ROBUST ESTIMATION AND OUTLIER DETECTION FOR REPEATED MEASURES EXPERIMENTS.

R.M. HUGGINS

1. INTRODUCTION.

Standard methods of testing for repeated measures experiments and other mixed linear models are based on the likelihood function or least squares methods. These methods are known to be sensitive to model misspecification, and can be adversely affected by the presence of outliers.

There have been several attempts at constructing robust estimates of the parameters associated with such models, eg Rocke (1982), Fellner (1986), Huggins (1991a,b) but only the methods of Huggins (1991a,b) based on M-estimators allow ready computation of standard errors of the resulting estimates and the construction of hypothesis tests. Note that the parametric form of the covariance matrix in repeated measures experiments allows the adoption of a simpler approach than that of Maronna (1976), see also Carroll (1978), who takes a different approach to constructing M-estimates.

The approach used here also differs from that of Fellner (1986) who considered the robust estimation of variance components in mixed models. He constructed estimating equations, which are then solved iteratively, by trimming large residuals rather than constructing a robustified likelihood as is done here. The methods based on the t-distribution, as in Lange et al (1989), can be difficult to interpret in practice. Huggins (1991a) gives an example where the inclusion of an outlier in a dataset causes the degrees of freedom of the t-distribution to fall from 4945 to 13, which gives a misleading impression of the distribution of the bulk of the data. Methods based on the t-distribution have the further disadvantage that they downweight entire vectors of observations when it may only be one of the components of the vector which is abnormal.

The approach here supposes that one is interested in the behaviour of the "centre" of the data and that the estimates, particularly those for variance components, should be able to be interpreted in a multivariate normal framework, see Huggins (1991b) for a discussion of this. In particular one would like to obtain estimates very close to the maximum likelihood estimates if there are no outliers.

2. ROBUST INFERENCE AND OUTLIER DETECTION.

We briefly outline and discuss the methods of Huggins (1991a,b,c).

For each individual let $X = (X_1, \ldots, X_t)^t$ be a vector of correlated continuous random variables. It is commonly supposed that $X \sim MVN(\mu, \Omega)$, where $\mu = Y\beta$ for some design matrix Y and vector of parameters β, and Ω is some nonsingular matrix which can be expressed as a function of some variance components $\sigma_1^2, \ldots, \sigma_m^2$. Suppose one then takes independent observations $X^{(1)}, \ldots, X^{(N)}$ and wishes to estimate β and $\sigma_1^2, \ldots, \sigma_m^2$.

If multivariate normality is assumed then the contribution of each individual to the likelihood is

$$-\frac{1}{2}(X-\mu)^t\Omega^{-1}(X-\mu) - \frac{1}{2}\ln(det(\Omega)).$$

Summing these contributions over the observed $X^{(j)}$, then maximising with respect to β and the σ_k^2 results in the maximum likelihood estimators.

In order to robustify these equations write $\Omega = A^{-t}A^{-1}$ and for each X let $Z = A^{-1}(X-\mu) = (Z_1, \ldots, Z_t)^t$. Then under the MVN assumption the Z_i are i.i.d. standard normal random variables and

$$(X-\mu)^t\Omega^{-1}(X-\mu) = Z^tZ = \sum_{i=1}^{t} Z_i^2.$$

That is the maximumn likelihood estimators minimize the sum over all the $X^{(j)}$ of

$$\frac{1}{2}\sum_{i=1}^{n} Z_i^2 + \frac{1}{2}\ln(det(\Omega)).$$

Now the Z_i^2 terms give excessive weight to large Z_i values and in particular to outliers. The approach to robust inference of Huggins (1991a,b) replaces Z_i^2 by $\rho(Z_i)$ where ρ gives less weight to large values of Z_i. A correction factor, K, must then be introduced so that the resulting estimators of the variance components are consistent and the robust estimators then arise from minimising,

$$\sum_{i=1}^{n} \rho(Z_i) + \frac{K}{2}\ln(det(\Omega)).$$

This results in M-estimators of the parameters associated with the model and the standard errrors etc. can be easily calculated using standard methods, as in Huggins (1991a). In the example below ρ corresponding to Tukey's bisquare was used.

The robust estimates are easily computed, as only a minor modification to programs which compute maximum likelihood estimates is required, and agree closely with the mle's if there are no outliers or contamination. A comparison of the resulting estimates with the mle's is often a good outlier diagnostic. The estimates are not perfect in that the influence function of the variance components is not bounded, Huggins (1991b). However, it is possible to modify the procedure so that the influence functions are bounded. In practice these estimators with bounded influence function are far more difficult to compute and for most purposes the estimates described above will suffice.

