Estimation of Secondary Parameters following Sequential Tests for Multivariate Data

Ying Yan[1] and Steve Coad[2]

[1] School of Mathematical Sciences,
Queen Mary, University of London,
Mile End Road,
London E1 4NS, UK
y.yan@qmul.ac.uk

[2] School of Mathematical Sciences,
Queen Mary, University of London,
Mile End Road,
London E1 4NS, UK
d.s.coad@qmul.ac.uk

Abstract. When a sequential clinical trial is carried out to compare two treatments in which primary response variables are monitored, there are often secondary response variables which are correlated with the primary ones. So far, most studies on secondary parameters have focused on a single primary parameter. However, sometimes more then one primary response variable is monitored. In the present work, the bias and variance of the maximum likelihood estimator of a single unknown secondary parameter are studied when there is more than one primary parameter, and approximations given. An approximate corrected confidence interval for the secondary parameter is also derived by constructing an approximately pivotal quantity. Simulations are carried out using *Matlab* in order to assess the accuracy of the approximations. We compare the approximate bias and variance with their simulated values. The corrected confidence intervals for the secondary parameter are investigated in terms of their coverage probabilities.

Keywords. approximate bias, approximately pivotal quantity, corrected confidence interval, maximum likelihood estimator, multivariate normal process, primary parameter.

1 Introduction

Suppose that we wish to use a sequential test to compare two treatments in a clinical trial. When the test is carried out in which primary response variables are monitored, there are often secondary response variables which are correlated with the primary ones. It follows that the usual estimators of the secondary parameters will be biased.

The method of maximum likelihood is often used to estimate unknown parameters following sequential tests. It is common that several primary parameters are to be estimated. One may also need to estimate some secondary parameters after the trial. So far, most studies on secondary parameters have focused on a single primary parameter. The following two cases have been well studied: (i) the estimation of a single primary parameter; (ii) the estimation of a secondary parameter when there is a single primary parameter. Whitehead (1986a) studied the bias of the maximum likelihood estimator of the primary parameter and Whitehead (1986b) showed how the bias of the estimator of the secondary parameter is related to that of the primary parameter for large samples. A method based on approximately normally distributed pivots was introduced for setting confidence intervals for the primary parameter by Woodroofe (1992). More recently, Whitehead et al. (2000) gave an approximate confidence region for a single primary parameter and a single secondary parameter for a bivariate normal process when the covariance matrix of the primary and secondary response variables is known.

However, in practical situations, there are often cases when several response variables are monitored during sequential experiments. Consequently, there might be more than one primary response variable. Our aim is to study this problem. More precisely, we consider the case of multivariate normal random variables. Woodroofe's (1992) results have been generalised to the multivariate normal case. The results of Whitehead (1986b) and Whitehead et al. (2000) are also extended to the case of more than one primary parameter and one secondary parameter.

2 Estimation of secondary parameters

Suppose that $\underline{X}_i = (X_{1i}, \ldots, X_{di})^T$ for $i = 1, 2, \ldots, n$ are multivariate normal random vectors with unknown mean vector $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^T = (\underline{\theta}^{(1)}, \theta_d)^T$ and known covariance matrix Σ, where

$$\Sigma = (\rho_{ij}\sigma_i\sigma_j) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

the ρ_{ij} for $i, j = 1, 2, \ldots, d$ are the correlation coefficients, the σ_i for $i = 1, 2, \ldots, d$ are the standard deviations and Σ_{11} is $(d-1) \times (d-1)$. If Σ is nonsingular, the maximum likelihood estimators of $\underline{\theta}^{(1)}$ and θ_d are $\hat{\underline{\theta}}^{(1)} = \underline{S}_n^{(1)}/n$ and $\hat{\theta}_d = S_{dn}/n$, where $\underline{S}_n^{(1)} = (S_{1n}, \ldots, S_{d-1,n})^T$ and $S_{kn} = \sum_{i=1}^n X_{ki}$ for $k = 1, 2, \ldots, d$. Clearly, the conditional mean and variance of S_{dn} given $\underline{S}_n^{(1)}$ are

