SELECTION INDEX ESTIMATION FROM PARTIAL MULTIVARIATE NORMAL DATA

W. B. SMITH
AND
R. C. PFAFFENBERGER

Reprinted from
BIOMETRICS
THE BIOMETRIC SOCIETY, Vol. 26, No. 4, December 1970

Reprinted from
BIOMETRICS
THE BIOMETRIC SOCIETY, Vol. 26, No. 4, December 1970

SELECTION INDEX ESTIMATION FROM PARTIAL MULTIVARIATE NORMAL DATA

W. B. SMITH AND R. C. PFAFFENBERGER
Institute of Statistics, Texas A&M University, College Station, Texas 77843, U. S. A.

SUMMARY

A technique is applied to the estimation of the selection index, when incomplete multivariate normal data records are available. The procedure utilizes all data available, both full and partial vectors, and represents an improvement in precision of the estimator over that found by using the full data vectors only. Estimates of phenotypic means and variance-covariance matrix also result from the procedure.

Individuals with partial data records are indexed by an extension of the technique, and this index is contrasted to that proposed by Henderson. The results of computer simulations are tabulated.

1. INTRODUCTION

The purpose of this research is to estimate the selection index from multivariate normal data which may have some observation vectors with missing elements. The selection index is a well-known procedure for grading individuals in large groups. For a thorough review of relevant work on this topic see Williams [1962]. Briefly we may note that Fairfield Smith [1936] originally developed the selection index in order to discriminate in selection programs among varieties of plants. Hazel [1943] initially applied the procedure to animal breeding programs. An extension of the technique was developed by Kempthorne and Nordskog [1959] in which the breeder could hold genetic values constant for some traits in the index as selection progressed. Henderson [1963] delineated the selection index procedure in the case where different information on individuals to be indexed was available (the phenotypic means assumed known), and indicated an alternative method for dealing with unequal information under these conditions. Williams [1962] devised a method for evaluating and comparing the four different indices he defined, giving particular attention to the estimation of the phenotypic and genotypic variance-covariance matrices in the commonly used estimated index.

Defining

$$I_i = \mathbf{b}'\mathbf{x}_i \,, \tag{1}$$

where I_i is a composite index value associated with the jth member of the population, $\mathbf{b}$ is an $n \times 1$ vector of unknown coefficients, and $\mathbf{x}_i$ is an $n \times 1$ vector of phenotypic values (observations) on the jth member of the population, Smith [1936] shows that

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\alpha. \tag{2}$$

In this context $\mathbf{P}$ is the $n \times n$ covariance matrix of phenotypic values, $\mathbf{G}$ is the $n \times n$ matrix of genotypic values, and $\boldsymbol{\alpha}$ is the $n \times 1$ vector of economic weights associated with the n traits.

The problem of estimating $\mathbf{b}$ has been attacked by many authors notably Henderson [1963]. This study estimates $\mathbf{b}$ using all information available (i.e. both full and incomplete data records) by estimating the elements of the matrix $\mathbf{P}$, assuming both $\mathbf{G}$ and $\boldsymbol{\alpha}$ known. The method of attack will be to apply and extend the technique of Hocking and Smith [1968]. This procedure was developed to estimate the parameters of a multivariate normal distribution in the presence of partial data, and the technique has been shown to lead to estimates which are consistent and asymptotically efficient. The method can be summarized as follows:

(1) Divide the data into groups according to which phenotypic values are missing.

(2) Within each group find the best estimate of all possible parameters.

(3) Sequentially adjoin, optimally, the information in each partial data group to the full data group.

In summary the above technique is applied to the selection index problem; the resulting estimates of $\mathbf{b}$, $\mathbf{P}$, and the vector, $\boldsymbol{\mu}$, of the phenotypic means are considerably 'better' than those using only the full data vectors. Extended Monte Carlo simulations are used to support this claim. In addition, the procedure is used in indexing individuals with partial data records, and contrasts with the method of Henderson [1963] are made.

2. ESTIMATION PROCEDURE

Assuming the $\mathbf{x}_i$ vectors are distributed as multivariate normal, that is, $\mathbf{x}_i \sim N_n(\boldsymbol{\mu}, \mathbf{P})$, we will now give a general procedure which utilizes all information arising from both complete and incomplete $\mathbf{x}_i$ vectors. As examples of the procedure consider two different cases: (1) some of the $\mathbf{x}_i$ vectors contain complete information on all n traits and (2) none of the $\mathbf{x}_i$ vectors contain complete information. In both cases, the improvement in the estimation of $\mathbf{b}$ is induced by improving the estimates of p_{ij}, the elements of the phenotypic covariance matrix, $\mathbf{P}$. Thus, in the process of estimating $\mathbf{b}$, improved estimates of the phenotypic variances and covariances follow directly. In addition, improved estimates of the phenotypic vector of means, $\boldsymbol{\mu}$, may be computed by the application of the procedure used to estimate the p_{ij}.

The procedure used to estimate p_{ij} optimally was developed by Hocking and Smith [1968] and will be briefly described by the examples below. The procedure is sequential, not iterative. These examples are given only to illustrate the technique; the general formulation is given in the Appendix.

2.1. *Case* 1—*Some complete* $\mathbf{x}_i$*-vectors*

We consider two examples, both involving 3 traits, x_1, x_2, and x_3. In Example 1, assume that we have n_1 complete vectors, n_2 vectors with information on x_1 and x_2, and n_3 vectors with information only on x_1. In Example

2, n_1 and n_2 are unchanged but we have n_3 vectors with information on x_2 and x_3 . Schematically, we have:

Example 1

Traits	Number of observations n_1	n_2	n_3
x_1	X	X	X
x_2	X	X	
x_3	X		

Example 2

Traits	Number of observations n_1	n_2	n_3
x_1	X	X	
x_2	X	X	X
x_3	X		X

The procedure will be described only for estimates of the phenotypic means, μ_1 , μ_2 , and μ_3 in Example 1. Generalizations to the p_{ij} estimates follow the same pattern.

The first step is to adjoin to the n_1 vectors of x_1 , x_2 , and x_3 the n_2 vectors on x_1 and x_2 as follows:

$$ {}_1\tilde{\mu}_i = {}_1\hat{\mu}_i + a_{1i}({}_1\hat{\mu}_1 - {}_2\hat{\mu}_1) + a_{2i}({}_1\hat{\mu}_2 - {}_2\hat{\mu}_2) \qquad i = 1, 2, 3, \tag{3} $$

where $\sim$ indicates the improved estimate and ${}_i\hat{\mu}_j$ denotes the estimate of μ_j obtained from the n_i data only.

