DISCOVERING MARKOV STRUCTURE IN GROUP SEQUENTIAL METHODS FOR LONGITUDINAL STUDIES

David M. Reboussin

Department of Human Oncology
and
Department of Statistics

UNIVERSITY OF WISCONSIN-MADISON
Discovering Markov Structure in Group Sequential Methods for Longitudinal Studies

by

David M. Reboussin
Department of Statistics
University of Wisconsin
Madison, WI 53706

Biostatistics Technical Report No. 61

November 18, 1990
Abstract

The covariance matrix for estimates of regression coefficients in group sequential testing of longitudinal or repeated measurements is shown to have a Markov structure. This means that variance of the statistics used for group sequential testing involve only the first previous, and not all previous, analyses. In addition, the computations required for Lan-DeMets boundaries involve only a modest change in existing software. The result is detailed for the normal theory model, and a sketch for generalized linear models is given.

**KEY WORDS:** Estimating Equations; Longitudinal data; Nuisance Parameter; Regression; Repeated measurements; Sequential Estimation

1 Introduction

Three recent papers address the problem of performing parametric interim analyses in clinical trial having repeated measurements: Lee and DeMets (1990), Wei, Su and Lachin (1990), and Wu and Lan (1990). This paper examines the covariance matrix for estimates in successive interim analyses and finds that when the covariance structure is known they will have a Markov structure; that is, when the covariance structure is known tests will depend only on the first preceding, and not all preceding, interim analyses. We will discuss only the first two of the above papers.

The methods these papers describe contain a conceptual dilemma. When individuals are measured only once, or when successive measurements on individuals are independent, the current interim analysis does not depend on previous analyses. This will be called the independence case. For longitudinal or repeated measurements which are *not* independent, there is some covariance structure. However, at the most abstract level, we might expect analyses to have the Markov property; that is, they may depend on the most recent previous analysis, but not on the entire sequence of past results. Imagine the second analysis: all available information is used and summarized, including any dependence on the first analysis. For the third analysis, new information collected since the second is incorporated, as well as the summary of information at the second. If the process is not Markovian, the summary and interpretation of the second and the first analysis may change: we are no longer updating and summarizing but completely supplanting previous conclusions. This implies a profound ignorance of the process.
2 Lee and DeMets (1990)

We will independently develop the estimates using our own notation. Readers desiring details of the model are referred to the original paper. We do not distinguish between treatment groups. As a convention, any quantity related specifically to the $k^{th}$ interim analysis will be denoted with a superscript in parentheses, e.g. $X_{i}^{(k)}$ denotes the quantity $X$ for the $i^{th}$ individual at the $k^{th}$ analysis (in the paper by Lee and DeMets, the interim analysis is sometimes denoted using subscripts). $n_{i}^{(k)}$ will denote the number of observations on the $i^{th}$ individual accumulated by the $k^{th}$ interim analysis. The actual covariate values on the $i^{th}$ individual accumulated by the $k^{th}$ interim analysis are denoted by the $n_{i}^{(k)} \times p$ matrix $X_{i}^{(k)}$, where $p$ is the number of covariates measured (plus one if there is an intercept term to be estimated). The matrix of covariate values measured on the $i^{th}$ individual at the end of the trial can be denoted simply $X_{i}$, but note that this is the same as $X_{i}^{(K)}$ if there are a total of $K$ interim analyses conducted. $X_{i}$ or $X_{i}^{(K)}$ has dimension $n_{i}^{(K)} \times p$, or $n_{i} \times p$, where $n_{i}$ denotes the total number of observations on the $i^{th}$ individual at the end of the trial.