$$\mathrm{E}\{S_{dn}|\underline{S}_n^{(1)}\} = n\theta_d + \Sigma_{21}\Sigma_{11}^{-1}\{\underline{S}_n^{(1)} - n\underline{\theta}^{(1)}\}$$

and

$$\mathrm{var}\{S_{dn}|\underline{S}_n^{(1)}\} = n\sigma_d^2 - n\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},$$

since $\underline{S}_n = (S_{1n}, \ldots, S_{dn})^T$ is multivariate normal with mean vector $n\underline{\theta}$ and covariance matrix $n\Sigma$.

Let N be the stopping time for the sequential test based on $\hat{\underline{\theta}}^{(1)}$. Then the bias and variance of the maximum likelihood estimator of the secondary parameter θ_d are

$$\mathrm{bias}(\hat{\theta}_d) = \Sigma_{21}\Sigma_{11}^{-1}\left[\mathrm{E}\{\hat{\underline{\theta}}^{(1)}\} - \underline{\theta}^{(1)}\right]$$

and

$$\mathrm{var}(\hat{\theta}_d) = \left(\sigma_d^2 - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)\mathrm{E}\left(N^{-1}\right) + \Sigma_{21}\Sigma_{11}^{-1}\mathrm{var}\{\hat{\underline{\theta}}^{(1)}\}\Sigma_{11}^{-1}\Sigma_{12}.$$

Similarly,

$$\mathrm{cov}\{\hat{\underline{\theta}}^{(1)}, \hat{\theta}_d\} = \Sigma_{21}\Sigma_{11}^{-1}\mathrm{var}\{\hat{\underline{\theta}}^{(1)}\}$$

is the covariance of the maximum likelihood estimators of $\underline{\theta}^{(1)}$ and θ_d.

We have also constructed an approximate pivot for θ_d. Let ν and ω^2 denote the respective mean and variance of the random variable $S'_{dN}(\theta_d) = (S_{dN} - N\theta_d)/(\sigma_d\sqrt{N})$, which is a first approximation to a pivot. Then these are determined by the primary data and need to be estimated. Define the renormalised pivot

$$S^\sharp_{dN}(\theta_d) = \frac{S'_{dN}(\theta_d) - \nu}{\omega} = \frac{S_{dN} - \nu\sigma_d\sqrt{N} - N\theta_d}{\omega\sigma_d\sqrt{N}}.$$

Then $S^\sharp_{dN}(\theta_d)$ is an approximate pivot for θ_d, and so an approximate $100(1-\alpha)\%$ confidence interval for θ_d is given by

$$\frac{s_{dN}}{N} - \frac{\nu\sigma_d}{\sqrt{N}} \mp \frac{\omega\sigma_d}{\sqrt{N}}\Phi^{-1}(1 - \frac{1}{2}\alpha),$$

where Φ denotes the standard normal distribution function. When Σ is unknown, σ_d is replaced by its maximum likelihood estimate and the standard normal values are replaced by values from a t distribution.

3 Mean and variance approximations

We now consider the primary parameter vector $\underline{\theta}^{(1)}$. Suppose that the stopping time for the sequential test is of the form

$$N = \inf\left\{n \geq m_0 : nq(\hat{\underline{\theta}}^{(1)}) \geq c\right\} \wedge m,$$

where $c > 0$ is a design parameter, m_0 is the minimum sample size and m is the maximum sample size. Let $\epsilon_0 = c/m_0$ and $\epsilon_1 = c/m$. Then $c/N \to g(\underline{\theta}^{(1)}) = \epsilon_0 \wedge q(\underline{\theta}^{(1)}) \vee \epsilon_1$ in probability as $c \to \infty$. Following Weng and Coad (2006), we take $q(\underline{\theta}^{(1)}) = \|\underline{\theta}^{(1)}\|^p$ for $p = 1, 2$, so that the stopping time becomes

$$N = \inf\left\{n \geq m_0 : \|\underline{S}_n^{(1)}\| \geq (cn^{p-1})^{1/p}\right\} \wedge m.$$