Notice that if a_{1i} and a_{2i} are constants,

$$ E[{}_1\tilde{\mu}_i] = E[{}_1\hat{\mu}_i] = \mu_i \, . \tag{4} $$

In order to find the coefficients a_{1i} and a_{2i} , the variance of ${}_1\mu_i$ is minimized with respect to them. The results are:

$$ i = 1, \qquad a_{1i} = -n_2/(n_1 + n_2) \qquad a_{2i} = 0; \tag{5} $$

$$ i = 2, \qquad a_{1i} = 0 \qquad a_{2i} = -n_2/(n_1 + n_2); \tag{6} $$

$$ \begin{aligned} i = 3, \qquad a_{1i} &= -n_2(p_{13}p_{22} - p_{12}p_{23})/(n_1 + n_2)(p_{11}p_{22} - p_{12}^2), \\ a_{2i} &= -n_2(p_{23}p_{21} - p_{12}p_{13})/(n_1 + n_2)(p_{11}p_{22} - p_{12}^2). \end{aligned} \tag{7} $$

Notice that ${}_1\tilde{\mu}_1$ and ${}_1\tilde{\mu}_2$ are the simple weighted means as one would expect.

The n_3 $\mathbf{x}_1$-vectors are incorporated as follows:

$$ {}_1\tilde{\mu}_i = {}_1\tilde{\mu}_i + a_i^*({}_1\tilde{\mu}_1 - {}_3\hat{\mu}_1). \tag{8} $$

As before, a_i^* arises from minimizing the variance of ${}_1\tilde{\mu}_i$. The general procedure to estimate the μ_i or p_{ij} is:

(i) For each type of vector n_i , compute the appropriate statistics as if n_i were a complete sample.

(ii) Adjoin sequentially to the complete vector n_1 the n_i , forming the linear combinations as illustrated above. In each step, the adjoined differences must have expectation zero, given that a_{1i} , a_{2i} , a_i^* are constants.

In general, the coefficients of the differences in the linear combinations are functions of the parameters (e.g. a_{1i} and a_{2i} in Example 1 for $i = 3$). Consequently, the parameters in these coefficients must be estimated. Hocking and Smith [1968] have shown that in the case of nested data the estimators, $\tilde{\mu}_i$ and $\tilde{p}_{ij}$, and the estimators, $\tilde{\mu}_i$ and $\tilde{p}_{ij}$, are invariant to whichever estimates of the parameters are selected in the coefficients; i.e. $\hat{\mu}_i$ and $\hat{p}_{ij}$, $\tilde{\mu}_i$ and $\tilde{p}_{ij}$, or $\tilde{\mu}_i$ and $\tilde{p}_{ij}$. In (7), for example, the $\hat{}$, $\sim$, and $\approx$ estimators could be used without altering the estimate given by (4). In general, resulting estimators are consistent and for small sample sizes, the bias appears to be slight (see Hocking and Smith [1968] p. 171).

2.2. *Case 2—No complete* $\mathbf{x}_i$*-vectors*

We consider only one example of this type: n_1 vectors with information on x_1 and x_2 , and n_2 vectors with information on x_1 and x_3 . Schematically, we have:

Example 3

Traits	Number of observations n_1	n_2
x_1	X	X
x_2	X	
x_3		X

For this procedure, an independent estimate of p_{13} is required since the data in the above configuration will not supply an estimate of this parameter. The procedure for deriving the estimators for Case 2 problems is indicated via Example 5 in the Appendix.

It is interesting to note, though somewhat difficult to prove in general, that, in both Cases 1 and 2, as the data sets n_i are adjoined sequentially to n_1 , the estimates of the b_i for the current sequence are the same as the $\hat{b}_i$ in the previous sequence for all those traits which are not common to n_i and n_{i-1} . That is, in Example 1, $\hat{b}_3 = \tilde{b}_3 = \approx{b}_3$ and $\tilde{b}_2 = \approx{b}_2$, and in Example 2, $\tilde{b}_1 = \approx{b}_1$ and $\hat{b}_3 = \tilde{b}_3$.

3. MONTE CARLO RESULTS AND THE INCOMPLETE INFORMATION CASE

3.1. *Demonstration of improved index estimates*

Hocking and Smith [1968] have shown analytically and by Monte Carlo studies that the estimates of the variances and covariances arising from the procedures described above using the incomplete data vectors are substantially more precise than the estimates based on the complete data only.

It will be shown that the same can be said about the estimation of $\mathbf{b}$ and, hence, the selection index I.

In the following examples, the means were assumed to be zero without loss of generality. In each example, 100 Monte Carlo trials were executed and the average estimate as well as the mean square error of the estimates were computed. Further, the generalized variance, $|\mathbf{V}(\hat{\mathbf{b}})|$, of the estimates was determined, $\mathbf{V}(\hat{\mathbf{b}})$ being the sample covariance matrix of the $\hat{\mathbf{b}}$-vector. For each estimate, the Cramér–Rao minimum variance bound (Smith and Hocking [1968]) was computed to give comparison with the mean square error.

In both examples, $\alpha' = (10, 10, 10)$; that is, equal economic weights were chosen for each trait. Experience with a variety of economic weight vectors α has shown that the efficiency of the estimate of $\mathbf{b}$ is dependent to some extent on the relative magnitude of the weights. That is, if the heaviest weight is given to the trait that has the least number of observations on it, the estimate is considerably more efficient than the complete data vector only estimate. However, a heavy weight on a trait that is observed in most, if not all, data vectors produces an estimate that is only slightly more efficient than the complete data estimate.

In Example 1, $n_1 = 100$ vectors were taken from $N_3(0, {}_3\mathbf{P})$, $n_2 = 50$ vectors from $N_2(0, {}_2\mathbf{P})$, and $n_3 = 25$ vectors from $N_1(0, {}_1\mathbf{P})$, where

$$ {}_3\mathbf{P} = \begin{bmatrix} 7 & 2 & 6 \\ 2 & 6 & 3.5 \\ 6 & 3.5 & 8 \end{bmatrix} \qquad {}_2\mathbf{P} = \begin{bmatrix} 7 & 2 \\ 2 & 6 \end{bmatrix} \qquad {}_1\mathbf{P} = [7]. $$

The results are given in Table 1, where data group 1 represents the complete data vector set. Notice that the bias, as well as the mean square error of the estimates, is measurably decreased as more information (n_2 , then n_3) is adjoined to n_1 . As remarked previously, if no further information is adjoined from one data group to the next on a specific trait, the b-coefficient associated with that trait, as well as its mean square error, remains unchanged. For instance, the average value of b_3 remains constant with the addition to n_1 of n_2 and then n_3 , since the n_2 and n_3 data vectors contain no information on x_3 .