Let $P_{i}^{(k)} = [I_{n_{i}^{(k)}} : 0_{n_{i} \times (n_{i} - n_{i}^{(k)})}]_{n_{i}^{(k)} \times n_{i}}$ so that $X_{i}^{(k)} = P_{i}^{(k)}X_{i}$. Notice that $P_{i}^{(k)}$ “slices off” the rows of $X_{i}$ not observed by the $k^{th}$ interim analysis. It can also be used in denoting the variance of the responses for the $i^{th}$ individual observed by the $k^{th}$ analysis in terms of the variances of all responses observed by the end of the trial. Let $y_{i} = y_{i}^{(K)}$ denote the responses on the $i^{th}$ individual accumulated by the end of the trial ($K^{th}$ analysis), and $\text{var}(y_{i}) = V_{i}$. Then $y_{i}^{(k)} = P_{i}^{(k)}y_{i}$ and $\text{var}(y_{i}^{(k)}) = V_{i}^{(k)} = P_{i}^{(k)}V_{i}P_{i}^{(k)'}$. Letting $W_{i} = V_{i}^{-1}$, we also have $V_{i}^{(k)-1} = W_{i}^{(k)} = (P_{i}^{(k)}V_{i}P_{i}^{(k)'})^{-1}$.

Now we have the variance matrix for the responses under the model, we can write down the variance matrix for the estimated coefficients. Let $\hat{\beta}$ be our (vector of) estimated regression coefficient(s) at the end of the trial, and $\hat{\beta}(k)$ be our (vector of) estimated regression coefficient(s) at the $k^{th}$ interim analysis. Then

$$
\hat{\beta}(k) = \left[ \sum_{i=1}^{n_{i}^{(k)}} X_{i}^{(k)}W_{i}^{(k)}X_{i}^{(k)} \right]^{-1} \left[ (X_{1}^{(k)'}W_{1}^{(k)}P_{1}^{(k)}y_{1}), \ldots, (X_{m(k)}^{(k)'}W_{m(k)}^{(k)}P_{m(k)}^{(k)}y_{m(k)}) \right]_{1 \times m(k)}
$$

where $m(k)$ is the total number of individuals entered in the study by the $k^{th}$ analysis. Denote

$$
U^{(k)} = \left[ \sum_{i=1}^{n_{i}^{(k)}} X_{i}^{(k)'}W_{i}^{(k)}X_{i}^{(k)} \right]^{-1}
$$

(it is symmetric), and modify the definition of $M$ given by Lee and DeMets:

$$
M(k)_{p \times \sum_{i=1}^{m} n_{i}} = U^{(k)} \left[ (X_{1}^{(k)'}W_{1}^{(k)}P_{1}^{(k)}), \ldots, (X_{m(k)}^{(k)'}W_{m(k)}^{(k)}P_{m(k)}^{(k)}), 0, \ldots, 0 \right]
$$
where $M(k)$ is the $k^{th}$ block of $p$ rows in the matrix $M$, and $m = m^{(K)}$ is the total number of individuals at the end of the trial.

At this point it may be helpful to think $p = 1$, but it is not necessary. We can write the matrix $\text{var}(\hat{\beta}^{(k)})$ as $M(k) \text{diag}(V_1, V_2, \ldots, V_m) M(k)'$ and the matrix $\text{var}(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \ldots, \hat{\beta}^{(K)})$ as

$$
\begin{pmatrix}
M(1) \\
M(2) \\
\vdots \\
M(K)
\end{pmatrix}
\text{diag}(V_1, V_2, \ldots, V_m) [M(1)', M(2)', \ldots, M(K)'].
$$

(2)

This, at long last, is the matrix we wish to simplify.