In order to detect the presence of outliers several methods may be used. The approach taken here is taken from Huggins (1991a,b) where earlier work of Hopper and Mathews (1982) concerning the detection of outliers in pedigree analysis using maximum likelihood estimates is extended to repeated measures and robust estimates. Firstly the Z_i may be computed using the estimated parameter values and any large values indicate the presence of outliers, but do not identify the outliers. Next outlying X can be detected by computing

$$Q = (X - \hat{\mu})^t \hat{\Omega}^{-1} (X - \hat{\mu}),$$

and then computing $Q^* = (2Q)^{1/2} - (2t-1)^{1/2}$ which has an approximate standard normal distribution, so that large values of Q^* for an observation X indicates that X is an outlier. In order to detect outlying observations within an observation $X = (X_1, \ldots, X_t)^t$ each observation X_j may be compared with its conditional distribution given the remaining X_i. For example for an observed x_1, let x_2 denote the remaining observations, $\hat{\mu}_1$ the estimated mean of x_1, $\hat{\mu}_2$ the estimated mean of x_2 and partition $\hat{\Omega}$ as

$$\hat{\Omega} = \begin{pmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} \\ \hat{\Omega}_{21} & \hat{\Omega}_{22} \end{pmatrix}.$$

Then under multivariate normality the distribution of

$$q_1 = \frac{x_1 - (\hat{\mu}_1 + \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}(x_2 - \hat{\mu}_2))}{\sqrt{\hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}}},$$

has approximately a standard normal distribution and this may be used to detect outliers. Note that the presence of outliers may mask the presence of other outliers or indeed detect spurious outliers through an abnormal value for x_2. To overcome this replace x_2 in q_1 by $\tilde{x}_2$ where $\tilde{x} = \hat{A}\psi(Z) + \hat{\mu}$ with $\psi(x) = x$ for $|x| < 2$ and $\psi(x) = 2$ if $|x| > 2$. This procedure has been used in the outlier detection tests below.

In order to conduct robust tests of hypotheses, the Wald and score tests of Basawa, Huggins and Staudte (1985) may be extended to this setting as in Huggins (1991c).

3. EXAMPLE.

The methodology is illustrated on the Guinea Pig dataset. It was expected that the animals would all grow at a similar rate until the start of week 5 at which time there would be different rates in the three groups.

Let $y_{ij}(t)$ denote the weight of the j *th* individual in group i at time t. The full model for the fixed effects is

$$E(Y_{ij}(t)) = \begin{cases} \mu + \alpha(t-1), & \text{if } t \leq 4; \\ \mu + 3\alpha + \beta_i(t-4) & \text{if } t \geq 5. \end{cases}$$

The null hypothesis is that the three treatment effects are the same, i.e. $H_0 : \beta_1 = \beta_2 = \beta_3$. The model for the covariance structure, based on the approach of Diggle (1988), considered here is $var(Y_{ij}(t)) = \sigma_e^2 + \sigma_r^2$ and $cov(Y_{ij}(t_k), Y_{ij}(t_l)) = \sigma_r^2 \exp(-\lambda|t_k - t_l|)$, where σ_e^2 corresponds to environmental variation and σ_r^2 to an autoregressive component of the variation.

Here individual 1 is identified as an outlier with a value of Q^* of 3.4 and five data points, three corresponding to individual 1 (weeks 3,6 and 7), one to individual 7 (week 5) and one to individual 8 (week 6), are identified as outliers. The parameter estimates and their associated standard errors, with the standard errors of the maximum likelihood estimates computed according to the approach of Royall (1986), for both the full data set and with individual 1 omitted are given in Table 1. It is clear from this table that the outlying individual has a large effect on the maximum likelihood estimate of β_1 whilst the robust estimate is noticeably less affected.

To conduct the testing we reparameterise so that $\beta_2 = \beta_1 + \gamma_2$ and $\beta_3 = \beta_1 + \gamma_3$. Thus we take $\theta^t = (\theta_1^t, \theta_2^t)$ where $\theta_1^t = (\gamma_2, \gamma_3)$ and $\theta_2^t = (\mu, \alpha, \beta, \sigma_e^2, \sigma_r^2, \lambda)$. The null hypothesis is now $H_0 : \gamma_2 = \gamma_3$.