Using Woodroofe's (1992) methods, approximations can be developed for the bias and variance of $\hat{\underline{\theta}}^{(1)}$, ν and ω in terms of $g(\underline{\theta}^{(1)})$.

Using a Taylor series expansion, we have

$$\text{bias}\{\hat{\underline{\theta}}^{(1)}\} \simeq \frac{1}{c}\Sigma_{11}\nabla g(\underline{\theta}^{(1)}).$$

Next, since $N \simeq c/g(\hat{\underline{\theta}}^{(1)})$ for large c, we obtain

$$\text{E}\left(\frac{1}{N}\right) \simeq \frac{g(\underline{\theta}^{(1)})}{c} + \frac{1}{c^2}\nabla g(\underline{\theta}^{(1)})^T\Sigma_{11}\nabla g(\underline{\theta}^{(1)}) + \frac{g(\underline{\theta}^{(1)})}{2c^2}\text{tr}\left\{\Sigma_{11}\nabla^2 g(\underline{\theta}^{(1)})\right\}$$

and

$$\text{var}\{\hat{\underline{\theta}}^{(1)}\} \simeq \frac{1}{c}\Sigma_{11}g(\underline{\theta}^{(1)}) + \frac{1}{c^2}g(\underline{\theta}^{(1)})\Sigma_{11}\nabla^2 g(\underline{\theta}^{(1)})\Sigma_{11}.$$

It then follows that approximations for the bias and variance of $\hat{\theta}_d$ are

$$\text{bias}(\hat{\theta}_d) \simeq \frac{1}{c}\Sigma_{21}\nabla g(\underline{\theta}^{(1)})$$

and

$$\begin{aligned}\text{var}(\hat{\theta}_d) \simeq{}& \frac{\sigma_d^2}{c}g(\underline{\theta}^{(1)}) + \frac{1}{c^2}g(\underline{\theta}^{(1)})\Sigma_{21}\nabla^2 g(\underline{\theta}^{(1)})\Sigma_{12} \\ &+\frac{1}{c^2}\left(\sigma_d^2 - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)\left[\nabla g(\underline{\theta}^{(1)})^T\Sigma_{11}\nabla g(\underline{\theta}^{(1)}) + \frac{1}{2}g(\underline{\theta}^{(1)})\text{tr}\left\{\Sigma_{11}\nabla^2 g(\underline{\theta}^{(1)})\right\}\right],\end{aligned}$$

by using the formulae in Section 2.

Similar calculations yield

$$\nu \simeq \frac{1}{\sqrt{c}}\frac{\Sigma_{21}}{\sigma_d}\nabla g^{\frac{1}{2}}(\underline{\theta}^{(1)})$$

and

$$\omega^2 \simeq 1 + \frac{1}{c}\frac{\Sigma_{21}\nabla g^{\frac{1}{2}}(\underline{\theta}^{(1)})\nabla g^{\frac{1}{2}}(\underline{\theta}^{(1)})^T\Sigma_{12}}{\sigma_d^2}.$$

For the case $d = 2$, see Weng and Coad (2006).

