The American Statistician* publishes announcements and selected reviews of teaching materials of general use to the statistical field. These may include (but will not necessarily be restricted to) curriculum material, collections of teaching examples or case studies, modular instructural material, transparency sets, films, filmstrips, videotapes, probability devices, audiotapes, slides, and data deck sets (with complete documentation).

Authors, producers, or distributors wishing to have such materials announced or reviewed should submit a single, complete copy of the product (three copies of printed material double-spaced) to Associate Editor Harry O. Posten, Statistics Department, University of Connecticut, Storrs, CT 06268. A statement of intention that the material will be available to all requestors for a minimum of a two-year period should be provided, along with information on the cost (including postage) and special features of the material. Information on classroom experience may also be included. All materials submitted must be of general use for teaching purposes in the area of probability and statistics.

Exploring Statistics With the IBM PC (Version 85.0). David P. Doane. Reading, MA: Addison-Wesley, 1985. Softbound manual, 1 program diskette, 1 database diskette. $39.95.

The heart of this statistical package is a single diskette containing a set of programs called the EXPLORE programs. These programs are designed to perform most of the statistical procedures used in a first or second course in statistical methods and are designed to be used with no computer background and negligible computer learning. The system operates under any of the disk operating systems DOS 1.0, 1.1, or 2.0.

The EXPLORE system is entirely menu or dialogue driven. The user need only answer questions or select from a menu on the