Table 1. *Maximum likelihood and robust estimates of the parameters associated with the full model.*

	All Data		Individual 1 Omitted	
Parameter	Likelihood	Robust	Likelihood	Robust
β_1	1.52 (6.00)	5.43 (4.83)	7.42 (9.09)	7.30 (5.99)
β_2	22.89 (4.51)	21.56 (9.87)	22.17 (5.08)	21.73 (9.26)
β_3	15.84 (2.72)	14.56 (3.66)	15.27 (4.23)	14.90 (4.82)
μ	484.93 (7.38)	489.47 (8.48)	487.57 (7.82)	488.56 (8.69)
α	26.75 (2.84)	27.96 (3.41)	27.25 (2.97)	27.91 (3.56)
σ_e^2	286.11 (121.20)	254.91 (122.91)	285.85 (139.78)	241.51 (144.45)
σ_r^2	1353.1 (461.43)	951.18 (419.49)	936.47 (613.03)	1004.6 (320.79)
λ	0.09 (0.09)	0.14 (0.10)	0.10 (0.13)	0.11 (0.16)

The fact that the outlier has a marked effect on the maximum likelihood estimates suggests that the likelihood ratio test, and related tests based on the maximum likelihood estimates, may be unreliable. The values of these statistics, with individual one included and excluded respectively were: likelihood ratio 11.1, 5.65, Wald statistic 7.39, 1.93, score test 7.56. 3.22. All these statistics have χ^2 distributions under H_0. However, the latter two are computed according to the methods of Kent (1982) and are robust against model misspecification. Thus one would reach different conclusions in the two cases. In the robust case only the Wald and score statistics are computed and the corresponding values are: Wald statistic 3.16, 1.94, score statistic 2.28, 1.85, so that the same conclusions would have been made in either case.

4. DISCUSSION

The example shows that robust techniques can be an important tool for the analysis of repeated measures experiments, both in the detection of outliers and in the estimation of parameters and testing of hypotheses. A major advantage in their use is that there is no need to discard an entire vector of observations if only one of the points in the vector is an outlier, and only this outlying point need be downweighted. Further, the initial stages of an analysis, and model building, can be simplified as one is less concerned that an observed effect may be due to only one or two points.

Robust techniques are not an excuse for poor modelling and the detection of outliers may merely mean the model being examined is incorrect. However, an examination of the residuals and any pattern in the outliers will often allow more realistic models to be constructed.

Care needs to be taken in the interpretation of the variance component estimates. However, this is also true for maximum likelihood estimates, and as discussed in Huggins (1991b), it is the distribution of the standardised residuals which determines the interpretation.

REFERENCES.

Basawa, I.V., Huggins, R.M. & Staudte, R.G. (1985) Robust tests for time series with an application to first-order autoregressive processes. *Biometrika* **72** 559-571.

Carroll, R.J. (1978) On the asymptotic distribution of multivariate M-estimates. *J. Mult. Anal.* **8.** 361-371.

Diggle, P. J. (1988) An approach to the analysis of repeated measures. *Biometrics* **44**, 959-971.

Fellner, W.H. (1986) Robust estimation of variance components. *Technometrics* **28**, 51-60.

Hopper, J.L. & Mathews, J.D. (1982) Extensions to multivariate normal models for pedigree analysis. *Ann. Hum. Genet.* **46** 373-383.

Huggins, R.M. (1991a) On the robust analysis of variance components models for pedigree data. La Trobe University Technical Report.

Huggins, R.M. (1991b) Robust Inference, outlier detection and prediction for inter-laboratory studies. La Trobe University Technical Report.

Huggins, R.M. (1991c) Robust tests for the analysos of repeated measures experiments. La Trobe University Technical Report.

Kent, J. T. (1982) Robust properties of likelihood ratio tests. *Biometrika* **69** 19-27.

Lange, K.L., Little, R.J.A. & Taylor, J.M.G. (1989) Robust statistical modeling using the t distribution. UCLA Statistical Series.

Maronna, R.A. (1976) Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51-67

Rocke, M.R. (1982) Robust statistical analysis of interlaboratory studies. *Biometrika* **70**, 421-431.

Royall, R.M . (1986) Model Robust confidence intervals using maximum likelihood estimators. *Int. Stat. Rev.* **54** 221-226.

Department of Statistics
La Trobe University
Bundoora, 3083, Australia.