4 Simulation results

A *Matlab* program has been developed for the simulations. This program works for unknown parameter vectors $\underline{\theta}$ of any dimension and any given number of replicates. We carried out the simulations for the case $d = 3$. For given values of $\underline{\theta} = (\theta_1, \theta_2, \theta_3)^T$ and given covariance matrices, the simulations are based on $10,000$ replicates. In each case, $c = 10$ and $m = 100$. For $p = 1$, the stopping time is given by $N = \inf\{n \geq m_0 : \|\underline{S}_n^{(1)}\| \geq c\} \wedge m$ and $m_0 = 3$, whereas for $p = 2$, we have the stopping time $N = \inf\{n \geq m_0 : \|\underline{S}_n^{(1)}\| \geq \sqrt{cn}\} \wedge m$ and $m_0 = 5$. For convenience, we assume that $\sigma_i^2 = 1$ for $i = 1, 2, 3$.

The simulated bias and variance are calculated by finding the bias and variance of the estimates over the replications. Three different situations for the approximate confidence intervals are considered. They are the cases of known covariance matrix, known variances and unknown correlation coefficients, and unknown covariance matrix. When the correlation coefficients are unknown, they are estimated by the sample correlation coefficients. When the covariance matrix is unknown, it is estimated by its maximum likelihood estimate given by $\hat{\Sigma} = (\hat{\rho}_{ij}\hat{\sigma}_i\hat{\sigma}_j)$, where the $\hat{\rho}_{ij}$ for $i, j = 1, 2, \ldots, d$ and $i \neq j$ are the sample correlation coefficients and the $\hat{\sigma}_i$ for $i = 1, 2, \ldots, d$ are the maximum likelihood estimates of the standard deviations. So the corrected confidence intervals now take the form

$$\frac{s_{dN}}{N} - \frac{\hat{\nu}\hat{\sigma}_d}{\sqrt{N}} \mp \frac{\hat{\omega}\hat{\sigma}_d}{\sqrt{N}} t_{d.f.,\alpha/2},$$

where $\hat{\nu}$ and $\hat{\omega}$ are estimates of ν and ω, and the unadjusted confidence intervals are

$$\frac{s_{dN}}{N} \mp \frac{\hat{\sigma}_d}{\sqrt{N}} t_{d.f.,\alpha/2},$$

where $t_{r,\gamma}$ denotes the upper $100\gamma\%$ point of the t distribution with r degrees of freedom. The two values for the degrees of freedom used in the simulations are the stopping time N for the sequential test and $c/g(\hat{\underline{\theta}}^{(1)})$.

The simulation results show that the approximate bias and variance are quite close to the simulated values when $p = 1$, but are less accurate when $p = 2$. The coverage probabilities for the corrected confidence intervals are much closer to the nominal values than those of the unadjusted confidence intervals when $p = 2$. For example, most of the coverage probabilities for the corrected confidence intervals are within two standard errors of the nominal values. There are only slight improvements in the coverage probabilities for the corrected confidence intervals when $p = 1$ and $\| \underline{\theta} \|$ is small. The results also show that most of the coverage probabilities are closer to the nominal values when $d.f. = c/g(\hat{\underline{\theta}}^{(1)})$ than when $d.f. = N$.

References

Weng, R. C. and Coad, D. S. (2006). Corrected Confidence Intervals for Secondary Parameters following Sequential Tests. In: *Recent Developments in Nonparametric Inference and Probability: Festschrift for Michael Woodroofe* (J. Sun, A. DasGupta, V. Melfi and C. Page, Eds). Institute of Mathematical Statistics, Hayward, California, 80–104.

Whitehead, J. (1986a). On the Bias of Maximum Likelihood Estimation following a Sequential Test. *Biometrika*, **73**, 573–581.

Whitehead, J. (1986b). Supplementary Analysis at the Conclusion of a Sequential Clinical Trial. *Biometrics*, **42**, 461–471.

Whitehead, J., Todd, S. and Hall, W. J. (2000). Confidence Intervals for Secondary Parameters following a Sequential Test. *Journal of the Royal Statistical Society Series B*, **62**, 731–745.

Woodroofe, M. (1992). Estimation after Sequential Testing: A Simple Approach for a Truncated Sequential Probability Ratio Test. *Biometrika*, **79**, 347–353.