Nonlinear component of variance models

BY P. J. SOLOMON

Department of Statistics, University of Adelaide, GPO Box 498, Adelaide SA 5001, Australia

AND D. R. COX

Nuffield College, Oxford OX1 1NF, U.K.

SUMMARY

General aspects of nonlinearity in the context of component of variance models are discussed, and two special topics are examined in detail. Firstly, simple procedures, both formal and informal, are proposed for describing departures from normal-theory linear models. Transformation models are shown to be a special case of a more general formulation, and data on blood pressure are analyzed in illustration. Secondly, an approximate likelihood is proposed and its accurate performance is examined numerically using examples of exponential regression and the analysis of several related 2×2 tables. In the latter example, the approximate score test has improved power over the Mantel-Haenszel test.

Some key words: Cumulant; Likelihood; Logit; Nonlinearity; Nonnormality; Score test; Transformation; Variance component.

1. INTRODUCTION

Nonlinearity is an important theme underlying many current developments in statistics and in applied mathematics. In this paper we discuss a few general aspects of nonlinearity in connexion with components of variance and then examine two special topics in more detail.

Our emphasis is on nonlinear models rather than on nonlinear methods of estimation, the latter being needed for efficient estimation even in unbalanced linear normal theory problems. Nonlinear models may be considered because they are of intrinsic interest or because they may be a convenient technical device for testing the adequacy of standard linear models. There is also a distinction between situations in which the variance components are of primary interest and those in which the components are required only for the assessment of the precision of contrasts.

If the amount of nonlinearity is small, local linearization will produce a model of linear form, typically unbalanced.

2. TYPES OF NONLINEARITY

We shall not attempt a careful classification of types of nonlinearity. In linear theory, the simplest and most general models are respectively

\[ Y_{is} = \mu + A_i + B_{is} \quad (i = 1, \ldots, m; \; s = 1, \ldots, r), \]

\[ Y = x\beta + aU. \]
In (1), \( A_1, A_2, \ldots, A_m \) are random variables of zero mean and variance \( \sigma^2_A \), independently normally distributed or, in second-order theory, uncorrelated and the \( B_{is} \) have analogous properties and variance \( \sigma^2_B \). In (2), \( Y \) is the full \( n \times 1 \) vector of observed response variables, the \( n \times q \) matrix \( x \) and the \( q \times 1 \) vector \( \beta \) of unknown parameters define the systematic part of the variation, whereas the matrix \( A \) and the vector \( U \) of random variables define the random structure; typically \( U \) contains components of \( k \) different types with variances \( \sigma^2_1, \sigma^2_2, \ldots, \sigma^2_k \), say.

Nonlinearity can arise from (1) and (2) in the following ways.

(i) The systematic part \( x\beta \) is replaced by a nonlinear form; for example, Rudemo, Ruppert & Streibig (1989) consider such models in application to bioassay data.

(ii) The random components, for example \( A_i \) and \( B_{is} \) in (1), combine nonlinearly; e.g. the nonlinearity may be modelled approximately as

\[
Y_{is} = \mu + A_i + B_{is} + \alpha_{20}A_i^2 + \alpha_{11}A_iB_{is} + \alpha_{02}B_{is}^2.
\]

(iii) The random and systematic parts in (2) combine nonlinearly, for example in an exponential growth model with random doubling time; i.e.

\[
Y_{is} = \exp\{\mu + (\beta + A_i)x_{is}\} + B_{is}.
\] (3)

Linear models combining random and systematic components, so-called regression models of the second kind, are presented in a broad context by Nelder (1977). See Racine-Poon (1985) for a Bayesian approach to estimation in such nonlinear models.

(iv) The essentially normal-theory based structure of (1) and (2) can be replaced by an analogous form for the exponential family. For the Poisson and binomial distributions with one overlaid component of variation this has a long history in the context of the negative binomial (Greenwood & Yule, 1920) and beta binomial (Skellam, 1948) distributions respectively. Much recent work has focused on the conceptual and numerical analytical problems of models for the canonical parameters of exponential family models of essentially linear structure, i.e. applying representations analogous to (2) not to observations but to canonical parameters.