In Example 2, $n_1 = 100$ vectors were taken from $N_3(0, {}_3\mathbf{P})$, $n_2 = 50$ vectors from $N_2(0, {}_2\mathbf{P})$, and $n_3 = 50$ vectors from $N_2(0, {}_4\mathbf{P})$, where

$$ {}_4\mathbf{P} = \begin{bmatrix} 6 & 3.5 \\ 3.5 & 8 \end{bmatrix}. $$

The results are given in Table 2. As in Example 1, the improvement in the estimation of $\mathbf{b}$ is apparent as the n_2 and n_3 information is adjoined to n_1 . The decrease in the generalized variance is even more pronounced than in Example 1. In both examples, $\alpha' = (1, 1, 1)$ and

$$ \mathbf{G} = \begin{bmatrix} 2 & 0.75 & 2 \\ 0.75 & 3 & 1.5 \\ 2 & 1.5 & 4 \end{bmatrix}. $$

TABLE 1

SUMMARY OF SIMULATION OF EXAMPLE 1 (100 RUNS)

Population value	Population parameter b_1 -2.486	b_2 4.058	b_3 9.464	Generalized variance
Data group 1 (n_1)				
Avg. Est.	-2.166	4.054	9.446	
MSE	4.032	2.169	4.092	14.949
CRLB	3.397	2.031	4.441	
Data groups 1 and 2 ($n_1 + n_2$)				
Avg. Est.	-2.262	4.005	9.446	
MSE	3.164	1.711	4.092	7.813
CRLB	3.006	1.490	4.441	
Data groups 1, 2, and 3 ($n_1 + n_2 + n_3$)				
Avg. Est.	-2.320	4.004	9.446	
MSE	2.903	1.711	4.092	6.687
CRLB	2.918	1.490	4.441	

MSE = Mean square error, CRLB = Cramér-Rao lower bound.

3.2. *Use of the improved index*

We now consider the task of evaluating an individual with only x_1 and x_2 information, say, given the index is trivariate, $I = b_1x_1 + b_2x_2 + b_3x_3$, and at least some of the $\mathbf{x}$-vectors observed contain complete information. Three methods have been considered for estimating the missing x_3 value:

(i) Use of $\bar{x}_3$ from the complete data set of n_1 observations.

(ii) Use of μ_3^*, where $\mu_3^* = \bar{\mu}_3 + d_1(x_1 - \bar{\mu}_1) + d_2(x_2 - \bar{\mu}_2)$ and d_1, d_2 arise through the minimization of the variance of μ_3^*.

(iii) Regression of x_3 on x_1 and x_2 in the complete data set; that is, $\hat{x}_3 = b_0' + b_1'x_1 + b_2'x_2$, where b_0', b_1', and b_2' are estimated by least squares.

Monte Carlo studies have indicated that method (iii) is superior to the others in terms of minimum mean square error. Our procedure for comparison among the methods is to generate a set of complete records, choose a subset in which one of the traits, say x_3, is discarded, use the above methods to estimate x_3 in this subset, compute the index values by using the $\tilde{\mathbf{b}}$-vector and the estimated x_3 values and, finally, compute the mean square errors of the indices arising from the use of the three methods. In Table 3, the respective mean square errors are given for an example where $n_1 = 100$ trivariate records are used with $n_2 = 50$ vectors on x_1 and x_2 for which the actual value of x_3 is known, but estimated by the above methods. For comparison, the mean square error using the full 150 complete vectors and the $\tilde{\mathbf{b}}$-estimates is

TABLE 2

SUMMARY OF SIMULATION OF EXAMPLE 2 (100 RUNS)

Population value	Population parameter b_1 -2.486	b_2 4.058	b_3 9.464	Generalized variance
Data group 1 (n_1)				
Avg. Est.	-2.388	4.467	9.482	
MSE	3.793	3.285	5.748	26.307
CRLB	3.397	2.031	4.441	
Data groups 1 and 2 ($n_1 + n_2$)				
Avg. Est.	-2.410	4.322	9.482	
MSE	3.341	2.356	5.748	11.645
CRLB	3.006	1.490	4.441	
Data groups 1, 2, and 3 ($n_1 + n_2 + n_3$)				
Avg. Est.	-2.408	4.226	9.468	
MSE	3.333	1.389	4.719	4.563
CRLB	2.982	1.111	3.779	

MSE = Mean square error, CRLB = Cramér-Rao lower bound.

given. Numerous studies have indicated that the regression technique is in general quite good. It has been found that if the variances are small in comparison with the means, the $\boldsymbol{\mu}^*$ estimator is usually the best. However, the regression technique is recommended since it is more easily used and can be extended to deal with two or more missing x_i values, as documented by Afifi and Elashoff [1967].

The philosophy behind estimating the missing phenotypic values is as follows: as much information as possible is drawn from the total set of data vectors resulting in optimal b-estimates before recourse is made to estimating the missing data. Hence, for individuals with complete data vectors, the

TABLE 3

SIMULATION RESULTS ON INDEXING PARTIAL DATA VECTORS
(ALL RESULTS USE $\tilde{\mathbf{b}}$ ESTIMATES OF b.)

	Mean square error (1000 Monte Carlo experiments)
True X_3	209.20
$\bar{X}_3$ Substitution	500.62
μ_3^* Substitution	394.34
Regression estimate	357.39

selection index arising from the b-estimates is an improvement over the index constructed only from the complete data set. For individuals with incomplete vectors, though index values are assigned by estimating the missing phenotypic values, the resulting index values are compensated for partially by the use of the optimal estimation of **P** and, hence, **b**. Further, by the use of this method of constructing the estimated selection index, improved estimates of the phenotypic variances and covariances follow immediately as pointed out previously.

