SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 1129

June 1995

# Distribution Theory for Group Sequential Analysis of General Linear Models[1]

by

Christopher Jennison[2] and
Bruce W. Turnbull

[1]This research was supported in part by Grant R01 CA 66218 from the U.S. National Institutes of Health.
[2]School of Mathematical Sciences, University of Bath, Bath BA2 7AY, U. K.

# Distribution Theory for Group Sequential Analysis of General Linear Models

Christopher Jennison
School of Mathematical Sciences,
University of Bath, Bath BA2 7AY, U. K.
and
Bruce W. Turnbull
School of Operations Research and Industrial Engineering
Cornell University, Ithaca, New York 14853-3801, U. S. A.

## 1 Introduction

In this paper, we derive the joint distribution of the sequence of estimates of the parameter vector $\theta$ in a normal general linear model when data accumulate over a series of analyses. This sequence of estimates has a remarkably simple covariance structure, even when observations are correlated, allowing standard group sequential tests to be applied in very general settings. If variances and covariances of the observations depend on an unknown scale factor $\sigma^2$, the joint distribution of the sequence of estimates of $\theta$ and $\sigma^2$ is required in order to construct sequential $t$-tests. We show that this joint distribution has a simple form, again even in the case of correlated observations, and a general treatment of group sequential $t$-tests can be obtained. Our results also provide a basis for group sequential $\chi^2$ and $F$-tests appropriate to the cases of known and unknown variance, respectively.

We consider the situation where univariate observations $Y_1, Y_2, \ldots$ are normally distributed with means depending on a parameter vector $\theta = (\theta_1, \ldots, \theta_p)^T$. In a group sequential study with a maximum of $K$ groups of observations, we denote the total number of observations in the first $k$ groups by $n_k$, $k = 1, \ldots, K$, thus $n_1 < \ldots < n_K$, and we denote by $Y^{(k)} = (Y_1, \ldots, Y_{n_k})^T$ the vector of observations available at the $k$th analysis. We assume that the full vector of $n_K$

observations, $Y^{(K)}$, has a multivariate normal distribution with design matrix $X^{(K)}$ and variance matrix $\Sigma^{(K)}\sigma^2$, where $X^{(K)}$ and $\Sigma^{(K)}$ are known. At each analysis, $k = 1, \ldots, K$, we observe $Y^{(k)} \sim N(X^{(k)}\theta, \Sigma^{(k)}\sigma^2)$, where $X^{(k)}$ and $\Sigma^{(k)}$ can be deduced from $X^{(K)}$ and $\Sigma^{(K)}$ by extracting the elements relating to the first $n_k$ components of $Y^{(K)}$.

If $\theta$ is estimable from $Y^{(k)}$, the maximum likelihood estimate based on $Y^{(k)}$ is the generalised least squares estimate

$$\hat{\theta}^{(k)} = (X^{(k)T}\Sigma^{(k)^{-1}}X^{(k)})^{-1}X^{(k)T}\Sigma^{(k)^{-1}}Y^{(k)}, \quad k = 1, \ldots, K. \tag{1}$$

If, in addition, $\sigma^2$ is unknown and $n_k > p$, the standard estimate of $\sigma^2$ based on the residuals at analysis $k$ is

$$\hat{\sigma}^{2(k)} = \frac{S^{(k)}}{n_k - p}$$

where

$$\begin{aligned} S^{(k)} &= (Y^{(k)} - X^{(k)}\hat{\theta}^{(k)})^T\Sigma^{(k)^{-1}}(Y^{(k)} - X^{(k)}\hat{\theta}^{(k)}) \\ &\sim \sigma^2\chi^2_{n_k-p}, \quad k = 1, \ldots, K. \end{aligned} \tag{2}$$

In Section 2 we derive the joint distribution of the sequence of estimates $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$. In Section 3 we consider the case of unknown variance and obtain the joint distribution of the sequence of pairs $(\hat{\theta}^{(k)}, \hat{\sigma}^{2(k)})$, $k = 1, \ldots, K$. Construction of a group sequential $t$-test of the hypothesis $c^T\theta = 0$, for a $p \times 1$ vector $c$, requires the joint distribution of the sequence of pairs $(c^T\hat{\theta}^{(k)}, \hat{\sigma}^{2(k)})$, $k = 1, \ldots, K$, and we derive this joint distribution in Section 4. In order to test the hypothesis $C^T\theta = 0$ for a $p \times m$ matrix $C$, a group sequential $\chi^2$ test is required if $\sigma^2$ is known and a group sequential $F$-test if $\sigma^2$ is unknown. We provide the theory underlying such tests in Sections 5 and 6. Note that all the joint distributions we consider are for full sequences of $K$ statistics. In the context of a group sequential test, our results are not conditional on the experiment continuing to each analysis, rather, they concern the values that statistics would have taken had the stopping rule not been applied. Nevertheless, these results provide the necessary basis for calculating properties of group sequential tests.

Our general result on the joint distribution of the sequence of estimates $\{\hat{\theta}^{(k)}; k = 1, \ldots, K\}$ provides a theoretical base for the sequential analysis of general linear models. It includes, as special cases, results for specific longitudinal data models obtained by Armitage, Stratton and Worthington (1985), Lee and DeMets (1991) and Reboussin, Lan and DeMets (1992). This result also enables us to extend the theory developed by Jennison and Turnbull (1991) for group sequential $\chi^2$ tests of $H_0$: $\theta = 0$ when observations are independent, multivariate $N(\theta, I_p)$ to tests of hypotheses of the form $C^T\theta = 0$ when observations follow a general linear model with parameter vector $\theta$.

Our results for the case of unknown $\sigma^2$ generalise the theory developed by Jennison and Turnbull (1991) for group sequential $t$-tests when observations are independent, univariate $N(\theta, \sigma^2)$ and for group sequential $F$-tests when observations are independent, multivariate $N(\theta, I_p\sigma^2)$. We show that their results extend, in very much the same form, to general linear models with correlated observations. We also explain the modifications that are needed when some elements of the parameter vector $\theta$ are unestimable at early analyses and when the variance matrix of the difference between successive estimates, $\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}$, is not of full rank.

These results are of fundamental importance to practical applications of sequential analysis. They provide a formal basis for sequential treatment comparisons adjusted for the effects of baseline covariates and of stratification factors; they allow the sequential analysis of a coefficient in a multiple linear regression model; they underlie the sequential analysis of general linear mixed models for longitudinal data, dealing automatically with staggered entry and the resulting variation between subjects in lengths of follow-up. Our theory for the case of unknown $\sigma^2$ provides the basis for a more accurate approach to the sequential analysis of normally distributed measurements than that based on methods derived for the case of known variance.

# 2 The joint distribution of $\{\hat{\theta}^{(k)}; k = 1, \ldots, K\}$

Since each $\hat{\theta}^{(k)}$ is a linear function of $Y^{(K)}$, the elements of the sequence of vectors $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(K)}$ have a multivariate normal joint distribution. Marginally, for each $k = 1, \ldots, K$,

$$E(\hat{\theta}^{(k)}) = \theta$$

and

$$Var(\hat{\theta}^{(k)}) = V_k\sigma^2,$$

where

$$V_k = (X^{(k)T}\Sigma^{(k)^{-1}}X^{(k)})^{-1}. \quad (3)$$

The covariance structure of the estimates $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$, is established by the following Theorem.

**Theorem 1** *Suppose* $Y^{(K)} = (Y_1, \ldots, Y_{n_K})^T \sim N(X^{(K)}\theta, \Sigma^{(K)}\sigma^2)$ *with non-singular variance matrix* $\Sigma^{(K)}\sigma^2$ *and the first* $n_k$ *elements of* $Y^{(K)}$ *are available at analyses* $k = 1, \ldots, K$. *Denote the generalised least squares estimate of* $\theta$ *at analysis* $k$ *by* $\hat{\theta}^{(k)} = (X^{(k)T}\Sigma^{(k)^{-1}}X^{(k)})^{-1}X^{(k)T}\Sigma^{(k)^{-1}}Y^{(k)}$. *Then, for* $1 \le k_1 \le k_2 \le K$,

$$Cov(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) = Var(\hat{\theta}^{(k_2)}). \quad (4)$$

**Proof**

Since the data available at analysis $k_1$ are a subset of the data available at analysis $k_2$, we can write

$$\hat{\theta}^{(k_1)} = M^T Y^{(k_2)}$$

for some $n_{k_2} \times p$ matrix $M$. As $\hat{\theta}^{(k_1)}$ is an unbiased estimate of $\theta$,

$$E(M^T Y^{(k_2)}) = M^T X^{(k_2)} \theta = \theta \quad \textit{for all} \;\; \theta$$

and we can deduce

$$M^T X^{(k_2)} = I_p,$$

where $I_p$ denotes the $p \times p$ identity matrix. Hence,

$$\begin{aligned}
Cov(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) &= Cov(\, M^T Y^{(k_2)},\, (X^{(k_2)^T} \Sigma^{(k_2)^{-1}} X^{(k_2)})^{-1} X^{(k_2)^T} \Sigma^{(k_2)^{-1}} Y^{(k_2)}\,) \\
&= M^T\, Var(Y^{(k_2)})\, \Sigma^{(k_2)^{-1}} X^{(k_2)} (X^{(k_2)^T} \Sigma^{(k_2)^{-1}} X^{(k_2)})^{-1} \\
&= (X^{(k_2)^T} \Sigma^{(k_2)^{-1}} X^{(k_2)})^{-1} \sigma^2 = Var(\hat{\theta}^{(k_2)}),
\end{aligned}$$

as required. □

Note that Equation (4) can be rewritten as

$$Cov(\hat{\theta}^{(k_2)}, \hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}) = 0. \tag{5}$$

showing that $\hat{\theta}^{(k_2)}$ and $\hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}$ are independent. We shall see the importance of this independence in deriving later results. It also follows from (4) that the sequence $\{V_k^{-1} \hat{\theta}^{(k)}; k = 1, \ldots, K\}$ has independent increments and, hence, that $\{\hat{\theta}^{(k)}; k = 1, \ldots, K\}$ is a Markov sequence.

# 3 The joint distribution of $\{(\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$

We start by developing theory for the case of independent observations $Y_i$, $i = 1, \ldots, n_K$, with common variance $\sigma^2$. We shall show later, in proving our Theorem 2, that the theory generalises to the case of correlated $Y_i$s. We consider analyses 1 and 2 of a group sequential test as an example of a general pair of successive analyses, $k$ and $k+1$, assuming for now that $\theta$ is estimable at analysis 1. This can be replaced by a weaker requirement that certain linear components of $\theta$ are estimable when considering specific sequential $t$ and $F$-tests.