In § 3 we examine the simplest model of type (ii) in which nonlinearity enters a one-way balanced structure. In § 4 we examine a technique for approximating the likelihoods for nonlinear variance component models.

3. Balanced structures

3-1. Formulation of model

Consider first a one-way arrangement in which \( r \) independent repeat continuous observations are obtained on each of \( m \) groups, the groups being regarded as a random sample from a population of groups, so that model (1) is potentially applicable. Now if \( \mu \) is defined as the overall population mean, and the mean for the \( i \)th group is defined to be \( \mu + A_i \), then the properties \( E(A_i) = 0, E(B_{is} | A_i = a) = 0 \) are matters of definition; all other properties of \( \{A_i\} \), \( \{B_{is}\} \) are to some extent capable of empirical test. Note that, in the absence of additional information about the groups, it is not meaningful to make, say, the properties of \( \{B_{is}\} \) depend on the label \( i \) as such. To represent local departures from the normal-theory version of (1) we shall investigate the representation

\[
Y_{is} = \mu + A_i + B_{is} + \alpha_{20}A_i^2 + \alpha_{11}A_iB_{is} + \alpha_{02}B_{is}^2,
\] (4)
where \(\{A_i\}\) and \(\{B_{is}\}\) are zero mean, independently normally distributed random variables with variances \(\sigma_{A_i}^2, \sigma_{B_{is}}^2\). We assume \((\sigma_{A_i}^2, \sigma_{B_{is}}^2)\) to be small and retain only first order terms; a formal asymptotic treatment would require them to be of order \(1/\sqrt{m}\) as \(m \to \infty\).

The second-order representation (4) models a quite general form of nonlinearity involving both skewness of the random effects and heterogeneity of the within-group variation. To interpret the parameters \(\sigma_{A_i}^2, \sigma_{B_{is}}^2\) and \(\sigma_{A_{i,s}}^2\) which capture these features, note first that we can write

\[
\mu + A_i + \sigma_{A_i}^2 A_i^2 = \mu' + A_i',
\]

where \(\mu' = \mu + \sigma_{A_i}^2\) and \(A_i'\) is a nonnormal random variable of zero mean, approximate variance \(\sigma_{A_i}^2\) and third moment \(9\sigma_{A_i}^2\), so that the standardized third cumulant of \(A'\) is

\[
\rho_A = 9\sigma_{A_i}^2
\]

with a similar interpretation for \(\rho_{B_s}\). Because \(\rho_A, \rho_{B_s}\) are dimensionless, they are more helpful for qualitative interpretation than \(\sigma_{A_i}^2, \sigma_{B_{is}}^2\).

Equation (5) can be regarded as two terms of the Fisher–Cornish inversion of the Edgeworth expansion of a nearly normal random variable \(A_i'\) and suffers from the usual disadvantage of such expansions of nonmonotonicity in \(A_i\). The importance of this depends on the context. Note that the nonmonotonicity with respect to \(A_i\), or to \(B_{is}\), but so far as we know not both simultaneously, can be removed by rewriting (5) in the form

\[
\mu + (2\sigma_{A_i}^2)^{-1}(e^{2\sigma_{A_i}^2 A_i} - 1),
\]

equivalent to the first order in \(\sigma_{A_i}^2\); we shall not do this here.

In some ways, however, the most interesting parameter is \(\sigma_{A_{i,s}}^2\). Note that given \(A_i = a\sigma_A\), the random variable attached to observation \((i, s)\) is \(B_{is}(1 + \sigma_{A_{i,s}} a\sigma_A) + \sigma_{B_{is}}^2\), which has approximate standard deviation \((1 + \sigma_{A_{i,s}} a\sigma_A)\sigma_B\); thus

\[
\rho_{i,s} = \sigma_{A_{i,s}} a\sigma_A
\]

is a dimensionless measure of the rate of change of the conditional standard deviation of within group variation with the group mean, thus completing the interpretation of the parameters in (4).

An important special case of (4) is the family of transformation models which is discussed in § 3–2 and exemplified in § 3–5.