Denote for the moment

$$
A_i^{(j)(k)} = U^{(j)}(X_i^{(j)}'W_i^{(j)}P_i^{(j)})V_i(X_i^{(k)}'W_i^{(k)}P_i^{(k)})'U^{(k)}
= U^{(j)}(X_i^{(j)}'W_i^{(j)}(P_i^{(j)}V_iP_i^{(k)})W_i^{(k)}X_i^{(k)})U^{(k)}.
$$

(3)

Then since individuals are independent (2) can be written

$$
\begin{bmatrix}
\sum_{i=1}^{m^{(1)}} A_i^{(1)(1)} & \sum_{i=1}^{m^{(1)}} A_i^{(1)(2)} & \cdots & \sum_{i=1}^{m^{(1)}} A_i^{(1)(K)} \\
\sum_{i=1}^{m^{(2)}} A_i^{(2)(1)} & \sum_{i=1}^{m^{(2)}} A_i^{(2)(2)} & \cdots & \sum_{i=1}^{m^{(2)}} A_i^{(2)(K)} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{m^{(K)}} A_i^{(K)(1)} & \sum_{i=1}^{m^{(K)}} A_i^{(K)(2)} & \cdots & \sum_{i=1}^{m^{(K)}} A_i^{(K)(K)}
\end{bmatrix}.
$$

What follows will establish that $\sum_{i=1}^{m^{(j)}} A_i^{(j)(k)} = \sum_{i=1}^{m^{(k)}} A_i^{(k)(k)} = U^{(k)}$ where $j \leq k$.

Equation (3) shows the form of the terms in the sums making up the variance matrix. It is the simplification of these terms we now undertake. First, consider $A_i^{(k)(k)}$. Since $V_i^{(k)} = P_i^{(k)} V_i P_i^{(k)}$, (3) becomes

$$
U^{(k)}X_i^{(k)'}(W_i^{(k)}V_i^{(k)}W_i^{(k)})X_i^{(k)}U^{(k)} = U^{(k)}X_i^{(k)'W_i^{(k)}X_i^{(k)}}U^{(k)}
$$

so that, recalling the definition of $U^{(k)}$ in (1),

$$
\sum_{i=1}^{m^{(k)}} A_i^{(k)(k)} = U^{(k)} \left[ \sum_{i=1}^{m^{(k)}} X_i^{(k)'W_i^{(k)}X_i^{(k)}} \right] U^{(k)} = U^{(k)}.
$$

Now consider $A_i^{(j)(k)}$. We have

$$
U^{(j)}X_i^{(j)'(W_i^{(j)}P_i^{(j)}V_iP_i^{(k)})X_i^{(k)}U^{(k)}}
= U^{(j)}X_i^{(j)'\left[W_i^{(j)}: 0_{n_i^{(j)} \times (n_i^{(k)} - n_i^{(j)})}\right]X_i^{(k)}U^{(k)}} = U^{(j)}X_i^{(j)'W_i^{(j)}X_i^{(j)}}U^{(k)}.
$$

(4)

(5)
so that
\[
\sum_{i=1}^{m^{(j)}} A_i^{(j)(k)} = U^{(j)} \left[ \sum_{i=1}^{m^{(j)}} X_i^{(j)} W_i^{(j)} X_i^{(j)} \right] U^{(k)} = U^{(k)}.
\]

The crucial step from (4) to (5) can be seen by rewriting
\[
(W_i^{(j)} (p_i^{(j)} v_i p_i^{(k)}) W_i^{(k)}) = \left( p_i^{(j)} v_i p_i^{(k)} \right)^{-1} (p_i^{(j)} v_i p_i^{(k)}) (p_i^{(k)} v_i p_i^{(k)})^{-1}.
\]

We have
\[
(p_i^{(j)} v_i p_i^{(j)})' = \begin{pmatrix} v_{11} & \cdots & v_{1n_j^{(j)}} \\ \vdots & \ddots & \vdots \\ v_{1n_j^{(j)}} & \cdots & v_{1n_j^{(k)}} \end{pmatrix} = B,
\]