3.3. *Henderson's method*

Henderson [1963] has given a procedure for constructing an index of form

$$I = b_1(y_1 - \mu_1) + b_2(y_2 - \mu_2) + \cdots + b_N(y_N - \mu_N), \tag{9}$$

where μ_i is the mean of the y_i phenotypic value, for a special case of the general missing data situation. In particular, he considers the case where the data vectors may be dichotomized into complete data subsets, so that for each subset an index can be computed in the usual manner. Let $I_A = \sum_{i=1}^{N} b_i(y_i - \mu_i)$ and $I_B = \sum_{i=1}^{N} b_i'(y_i - \mu_i)$ represent these two indices, where some, if not all, of the b_i and b_i' are zero. Now, Henderson considers the index

$$I_D = \sum_{i=1}^{N} (b_i - b_i')y_i = \mathbf{b}^{*\prime}\mathbf{y}$$

and shows that $\hat{\mathbf{b}}^* = \mathbf{P}^{-1}(\hat{\mathbf{t}}_A - \hat{\mathbf{t}}_B)$, where $\mathbf{t}_A$ and $\mathbf{t}_B$ are the covariance vectors of the phenotypic values $\mathbf{y}$ and the breeding values H_A and H_B, respectively (see Henderson [1963]).

It should be noted that Henderson, as well as most other workers in this field, has assumed **P**, **G**, and $\boldsymbol{\alpha}$ are known, whereas, this paper only assumes **G** and $\boldsymbol{\alpha}$ known. In some applications there is sufficient prior data to make the assumptions on **P**, **G**, and $\boldsymbol{\alpha}$ tenable, in which case the procedures of Henderson [1963] are optimal. Under these conditions Henderson's method will simultaneously give estimates of the mean $\boldsymbol{\mu}$ and the selection index.

Our procedure is designed expressly for the missing value situation where at least some of the records are complete and **P** is unknown. Though the same estimation procedure on **P** as described in the text could be used if some of the information is overlapping between the two groups, it is not recommended, for independent estimates of certain covariances will still be required.

4. CONCLUSIONS

Using all multivariate normal data records, both the full and partial vectors, a technique is developed for improved estimation on the selection index. In this process, improved estimates of the phenotypic mean vector and covariance matrix are immediate.

The procedure is extended to the indexing of partial data records, and when applicable, a contrast is made to the Henderson [1963] method. Monte Carlo simulation studies are tabulated.

The Appendix spells out in considerable detail the general formulation of the estimation procedure. Computer programs to achieve this estimation can be written easily using the matrix notation of the Appendix.

Several extensions of the above procedure are currently under investigation by the authors; in particular, a case of a mixed distribution (continuous and discrete variates) is currently being prepared for submission. In addition, a complete generalization of the procedure is being studied.

ACKNOWLEDGMENTS

The authors wish to thank the referees and the *Biometrics* editorial staff for their helpful comments and criticisms. During the course of this research the authors were partially supported by the following grants: Smith—NASA NGR 44-001-097 and Pfaffenberger—NIH S-T1GM-1321-06.

ESTIMATION DE L'INDEX DE SELECTION A PARTIR DE DONNEES MULTIVARIATES NORMALES INCOMPLETES

RESUME

On applique une technique à l'estimation de l'index de sélection, au cas où l'on dispose de données multivariates normales incomplètes. Le procédé utilise toutes les données disponibles, à la fois les vecteurs complets et partiels, et représente une amélioration en précision de l'estimateur relativement à celui obtenu en utilisant seulement les vecteurs-données complets. On obtient aussi à partir de ce procédé des estimateurs des moyennes phénotypiques et de la matrice variance-covariance.

Les individus ayant des données incomplètes sont indexés par une extension de la technique, et cet index est opposé à celui proposé par Henderson. Les résultats de simulation obtenus sur ordinateur sont tabulés.

REFERENCES

Afifi, A. A. and Elashoff, R. M. [1967]. Missing observations in multivariate statistics II. Point estimation in simple linear regression. *J. Amer. Statist. Ass. 62*, 10–29.

Hazel, L. N. [1943]. Genetic basis for selection indices. *Genetics 28*, 476–90.

Henderson, C. R. [1963]. Selection index and expected genetic advance. *NAS-NRC Publ. 982*, 141–63.

Hocking, R. R. and Smith, W. B. [1968]. Estimation of parameters in the multivariate normal distribution with missing observations. *J. Amer. Statist. Ass. 63*, 159–73.

Hocking, R. R., Smith, W. B., Waldron, B. R., and Oxspring, H. H. [1969a]. Estimation of parameters with incomplete data. Technical Report No. 12, Project Themis Contract. N00014-68-A-0140, Project NRO 47-700.

Hocking, R. R., Oxspring, H. H., and Waldron, B. R. [1969b]. Maximum likelihood estimation with incomplete normal data, Part I. (Submitted for publication.)

Kempthorne, O. and Nordskog, A. W. [1959]. Restricted selection indices. *Biometrics 15*, 10–19.

Smith, H. F. [1936]. A discriminant function for plant selection. *Ann. Eugen. 7*, 240–50.

Smith, W. B. and Hocking, R. R. [1968]. A simple method for obtaining the information matrix for a multivariate normal distribution. *Amer. Statistican 22*, No. 1, 18–20.

Williams, J. S. [1962]. The evaluation of a selection index. *Biometrics 18*, 375–93.

APPENDIX

General expressions for the estimators

The general expressions for the estimators of the means and variances are given for the case where at least some vectors of observations are complete. As will become clear in the development of the estimators, their virtue is the relative ease by which they may be programmed for computer computation.

In order to specify the general expressions for the estimators, it is necessary to develop some new notation (see Hocking *et al.* [1969a] and [1969b] for the complete development). Let q represent the dimensionality of the multivariate normal population from which the observations are sampled. Assume that n_k observations are recorded on the kth marginal of the q-variate distribution, where the likelihood function of the kth marginal is given by L_k . It follows that the complete likelihood is given by $L = \Pi_{k=1}^{T} L_k$, where T represents the number of marginal distributions sampled. Let $\mathfrak{y}_k^*$ and $\mathfrak{p}_k^*$ represent the column vectors of the parameters from the kth marginal distribution, where $\mathfrak{p}_k^*$ is the column ordering from the $\mathbf{P}_k$ covariance matrix. When $k = 1$, $\mathfrak{y}_k^*$ and $\mathfrak{p}_k^*$ are always taken to be the vectors associated with the q-variate distribution whether or not any q-variate observations are actually recorded; i.e., $\mathfrak{y}_1^{*\prime} = (\mu_1 , \mu_2 , \cdots , \mu_p)$ and

$$\mathfrak{p}_1^{*\prime} = (p_{11} , p_{12} , p_{22} , p_{13} , p_{23} , p_{33} , \cdots , p_{qq}).$$

Further let $\mathfrak{p}^{-1\prime} = (p^{11}, p^{12}, \cdots , p^{qq})$, where p^{ij} is the (i, j) element of $\mathbf{P}_k^{-1}$. Denote by $\hat{\mathfrak{y}}_k$ and $\hat{\mathfrak{p}}_k$ the estimates of $\mathfrak{y}_k^*$ and $\mathfrak{p}_k^*$ from the kth data set associated with the kth marginal distribution.