The least squares estimate of $\theta$ at analysis 1 is

$$\hat{\theta}^{(1)} = (X^{(1)T} X^{(1)})^{-1} X^{(1)T} Y^{(1)} = A^T Y^{(2)},$$

where

$$A = \begin{pmatrix} X^{(1)}(X^{(1)T}X^{(1)})^{-1} \\ 0_{(n_2-n_1)\times p} \end{pmatrix}$$

in the usual notation for a partitioned matrix and with $0_{n\times m}$ denoting an $n \times m$ matrix of zeroes. At analysis 2,

$$\hat{\theta}^{(2)} = (X^{(2)T}X^{(2)})^{-1}X^{(2)T}Y^{(2)} = B^TY^{(2)},$$

where $B = X^{(2)}(X^{(2)T}X^{(2)})^{-1}$.

We shall decompose $Y^{(1)}$ and $Y^{(2)}$ into sums of projections onto subspaces associated with the estimates $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ and various components of error sums of squares. We first define the projection which takes $Y^{(2)}$ onto the space spanned by the columns of $A$,

$$P_1 = A(A^TA)^{-1}A^T.$$

Although this is a projection of $Y^{(2)}$, it gives zero weight to the observations $Y_{n_1+1},\ldots,Y_{n_2}$ which only become available at analysis 2, thus the top left $n_1\times n_1$ submatrix of $P_1$ is the projection matrix for $Y^{(1)}$ onto the model space at analysis 1. Note that $P_1$ satisfies the determining properties of a projection matrix in that it is symmetric, $P_1^T = P_1$, and idempotent, $P_1^2 = P_1$. It is easily checked that $P_1Y = Y$ if $Y$ is in the column space of $A$, and $P_1Y = 0$ if $Y$ is orthogonal to this space. Recall that a general $n \times n$ projection matrix $P$ of rank $\nu$ has $\nu$ eigen-values equal to 1 and $n-\nu$ eigen-values equal to zero. If $Y$ is multivariate normal with $Var(Y) = \sigma^2 I_n$, then $\|PY\|^2/\sigma^2$ has a non-central $\chi^2_\nu$ distribution or, if $E(PY) = 0$, a central $\chi^2_\nu$ distribution. (See Seber, 1980, Chs 1–3 for a detailed discussion of linear models, projection matrices and the multivariate normal distribution.) We define the projection associated with $\hat{\theta}^{(2)}$ in the same way,

$$P_2 = B(B^TB)^{-1}B^T.$$

The following Lemma is useful in establishing independence between projections. Since the proof is no harder, we prove this for the case of general $Var(Y)$.

**Lemma 1** *Suppose $Y \sim N(X\theta, \Sigma\sigma^2)$ and $\hat{\theta}$ is the minimum variance linear unbiased estimate of $\theta$ based on $Y$. If $Q$ is an $r \times n$ matrix such that $E(QY) = 0$ for all $\theta$, then*

$$Cov(\hat{\theta}, QY) = 0.$$

**Proof**

The Gauss-Markov theorem (see, for example, Seber, 1980, p. 18) establishes the existence of a minimum variance linear unbiased estimate of $\theta$, *i.e.*, an estimate which minimises $Var(c^T\hat{\theta})$ within this class for all $p$-vectors $c$. Suppose $\hat{\theta}$ is this minimum variance estimate and $Cov(\hat{\theta}_i, q^T Y) \neq 0$ for some $1 \leq i \leq p$ and row $q^T$ of $Q$. Let $\tilde{\theta}_j = \hat{\theta}_j$ for $j \neq i$ and

$$\tilde{\theta}_i = \hat{\theta}_i + \epsilon\, q^T Y.$$

Then $E(\tilde{\theta}) = \theta$ and

$$Var(\tilde{\theta}_i) = Var(\hat{\theta}_i) + 2\epsilon\, Cov(\hat{\theta}_i, q^T Y) + \epsilon^2 q^T Var(Y) q.$$

For $\epsilon$ sufficiently close to zero and of opposite sign to $Cov\ (\hat{\theta}_i, q^T Y)$ this gives $Var(\tilde{\theta}_i) < Var(\hat{\theta}_i)$, contradicting the minimum variance property of $\hat{\theta}$. $\square$

This Lemma offers an alternative proof of the result in Section 2 that $Cov(\hat{\theta}^{(k_2)}, \hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}) = 0$ since $\hat{\theta}^{(k_2)}$ is the minimum variance unbiased estimate of $\theta$ based on $Y^{(k_2)}$, $\hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}$ is of the form $QY^{(k_2)}$ and $E(\hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}) = 0$ for all $\theta$.

If $Y$ is a vector random variable with $Var(Y) = \sigma^2 I_n$ and $\sigma^2 > 0$, two $n \times n$ projection matrices $P_i$ and $P_j$ satisfy $P_i P_j = 0$ if and only if $Cov(P_i Y, P_j Y) = 0$. We then say that $P_i$ and $P_j$ are orthogonal projections and write $P_i \perp P_j$.

The residual sum of squares at analysis 1 is the squared norm of the projection of $Y^{(1)}$ onto the space orthogonal to the column space of $X^{(1)}$, *i.e.*, $\|P_3 Y^{(2)}\|^2$ where

$$P_3 Y^{(2)} = \begin{pmatrix} Y^{(1)} \\ 0_{(n_2-n_1)\times 1} \end{pmatrix} - P_1 Y^{(2)} = \begin{pmatrix} I_{n_1 \times n_1} & 0_{n_1 \times (n_2-n_1)} \\ 0_{(n_2-n_1)\times n_1} & 0_{(n_2-n_1)\times(n_2-n_1)} \end{pmatrix} Y^{(2)} - P_1 Y^{(2)}. \qquad (6)$$

Since the last $n_2 - n_1$ columns of $A^T$ and of $P_3$ contain only zeroes, $\hat{\theta}^{(1)}$ and $P_3 Y^{(2)}$ can be expressed as matrix multiples of $Y^{(1)}$. Also, $E(P_3 Y^{(2)}) = 0$ for all $\theta$. Hence, applying Lemma 1 at analysis 1 with $Y = Y^{(1)}$ gives $Cov(\hat{\theta}^{(1)}, P_3 Y^{(2)}) = 0$. Since $P_1 Y^{(2)} = A(A^T A)^{-1}\hat{\theta}^{(1)}$, we have $Cov(P_1 Y^{(2)}, P_3 Y^{(2)}) = 0$ and $P_1 \perp P_3$. This is, of course, the standard result concerning the independence of parameter estimates and residuals in a linear model.

We now show that the residual sum of squares at analysis 2 can be decomposed into the residual sum of squares at analysis 1, a component associated with $\hat{\theta}^{(2)} - \hat{\theta}^{(1)}$, and further "pure error" in the additional observations at analysis 2. The difference between parameter estimates at analyses 1 and 2 is

$$\hat{\theta}^{(2)} - \hat{\theta}^{(1)} = (B - A)^T Y^{(2)}.$$

In general, $B - A$ may have less than full column rank, $p$. If $B - A$ has column rank $\tilde{p}$, we remove $p - \tilde{p}$ columns from $B - A$ which are linearly dependent on the remaining columns to create an $n_2 \times \tilde{p}$ matrix $D$ of full column rank and with the same column space as $B - A$. Thus,

$$D = (B - A)E,$$

where $E$ is a $p \times \tilde{p}$ matrix with a single one and $p-1$ zeroes in each column, and

$$D^T Y^{(2)} = E^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}).$$

Let $F$ be the $\tilde{p} \times p$ matrix such that

$$\hat{\theta}^{(2)} - \hat{\theta}^{(1)} = F^T D^T Y^{(2)}.$$

Note that if $B - A$ has full column rank, $E = F = I_p$ and $D$ is simply $B - A$. We define the projection associated with $\hat{\theta}^{(2)} - \hat{\theta}^{(1)}$ to be

$$P_4 = D(D^T D)^{-1} D^T.$$

The following orthogonalities can now be deduced.

$P_2 \perp P_3$ : Apply Lemma 1 to $\hat{\theta}^{(2)}$ with $Y = Y^{(2)}$ and $Q = P_3$.

$P_2 \perp P_4$ : Note that $P_2 Y^{(2)} = B(B^T B)^{-1}\hat{\theta}^{(2)}$, $P_4 Y^{(2)} = D(D^T D)^{-1} E^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})$ and, by Theorem 1, $Cov(\hat{\theta}^{(2)}, \hat{\theta}^{(2)} - \hat{\theta}^{(1)}) = 0$. Hence, $Cov(P_2 Y^{(2)}, P_4 Y^{(2)}) = 0$.

$P_3 \perp P_4$ : Note that $P_4 = D(D^T D)^{-1} E^T(B^T - A^T) = D(D^T D)^{-1} E^T(B^T P_2 - A^T P_1)$ and we already have $P_3 \perp P_2$ and $P_3 \perp P_1$.

Since $P_2$, $P_3$ and $P_4$ are mutually orthogonal, it is easily checked that

$$P_5 = I - P_2 - P_3 - P_4$$

is a projection orthogonal to $P_2$, $P_3$ and $P_4$. This projection represents the "pure error" in observations recorded between analysis 1 and analysis 2. Now, $\hat{\theta}^{(1)} = \hat{\theta}^{(2)} - (\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) = B^T P_2 Y^{(2)} - F^T D^T P_4 Y^{(2)}$. Hence,

$$P_1 Y^{(2)} = A(A^T A)^{-1}\hat{\theta}^{(1)} = A(A^T A)^{-1}(B^T P_2 - F^T D^T P_4) Y^{(2)}$$

and

$$P_1 P_5 = A(A^T A)^{-1}(B^T P_2 - F^T D^T P_4) P_5 = 0,$$

since $P_2 P_5 = P_4 P_5 = 0$. Thus, we have the further orthogonality $P_1 \perp P_5$. The projection matrices $P_2$, $P_3$ and $P_4$ have ranks $p$, $n_1 - p$ and $\tilde{p}$ respectively, hence $P_5$ has rank $n_2 - (p + n_1 - p + \tilde{p}) = n_2 - n_1 - \tilde{p}$.