say, and, using the symmetry of $V_i$,
\[
(p_i^{(j)} v_i p_i^{(k)}) = \begin{pmatrix} v_{11} & \cdots & v_{1n_j^{(j)}} & v_{1n_j^{(j)+1}} & \cdots & v_{1n_j^{(k)}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ v_{1n_j^{(j)}} & \cdots & v_{1n_j^{(j)+1}} & v_{1n_j^{(j)+2}} & \cdots & v_{1n_j^{(k)}} \end{pmatrix} = [B : C],
\]
\[
(p_i^{(k)} v_i p_i^{(k)}) = \begin{pmatrix} v_{11} & \cdots & v_{1n_j^{(j)}} & v_{1n_j^{(j)+1}} & \cdots & v_{1n_j^{(k)}} \\ v_{1n_j^{(j)+1}} & \cdots & v_{1n_j^{(j)+1}} & v_{1n_j^{(j)+2}} & \cdots & v_{1n_j^{(k)}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ v_{1n_j^{(k)}} & \cdots & v_{1n_j^{(k)}} & v_{1n_j^{(k)+1}} & \cdots & v_{1n_j^{(k)}} \end{pmatrix} = \begin{bmatrix} B & C \\ C' & D \end{bmatrix}
\]

Using the standard formula for inverting partitioned matrices (e.g. in Rao, 1973), it is easy to show
\[
B^{-1} [B : C] \begin{bmatrix} B & C \\ C' & D \end{bmatrix}^{-1} = [I_{n_j^{(j)}} : B^{-1} C] \begin{bmatrix} B & C \\ C' & D \end{bmatrix}^{-1} = [B^{-1} : 0_{n_j^{(j)} \times (n_i^{(k)} - n_j^{(j)})}]
\]

which, recalling (6), gives the result above.

Thus the variance-covariance matrix for the sequential regression coefficients may be written
\[
\begin{bmatrix} U^{(1)} & U^{(2)} & \cdots & U^{(K)} \\ U^{(2)} & U^{(2)} & \cdots & U^{(K)} \\ \vdots & \vdots & \ddots & \vdots \\ U^{(K)} & U^{(K)} & \cdots & U^{(K)} \end{bmatrix}.
\]
Consider now the variance-covariance matrix for the differences in a single sequential regression coefficients between the control and treatment groups. It clearly has the same form: write it

\[
\begin{bmatrix}
\sigma_1^2 & \sigma_2^2 & \ldots & \sigma_K^2 \\
\sigma_2^2 & \sigma_2^2 & \ldots & \sigma_K^2 \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_K^2 & \sigma_K^2 & \ldots & \sigma_K^2 \\
\end{bmatrix}
\]

and denote the vector of differences itself \((\hat{\delta}_1, \ldots, \hat{\delta}_K)\). This vector of differences is (under the null hypothesis) normally distributed with mean 0 and the variance described above. What are the conditional variances of the differences as the trial progresses? We have \(\text{var}(\hat{\delta}_2 | \hat{\delta}_1) = \sigma_2^2 \left(1 - \frac{\sigma_2^2}{\sigma_1^2}\right)\), and

\[
\text{var}(\hat{\delta}_3 | \hat{\delta}_1, \hat{\delta}_2) = \sigma_3^2 - \left(\sigma_3^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \sigma_2^2 \right)^{-1} \left(\sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \sigma_3^2 \right) = \sigma_3^2 \left(1 - \frac{\sigma_3^2}{\sigma_2^2}\right),
\]

which is just \(\text{var}(\hat{\delta}_3 | \hat{\delta}_2)\). The conditional variance only depends on the most recently past interim analysis, and not on the entire history of the trial.

This is the Markov property, at least in the normal case. The example in Lee and DeMets (1990) has confirmed this result numerically. That means the only computational difference between the spending function approach of Lan and DeMets (1983) and this one is an adjustment to the variances: a very simple modification of the programs already written (see DeMets, Kim, Lan, and Reboussin, 1990). Also, when \(V_i\) is the identity matrix, this result reduces to the independence case.