### 3.2. A transformation model

The model (4) with three additional parameters can be contrasted with a transformation model (Solomon, 1985) with one additional parameter, namely that for some \(\kappa\)

\[
Y_{is}^{1/\kappa} = \mu + A_i^\kappa + B_{is}^\kappa,
\]

the right-hand side being a normal-theory representation. Provided that \(\mu\) is very much greater than \(\sqrt{(\sigma_{A_i}^2 + \sigma_{B_{is}}^2)}\) we have an expansion that, to quadratic terms,

\[
Y_{is} = \mu^\kappa \{1 + (A_i^\kappa + B_{is}^\kappa)/\mu\}^\kappa
= \mu^\kappa + \kappa \mu^{\kappa - 1} A_i^\kappa + \kappa \mu^{\kappa - 1} B_{is}^\kappa + \frac{1}{2} \kappa (\kappa - 1) \mu^{\kappa - 2}(A_i^{2\kappa} + 2 A_i^\kappa B_{is}^\kappa + B_{is}^{2\kappa}),
\]

which is of the form (4) with, however, the special feature that \(\sigma_{A_i}^2 = \sigma_{A_{i,s}}^2 = \frac{1}{2} \sigma_{A_{i,s}}^2\).

The transformation model (9) is clearly much more restrictive than the general model (4); however transformation models are widely used in practice and often provide a
simple basis for statistical analysis and interpretation. Choice between the models will depend on the context, as well as on practical considerations. Estimation based on (4) separates three distinct ways in which the statistical model may fail, and thus, in particular, throws light on what is achieved by transformation.

By way of example, and comparison, §3.5 discusses an analysis of data on blood pressure using both models.

3.3. Statistical analysis

One possible approach to statistical analysis based on (4), especially for testing the null hypothesis \( \alpha_{20} = \alpha_{11} = \alpha_{02} = 0 \), is to compute log likelihood derivatives. We shall follow the equivalent but less formal route of considering, together with the second-order statistics, the cubic statistics

\[
S_{30} = \sum_{i} (\bar{Y}_i - \bar{Y}_.)^3, \quad S_{21} = \sum_{i} (\bar{Y}_i - \bar{Y}_.)^2(\bar{Y}_{2i} - \bar{Y}_i),
\]

\[
S_{12} = \sum_{i} (\bar{Y}_i - \bar{Y}_.)^2(\bar{Y}_{3i} - \bar{Y}_i)^2, \quad S_{03} = \sum_{i} (\bar{Y}_{3i} - \bar{Y}_i)^3,
\]

of which the second is identically zero. Direct evaluation under (4) gives

\[
E(S_{30}) = \alpha_{20} \sigma_A^6 r \frac{m}{m-1}(m-2) - \alpha_{11} \sigma_A^2 \sigma_B^2 6 \frac{m}{m-1}(m-2) + \alpha_{02} \sigma_B^4 6 \frac{1}{m} (m-1)(m-2),
\]

\[
E(S_{12}) = \alpha_{11} \sigma_A^2 \sigma_B^2 2(m-1)(r-1) + \alpha_{02} \sigma_B^4 \frac{r}{r} (m-1)(r-1),
\]

\[
E(S_{03}) = \alpha_{02} \sigma_B^4 \frac{r}{r} (r-1)(r-2).
\]

Thus unbiased estimates of \( \alpha_{20}, \alpha_{11}, \alpha_{02} \) can be found if \( \sigma_A^2, \sigma_B^2 \) are known. If, as would typically be the case, \( \sigma_A^2, \sigma_B^2 \) are unknown and replaced by the usual estimates, we obtain as consistent estimates of the \( \alpha \)'s

\[
\hat{\alpha}_{20} = \frac{m}{\sigma_A^6} \frac{r}{m} (m-1)(m-2) \{ S_{30} + \frac{3}{m(r-1)} S_{12} - \frac{(m-1)(m+1)}{m^2(r-1)(r-2)} S_{03} \},
\]

\[
\hat{\alpha}_{11} = \frac{1}{\sigma_A^2 \sigma_B^2} 2(m-1)(r-1) \{ S_{12} - \frac{(m-1)}{m(r-2)} S_{03} \},
\]

\[
\hat{\alpha}_{02} = \frac{r}{\sigma_B^4 6 m(r-1)(r-2)} S_{03}.
\]

These are converted into estimates \( \hat{\rho}_A, \hat{\rho}_{11}, \hat{\rho}_B \) of the dimensionless parameters via the standard estimates of the components of variance \( \sigma_A^2, \sigma_B^2 \).