The projections $P_2$, $P_3$, $P_4$ and $P_5$ decompose $Y^{(2)}$ into four independent components,

$$Y^{(2)} = P_2Y^{(2)} + P_3Y^{(2)} + P_4Y^{(2)} + P_5Y^{(2)}.$$

The first component, $P_2Y^{(2)}$, is associated with the fitted model at analysis 2, $\hat{\theta}^{(2)} = B^T P_2 Y^{(2)}$, and the latter three with the residual sum of squares about this fitted model,

$$S^{(2)} = \|P_3Y^{(2)}\|^2 + \|P_4Y^{(2)}\|^2 + \|P_5Y^{(2)}\|^2. \tag{7}$$

The conditional distribution of $\hat{\theta}^{(2)}|\hat{\theta}^{(1)}$ can be obtained from Theorem 1. In (7), $\|P_3Y^{(2)}\|^2 = S^{(1)}$ and $\|P_5Y^{(2)}\|^2$ is distributed as $\sigma^2$ times a $\chi^2_{n_2-n_1-\tilde{p}}$ random variable, independently of everything else. The remaining term in (7) is determined by $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$, and can be written as

$$\begin{aligned}\|P_4Y^{(2)}\|^2 &= Y^{(2)T}D(D^TD)^{-1}D^TY^{(2)} \\ &= \{E^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}^T[\sigma^{-2}Var\{E^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}]^{-1}E^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}),\end{aligned}$$

since $E^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) = D^TY^{(2)}$ and $Var(D^TY^{(2)}) = \sigma^2 D^TD$. An alternative expression is possible in terms of the generalised inverse of $Var(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})$. A generalised inverse of a matrix $A$ is a matrix $A^-$ satisfying $AA^-A = A$ (see Seber, 1980, p17). Since $\hat{\theta}^{(2)} - \hat{\theta}^{(1)} = F^TD^TY^{(2)}$, where $F$ has full row rank $\tilde{p}$,

$$\begin{aligned}&(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})^T\{\sigma^{-2}Var(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}^-(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) \\ &= (F^TD^TY^{(2)})^T\{\sigma^{-2}Var(F^TD^TY^{(2)})\}^-(F^TD^TY^{(2)}) \\ &= Y^{(2)T}DF(F^TD^TDF)^-F^TD^TY^{(2)} \\ &= Y^{(2)T}D(D^TD)^{-1}(FF^T)^{-1}FF^TD^TDF(F^TD^TDF)^-F^TD^TDFF^T(FF^T)^{-1}(D^TD)^{-1}D^TY^{(2)} \\ &= Y^{(2)T}D(D^TD)^{-1}(FF^T)^{-1}FF^TD^TDFF^T(FF^T)^{-1}(D^TD)^{-1}D^TY^{(2)} \\ &= Y^{(2)T}D(D^TD)^{-1}D^TY^{(2)} \\ &= \|P_4Y^{(2)}\|^2.\end{aligned}$$

The preceding results will allow us to deduce the conditional distributions of successive pairs of sufficient statistics, $(\hat{\theta}^{(k)}, S^{(k)})$, given their predecessors. To establish the Markov nature of this sequence, we examine pairs of successive analyses. Consider first the conditional distribution of $(\hat{\theta}^{(2)}, S^{(2)})|Y^{(1)}$. Now, $\hat{\theta}^{(2)} = B^TY^{(2)} = B^TP_2Y^{(2)}$ and $P_2 \perp P_3$, so $\hat{\theta}^{(2)}$ is independent of $P_3Y^{(2)}$ and, therefore, of $S^{(1)}$. Since $Y^{(1)}$ is determined by $P_1Y^{(2)}$ and $P_3Y^{(2)}$, $\hat{\theta}^{(2)}$ depends on the whole of $Y^{(1)}$ only through $P_1Y^{(2)}$, *i.e.*, through $\hat{\theta}^{(1)}$ since $P_1Y^{(2)} = A(A^TA)^{-1}\hat{\theta}^{(1)}$. Note that in the decomposition of $S^{(2)}$,

$$S^{(2)} = S^{(1)} + (\hat{\theta}^{(2)} - \hat{\theta}^{(1)})^T\{\sigma^{-2}Var(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}^-(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) + \|P_5Y^{(2)}\|^2, \tag{8}$$

$\hat\theta^{(2)}$ and $\hat\theta^{(1)}$ are independent of $P_3Y^{(2)}$ and so the second term depends on $Y^{(1)}$ only through $\hat\theta^{(1)}$. Also, $P_5 \perp P_1$ and $P_5 \perp P_3$ implies that $P_5Y^{(2)}$ is completely independent of $Y^{(1)}$. Thus, $S^{(2)}$ depends on $Y^{(1)}$ only through $\hat\theta^{(1)}$ and $S^{(1)}$ and, therefore, the pair $(\hat\theta^{(2)}, S^{(2)})$ depends on $Y^{(1)}$ only through $(\hat\theta^{(1)}, S^{(1)})$. Similarly, examining analyses $k$ and $k+1$ for general $k$, we can deduce that $(\hat\theta^{(k+1)}, S^{(k+1)})$ depends on all of $Y^{(k)}$ only through $\hat\theta^{(k)}$ and $S^{(k)}$, in particular, there is no additional dependence on $(\hat\theta^{(1)}, \dots, \hat\theta^{(k-1)})$ and $(S^{(1)}, \dots, S^{(k-1)})$ and so the sequence $\{(\hat\theta^{(k)}, S^{(k)}); k = 1, \dots, K\}$ is Markov. It is also seen that $\hat\theta^{(k+1)}$ is independent of $S^{(k)}$ and, thus, transition probabilities for the sequence $(\hat\theta^{(k)}, S^{(k)})$ are conveniently described by the conditional distributions of $\hat\theta^{(k+1)}$ given $\hat\theta^{(k)}$ and of $S^{(k+1)}$ given $\hat\theta^{(k+1)}$, $\hat\theta^{(k)}$ and $S^{(k)}$.

We are now in a position to prove the general Theorem.

**Theorem 2** *Suppose $Y^{(K)} = (Y_1, \dots, Y_{n_K})^T \sim N(X^{(K)}\theta, \Sigma^{(K)}\sigma^2)$ with non-singular variance matrix $\Sigma^{(K)}\sigma^2$ and the first $n_k$ elements of $Y^{(K)}$ are available at analyses $k = 1, \dots, K$. Suppose $\theta$ is estimable from $Y^{(1)} = (Y_1, \dots, Y_{n_1})^T$ and let $\hat\theta^{(k)}$, $S^{(k)}$ and $V_k = Var(\hat\theta^{(k)})/\sigma^2$, $k = 1, \dots, K$, be as defined by (1), (2) and (3).*

*Then, the sequence $\{(\hat\theta^{(k)}, S^{(k)}); k = 1, \dots, K\}$ is Markov, $\hat\theta^{(1)} \sim N(\theta, V_1\sigma^2)$, $S^{(1)} \sim \sigma^2\chi^2_{n_1-p}$ and, for $k = 1, \dots, K-1$,*

$$\hat\theta^{(k+1)} | \hat\theta^{(k)}, S^{(k)} \sim N(\theta + V_{k+1}V_k^{-1}(\hat\theta^{(k)} - \theta), V_{k+1}\sigma^2 - V_{k+1}V_k^{-1}V_{k+1}\sigma^2) \tag{9}$$

*and*

$$S^{(k+1)} | \hat\theta^{(k+1)}, \hat\theta^{(k)}, S^{(k)} \sim S^{(k)} + (\hat\theta^{(k+1)} - \hat\theta^{(k)})^T(V_k - V_{k+1})^-(\hat\theta^{(k+1)} - \hat\theta^{(k)}) + \sigma^2\chi^2_{n_{k+1}-n_k-\tilde p_{k+1}} \tag{10}$$

*where $\tilde p_{k+1}$ is the rank of $(V_k - V_{k+1})$.*

**Proof**

For the case of uncorrelated observations with $\Sigma^{(K)} = I_{n_K}$, we have established the Markov property of the sequence of pairs $(\hat\theta^{(k)}, S^{(k)})$ and the independence of $\hat\theta^{(k+1)}$ and $S^{(k)}$. The conditional distribution of $\hat\theta^{(k+1)}|\hat\theta^{(k)}$ follows from Theorem 1. The conditional distribution of $S^{(k+1)}|\hat\theta^{(k+1)}, \hat\theta^{(k)}, S^{(k)}$ is derived in the same way that (8) was obtained for the case $k = 1$ but we have simplified the second term using the relation

$$\sigma^{-2}Var(\hat\theta^{(k+1)} - \hat\theta^{(k)}) = \sigma^{-2}\{Var(\hat\theta^{(k)}) - Var(\hat\theta^{(k+1)})\} = V_k - V_{k+1},$$

which follows from (4). Note that $\tilde p_{k+1} = p$ if $\theta$ is estimable from the observations recorded between analyses $k$ and $k+1$. If not, $\tilde p_{k+1}$ is the rank of the projection matrix associated with $(\hat\theta^{(k+1)} - \hat\theta^{(k)})$, *i.e.*, the number of linearly independent components of $\theta$ estimable from the data accrued between analyses $k$ and $k+1$.

Suppose now that observations $Y_1, \ldots, Y_{n_K}$ are correlated and $Var(Y^{(K)}) = \Sigma^{(K)}\sigma^2$. (The Theorem assumes that $\Sigma^{(K)}\sigma^2$ is non-singular but note that if $Var(Y^{(K)})$ were singular, one could simply remove those $Y_i$ expressible as linear combinations of previous elements of $Y$ to achieve this.) Define the sequence of transformed observations $Z_1, \ldots, Z_{n_K}$ by

$$Z_i = \gamma_i(Y_i - \sum_{j=1}^{i-1} \lambda_{ij} Z_j),$$

where $\lambda_{ij} = \sigma^{-2} Cov(Y_i, Z_j)$ and

$$\gamma_i = \{\sigma^{-2} Var(Y_i) - \sum_{j=1}^{i-1} \lambda_{ij}^2\}^{-1/2}, \quad i = 1, \ldots, n_K.$$

It is easily verified that the $Z_i$, $i = 1, \ldots, n_K$, are independent and each has variance $\sigma^2$. Moreover, $Z^{(k)} = (Z_1, \ldots, Z_{n_k})$ is a non-singular linear transformation of $Y^{(k)} = (Y_1, \ldots, Y_{n_k})$ so either provides a complete representation of the data available at analysis $k$.

Let $Q^{(k)}$ be the transformation matrix such that

$$Z^{(k)} = Q^{(k)} Y^{(k)} \sim N(Q^{(k)} X^{(k)} \theta, I_{n_k} \sigma^2).$$

Let $\tilde{X}^{(k)} = Q^{(k)} X^{(k)}$ and define

$$\tilde{\theta}^{(k)} = (\tilde{X}^{(k)T} \tilde{X}^{(k)})^{-1} \tilde{X}^{(k)T} Z^{(k)},$$

$$\tilde{S}^{(k)} = (Z^{(k)} - \tilde{X}^{(k)} \tilde{\theta}^{(k)})^T (Z^{(k)} - \tilde{X}^{(k)} \tilde{\theta}^{(k)})$$

and

$$\tilde{V}_k = Var(\tilde{\theta}^{(k)})/\sigma^2, \quad k = 1, \ldots, K.$$

Using the relation $Q^{(k)T} Q^{(k)} = \Sigma^{(k)^{-1}}$, it is seen that $\tilde{\theta}^{(k)} = \hat{\theta}^{(k)}$, $\tilde{S}^{(k)} = S^{(k)}$ and $\tilde{V}_k = V_k$ where $\hat{\theta}^{(k)}$ and $S^{(k)}$ are obtained from $Y^{(k)}$ according to the usual formulae, (1) and (2), and $V_k = Var(\hat{\theta}^{(k)})/\sigma^2$. This is in keeping with the fact that linear transformation of an observation vector does not affect generalised least squares parameter estimates. The results for the case of uncorrelated observations hold for the sequence of estimates, $\{(\tilde{\theta}^{(k)}, \tilde{S}^{(k)}); k = 1, \ldots, K\}$, obtained from the transformed vector, $Z^{(K)}$, thus, the identical sequence $\{(\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ is also Markov with the transition probabilities specified by (9) and (10). This establishes the Theorem for the case of correlated observations.