Notice that we have \(\text{var}(\hat{\beta})\), which depends on the unknown within individual variance matrices \(V_i\). To use this, some estimate of \(V_i\) must be used, and this estimate may change from earlier to later interim analyses. This difficulty, sequential estimation of a nuisance parameter, is endemic to early stopping designs, even in the independence case. In one sense, little has been gained: the covariance between the first and second analyses may change at the third or later. This means that tests at earlier times may have been at different \(\alpha\) than supposed. But if errors in the estimates of nuisance parameters can be ignored in the independence case (by not recomputing past tests), they can also be ignored here. The underlying Markov structure is useful to know.

3 Wei, Su and Lachin (1990)

For this section we use notation conforming (more or less) to that of Liang and Zeger (1986). Let \(Y_i = (y_{i1}, \ldots, y_{im})^T\) be the outcomes observed on the \(i^{th}\) subject with an \(n_i \times p\) covariate
matrix \( X_i = (x_{i1}^T, \ldots, x_{im}^T)^T \), where \( i = 1, \ldots, m \). The marginal density of \( y_{ij} \) is assumed to be

\[
f(y_{ij}) = \exp\{y_{ij} \theta_{ij} - a(\theta_{ij}) + b(y_{ij})\} \phi
\]

where \( \theta_{ij} = h(\eta_{ij}), \eta_{ij} = x_{ij} \beta \) so that \( E(y_{ij}) = a'(\theta_{ij}) \), and \( \text{var}(y_{ij}) = a''(\theta_{ij}) \phi \). We henceforth assume \( \phi = 1 \) for added simplicity.

The estimate for \( \beta \) is defined using estimating equations involving the score function. Let \( a' = (a'(\theta_{i1}), \ldots, a'(\theta_{ij}))^T \) and \( S_i = Y_i - a'(\theta_i) \), a residual-like vector with dimension \( n_i \times 1 \) and elements denoted \( S_{ij} \). Also let \( \Delta_i = \text{diag}[d\theta_{ij}/d\eta_{ij}] \) with dimension \( n_i \times n_i \). Liang and Zeger take as their estimate in the independence-within-subjects case the value of \( \beta \), say \( \hat{\beta} \), satisfying \( U(\beta) = \sum_{i=1}^m X_i^T \Delta_i S_i = 0 \).

Now for large \( m \), we can employ the approximation \( (\hat{\beta} - \beta) \approx (-\partial U(\beta)/\partial \beta)^{-1} U(\beta) \).

When expanding the partial derivative, we will need the \( n_i \times n_i \) matrix \( A_i = \text{diag}(a''(\theta_{ij})) \). Notice that under the independence model and up to a constant which is set to one, \( A_i = \text{cov}(Y_i) \). We can now write down an approximate variance for \( \hat{\beta} \), viz.

\[
\left(-\frac{\partial}{\partial \beta} U(\beta)\right)^{-1} U(\beta) U(\beta)^T \left(-\frac{\partial}{\partial \beta} U(\beta)\right)^{-1}
\]

Expanding the partial derivative, we have

\[
\frac{\partial}{\partial \beta} U(\beta) = \frac{\partial}{\partial \beta} \sum_{i=1}^m X_i^T \Delta_i(\beta) S_i(\beta)
\]

\[
= \sum_{i=1}^m \sum_{j=1}^{n_i} X_{ij}^T \left[ \frac{\partial}{\partial \beta} \left( \Delta_i(\beta) S_{ij}(\beta) \right) \right]
\]

\[
= \sum_{i=1}^m \sum_{j=1}^{n_i} X_{ij}^T \left[ h''(\eta_{ij}) S_{ij} - \Delta_{ij}^2 a''(\theta_{ij}) \right] X_{ij}
\]

\[
= \sum_{i=1}^m \sum_{j=1}^{n_i} X_{ij}^T h''(\eta_{ij}) S_{ij} X_{ij} - \sum_{i=1}^m X_i^T \Delta_i A_i \Delta_i X_i
\]

with some details of the differentiation left out. Noticing that \( \sum_{i=1}^m S_i/m \to 0 \) leads quickly to the result in Liang and Zeger, p. 15:

\[
\text{var}(\hat{\beta}) \approx \left[ \sum_{i=1}^m X_i^T \Delta_i A_i \Delta_i X_i \right]^{-1} \left( \sum_{i=1}^m X_i^T \text{cov}(Y_i) \Delta_i X_i \right) \left[ \sum_{i=1}^m X_i^T \Delta_i A_i \Delta_i X_i \right]^{-1}.
\]