In this preliminary study we shall not develop in detail the sampling properties of the resulting estimates. Under the null hypothesis of normality, for large \( r \) and \( m \), the standard error of \( \hat{\rho}_B \) is \( \sqrt{6/(rm)} \); the corresponding standard error for \( \hat{\rho}_A \) is \( \sqrt{6/m} \) if the "correction terms" are small, i.e. if the estimate is essentially the standardized third cumulant of the \( m \) sample means.

Note that a very general mathematical development for calculating cumulants in component of variance models is given by Speed in a series of papers that appeared in the 1980's; see, e.g. Speed & Silcock (1988).

There are numerous ways of testing whether \( \alpha_{11} = 0 \). A graphical method closely corresponding to the estimation of \( \alpha_{11} \) is to plot the within group sum of squares for the \( i \)th group against the group mean. Indeed because under the null hypothesis \( \alpha_{20} = \alpha_{11} = \alpha_{02} = 0 \) the quantities plotted are independent, an 'exact' permutation test of regression
Nonlinear component of variance models

is available. This could take various forms depending on the choice of test statistic. The most effective practical procedure is likely to be to regress the log mean square within groups on the sample mean applying standard least squares methods. In doing this the residual mean square has a known theoretical value under normality and a permutation test could be applied in those cases where careful, cautious assessment of significance is important.

3.4. Generalizations

While the above discussion has been for the one-way arrangement, essentially similar arguments apply to general balanced cross-classified and nested arrangements. For example, the hierarchical model with three nested effects has a representation analogous to (4),

$$Y_{ist} = \mu + A_i + B_{is} + C_{ist} + \alpha_{200}A_i^2 + \alpha_{020}B_{is}^2 + \alpha_{002}C_{ist}^2 + \alpha_{110}A_iB_{is} + \alpha_{101}A_iC_{ist} + \alpha_{011}B_{is}C_{ist},$$

and straightforward calculation shows that the cubic statistics, again defined analogously to (11), have expectations

$$E(S_{003}) = \alpha_{002}\sigma_C^4 6m \frac{r_1}{r_2} (r_2 - 1)(r_2 - 2),$$

$$E(S_{030}) = \alpha_{020}\sigma_B^4 6m \frac{r_2}{r_1} (r_1 - 1)(r_1 - 2) - \alpha_{011}\sigma_B^2 \sigma_C^2 6 \frac{m}{r_1} (r_1 - 1) + \alpha_{002}\sigma_C^4 6 \frac{m}{r_2} (r_1 - 1),$$

and so on.

More importantly, under the transformation model,

$$\alpha_{200} = \alpha_{020} = \alpha_{002} = \frac{1}{2}\alpha_{110} = \frac{1}{2}\alpha_{101} = \frac{1}{2}\alpha_{011},$$

which is simply a generalization of the special feature of (4), and similarly for general balanced cross-classified arrangements.

3.5. An example

To illustrate the above discussion, we analyze data on the blood pressure of hypertensive males obtained in the International Prospective Primary Prevention Study in Hypertension (Solomon, 1985). The groups are $m = 25$ individual patients and the repeat observations refer to the $r = 16$ distinct measurements per patient; of course we are here ignoring the time series structure of the data. Table 1 gives the main statistics computed from diastolic and systolic blood pressures (mmHg) and various transforms of these.

Table 1. Estimates of the $\alpha$ parameters and their dimensionless equivalents for data on blood pressure for 25 hypertensive males from the International Prospective Primary Prevention Study in Hypertension