We would stress that the transformation to uncorrelated observations is only introduced as a device to prove this result. There is no need to transform data in practice: $\hat{\theta}^{(k)}$, $S^{(k)}$ and $V_k$ can, and should, be obtained directly from the original data, $Y_1, \ldots, Y_{n_K}$. □

# 4 Group sequential $t$-tests

The results of Section 3 concern the joint distribution of the sequence of pairs $\{(\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$, where the $\hat{\theta}^{(k)}$ are $p$-dimensional. Sequential $t$-tests involve only a one-dimensional component of $\hat{\theta}^{(k)}$ at each analysis. For example, in testing the hypothesis $H_0 : c^T\theta = \gamma$, the $t$-statistic at analysis $k$ is

$$T^{(k)} = \frac{c^T\hat{\theta}^{(k)} - \gamma}{\sqrt{\{c^T V_k c S^{(k)}/(n_k - p)\}}}, \quad k = 1, \ldots, K,$$

and a typical two-sided group sequential test stops to reject $H_0$ at analysis $k$ if $|T^{(k)}| > a_k$ for some sequence of critical values $a_1, \ldots, a_K$. Since the sequence of $t$-statistics, $\{T^{(k)}; k = 1, \ldots, K\}$ is *not* Markov, it is useful in sequential calculations to consider a sequence of bivariate statistics which is Markov and the sequence of pairs $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ is the natural choice.

To see that the sequence of $t$-statistics is not Markov, compare the conditional distributions of $T^{(3)}|T^{(2)} = 0$ and of $T^{(3)}|T^{(1)} = 0, T^{(2)} = 0$ when $Var(Y^{(K)}) = I_{n_K}\sigma^2$, as in the development of Section 3. In the first case we can deduce that $c^T\hat{\theta}^{(2)} = \gamma$, so $\hat{\theta}^{(2)}$ follows the appropriate multivariate normal distribution conditional on $c^T\hat{\theta}^{(2)} = \gamma$, and $S^{(2)} \sim \sigma^2\chi^2_{n_2-p}$. In the second case we can deduce that $c^T\hat{\theta}^{(1)} = \gamma$ and $c^T\hat{\theta}^{(2)} = \gamma$, so we have the additional information $c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) = 0$. The estimate $\hat{\theta}^{(2)}$ follows the same multivariate normal distribution conditional on $c^T\hat{\theta}^{(2)} = \gamma$ as in the first case since, by (5), its distribution is independent of $c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})$. However, since $c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})$ is known to be zero, one component of the second term in the right hand side of (8) will be smaller than average, with the implication that the denominator of $T^{(3)}$ will tend to be smaller too. In fact, it follows from equation (12), which we derive later in this Section, that conditionally, $S^{(2)} \sim \sigma^2\chi^2_{n_2-p-1}$. Thus, the numerator of $T^{(3)}$ has the same conditional distribution in both situations but the denominator conditional on $T^{(2)} = 0$ is stochastically larger than that conditional on $T^{(1)} = 0$ and $T^{(2)} = 0$. More generally, we shall see that fluctuations in $c^T\hat{\theta}^{(k)}$ contribute to the residual sum of squares and thus, if the same value $T^{(k)}$ occurs in two sequences of $t$-statistics, it is likely that a larger residual sums of squares will be associated with the sequence which has shown greater variation in reaching its current value.

In order to show that the sequence of pairs $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ is Markov and to obtain its joint distribution we must refine our results concerning the sequence $\{(\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$. To do this, we decompose observation vectors further into yet more orthogonal components. As in Section 3, we first consider the case of uncorrelated observations, each of variance $\sigma^2$, and then generalise the results to correlated observations. Thus, we suppose observation vectors $Y^{(k)} \sim N(X^{(k)}\theta, I_{n_k}\sigma^2)$ are available at analyses $k = 1, \ldots, K$ and $\hat{\theta}^{(k)}$, $S^{(k)}$ and $V_k = Var(\hat{\theta}^{(k)})/\sigma^2$ are as defined by (1), (2) and (3) with $\Sigma^{(k)} = I_{n_k}$. Although we assume here that $\theta$ is estimable at analysis 1, we shall explain at the end of this Section how this assumption can be relaxed. As in Section 3, we describe results for analyses 1 and 2 as an example of a general pair of analyses, retaining the same definitions of projection matrices $P_1$ to $P_5$.

We decompose $P_1$, the projection associated with $\hat{\theta}^{(1)}$, into projections associated with $c^T\hat{\theta}^{(1)}$ and with its orthogonal complement. Let $d = Ac$, so

$$c^T\hat{\theta}^{(1)} = c^T A^T Y^{(2)} = d^T Y^{(2)},$$

and define $P_6$ and $P_7$ by

$$P_6 = d(d^T d)^{-1} d^T$$

and

$$P_7 = P_1 - P_6.$$

Note that we can write $P_6 = d(d^T d)^{-1} c^T A^T = d(d^T d)^{-1} c^T A^T P_1$, which shows that $P_6$ projects onto a subspace of the range space of $P_1$. Hence, $P_6 P_1 = P_6$, $P_7 = (I - P_6)P_1$ and $P_1 \perp P_j$ implies $P_6 \perp P_j$ and $P_7 \perp P_j$ for any projection $P_j$. Thus, $P_1 \perp P_3$ implies $P_6 \perp P_3$.

We decompose $P_2$ similarly. Setting $e = Bc$, so

$$c^T\hat{\theta}^{(2)} = c^T B^T Y^{(2)} = e^T Y^{(2)},$$

we define $P_8$ and $P_9$ by

$$P_8 = e(e^T e)^{-1} e^T$$

and

$$P_9 = P_2 - P_8.$$

Since $P_8 = e(e^T e)^{-1} c^T B^T = e(e^T e)^{-1} c^T B^T P_2$, $P_8 P_2 = P_8$, $P_9 = (I - P_8)P_2$ and $P_2 \perp P_j$ implies $P_8 \perp P_j$ and $P_9 \perp P_j$ for any projection $P_j$.

Finally we decompose $P_4$ into projections associated with $c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})$ and with the components of $\hat{\theta}^{(2)} - \hat{\theta}^{(1)}$ orthogonal to this, assuming for the moment that $c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})$ is non-degenerate, a necessary and sufficient condition for which is that $c^T\theta$ be estimable from the observations recorded between analyses $k$ and $k+1$. Let $f = DFc$, so

$$c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) = c^T F^T D^T Y^{(2)} = f^T Y^{(2)}$$

and define $P_{10}$ and $P_{11}$ by

$$P_{10} = f(f^T f)^{-1} f^T$$

and

$$P_{11} = P_4 - P_{10}.$$

Since $P_{10} = f(f^T f)^{-1} c^T F^T D^T = f(f^T f)^{-1} c^T F^T D^T P_4$, we see that $P_{10}$ projects onto a subspace of the range space of $P_4$, $P_{10} P_4 = P_{10}$ and $P_{11} = (I - P_{10}) P_4$. It follows immediately that $P_{11} \perp P_{10}$ and that $P_4 \perp P_j$ implies $P_{10} \perp P_j$ and $P_{11} \perp P_j$ for any projection $P_j$.

The elements of $Y^{(2)}$ observable at analysis 1 have now been decomposed into orthogonal projections $P_6 Y^{(2)}$, $P_7 Y^{(2)}$ and $P_3 Y^{(2)}$ associated with $c^T \hat{\theta}^{(1)}$, elements of $\hat{\theta}^{(1)}$ uncorrelated with $c^T \hat{\theta}^{(1)}$ and residual error, respectively. At analysis 2, we have

$$Y^{(2)} = P_8 Y^{(2)} + P_9 Y^{(2)} + P_3 Y^{(2)} + P_{10} Y^{(2)} + P_{11} Y^{(2)} + P_5 Y^{(2)},$$

where $P_8 + P_9 = P_2$ and $P_{10} + P_{11} = P_4$. The $t$-statistic at analysis 1 involves $P_6 Y^{(2)}$ and $P_3 Y^{(2)}$ but not $P_7 Y^{(2)}$ and the $t$-statistic at analysis 2 involves $P_8 Y^{(2)}$, $P_3 Y^{(2)}$, $P_{10} Y^{(2)}$, $P_{11} Y^{(2)}$ and $P_5 Y^{(2)}$ but not $P_9 Y^{(2)}$, the projection associated with elements of $\hat{\theta}^{(2)}$ uncorrelated with $c^T \hat{\theta}^{(2)}$. Thus, in computations for group sequential $t$-tests, we need to know the conditional distribution of $(P_8 Y^{(2)}, P_3 Y^{(2)}, P_{10} Y^{(2)}, P_{11} Y^{(2)}, P_5 Y^{(2)})$ given $(P_6 Y^{(2)}, P_3 Y^{(2)})$.

Note first that $P_2 \perp P_3$ implies $P_8 \perp P_3$ and so $c^T \hat{\theta}^{(2)}$ is independent of $S^{(1)}$. Also, $P_4 \perp P_3$ implies $P_{10} \perp P_3$ and $P_{11} \perp P_3$, we have already established $P_5 \perp P_3$, and the conditional distribution of $P_3 Y^{(2)}$ given $P_3 Y^{(2)}$ is, of course, deterministic. Thus, all projections of interest at analysis 2 are independent of $P_3 Y^{(2)}$ and it remains to find the conditional distribution of $(P_8 Y^{(2)}, P_{10} Y^{(2)}, P_{11} Y^{(2)}, P_5 Y^{(2)})$ given $P_6 Y^{(2)}$. We do this by noting the conditional distribution of $P_8 Y^{(2)}$ given $P_6 Y^{(2)}$, showing that $P_{10} Y^{(2)}$ is determined by $P_8 Y^{(2)}$ and $P_6 Y^{(2)}$, and establishing that $P_{11} Y^{(2)}$ and $P_5 Y^{(2)}$ are independent of each other and of $P_{10} Y^{(2)}$, $P_8 Y^{(2)}$ and $P_6 Y^{(2)}$.

The conditional distribution of $P_8 Y^{(2)}$ given $P_6 Y^{(2)}$ follows from that of $c^T \hat{\theta}^{(2)}$ given $c^T \hat{\theta}^{(1)}$, since $P_8 Y^{(2)} = e(e^T e)^{-1} c^T \hat{\theta}^{(2)}$ and $c^T \hat{\theta}^{(1)} = d^T P_6 Y^{(2)}$. From Theorem 1, we have

$$c^T \hat{\theta}^{(2)} | c^T \hat{\theta}^{(1)} \sim N\left( c^T \theta + \frac{c^T V_2 c}{c^T V_1 c} c^T (\hat{\theta}^{(1)} - \theta), c^T V_2 c \sigma^2 - \frac{(c^T V_2 c)^2}{c^T V_1 c} \sigma^2 \right).$$

Now,

$$P_{10} Y^{(2)} = f(f^T f)^{-1} c^T (\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) = f(f^T f)^{-1} (e^T P_8 Y^{(2)} - d^T P_6 Y^{(2)})$$

is determined by $P_8 Y^{(2)}$ and $P_6 Y^{(2)}$. Alternatively,

$$\begin{aligned} \|P_{10} Y^{(2)}\|^2 &= \|f(f^T f)^{-1} c^T (\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\|^2 \\ &= (\hat{\theta}^{(2)} - \hat{\theta}^{(1)})^T c (f^T f)^{-1} c^T (\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) \\ &= \{c^T (\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}^2 [\{\sigma^{-2} Var\{c^T (\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}]^{-1}, \end{aligned} \tag{11}$$

since $Var\{c^T (\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\} = Var(f^T Y^{(2)}) = \sigma^2 f^T f$.