Of course, when \( \text{cov}(Y_i) = A_i \) this reduces to \( \left[ \sum_{i=1}^m X_i^T \Delta_i A_i \Delta_i X_i \right]^{-1} \).

Wei, Su and Lachin develop the above with notation appropriate for sequential testing. This section is only a sketch of the application of ideas from the previous section to this setting, so the placement of superscripts in parentheses is a shorthand for more complicated
expressions involving matrices like $P_i^{(k)}$ earlier. Thus

$$ \hat{\beta}^{(k)} - \beta \approx \left( -\frac{\partial}{\partial \beta} U^{(k)}(\beta) \right)^{-1} U^{(k)}(\beta) $$

$$ = \left[ \frac{\partial}{\partial \beta} \sum_{i=1}^{m} X_i^{(k)^T} \Delta_i^{(k)}(\beta) S_i^{(k)}(\beta) \right]^{-1} \sum_{i=1}^{m} X_i^{(k)^T} \Delta_i^{(k)} S_i^{(k)} $$

and since subjects are independent, the covariance matrix for $\hat{\beta}^{(k)}$ and $\hat{\beta}^{(l)}$, where $k \leq l$, as reported by Wei, Su and Lachin may be written

$$ \left( -\frac{\partial}{\partial \beta} U^{(k)}(\hat{\beta}^{(k)}) \right)^{-1} \sum_{i=1}^{m} X_i^{(k)^T} \Delta_i^{(k)} S_i^{(k)^T} \Delta_i^{(l)} X_i^{(l)} \left( -\frac{\partial}{\partial \beta} U^{(l)}(\hat{\beta}^{(l)}) \right)^{-1}. $$

This estimates the asymptotic variance matrix

$$ \left( -\frac{\partial}{\partial \beta} U^{(k)}(\beta) \right)^{-1} \sum_{i=1}^{m} X_i^{(k)^T} \Delta_i^{(k)} \text{cov}(Y_i^{(k)}, Y_i^{(l)}) \Delta_i^{(l)} X_i^{(l)} \left( -\frac{\partial}{\partial \beta} U^{(l)}(\beta) \right)^{-1}. $$

Now

$$ \text{cov}(Y_i^{(k)}, Y_i^{(l)}) = \begin{bmatrix} \text{var}(Y_i^{(k)}) & 0 \\ 0 & 0 \end{bmatrix} $$

since we are working with the assumption that repeated observations on individuals are independent. Furthermore we can make use of the asymptotic approximation of the previous paragraph and simplify:

$$ \left[ \sum_{i=1}^{m} X_i^{T} \Delta_i A_i \Delta_i X_i \right]^{-1} \left[ \sum_{i=1}^{m} X_i^{(k)^T} \Delta_i^{(k)} \text{cov}(Y_i^{(k)}, \Delta_i^{(k)} X_i^{(k)} \right] \left[ \sum_{i=1}^{l} X_i^{T} \Delta_i A_i \Delta_i X_i \right]^{-1} $$

$$ = \left[ \sum_{i=1}^{l} X_i^{T} \Delta_i A_i \Delta_i X_i \right]^{-1} $$

if $\text{cov}(Y_i^{(k)}) = A_i^{(k)}$. This is the same result as the previous section, and the difficulties of sequentially estimating a nuisance parameter are similarly ignored. There are also two approximations involved here, one asymptotic and one assuming estimates are close to parameters. The work of the previous section also required variances to be known, but no asymptotics were involved.