<table>
<thead>
<tr>
<th>Blood pressure</th>
<th>$\hat{\alpha}_{20}$</th>
<th>$\hat{\alpha}_{02}$</th>
<th>$\hat{\alpha}_{11}$</th>
<th>$\hat{\rho}^{A}_{2}$</th>
<th>$\hat{\rho}^{B}_{2}$</th>
<th>$\hat{\rho}_{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>diastolic</td>
<td>0.0166</td>
<td>0.0114</td>
<td>0.0248</td>
<td>0.7565</td>
<td>0.7150</td>
<td>0.1254</td>
</tr>
<tr>
<td>$\sqrt{\text{diastolic}}$</td>
<td>0.2851</td>
<td>0.0924</td>
<td>0.2908</td>
<td>0.6686</td>
<td>0.3113</td>
<td>0.0758</td>
</tr>
<tr>
<td>log diastolic</td>
<td>1.1656</td>
<td>0.0410</td>
<td>0.7568</td>
<td>0.5703</td>
<td>0.0276</td>
<td>0.0411</td>
</tr>
<tr>
<td>systolic</td>
<td>0.0992</td>
<td>0.0043</td>
<td>0.0121</td>
<td>1.1428</td>
<td>0.5026</td>
<td>0.1669</td>
</tr>
<tr>
<td>$\sqrt{\text{systolic}}$</td>
<td>0.1915</td>
<td>0.0455</td>
<td>0.1773</td>
<td>0.9587</td>
<td>0.2251</td>
<td>0.0986</td>
</tr>
<tr>
<td>log systolic</td>
<td>0.9409</td>
<td>0.0151</td>
<td>0.5308</td>
<td>0.7710</td>
<td>0.0117</td>
<td>0.0479</td>
</tr>
</tbody>
</table>
Under the transformation model, we expect $a_{20} = a_{02} = \frac{1}{2}a_{11}$. The corresponding estimates of the parameters for both diastolic (0.0166, 0.0114, 0.0248) and systolic (0.0092, 0.0043, 0.0121) pressures reasonably approximate this expected relationship. The estimates suggest therefore that the transformation model, which we have noted as being really quite restrictive, is appropriate. Indeed, previous analysis (Solomon, 1985) showed that, assuming a common transformation in a bivariate model, log transformation for both diastolic and systolic blood pressures is close to maximizing the likelihood.

The moment calculations presented in Table 1 do more than validate the application of the transformation model. The parameter estimates provide greater insight into the underlying relationships in the data, as well as into what the transformation model is doing. The parameters $\rho_3^A$ and $\rho_3^B$ represent the degree of skewness of the between- and within-patient effects. For diastolic blood pressure, $\hat{\rho}_3^A = 0.7565$ and $\hat{\rho}_3^B = 0.7150$ are of moderate and comparable skewness on the original scale, but the distribution of the between-patient effects for systolic pressure has a relatively longer tail ($\hat{\rho}_3^B = 1.1428$). This difference between the blood pressures is likely to be a reflection of the fact that diastolic, but not systolic, pressure was subject to a treatment target level of 95 mmHg. Note, however, that log transformation virtually eliminates the within-patient skewness, and reduces the between-patient skewness.

The estimated rate of change of the conditional standard deviation of the within-patient variation with the patient mean, $\hat{\rho}_{11}$, is 16.7% for systolic, and 12.5% for diastolic pressure. Thus a moderate proportion of the within-patient variation is attributable to the fractional change of the standard deviation within-patients. In practice, such variability may well be important biologically. For instance, published standards of treatment dosage for hypertensives typically assume that the relationship between the ‘true’ patient mean and the ‘true’ within-patient variation is known, and that the conditional standard deviation of the within-patient variation does not change with changes in the mean; clearly such assumptions may not be appropriate and the example illustrates that superficial understanding of the underlying relationships in the data on blood pressure may well lead to inappropriate treatment. Again, note that both square root and log transformation reduce this change in the dependency between the patient mean and within-patient variation.

4. Likelihoods for nonlinear variance component models

4.1. An approximate likelihood

We now turn to a quite different issue, namely the approximate evaluation of likelihoods for models containing at least one additional level of random variation entering nonlinearly into a likelihood of given form. An example is the likelihood for the exponential regression model (3) with random doubling times. We propose one such approximation.