We have already noted that $P_{11} \perp P_{10}$. Also, $P_4 \perp P_2$ implies $P_4 \perp P_8$ and, hence, $P_{11} \perp P_8$. Since

$$\begin{aligned} P_6 Y^{(2)} &= d(d^T d)^{-1}\{c^T\hat{\theta}^{(2)} - (c^T\hat{\theta}^{(2)} - c^T\hat{\theta}^{(1)})\} \\ &= d(d^T d)^{-1}\{e^T P_8 Y^{(2)} - f^T P_{10} Y^{(2)}\}, \end{aligned}$$

we see $P_{11} \perp P_6$. Finally, $P_5 \perp P_1$ implies $P_5 \perp P_6$, $P_5 \perp P_2$ implies $P_5 \perp P_8$ and $P_5 \perp P_4$ implies $P_5 \perp P_{10}$ and $P_5 \perp P_{11}$.

We can now draw conclusions concerning the conditional distribution of $S^{(2)}$ given $S^{(1)}$, $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$. We have

$$\begin{aligned} S^{(2)} &= \|P_3 Y^{(2)} + P_{10} Y^{(2)} + P_{11} Y^{(2)} + P_5 Y^{(2)}\|^2 \\ &= \|P_3 Y^{(2)}\|^2 + \|P_{10} Y^{(2)}\|^2 + \|P_{11} Y^{(2)}\|^2 + \|P_5 Y^{(2)}\|^2 \\ &= S^{(1)} + \{c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}^2[\sigma^{-2} Var\{c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}]^{-1} + \|P_{11} Y^{(2)}\|^2 + \|P_5 Y^{(2)}\|^2 \end{aligned} \tag{12}$$

where $\|P_{11} Y^{(2)}\|^2 + \|P_5 Y^{(2)}\|^2 \sim \sigma^2 \chi^2_{n_2 - n_1 - 1}$ as long as the projection matrix $P_{10}$ has rank 1. If, however, $c^T\theta$ is not estimable from the observations recorded between analyses 1 and 2, so $c^T\hat{\theta}^{(2)} = c^T\hat{\theta}^{(1)}$ automatically, there is no need to introduce $P_{10}$, we set $P_{11} = P_4$ and

$$S^{(2)} \sim S^{(1)} + \sigma^2 \chi^2_{n_2 - n_1}.$$

In establishing the Markov nature of $\{(\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ in Section 3, we showed that $(\hat{\theta}^{(k+1)}, S^{(k+1)})$ depends on all of $Y^{(k)}$ only through $\hat{\theta}^{(k)}$ and $S^{(k)}$ and, hence, there could be no additional dependence on $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(k-1)})$ and $(S^{(1)}, \ldots, S^{(k-1)})$, which are determined by $Y^{(k)}$. We shall use a similar approach to prove that the sequence $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ is Markov but here we consider the dependence of $(c^T\hat{\theta}^{(k+1)}, S^{(k+1)})$ only on the projection of $Y^{(k)}$ which contributes to $\hat{\theta}^{(k)}$ and $S^{(k)}$. This is because components of $\hat{\theta}^{(k)}$ orthogonal to $c^T\hat{\theta}^{(k)}$ affect $S^{(k+1)}$, for example, for $k = 1$, these appear in the term $\|P_{11} Y^{(2)}\|^2$ in the decomposition of $S^{(2)}$ in Equation (12).

Let $\tilde{P}_k$ be the $n_k \times n_k$ projection matrix for $Y^{(k)}$ associated with those components of $\hat{\theta}^{(k)}$ orthogonal to $c^T\hat{\theta}^{(k)}$. Thus, for $k = 2$, $\tilde{P}_k$ is the projection matrix $P_9$ and, for $k = 1$, $\tilde{P}_k$ is the top left $n_1 \times n_1$ submatrix of $P_7$. In general $\tilde{P}_k Y^{(k)}$ represents components of $Y^{(k)}$ that are orthogonal both to $c^T\hat{\theta}^{(k)}$ and to the residuals $Y^{(k)} - X^{(k)}\hat{\theta}^{(k)}$ that determine $S^{(k)}$.

Consider first the conditional distribution of $(c^T\hat{\theta}^{(2)}, S^{(2)})$ given $(I_{n_1} - \tilde{P}_1)Y^{(1)}$ or, equivalently, given $(I_{n_2} - P_7)Y^{(2)}$, which is also equal to $(P_6 + P_3)Y^{(2)}$. Since $c^T\hat{\theta}^{(2)} = e^T P_8 Y^{(2)}$ and $P_8 \perp P_3$, $c^T\hat{\theta}^{(2)}$ depends on $(I_{n_1} - \tilde{P}_1)Y^{(1)}$ only through $P_6 Y^{(2)}$, *i.e.*, through $c^T\hat{\theta}^{(1)}$. The dependence of $S^{(2)}$ on $(I_{n_1} - \tilde{P}_1)Y^{(1)}$ can be determined from (12): we have already shown that $c^T\hat{\theta}^{(2)}$ depends on $(I_{n_1} - \tilde{P}_1)Y^{(1)}$ only through $c^T\hat{\theta}^{(1)}$ and, since $P_{11} \perp P_3$ and $P_5 \perp P_3$, $S^{(2)}$ depends on $(I_{n_1} - \tilde{P}_1)Y^{(1)}$ only through $(c^T\hat{\theta}^{(1)}, S^{(1)})$. Similar examination of analyses $k$ and $k+1$ for general $k$, shows that $(c^T\hat{\theta}^{(k+1)}, S^{(k+1)})$ depends on $(I_{n_k} - \tilde{P}_k)Y^{(k)}$ only through $(c^T\hat{\theta}^{(k)}, S^{(k)})$ and, since $(c^T\hat{\theta}^{(1)}, \ldots, c^T\hat{\theta}^{(k-1)})$ and $(S^{(1)}, \ldots, S^{(k-1)})$ are determined by $(I_{n_k} - \tilde{P}_k)Y^{(k)}$, the sequence $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\})$ is Markov. We are now in a position to prove the following Theorem.

**Theorem 3** *Suppose $Y^{(K)} = (Y_1, \ldots, Y_{n_K})^T \sim N(X^{(K)}\theta, \Sigma^{(K)}\sigma^2)$ with non-singular variance matrix $\Sigma^{(K)}\sigma^2$ and the first $n_k$ elements of $Y^{(K)}$ are available at analyses $k = 1, \ldots, K$. Suppose $\theta$ is estimable from $Y^{(1)} = (Y_1, \ldots, Y_{n_1})^T$ and let $\hat{\theta}^{(k)}$, $S^{(k)}$ and $V_k = Var(\hat{\theta}^{(k)})/\sigma^2$, $k = 1, \ldots, K$, be as defined by (1), (2) and (3).*

*Then, the sequence $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ is Markov, $c^T\hat{\theta}^{(1)} \sim N(c^T\theta, c^T V_1 c\sigma^2)$, $S^{(1)} \sim \sigma^2\chi^2_{n_1 - p}$ and, for $k = 1, \ldots, K-1$,*

$$c^T\hat{\theta}^{(k+1)} | c^T\hat{\theta}^{(k)}, S^{(k)} \sim N(c^T\theta + \frac{c^T V_{k+1} c}{c^T V_k c} c^T(\hat{\theta}^{(k)} - \theta), c^T V_{k+1} c\sigma^2 - \frac{(c^T V_{k+1} c)^2}{c^T V_k c}\sigma^2) \tag{13}$$

*and*

$$S^{(k+1)} | c^T\hat{\theta}^{(k+1)}, c^T\hat{\theta}^{(k)}, S^{(k)} \sim S^{(k)} + \{c^T(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)})\}^2 \{c^T(V_k - V_{k+1})c\}^{-1} + \sigma^2\chi^2_{n_{k+1} - n_k - 1} \tag{14}$$

*as long as $\{c^T(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)})\}$ is non-degenerate but, if $c^T\theta$ is not estimable from observations recorded between analyses $k$ and $k+1$,*

$$c^T\hat{\theta}^{(k+1)} = c^T\hat{\theta}^{(k)}$$

*and*

$$S^{(k+1)} \sim S^{(k)} + \sigma^2\chi^2_{n_{k+1} - n_k}.$$

**Proof**

We have established the Markov property of $\{(\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ for the case of uncorrelated observations with $\Sigma^{(K)} = I_{n_K}$. The above discussion of analyses 1 and 2 generalises to analyses $k$ and $k+1$ in a straightforward manner, hence $c^T\hat{\theta}^{(k+1)}$ is independent of $S^{(k)}$ and (13) follows from Theorem 1. The conditional distribution of $S^{(k+1)} | \hat{\theta}^{(k+1)}, \hat{\theta}^{(k)}, S^{(k)}$ is the generalisation of (12) with a simplification of the second term following from the relation

$$\sigma^{-2} Var\{c^T(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)})\} = \sigma^{-2} c^T\{Var(\hat{\theta}^{(k)}) - Var(\hat{\theta}^{(k+1)})\}c = c^T(V_k - V_{k+1})c.$$

If observations $Y_1, \ldots, Y_{n_K}$ are correlated and $Var(Y^{(K)}) = \Sigma^{(K)}\sigma^2$, the same argument can be used as in Theorem 2. Uncorrelated observations of variance $\sigma^2$ can be constructed from the $Y_i$s and the sequence $\{(c^T\tilde{\theta}^{(k)}, \tilde{S}^{(k)}); k = 1, \ldots, K\}$ based on these observations will have the properties stated in the Theorem. However, the sequence $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \dots, K\}$ based on the original observations, $Y_1, \dots, Y_{n_K}$, is exactly the same as the sequence based on the transformed variables and, therefore, has the same properties. $\square$

Note that $\sigma^2$ appears as a scale factor in the variance of each $c^T\hat{\theta}^{(k)}$ and in the distribution of each $S^{(k)}$, $k = 1, \dots, K$. Thus, the distribution of the sequence $\{T^{(k)}; k = 1, \dots, K\}$ is independent of $\sigma^2$ under $H_0$: $c^T\theta = \gamma$ and any convenient value of $\sigma^2$ can be used in calculating properties of a group sequential $t$-test under the null hypothesis.