As given by Hocking and Smith [1968], let $\mathbf{\Delta}_1$ be the following $\frac{1}{2}q\ (q + 1)$ square matrix:

$$\mathbf{\Delta}_1 = \frac{\partial \mathfrak{p}'}{\partial \mathfrak{p}^{-1}} = \begin{bmatrix} \frac{\partial p_{11}}{\partial p^{11}} , & \frac{\partial p_{12}}{\partial p^{11}} , & \cdots & \frac{\partial p_{qq}}{\partial p^{11}} \\ \frac{\partial p_{11}}{\partial p^{12}} , & & & \\ \vdots & & & \vdots \\ \frac{\partial p_{11}}{\partial p^{qq}} , & & \cdots & \frac{\partial p_{qq}}{\partial p^{qq}} \end{bmatrix}, \tag{A1}$$

where

$$\frac{\partial p_{ij}}{\partial p^{uv}} = \frac{-1}{K_{uv}} (p_{iu}p_{jv} + p_{iv}p_{ju}), \qquad K_{uv} = \begin{matrix} 2 & \text{if} & u = v \\ 1 & \text{if} & u \neq v \end{matrix}.$$

Let $\mathbf{M}$ be the diagonal matrix of dimension $\frac{1}{2}q\ (q + 1)$ such that the diagonal elements are 2 in rows $\frac{1}{2}r\ (r + 1)$ for $r = 1, 2, \cdots , q$, and 1 elsewhere. Let $\mathbf{U}_1 = \mathbf{M}\mathbf{\Delta}_1$. It now follows that the large sample covariance matrices of $\hat{\mathfrak{y}}_1^*$ and $\hat{\mathfrak{p}}_1^*$, denoted by ${}_1\mathbf{V}_\mu$ and ${}_1\mathbf{V}_p$ respectively, are:

$$ {}_1\mathbf{V}_\mu = \frac{1}{n_1}\mathbf{P}_1 \quad \text{and} \quad {}_1\mathbf{V}_p = -\frac{1}{n_1}\mathbf{U}_1 , \tag{A2} $$

where $\mathbf{P}_1$ is the covariance matrix associated with the q-variate distribution and $\mathbf{U}_1$ is as given above. The covariance matrix of the kth marginal is

$$ {}_k\mathbf{W} = \begin{pmatrix} {}_k\mathbf{W}_\mu & 0 \\ 0 & {}_k\mathbf{W}_p \end{pmatrix} $$

when ${}_k\mathbf{W}_\mu = \mathbf{P}_k/n_k$ and ${}_k\mathbf{W}_p = -\mathbf{U}_k/n_k$, $\mathbf{U}_k = \mathbf{M}\boldsymbol{\Delta}_k$, and $\boldsymbol{\Delta}_k$ is given by (A1) with the elements of $\mathbf{P}_k$ used rather than those of $\mathbf{P}_1$. Notice that if the kth marginal has dimensionality s, $\boldsymbol{\Delta}_k$ is a $\frac{1}{2}s\,(s+1)$ square matrix. As in Hocking *et al.* [1969a] the joint covariance matrix of the first k marginals is block diagonal with elements

$$ {}_k\mathbf{V}_\mu = \prod_{j=k}^{2} (\mathbf{I} + \mathbf{B}_j\mathbf{D}_j){}_1\mathbf{V}_\mu , \tag{A3} $$

$$ {}_k\mathbf{V}_p = \prod_{j=k}^{2} (\mathbf{I} + \mathbf{A}_j\mathbf{C}_j){}_1\mathbf{V}_p , \tag{A4} $$

where $\mathbf{A}_j$, $\mathbf{B}_j$, $\mathbf{C}_j$, and $\mathbf{D}_j$ are defined below.

Finally, let $\mathbf{D}_k$ and $\mathbf{C}_k$ be indicator matrices used to relate the parameters associated with the kth marginal and those of the q-variate distribution. For example, if we have n_1 observations from $N_q(\boldsymbol{\mu}_1^*, \mathbf{P}_1)$, n_2 observations from $N_r(\boldsymbol{\mu}_2^*, \mathbf{P}_2)$ and n_3 observations from $N_s(\boldsymbol{\mu}_3^*, \mathbf{P}_3)$, $q > r$, $q > s$, then

$$ \boldsymbol{\mu}_2^* = \mathbf{D}_2\boldsymbol{\mu}_1^* , \qquad \mathbf{p}_2^* = \mathbf{C}_2\mathbf{p}_1^* , \qquad \boldsymbol{\mu}_3^* = \mathbf{D}_3\boldsymbol{\mu}_1^* , \qquad \mathbf{p}_3^* = \mathbf{C}_3\mathbf{p}_1^* . \tag{A5} $$

The general expressions for the estimators of the means and variances are given by:

$$ {}_T\hat{\boldsymbol{\mu}}_1 = \prod_{k=T}^{2} (\mathbf{I} + \hat{\mathbf{B}}_k\mathbf{D}_k)\hat{\boldsymbol{\mu}}_1^* - \sum_{j=2}^{T-1} \prod_{k=T}^{j+1} (\mathbf{I} + \hat{\mathbf{B}}_k\mathbf{D}_k)\hat{\mathbf{B}}_j\hat{\boldsymbol{\mu}}_j^* - \hat{\mathbf{B}}_T\hat{\boldsymbol{\mu}}_T^* \tag{A6} $$

$$ {}_T\hat{\mathbf{p}}_1 = \prod_{k=T}^{2} (\mathbf{I} + \hat{\mathbf{A}}_k\mathbf{C}_k)\hat{\mathbf{p}}_1^* - \sum_{j=2}^{T-1} \prod_{k=T}^{j+1} (\mathbf{I} + \hat{\mathbf{A}}_k\mathbf{C}_k)\hat{\mathbf{A}}_j\hat{\mathbf{p}}_j^* - \hat{\mathbf{A}}_T\hat{\mathbf{p}}_T^* , \tag{A7} $$

where

$$ \prod_{k=T}^{2} (\mathbf{I} + \hat{\mathbf{B}}_k\mathbf{D}_k) = (\mathbf{I} + \hat{\mathbf{B}}_T\mathbf{D}_T)(\mathbf{I} + \hat{\mathbf{B}}_{T-1}\mathbf{D}_{T-1}) \cdots (\mathbf{I} + \hat{\mathbf{B}}_2\mathbf{D}_2) $$