4 Extension to the general case

Extending the results to Liang and Zeger's general case is of little practical importance, but it is logically a further step. We ignore the many difficulties of estimation in section
3 of Liang and Zeger. These notwithstanding, the estimating equations are now
\[ U(\beta) = \sum_{i=1}^{m} D_i^T V_i^{-1} S_i = 0 \]
where \( D_i = A_i \Delta_i X_i \), and \( V_i = A_i^T R(\alpha) A_i / \phi \). The useful approximation is \( (\hat{\beta} - \beta) \approx \left( -\frac{\partial}{\partial \beta} U(\beta) \right)^{-1} U(\beta) \). and the asymptotic variance of \( \hat{\beta} \) is

\[
\text{var}(\hat{\beta}) \approx \left[ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right]^{-1} \left( \sum_{i=1}^{m} D_i^T V_i^{-1} \text{cov}(Y_i) V_i^{-1} D_i \right) \left( \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right)^{-1}.
\]

We omit details for adding superscripts in parentheses and proceed as before: the variance for \( \hat{\beta}^{(k)} \) is

\[
\left[ \sum_{i=1}^{m_k} D_i^{(k)T} V_i^{(k)-1} D_i^{(k)} \right]^{-1} \left( \sum_{i=1}^{m_k} D_i^{(k)T} V_i^{(k)-1} \text{cov} \left( Y_i^{(k)} \right) V_i^{(k)-1} D_i^{(k)} \right) \left[ \sum_{i=1}^{m_k} D_i^{(k)T} V_i^{(k)-1} D_i^{(k)} \right]^{-1}
\]

And if that weren’t easy enough, take a look at that covariance matrix in the middle: the same calculation as the first section delivers the covariance matrix for \( \hat{\beta}^{(k)} \) and \( \hat{\beta}^{(l)} \)

\[
\left[ \sum_{i=1}^{m_l} D_i^{(l)T} V_i^{(l)-1} D_i^{(l)} \right]^{-1} \left( \sum_{i=1}^{m_l} D_i^{(l)T} V_i^{(l)-1} \text{cov} \left( Y_i^{(l)}, Y_i^{(l)} \right) V_i^{(l)-1} D_i^{(l)} \right) \left[ \sum_{i=1}^{m_l} D_i^{(l)T} V_i^{(l)-1} D_i^{(l)} \right]^{-1}
\]

Again, if \( \text{cov}(Y_i) \) is correctly specified by \( V_i \), there is a simplification, and the covariance matrix for the sequential regression coefficients has the special structure noticed in the first section.

5 Is there an underlying correlated Gaussian process?

In the independence case, there is an underlying Gaussian process from which the covariance matrix can be derived. It is natural to ask whether there is a parallel, correlated Gaussian process underlying the repeated measures case which yields a covariance matrix having the structure described earlier.

Since for standard Brownian Motion, say \( W(t) \), we have \( E[W(s), W(t)] = \min(s, t) \), perhaps there is a stochastic process with normally distributed marginals having \( E[X(s), X(t)] = \max(s, t) \). Unfortunately, this is not a positive definite function. For example, if \( s < t \)

\[
\text{Var}[X(t) - X(s)] = \text{Var}[X(t)] + \text{Var}[X(s)] - 2 \cdot \text{cov}[X(s), X(t)]
\]

\[
= t + s - 2 \cdot t = s - t < 0
\]

which is negative.

It may also be worth noting that for \( s < t \), \( \text{Cov}(X(t) - X(s), X(s)) = t - s > 0 \). However, for \( s < t < u \), \( \text{Cov}(X(u) - X(t), X(t) - X(s)) = u - t - u + t = 0 \).
But the foregoing is a little wrong-headed. For the independence case the observations are independent, and this carries over to the covariances of estimates. We are here considering the regression coefficients. The observations on individuals over time may well represent some process, but it does not follow that the regression coefficients must also follow a process.

Acknowledgement

This work was supported in part by NIH Training Grant T32-CA-09-565.

References