It remains to discuss the case where $c^T\theta$ is estimable but the whole of $\theta$ cannot be estimated at the earliest analyses. This is likely to occur in longitudinal studies in which initial analyses take place before accrual is complete and subject effects are not estimable for those individuals who are yet to enter the study. To review the standard theory for this situation, consider the general linear model $Y \sim N(X\theta, \Sigma\sigma^2)$ where $Y$ is a vector of length $n$ and $\theta$ a parameter vector of length $p$, and suppose $\theta$ can be decomposed as

$$\theta = G\phi + H\psi$$

where $\phi$ and $\psi$ are vectors of length $r$ and $p - r$, $G$ and $H$ are $p \times r$ and $p \times (p - r)$ matrices, $XG$ is of rank $r$ and $XH = 0$. Then

$$E(Y) = X\theta = XG\phi$$

and $\phi$ is estimable but $\psi$ is not. The generalised least squares estimate of $\phi$ is

$$\hat{\phi} = (G^T X^T \Sigma^{-1} XG)^{-1} G^T X^T \Sigma^{-1} Y \sim N(\phi, (G^T X^T \Sigma^{-1} XG)^{-1}\sigma^2)$$

and $\sigma^2$ can be estimated from

$$S = (Y - XG\hat{\phi})^T \Sigma^{-1} (Y - XG\hat{\phi}) \sim \sigma^2 \chi^2_{n-r}.$$

Since $c^T\theta = c^T G\phi + c^T H\psi$, $c^T\theta$ is estimated by

$$c^T G\hat{\phi} \sim N(c^T\theta, c^T G(G^T X^T \Sigma^{-1} XG)^{-1} G^T c\sigma^2)$$

for any $c$ satisfying $c^T H = 0$, *i.e.*, for any $c^T$ in the row space of $X$. We shall write $c^T\hat{\theta}$ to denote the above estimate of $c^T\theta$, even though $\hat{\theta}$ itself is not properly defined.

In a group sequential study where $\theta$ is not completely estimable at the early analyses but $c^T\theta$ is estimable, we can define estimates $c^T\hat{\theta}^{(k)}$ in the above manner. To determine the joint distribution of $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \dots, K\}$, we redefine the projections introduced in Section 3 and earlier in this Section. Let $r_k$ denote the number of linearly independent components of $\theta$ estimable at analysis $k$, $k = 1, \dots, K$. In studying the first two analyses, we decompose $\theta$ as

$$\theta = G\phi + H\psi + J\xi$$

where $\phi$, $\psi$ and $\xi$ are vectors of length $r_1$, $r_2-r_1$ and $p-r_2$ respectively, such that only $\phi$ is estimable at analysis 1 and $\phi$ and $\psi$, but not $\xi$, are estimable at analysis 2. We denote the corresponding estimates by $\hat{\phi}^{(1)}$, $\hat{\phi}^{(2)}$ and $\hat{\psi}^{(2)}$. Since $c^T\theta$ is estimable at analysis 1, $c^T\theta = c^T G\phi$, $c^T\hat{\theta}^{(1)} = c^T G\hat{\phi}^{(1)}$ and $c^T\hat{\theta}^{(2)} = c^T G\hat{\phi}^{(2)}$. We define $P_1$ to be the rank $r_1$ projection of $Y^{(2)}$ associated with $\hat{\phi}^{(1)}$, $P_2$ to be the rank $r_2$ projection associated with $(\hat{\phi}^{(2)T}, \hat{\psi}^{(2)T})^T$ and $P_4$ to be the rank $r_1$ projection associated with $(\hat{\phi}^{(2)} - \hat{\phi}^{(1)})$. Other projections are as defined previously, although the ranks of some are now changed. The projection $P_3$ defined by (6) has rank $n_1 - r_1$ and $P_5 = I - P_2 - P_3 - P_4$ has rank $n_2 - r_2 - (n_1 - r_1) - r_1 = n_2 - n_1 - r_2$. We retain $P_6$ and $P_8$ as the projections associated with $c^T\hat{\theta}^{(1)} = c^T G\hat{\phi}^{(1)}$ and $c^T\hat{\theta}^{(2)} = c^T G\hat{\phi}^{(2)}$ and set $P_7 = P_1 - P_6$ and $P_9 = P_2 - P_8$. Likewise, $P_{10}$ is the projection associated with $c^T\,(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}) = c^T G(\hat{\phi}^{(2)} - \hat{\phi}^{(1)})$ and $P_{11} = P_4 - P_{10}$ now has rank $r_1 - 1$. It is easily checked that the arguments used to prove orthogonalities between projections in the case where $\theta$ is estimable at analysis 1 still hold and, hence, $c^T\hat{\theta}^{(2)}$ is independent of $S^{(1)}$ and we can decompose $S^{(2)}$ into independent terms as

$$S^{(2)} = S^{(1)} + \{c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}^2[\sigma^{-2}Var\{c^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})\}]^{-1} + \|P_{11}Y^{(2)}\|^2 + \|P_5 Y^{(2)}\|^2,$$

where the sum of the ranks of $P_{11}$ and $P_5$ is $n_2 - n_1 - (r_2 - r_1) - 1$. Arguments analogous to those used in proving Theorem 3 give the following Corollary.

**Corollary 1** *Suppose $Y^{(K)} = (Y_1, \ldots, Y_{n_K})^T \sim N(X^{(K)}\theta, \Sigma^{(K)}\sigma^2)$ with non-singular variance matrix $\Sigma^{(K)}\sigma^2$ and the first $n_k$ elements of $Y^{(K)}$ are available at analyses $k = 1, \ldots, K$. Suppose $r_k$ linearly independent components of $\theta$ are estimable at analysis $k$, $k = 1, \ldots, K$, and $E(Y^{(k)}) = X^{(k)}\theta$ can be represented as $X^{(k)}R^{(k)}\eta^{(k)}$, where $R^{(k)}$ is a $p \times r_k$ matrix and $\eta^{(k)}$ an $r_k \times 1$ vector. If $c^T$ is in the row space of $X^{(1)}$, $c^T\theta$ is estimable from $Y^{(1)}$ and, therefore, at all analyses $k = 1, \ldots, K$. Denote the generalised least squares estimate of $c^T\theta$ at analysis $k$ by $c^T\hat{\theta}^{(k)} = c^T R^{(k)}\hat{\eta}^{(k)}$ and define $\Gamma_k = Var(c^T\hat{\theta}^{(k)})/\sigma^2$ and $S^{(k)} = (Y^{(k)} - X^{(k)}R^{(k)}\hat{\eta}^{(k)})^T\Sigma^{-1}(Y^{(k)} - X^{(k)}R^{(k)}\hat{\eta}^{(k)})$, $k = 1, \ldots, K$.*

*Then, the sequence $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ is Markov, $c^T\hat{\theta}^{(1)} \sim N(c^T\theta, \Gamma_1\sigma^2)$, $S^{(1)} \sim \sigma^2\chi^2_{n_1 - r_1}$ and, for $k = 1, \ldots, K-1$,*

$$c^T\hat{\theta}^{(k+1)}|c^T\hat{\theta}^{(k)}, S^{(k)} \sim N(c^T\theta + \frac{\Gamma_{k+1}}{\Gamma_k}c^T(\hat{\theta}^{(k)} - \theta), \Gamma_{k+1}\sigma^2 - \frac{\Gamma_{k+1}^2}{\Gamma_k}\sigma^2)$$

*and*

$$S^{(k+1)}|c^T\hat{\theta}^{(k+1)}, c^T\hat{\theta}^{(k)}, S^{(k)} \sim S^{(k)} + \{c^T(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)})\}^2(\Gamma_k - \Gamma_{k+1})^{-1} + \sigma^2\chi^2_{n_{k+1} - n_k - (r_{k+1} - r_k) - 1}$$

*as long as $\{c^T(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)})\}$ is non-degenerate but, if $c^T\theta$ is not estimable from observations recorded between analyses $k$ and $k+1$,*

$$c^T\hat{\theta}^{(k+1)} = c^T\hat{\theta}^{(k)}$$

*and*

$$S^{(k+1)} \sim S^{(k)} + \sigma^2 \chi^2_{n_{k+1}-n_k-(r_{k+1}-r_k)}.$$

The results of this Section generalise those of Jennison and Turnbull (1991), which were obtained for the case of independent univariate $N(\theta, \sigma^2)$ observations with a scalar parameter $\theta$; their results for the sequential $t$-test can be obtained from Theorem 3, by substituting $\theta$ for $c^T\theta$ and $\hat{\theta}^{(j)}$ for $c^T\hat{\theta}^{(j)}$ and $n_j^{-1}$ for $c^T V_j c$, for $j = k$ and $k+1$, in Equations (13) and (14). Jennison and Turnbull (1991) used their results in calculating boundaries for group sequential $t$-tests. The same computational methods can be used in the general case. However, if the sequence $\{(c^T V_k c)^{-1}; k = 1, \ldots, K\}$ has unequal increments, and especially if this sequence is unpredictable, it would be better to define tests using an "error spending" approach, as described by Lan and DeMets (1983) or Slud and Wei (1982).

# 5 Group sequential $\chi^2$ tests

We now consider group sequential tests of the hypothesis $C^T\theta = 0$ where $C$ is a $q \times p$ matrix of full row rank, $q$. As before, we assume the full data vector $Y^{(K)} \sim N(X^{(K)}\theta, \Sigma^{(K)}\sigma^2)$ and the first $n_k$ elements of $Y^{(K)}$ are available at analyses $k = 1, \ldots, K$. We restrict attention in this Section to the case of known $\sigma^2$ and where $\theta$ is fully estimable at each analysis. Our results can be extended to situations where only certain components of $\theta$ are estimable at early analyses as long as these include $C^T\theta$, the treatment following that presented at the end of Section 4.

Since each $C^T\hat{\theta}^{(k)}$ is a linear function of $Y^{(K)}$, the elements of the sequence of vectors $C^T\hat{\theta}^{(1)}, \ldots, C^T\hat{\theta}^{(K)}$ have a multivariate normal joint distribution. Marginally, for each $k = 1, \ldots, K$,

$$C^T\hat{\theta}^{(k)} \sim N(C^T\theta, C^T V_k C\sigma^2)$$

where, as before,

$$V_k = (X^{(k)T}\Sigma^{(k)^{-1}}X^{(k)})^{-1} = Var(\hat{\theta}^{(k)})/\sigma^2.$$

It follows from Theorem 1 that $Cov(C^T\hat{\theta}^{(k_1)}, C^T\hat{\theta}^{(k_2)}) = Var(C^T\hat{\theta}^{(k_2)})$ for $1 \leq k_1 \leq k_2 \leq K$ and, since the process $\{(C^T V_k C)^{-1} C^T\hat{\theta}^{(k)}; k = 1, \ldots, K\}$ has independent increments, the sequence $\{C^T\hat{\theta}^{(k)}; k = 1, \ldots, K\}$ is Markov. Further, the conditional distribution of $C^T\hat{\theta}^{(k+1)}$ given $C^T\hat{\theta}^{(k)}$ is

$$N(C^T\theta + C^T V_{k+1} C (C^T V_k C)^{-1} C^T(\hat{\theta}^{(k)} - \theta), C^T V_{k+1} C\sigma^2 - C^T V_{k+1} C (C^T V_k C)^{-1} C^T V_{k+1} C\sigma^2). \quad (15)$$

The $\chi^2$ test of $H_0$: $C^T\theta = 0$ at analysis $k$ is based on the statistic

$$\Omega^{(k)} = (C^T\hat{\theta}^{(k)})^T (C^T V_k C\sigma^2)^{-1} C^T\hat{\theta}^{(k)} \quad (16)$$

which has a $\chi_q^2$ distribution if $H_0$ is true. Jennison and Turnbull (1991) showed that, for the case of independent multivariate $N(\theta, \Sigma\sigma^2)$ observations, the sequence of $\chi^2$ statistics for testing $\theta = 0$ is Markov and successive $\chi^2$ statistics have a simple form of conditional distribution. We shall see that similar properties hold more generally if certain conditions are satisfied.