$$ \prod_{k=T}^{j+1} (\mathbf{I} + \hat{\mathbf{B}}_k\mathbf{D}_k) = (\mathbf{I} + \hat{\mathbf{B}}_T\mathbf{D}_T)(\mathbf{I} + \hat{\mathbf{B}}_{T-1}\mathbf{D}_{T-1}) \cdots (\mathbf{I} + \hat{\mathbf{B}}_{j+1}\mathbf{D}_{j+1}) $$

$$ \prod_{k=T}^{2} (\mathbf{I} + \hat{\mathbf{A}}_k\mathbf{C}_k) = (\mathbf{I} + \hat{\mathbf{A}}_T\mathbf{C}_T)(\mathbf{I} + \hat{\mathbf{A}}_{T-1}\mathbf{C}_{T-1}) \cdots (\mathbf{I} + \hat{\mathbf{A}}_2\mathbf{C}_2) $$

$$ \prod_{k=T}^{j+1} (\mathbf{I} + \hat{\mathbf{A}}_k\mathbf{C}_k) = (\mathbf{I} + \hat{\mathbf{A}}_T\mathbf{C}_T)(\mathbf{I} + \hat{\mathbf{A}}_{T-1}\mathbf{C}_{T-1}) \cdots (\mathbf{I} + \hat{\mathbf{A}}_{j+1}\mathbf{C}_{j+1}) $$

and, for $t = 2, \cdots, T$,

$$ \mathbf{B}_t = -{}_{t-1}\mathbf{V}_\mu\mathbf{D}_t'\left[\mathbf{D}_t\left({}_{t-1}\mathbf{V}_\mu + \frac{1}{n_t}\mathbf{P}_1\right)\mathbf{D}_t'\right]^{-1} \tag{A8} $$

$$\mathbf{A}_t = -{}_{t-1}\mathbf{V}_p\mathbf{C}_t'\left[\mathbf{C}_t\left({}_{t-1}\mathbf{V}_p - \frac{1}{n_t}\mathbf{U}_1\right)\mathbf{C}_t'\right]^{-1}. \tag{A9}$$

Notice that parts of (A8) and (A9) must be estimated; that is, $\mathbf{B}_t$ and $\mathbf{A}_t$ are functions of the parameters $\mathfrak{y}_k^*$ and $\mathbf{p}_k^*$. It is recommended that the $\mathbf{p}_k^*$ parameters in (A8) and (A9) be estimated by using the elements of $\hat{\mathbf{p}}_1$ from the complete sample for $t = 2, \cdots, T$. It is worthy to note that the estimators given by (A6) and (A7) are relatively easy to program on a computer such that for each data set k, only $\mathbf{C}_k$, $\mathbf{D}_k$, and the appropriate data are required by the program.

Example 4: Let $q = 2$ with n_1 observations from $N_2(\mathfrak{y}_1^*, \mathbf{P}_1)$, n_2 , and n_3 observations from the univariate marginals, $N_1(\mu_1, p_{11})$ and $N_1(\mu_2, p_{22})$, respectively. From (1), it follows that

$$\mathbf{U}_1 = -\begin{bmatrix} 2p_{11}^2 & 2p_{11}p_{12} & 2p_{12}^2 \\ & p_{11}p_{22} + p_{12}^2 & 2p_{12}p_{22} \\ & & 2p_{22}^2 \end{bmatrix}.$$

Combining $N_1(\mu_1, p_{11})$ with $N_2(\mathfrak{y}_1^*, \mathbf{P}_1)$, we have $\mathbf{D}_2 = (1, 0)$, $\mathbf{C}_2 = (1, 0, 0)$ resulting in

$$\mathbf{B}_2 = -\frac{1}{n_1}\begin{pmatrix} p_{11} \\ p_{21} \end{pmatrix}\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)p_{11}\right]^{-1} = \frac{-n_2}{n_1 + n_2}\cdot\frac{p_{12}}{p_{11}},$$

$$\mathbf{A}_2 = -\frac{1}{n_1}\begin{bmatrix} 2p_{11}^2 \\ 2p_{11}p_{12} \\ 2p_{12}^2 \end{bmatrix}\left(\frac{n_1 + n_2}{n_1 n_2}\,2p_{11}^2\right)^{-1} = \frac{-n_2}{n_1 + n_2}\begin{bmatrix} 1 \\ p_{12}/p_{11} \\ p_{12}^2/p_{11}^2 \end{bmatrix}.$$

Further,

$${}_2\mathbf{V}_\mu = \frac{1}{n_1}\left[\mathbf{I} - \frac{n_2}{n_1 + n_2}\begin{bmatrix} 1 & 0 \\ p_{12}/p_{11} & 0 \end{bmatrix}\right]\mathbf{P}_1,$$

$${}_2\mathbf{V}_p = -\frac{1}{n_1}\left\{\mathbf{I} - \frac{n_2}{n_1 + n_2}\begin{bmatrix} 1 & 0 & 0 \\ p_{12}/p_{11} & 0 & 0 \\ (p_{12}/p_{11})^2 & 0 & 0 \end{bmatrix}\right\}\mathbf{U}_1.$$

The resulting estimators from (A6) and (A7) are:

$${}_2\tilde{\mathfrak{y}}_1 = \begin{bmatrix} {}_1\hat{\mu}_1 \\ {}_1\hat{\mu}_2 \end{bmatrix} - \frac{n_2}{n_1 + n_2}\begin{bmatrix} 1 \\ \hat{p}_{12}/\hat{p}_{11} \end{bmatrix}({}_1\hat{\mu}_1 - {}_2\hat{\mu}_1),$$

where ${}_i\hat{\mu}_j$ is the estimator of the jth mean from the ith likelihood, and

$${}_2\tilde{\mathbf{p}}_1 = \begin{bmatrix} {}_1\hat{p}_{11} \\ {}_1\hat{p}_{12} \\ {}_1\hat{p}_{22} \end{bmatrix} - \frac{n_2}{n_1 + n_2}\begin{bmatrix} 1 \\ \hat{p}_{12}/\hat{p}_{11} \\ (\hat{p}_{12}/\hat{p}_{11})^2 \end{bmatrix}({}_1\hat{p}_{11} - {}_2\hat{p}_{11}),$$

where ${}_k\hat{p}_{ij}$ is the estimator of p_{ij} from the kth likelihood. Combining $N_1(\mu_3 , p_{22})$ to the above results, we have $\mathbf{D}_3 = (0, 1)$, $\mathbf{C}_3 = (0, 0, 1)$, which result in

$$\mathbf{B}_3 = -\frac{1}{n_1}\left[\begin{pmatrix} p_{12} \\ p_{22} \end{pmatrix} - \frac{n_2}{n_1 + n_2}\begin{pmatrix} p_{12} \\ p_{12}^2/p_{11} \end{pmatrix}\right]\left[\frac{n_1 + n_3}{n_1 n_3} p_{22} - \frac{n_2}{n_1(n_1 + n_2)}\frac{p_{12}^2}{p_{11}}\right]^{-1},$$

$$\mathbf{A}_3 = -\frac{1}{n_1}\left[\begin{pmatrix} p_{12}^2 \\ p_{12}p_{22} \\ p_{22}^2 \end{pmatrix} - \frac{n_2}{n_1 + n_2}\begin{pmatrix} p_{12}^2 \\ p_{12}^3/p_{11} \\ p_{12}^4/p_{11}^2 \end{pmatrix}\right]\left[\frac{n_1 + n_3}{n_1 n_3} p_{22}^2 - \frac{n_2}{n_1(n_1 + n_3)}\frac{p_{12}^4}{p_{11}^2}\right]^{-1}.$$

and so forth.

In order to estimate the parameters p_{ij} arising in $\mathbf{A}_2$, $\mathbf{B}_2$ and $\mathbf{A}_3$, $\mathbf{B}_3$, the estimates from $\hat{\mathbf{P}}_1$ would be used as indicated previously. Notice that it is not necessary to spell out $\mathbf{B}_t$ and $\mathbf{A}_t$ as we have done here for illustrative purposes. Rather, $\mathbf{B}_t$ and $\mathbf{A}_t$ are estimated in the forms (A8) and (A9) by using $\hat{\mathbf{P}}_1$, and the resulting numerical matrices are used to find ${}_T\tilde{\mathbf{u}}_1$ and ${}_T\tilde{\mathbf{p}}_1$, respectively.

To illustrate numerically the use of the general formulae, we consider the following set of data:

Traits	X_1	X_2
Observations	12	9
	9	3
	6	3
	9	
	5	
		16
		12

The data are divisible into three sets with $n_1 = 3$, $\hat{\mathbf{u}}_1^* = (9, 5)$, $\hat{\mathbf{p}}_1^* = (9, 9, 12)$; $n_2 = 2$, $\hat{\mathbf{u}}_2^* = (7)$, $\hat{\mathbf{p}}_2^* = (8)$ and $n_3 = 2$, $\hat{\mathbf{u}}_3^* = (14)$, $\hat{\mathbf{p}}_3^* = (8)$. As noted above, $\mathbf{D}_2 = (1, 0)$, $\mathbf{C}_2 = (1, 0, 0)$, $\mathbf{D}_3 = (0, 1)$, and $\mathbf{C}_3 = (0, 0, 1)$. By (A1) and (A2),

$$U_1 = -\begin{bmatrix} 162 & 162 & 162 \\ 162 & 189 & 216 \\ 162 & 216 & 288 \end{bmatrix}, \qquad {}_1\mathbf{V}_\mu = \begin{bmatrix} 3 & 3 \\ 3 & 4 \end{bmatrix},$$

$${}_1\mathbf{V}_p = \begin{bmatrix} 54 & 54 & 54 \\ 54 & 63 & 72 \\ 54 & 72 & 96 \end{bmatrix}.$$

From (A8) and (A9),

$$\hat{\mathbf{B}}_2 = -\begin{bmatrix} 0.4 \\ 0.4 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{A}}_2 = -\begin{bmatrix} 0.4 \\ 0.4 \\ 0.4 \end{bmatrix}, \quad \text{respectively.}$$

Now

$$_2\tilde{\mathbf{y}}_1 = (\mathbf{I} + \hat{\mathbf{B}}_2\mathbf{D}_2)\hat{\mathbf{y}}_1^* - \hat{\mathbf{B}}_2\hat{\mathbf{y}}_2^* = \begin{bmatrix} 8.2 \\ 4.2 \end{bmatrix}$$

and

$$_2\tilde{\mathbf{p}}_1 = (\mathbf{I} + \hat{\mathbf{A}}_2\mathbf{C}_2)\hat{\mathbf{p}}_1^* - \hat{\mathbf{A}}_2\hat{\mathbf{p}}_2^* = \begin{bmatrix} 8.6 \\ 8.6 \\ 11.6 \end{bmatrix}.$$

To adjoin the third data set, by (A3) and (A4), we have

$$_2\mathbf{V}_\mu = (\mathbf{I} + \hat{\mathbf{B}}_2\mathbf{D}_2)_1\mathbf{V}_\mu = \begin{bmatrix} 1.8 & 1.8 \\ 1.8 & 2.8 \end{bmatrix},$$

$$_2\mathbf{V}_p = (\mathbf{I} + \hat{\mathbf{A}}_2\mathbf{C}_2)_1\mathbf{V}_p = \begin{bmatrix} 32.4 & 32.4 & 32.4 \\ 32.4 & 41.4 & 50.4 \\ 32.4 & 50.4 & 74.4 \end{bmatrix}, \quad \text{respectively.}$$

From a reapplication of (A8) and (A9)

$$\hat{\mathbf{B}}_3 = -\begin{bmatrix} 0.205 \\ 0.318 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{A}}_3 = -\begin{bmatrix} 0.148 \\ 0.231 \\ 0.341 \end{bmatrix}.$$

Now

$$_3\tilde{\mathbf{y}}_1 = (\mathbf{I} + \hat{\mathbf{B}}_3\mathbf{D}_3)(\mathbf{I} + \hat{\mathbf{B}}_2\mathbf{D}_2)\tilde{\mathbf{y}}_1^* - (\mathbf{I} + \hat{\mathbf{B}}_3\mathbf{D}_3)\hat{\mathbf{B}}_2\hat{\mathbf{y}}_2^* - \hat{\mathbf{B}}_3\hat{\mathbf{y}}_3^* = \begin{bmatrix} 10.21 \\ 7.32 \end{bmatrix}$$

and

$$_3\tilde{\mathbf{p}}_1 = (\mathbf{I} + \hat{\mathbf{A}}_3\mathbf{C}_3)(\mathbf{I} + \hat{\mathbf{A}}_2\mathbf{C}_2)\hat{\mathbf{p}}_1^* - (\mathbf{I} + \hat{\mathbf{A}}_3\mathbf{C}_3)\hat{\mathbf{A}}_2\hat{\mathbf{p}}_2^* - \hat{\mathbf{A}}_3\hat{\mathbf{p}}_3^* = \begin{bmatrix} 8.07 \\ 7.77 \\ 10.37 \end{bmatrix}.$$