Suppose the variances $Var(C^T\hat{\theta}^{(k)}) = C^T V_k C\sigma^2$ are related by

$$C^T V_k C = \gamma_k \Lambda, \quad k = 1, \dots, K, \tag{17}$$

where $\Lambda$ is a $q \times q$ positive definite symmetric matrix and $\gamma_1, \dots, \gamma_K$ a decreasing sequence of scalars and let

$$\rho_k = \frac{\gamma_{k+1}}{\gamma_k}, \quad k = 1, \dots, K-1.$$

Define

$$W^{(k)} = \gamma_k^{-1/2}\Lambda^{-1/2}C^T\hat{\theta}^{(k)} \sim N(\gamma_k^{-1/2}\Lambda^{-1/2}C^T\theta, I_q\sigma^2), \quad k = 1, \dots, K,$$

where $\Lambda^{-1/2}$ is a symmetric matrix satisfying $\Lambda^{-1/2}\Lambda\Lambda^{-1/2} = I_q$. Thus, $\Omega^{(k)} = \|W^{(k)}\|^2/\sigma^2$.

The $W^{(k)}$ are invertible linear transforms of the $C^T\hat{\theta}^{(k)}$ and it follows that they are jointly multivariate normal and the sequence $\{W^{(k)}; k = 1, \dots, K\}$ is Markov. The conditional distribution under $H_0$: $C^T\theta = 0$ of $W^{(k+1)}$ given $W^{(k)}$ follows from that of $C^T\hat{\theta}^{(k+1)}|C^T\hat{\theta}^{(k)}$, given by (15), and is

$$W^{(k+1)}|W^{(k)} \sim N(\rho_k^{1/2}W^{(k)}, (1-\rho_k)I_q\sigma^2), \quad k = 1, \dots, K-1. \tag{18}$$

Hence, under $H_0$, the conditional distribution of $\|(1-\rho_k)^{-1/2}\sigma^{-1}W^{(k+1)}\|^2$ given $W^{(k)}$ is non-central $\chi^2$ distribution with non-centrality $\{\rho_k/(1-\rho_k)\}\|W^{(k)}\|^2/\sigma^2$. We state the implications for the sequence of $\chi^2$ statistics, $\{\Omega^{(k)}; k = 1, \dots, K\}$, in the following Theorem.

**Theorem 4** *Suppose $Y^{(K)} = (Y_1, \dots, Y_{n_K})^T \sim N(X^{(K)}\theta, \Sigma^{(K)}\sigma^2)$ with non-singular variance matrix $\Sigma^{(K)}\sigma^2$ and the first $n_k$ elements of $Y^{(K)}$ are available at analyses $k = 1, \dots, K$. Suppose $\theta$ is estimable from $Y^{(1)}$ and let $\hat{\theta}^{(k)}$, $S^{(k)}$ and $V_k = Var(\hat{\theta}^{(k)})/\sigma^2$, $k = 1, \dots, K$, be as defined by (1), (2) and (3). Let $C$ be a $q \times r$ matrix such that the variances $Var(C^T\hat{\theta}^{(k)}) = C^T V_k C\sigma^2$ are related by*

$$C^T V_k C = \gamma_k \Lambda, \quad k = 1, \dots, K,$$

*where $\Lambda$ is a $q \times q$ positive definite symmetric matrix and $\gamma_1, \dots, \gamma_K$ a decreasing sequence of scalars and define $\rho_k = \gamma_{k+1}/\gamma_k$, $k = 1, \dots, K-1$. Denote the sequence of $\chi^2$ statistics for testing $H_0$: $C^T\theta = 0$ by*

$$\Omega^{(k)} = (C^T\hat{\theta}^{(k)})^T(C^T V_k C\sigma^2)^{-1}C^T\hat{\theta}^{(k)}, \quad k = 1, \dots, K.$$

*Then, under $H_0$, $\{\Omega^{(k)}; k = 1, \dots, K\}$ is a Markov sequence, $\Omega^{(1)} \sim \chi_q^2$ and, for $k = 1, \dots, K-1$,*

$$\Omega^{(k+1)}|\Omega^{(k)} \sim (1-\rho_k)\chi_q^2\left(\frac{\rho_k}{1-\rho_k}\Omega^{(k)}\right). \tag{19}$$

The spherical symmetry of the conditional variance of $W^{(k+1)}$ is crucial to this result as it ensures that $\|W^{(k+1)}\|^2$ depends on $W^{(k)}$ only through $\|W^{(k)}\|^2$ and, thus, the sequence of $\chi^2$ statistics, $\{\|W^{(k)}\|^2/\sigma^2; k = 1, \dots, K\}$, is Markov. The sufficient condition for this, (17), will hold in "balanced" experiments where new data arrive as additional replicates of a certain experimental design. For example, (17) would apply in a two-way Analysis of Variance with equal numbers of observations in each cell of the two-way table at each analysis $k = 1, \dots, K$. Slight imbalances should cause only minor departures from (17) but highly differential rates of accrual of information on some components of $C^T\theta$ at different times during the study could have more serious effects.

Theorem 4 generalises the results obtained by Jennison and Turnbull (1991) for independent $N(\theta, \Sigma\sigma^2)$ observations, the factors $\gamma_k$ replacing the cumulative sample sizes $n_k$ in the extension to the more general case. Equation (19) can be used in computing boundaries of group sequential $\chi^2$ tests following exactly the approach adopted by Jennison and Turnbull (1991) for the case of identically distributed observations.

Jennison and Turnbull (1991) also derive the joint distribution of the $\Omega^{(k)}$s in the non-null case for use in power calculations. Suppose $C^T\theta = \phi$, so $\Lambda^{-1/2}C^T\theta = \Lambda^{-1/2}\phi$ and

$$W^{(k)} \sim N(\gamma_k^{-1/2}\Lambda^{-1/2}\phi, I_q\sigma^2), \quad k = 1, \dots, K.$$

Let $R$ be a rotation matrix such that $R\Lambda^{-1/2}\phi$ lies in the direction of $(1, 0, \dots, 0)^T$. Since $\|W^{(k)}\|^2 = \|RW^{(k)}\|^2$, the joint distribution of $\{\|W^{(k)}\|^2; k = 1, \dots, K\}$ is equal to that of $\{\|RW^{(k)}\|^2; k = 1, \dots, K\}$ and, thus, there is no loss of generality in supposing $\Lambda^{-1/2}\phi = \psi(1, 0, \dots, 0)^T$ where $\psi = \|\Lambda^{-1/2}\phi\|$. For $k = 1, \dots, K$, we define the bivariate statistic $(U^{(k)}, V^{(k)})$, where $U^{(k)} = W_1^{(k)}$ is a scalar random variable and $V^{(k)} = (W_2^{(k)}, \dots, W_q^{(k)})^T$ is a $(q-1)$-vector. Thus

$$\Omega^{(k)} = \frac{U^{(k)2} + \|V^{(k)}\|^2}{\sigma^2}.$$

The sequences $\{U^{(k)}; k = 1, \dots, K\}$ and $\{V^{(k)}; k = 1, \dots, K\}$ are independent of each other, $U^{(1)} \sim N(\gamma_1^{-1/2}\psi, \sigma^2)$, $\|V^{(1)}\|^2 \sim \sigma^2\chi^2_{q-1}$ and, for $k = 1, \dots, K-1$,

$$U^{(k+1)}|U^{(k)} \sim N(\gamma_{k+1}^{-1/2}(1-\rho_k)\psi + \rho_k^{1/2}U^{(k)}, (1-\rho_k)\sigma^2)$$

and

$$\|V^{(k+1)}\|^2 \mid \|V^{(k)}\|^2 \sim \sigma^2(1-\rho_k)\chi^2_{q-1}\left(\frac{\rho_k}{1-\rho_k}\frac{\|V^{(k)}\|^2}{\sigma^2}\right).$$

Writing $A^{(k)} = \sigma^{-1}U^{(k)}$ and $B^{(k)} = \|V^{(k+1)}\|^2/\sigma^2$, $k = 1, \dots, K$, we obtain the following Theorem.

**Theorem 5** *With the definitions and assumptions of Theorem 4, the joint distribution of the sequence of $\chi^2$ statistics for testing $H_0$: $C^T\theta = 0$ depends on the parameter $\psi = \|\Lambda^{-1/2}C^T\theta\|$. The $\chi^2$ statistic at analysis $k$ can be expressed as*

$$\Omega^{(k)} = A^{(k)2} + B^{(k)}, \quad k = 1, \ldots, K,$$

*where $\{A^{(k)}; k = 1, \ldots, K\}$ and $\{B^{(k)}; k = 1, \ldots, K\}$ are independent Markov sequences, $A^{(1)} \sim N(\sigma^{-1}\gamma_1^{-1/2}\psi, 1)$, $B^{(1)} \sim \chi^2_{q-1}$ and, for $k = 1, \ldots, K-1$,*

$$A^{(k+1)}|A^{(k)} \sim N(\sigma^{-1}\gamma_{k+1}^{-1/2}(1-\rho_k)\psi + \rho_k^{1/2}A^{(k)}, 1-\rho_k)$$

*and*

$$B^{(k+1)}|B^{(k)} \sim (1-\rho_k)\chi^2_{q-1}\left(\frac{\rho_k}{1-\rho_k}B^{(k)}\right).$$

# 6 Group sequential $F$-tests

We consider the problem described in Section 5 of testing the hypothesis $C^T\theta = 0$ where $C$ is a $q \times p$ matrix but now we do not assume $\sigma^2$ to be known and so a group sequential $F$-test is required. Again we restrict attention to cases where $\theta$ is fully estimable at each analysis, noting that our results can be extended to other situations using the methods presented at the end of Section 4.