Thus, the phenotypic covariance matrix to be used in $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$ is:

$$\mathbf{P} = \begin{bmatrix} 8.07 & 7.77 \\ 7.77 & 10.37 \end{bmatrix}.$$

In order to rank the last 4 observation vectors, the missing x_1 and x_2 values would be estimated from the two regression equations, computed from the 3 complete vectors, $x_1 = 5.25 + 0.75\,x_2$ and $x_2 = -4 + x_1$, respectively. Thus the selection index for each observation could be easily found by inverting $\mathbf{P}$ and multiplying it by appropriate $\mathbf{G}$ and $\boldsymbol{\alpha}$ matrices.

Example 5: Suppose we have a sample from $N_2(\mathfrak{y}_2^*, \mathbf{P}_2)$ and $N_2(\mathfrak{y}_3^*, \mathbf{P}_3)$, both of which are marginals of $N_3(\mathfrak{y}_1^*, \mathbf{P}_1)$. Let $\mathbf{D}_{ij}$ and $\mathbf{C}_{ij}$ be the nonzero rows of $\mathbf{D}_i\mathbf{D}_j'$ and $\mathbf{C}_i\mathbf{C}_j'$ respectively. Consider the following estimators:

$$ {}_3\tilde{\mathfrak{y}}_2 = \hat{\mathfrak{y}}_2^* + \hat{\mathbf{B}}_{23}(\mathbf{D}_{32}\hat{\mathfrak{y}}_2^* - \mathbf{D}_{23}\hat{\mathfrak{y}}_3^*), \qquad {}_2\tilde{\mathfrak{y}}_3 = \hat{\mathfrak{y}}_3^* + \hat{\mathbf{B}}_{32}(\mathbf{D}_{23}\hat{\mathfrak{y}}_3^* - \mathbf{D}_{32}\hat{\mathfrak{y}}_2^*), \tag{A10} $$

$$ {}_3\tilde{\mathbf{p}}_2 = \hat{\mathbf{p}}_2^* + \hat{\mathbf{A}}_{23}(\mathbf{C}_{32}\hat{\mathbf{p}}_2^* - \mathbf{C}_{23}\hat{\mathbf{p}}_3^*), \qquad {}_2\tilde{\mathbf{p}}_3 = \hat{\mathbf{p}}_3^* + \hat{\mathbf{A}}_{32}(\mathbf{C}_{23}\hat{\mathbf{p}}_3^* - \mathbf{C}_{32}\hat{\mathbf{p}}_2^*). $$

Notice that $\mathbf{D}_{32}\hat{\mathfrak{y}}_2^* - \mathbf{D}_{23}\hat{\mathfrak{y}}_3^*$, for instance, is nothing more than the vector differences of the estimates of the common parameters. Indeed, when one marginal is not a marginal of another, the procedure is to use a 'double application' of the method described in the text. By minimizing the variance of the estimators given in (A10), it follows that

$$ \begin{aligned} \mathbf{B}_{23} &= -\mathbf{V}_{\mu 2}\mathbf{D}_{32}'(\mathbf{D}_{32}\mathbf{V}_{\mu 2}\mathbf{D}_{32}' + \mathbf{D}_{23}\mathbf{V}_{\mu 3}\mathbf{D}_{23}')^{-1}, \\ \mathbf{B}_{32} &= -\mathbf{V}_{\mu 3}\mathbf{D}_{23}'(\mathbf{D}_{32}\mathbf{V}_{\mu 2}\mathbf{D}_{32}' + \mathbf{D}_{23}\mathbf{V}_{\mu 3}\mathbf{D}_{23}')^{-1}, \\ \mathbf{A}_{23} &= -\mathbf{V}_{p 2}\mathbf{C}_{32}'(\mathbf{C}_{32}\mathbf{V}_{p 2}\mathbf{C}_{32}' + \mathbf{C}_{23}\mathbf{V}_{p 3}\mathbf{C}_{23}')^{-1}, \\ \mathbf{A}_{32} &= -\mathbf{V}_{p 3}\mathbf{C}_{23}'(\mathbf{C}_{23}\mathbf{V}_{p 2}\mathbf{C}_{32}' + \mathbf{C}_{23}\mathbf{V}_{p 3}\mathbf{C}_{23}')^{-1}, \end{aligned} $$

where, for instance,

$$ \mathbf{V}_{\mu 2} = \frac{1}{n_2}\mathbf{D}_2'\boldsymbol{\Sigma}_1\mathbf{D}_2 \quad \text{and} \quad \mathbf{V}_{p 2} = -\frac{1}{n_2}\mathbf{C}_2\mathbf{U}_1\mathbf{C}_2' . $$

The estimators ${}_3\tilde{\mathfrak{y}}_2$, ${}_2\tilde{\mathfrak{y}}_3$ and ${}_3\tilde{\mathbf{p}}_2$, ${}_2\tilde{\mathbf{p}}_3$ are combined to give the following complete estimators:

$$ {}_3\tilde{\mathfrak{y}}_1 = \mathbf{D}_2'\,{}_3\hat{\mathfrak{y}}_2 + (\mathbf{I} - \mathbf{D}_2'\mathbf{D}_2)\mathbf{D}_3'\,{}_2\hat{\mathfrak{y}}_3, \quad {}_3\tilde{\mathbf{p}}_1 = \mathbf{C}_2'\,{}_3\hat{\mathbf{p}}_2 + (\mathbf{I} - \mathbf{C}_2'\mathbf{C}_2)\mathbf{C}_3'\,{}_2\hat{\mathbf{p}}_3 . \tag{A11} $$

To adjoin additional information, given that the parameters of the new likelihood are not a subset of those previously considered, the expressions given by (A10) are again used with ${}_3\hat{\mathfrak{y}}_1$, ${}_3\hat{\mathbf{p}}_1$ from (A11) adjoined with the parameters of the new likelihood in the same 'double application' manner.

By substituting the expressions given by (A10) into (A11), it is possible to arrive at equations analogous to (A6) and (A7). However, no general expressions follow for the case in which $n_1 = 0$ as they do when $n_1 > 0$.

Received October 1969, *Revised February* 1970