At analysis $k$, the $F$-statistic for testing $H_0$: $C^T\theta = 0$ is

$$F^{(k)} = \frac{(C^T\hat{\theta}^{(k)})^T(C^TV_kC)^{-1}C^T\hat{\theta}^{(k)}/q}{S^{(k)}/(n_k - p)}, \quad k = 1, \ldots, K,$$

where $S^{(k)}$ is as defined by (2). The joint distribution of the sequence of $F$-statistics is determined by that of $\{(C^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$. This can be derived in the same way that the joint distribution of the sequence $\{(c^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ was derived in Section 4, the projections $P_6$, $P_8$ and $P_{10}$ being modified so that they are associated with $C^T\hat{\theta}^{(1)}$, $C^T\hat{\theta}^{(2)}$ and $C^T(\hat{\theta}^{(2)} - \hat{\theta}^{(1)})$ respectively. It is then seen that the sequence $\{(C^T\hat{\theta}^{(k)}, S^{(k)}); k = 1, \ldots, K\}$ is Markov, $C^T\hat{\theta}^{(1)} \sim N(C^T\theta, C^TV_1C\sigma^2)$ and $S^{(1)} \sim \sigma^2\chi^2_{n_1-p}$. Further, for $k = 1, \ldots, K-1$, the conditional distribution of $\hat{C^T\theta}^{(k+1)}$ given $C^T\hat{\theta}^{(k)}$ and $S^{(k)}$ is independent of $S^{(k)}$ and follows (15), and

$$S^{(k+1)}|C^T\hat{\theta}^{(k+1)}, C^T\hat{\theta}^{(k)}, S^{(k)} \sim$$
$$S^{(k)} + \{C^T(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)})\}^T\{C^T(V_k - V_{k+1})C\}^{-1}\{C^T(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)})\} + \sigma^2\chi^2_{n_{k+1}-n_k-\tilde{q}_{k+1}} \tag{20}$$

where $\tilde{q}_{k+1}$ is the rank of $C^T(V_k - V_{k+1})C$.

As for the sequential $\chi^2$ test, the joint distribution of the sequence of $F$-statistics, $F^{(1)}, \ldots, F^{(K)}$, simplifies when the variances $Var(C^T\hat{\theta}^{(k)}) = C^T V_k C \sigma^2$ are related by

$$C^T V_k C = \gamma_k \Lambda, \quad k = 1, \ldots, K,$$

where $\Lambda$ is a $q \times q$ positive definite symmetric matrix and $\gamma_1, \ldots, \gamma_K$ a decreasing sequence of scalars. As before, we set

$$\rho_k = \frac{\gamma_{k+1}}{\gamma_k}, \quad k = 1, \ldots, K-1.$$

With $W^{(k)} = \gamma_k^{-1/2} \Lambda^{-1/2} C^T \hat{\theta}^{(k)}$, as before,

$$F^{(k)} = \frac{\|W^{(k)}\|^2/q}{S^{(k)}/(n_k - p)}, \quad k = 1, \ldots, K.$$

Under $H_0$ the conditional distribution of $W^{(k+1)}$ given $W^{(k)}$ is as stated in (18) and we can write

$$W^{(k+1)} = \rho_k^{1/2} W^{(k)} + (1-\rho_k)^{1/2} \sigma \epsilon,$$

where $\epsilon \sim N(0, I_q)$. Note that the second term in the expansion (20) of $S^{(k+1)}$ can be written as

$$\begin{aligned}
&(\gamma_{k+1}^{1/2}\Lambda^{1/2}W^{(k+1)} - \gamma_k^{1/2}\Lambda^{1/2}W^{(k)})^T(\gamma_k\Lambda - \gamma_{k+1}\Lambda)^{-1}(\gamma_{k+1}^{1/2}\Lambda^{1/2}W^{(k+1)} - \gamma_k^{1/2}\Lambda^{1/2}W^{(k)}) \\
&\qquad = (\gamma_k - \gamma_{k+1})^{-1}\|\gamma_{k+1}^{1/2}W^{(k+1)} - \gamma_k^{1/2}W^{(k)}\|^2.
\end{aligned}$$

Let $1_{W^{(k)}} = W^{(k)}/\|W^{(k)}\|$ denote the unit vector in the direction of $W^{(k)}$ and write $\epsilon = \lambda 1_{W^{(k)}} + \nu$ where the scalar variable $\lambda \sim N(0,1)$ and the $q$-vector $\nu$ is the projection of $\epsilon$ orthogonal to $1_{W^{(k)}}$. Then

$$\begin{aligned}
&(\gamma_k - \gamma_{k+1})^{-1}\|\gamma_{k+1}^{1/2}W^{(k+1)} - \gamma_k^{1/2}W^{(k)}\|^2 \\
&= (\gamma_k - \gamma_{k+1})^{-1}\|(\rho_k - 1)\gamma_k^{1/2}W^{(k)} + \gamma_{k+1}^{1/2}(1-\rho_k)^{1/2}\sigma(\lambda 1_{W^{(k)}} + \nu)\|^2 \\
&= \gamma_k^{-1}(1-\rho_k)^{-1}\|\{\gamma_{k+1}^{1/2}(1-\rho_k)^{1/2}\sigma\lambda - (1-\rho_k)\gamma_k^{1/2}\|W^{(k)}\|\}1_{W^{(k)}} + \gamma_{k+1}^{1/2}(1-\rho_k)^{1/2}\sigma\nu\|^2 \\
&= \{\rho_k^{1/2}\sigma\lambda - (1-\rho_k)^{1/2}\|W^{(k)}\|\}^2 + \rho_k\sigma^2\|\nu\|^2.
\end{aligned}$$

Still under $H_0$, and with the same definitions of $\lambda$ and $\nu$,

$$W^{k+1} = \{\rho_k^{1/2}\|W^{(k)}\| + (1-\rho_k)^{1/2}\sigma\lambda\}1_{W^{(k)}} + (1-\rho_k)^{1/2}\sigma\nu$$

and, hence,

$$\|W^{k+1}\|^2 = \{\rho_k^{1/2}\|W^{(k)}\| + (1-\rho_k)^{1/2}\sigma\lambda\}^2 + (1-\rho_k)\sigma^2\|\nu\|^2.$$

These results lead to the following Theorem.

**Theorem 6** *Suppose $Y^{(K)} = (Y_1, \ldots, Y_{n_K})^T \sim N(X^{(K)}\theta, \Sigma^{(K)}\sigma^2)$ with non-singular variance matrix $\Sigma^{(K)}\sigma^2$ and the first $n_k$ elements of $Y^{(K)}$ are available at analyses $k = 1, \ldots, K$. Suppose $\theta$ is estimable from $Y^{(1)}$ and let $\hat{\theta}^{(k)}$, $S^{(k)}$ and $V_k = Var(\hat{\theta}^{(k)})/\sigma^2$, $k = 1, \ldots, K$, be as defined by (1), (2) and (3). Let $C$ be a $q \times r$ matrix such that the variances $Var(C^T\hat{\theta}^{(k)}) = C^T V_k C \sigma^2$ are related by*

$$C^T V_k C = \gamma_k \Lambda, \quad k = 1, \ldots, K,$$

*where $\Lambda$ is a $q \times q$ positive definite symmetric matrix and $\gamma_1, \ldots, \gamma_K$ a decreasing sequence of scalars and define $\rho_k = \gamma_{k+1}/\gamma_k$, $k = 1, \ldots, K-1$. Denote the sequence of $F$-statistics for testing $H_0$: $C^T\theta = 0$ by*

$$F^{(k)} = \frac{(C^T\hat{\theta}^{(k)})^T (C^T V_k C)^{-1} C^T \hat{\theta}^{(k)}/q}{S^{(k)}/(n_k - p)} = \frac{\|W^{(k)}\|^2/q}{S^{(k)}/(n_k - p)}, \quad k = 1, \ldots, K,$$

*where $W^{(k)} = \gamma_k^{-1/2}\Lambda^{-1/2}C^T\hat{\theta}^{(k)}$.*

*Then, under $H_0$, $\{(\|W^{(k)}\|^2, S^{(k)}); k = 1, \ldots, K\}$ is a Markov sequence, $\|W^{(1)}\|^2$ and $S^{(1)}$ are independent with $\|W^{(1)}\|^2 \sim \sigma^2\chi^2_q$ and $S^{(1)} \sim \sigma^2\chi^2_{n_1-p}$ and, for $k = 1, \ldots, K-1$, the joint distribution of $(\|W^{(k+1)}\|^2, S^{(k+1)})$ given $(\|W^{(k)}\|^2, S^{(k)})$ is determined by the relations*

$$\|W^{(k+1)}\|^2 = \{\rho_k^{1/2}\|W^{(k)}\| + (1-\rho_k)^{1/2}\sigma\lambda\}^2 + (1-\rho_k)\sigma^2\nu^2$$

*and*

$$S^{(k+1)} = S^{(k)} + \{\rho_k^{1/2}\sigma\lambda - (1-\rho_k)^{1/2}\|W^{(k)}\|\}^2 + \rho_k\sigma^2\nu^2 + \sigma^2\chi^2_{n_{k+1}-n_k-q},$$

*where $\lambda \sim N(0,1)$ and $\nu^2 \sim \chi^2_{q-1}$ are random variables, independent of each other and of $\|W^{(k)}\|^2$ and $S^{(k)}$.*

Note that $\sigma^2$ appears as a scale factor in the distributions of each $\|W^{(k)}\|^2$ and $S^{(k)}$, $k = 1, \ldots, K$. Thus, the distribution of the sequence $\{F^{(k)}; k = 1, \ldots, K\}$ is independent of $\sigma^2$ under $H_0$: $C^T\theta = 0$ and any convenient value of $\sigma^2$ can be used in calculating properties of a group sequential $F$-test under the null hypothesis.

Once more, our Theorem generalises the results obtained by Jennison and Turnbull (1991) for independent $N(\theta, \Sigma\sigma^2)$ observations, the factors $\gamma_k$ replacing the cumulative sample sizes $n_k$ in the extension to the more general case. In order for the joint distribution of the sequence of $F$-statistics to simplify in the way described in Theorem 6, the variances $Var(C^T\hat{\theta}^{(k)})$ must satisfy the same condition as was required for sequential $\chi^2$ tests and, again, this condition will hold in balanced experiments where new data arrive as additional replicates of a certain experimental design.

# References


Armitage, P., Stratton, I.M. & Worthington, H.V. (1985). Repeated significance tests for clinical trials with a fixed number of patients and variable follow-up. *Biometrics* **41**, 353-359.

Jennison, C. & Turnbull, B.W. (1991). Exact calculations for sequential $t$, $\chi^2$ and $F$ tests. *Biometrika* **78**, 133-141.

Lan, K.K.G. & DeMets, D.L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659-663.

Lee, J.W. & DeMets, D.L. (1991). Sequential comparison of changes with repeated measurements data. *J. Am. Statist. Assoc.* **86**, 757-762.

Reboussin, D.M., Lan, K.K.G. & DeMets, D.L. (1992). Group sequential testing of longitudinal data. *Univ. of Wisconsin, Dept. of Biostatistics, Tech. Report 72.*

Seber, G. A. F. (1980). *The Linear Hypothesis: A General Theory*, 2nd edition. London: Griffin and Company.

Slud, E.V. & Wei, L.J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *J. Am. Statist. Assoc.* **77**, 862-868.