Let $A_1, \ldots, A_m$ be independently, identically normally distributed with mean $\mu_A$ and variance $\sigma_A^2$, the variance being in some sense relatively small. Suppose that for fixed $A_1, \ldots, A_m$ the likelihood has the form $\Pi L_j(\theta, A_j)$, where $\theta$ is a parameter which may or may not be of primary interest. Then the required likelihood is $\Pi E\{L_j(\theta, A_j)\}$, where the expectation is over the distribution of $A_j$. Thus we consider approximations to $E\{L(\theta, A)\}$, dropping the suffix $j$ temporarily. Write

$$l(\theta, A) = \log L(\theta, A), \quad l^{(r)} = [\frac{\partial^2 l(\theta, A)}{\partial A^2}]_{A = \mu_A},$$

and define $A^* = (A - \mu_A)/\sigma_A$, which has a standard normal distribution.
Nonlinear component of variance models

Then

\[ L(\theta, A) = L(\theta, \mu_A) \exp \left( \sigma_A l^{(1)} A^* + \frac{1}{2} \sigma_A^2 l^{(2)} A^{*2} \right) (1 + \frac{1}{6} \sigma_A^3 l^{(3)} A^{*3} + \frac{1}{24} \sigma_A^4 l^{(4)} A^{*4}) \]  \tag{12}

We now take expectations over the distribution of \( A^* \). The first term gives

\[ \frac{1}{\sqrt{1 - \sigma_A^2 l^{(2)}}} \exp \left\{ \frac{\sigma_A^2 l^{(1)}^2}{2(1 - \sigma_A^2 l^{(2)})} \right\}, \]  \tag{13}

the leading term of a Laplace expansion. It is exact for a one-way normal theory arrangement. Inclusion of further terms in (12) multiplies (13) by

\[ 1 + \left( \frac{1}{6} l^{(3)} + \frac{1}{24} l^{(4)} \right) \sigma_A^4 + o(\sigma_A^4), \]

which to ensure positivity we rewrite in exponentiated form as

\[ \exp \left\{ \left( \frac{1}{6} l^{(3)} + \frac{1}{24} l^{(4)} \right) \sigma_A^4 \right\} \{ 1 + o(\sigma_A^4) \}. \]  \tag{14}

Thus the final approximation to the full likelihood is

\[ \prod L_j(\theta, \mu_A) \left\{ (1 - \sigma_A^2 l^{(2)}) \right\}^{-\frac{1}{2}} \exp \left\{ \frac{\sigma_A^2 l^{(1)}^2}{2(1 - \sigma_A^2 l^{(2)})} + \frac{1}{6} (l^{(3)} + \frac{1}{24} l^{(4)}) \sigma_A^4 \right\} \{ 1 + o(\sigma_A^4) \}. \]  \tag{15}

Of course, were the final term in (14) and (15) appreciable, this would be a warning of the probable failure of the expansions employed. We point out that large values of the \( \alpha \)'s defined in § 3 take us outside the range of the approximation, but point qualitatively in the right direction.

For example, if the random variables \( Y_{js} \) (\( j = 1, \ldots, m; s = 1, \ldots, r \)) have the structure

\[ Y_{js} = \exp \left\{ (\theta + A_j) x_{js} \right\} + B_{js}, \]  \tag{16}

where \( \{B_{js}\}, \{A_j\} \) are independently normal with zero mean and variances \( \sigma_B^2, \sigma_A^2 \), and the \( \{x_{js}\} \) are fixed constants, we may apply the above results taking \( L_j \) to refer to the set \( \{ Y_{j1}, \ldots, Y_{jr} \} \). There is a slight gain in simplicity by considering the special case in which \( x_{js} = x_j \) (\( j = 1, \ldots, m \)).

Let \( T_j(\theta) = \sum Y_{js} x_j e^{\theta x_j}, m_r(\theta) = 2^{(r-1)} \sum x_j e^{\theta x_j}. \) Then

\[ l^{(r)}_r = \{ T_j(\theta) - m_r(\theta) \}/\sigma_B^2, \]  \tag{17}

so that the contribution to the likelihood from the \( j \)th sample is, from (15), approximately

\[ \frac{1 - \sigma_A^2 \{ T_2(\theta) - m_2(\theta) \}/\sigma_B^2}{(2 \pi)^{\frac{1}{2}}} \exp \left\{ -\frac{\sigma_A^4 \{ T_1(\theta) - m_1(\theta) \}^2}{2 \sigma_B^2} \right\} \exp \left\{ \frac{1}{2 \sigma_A^4} \left[ \frac{T_3(\theta) - m_3(\theta)}{\sigma_B^2} \right] \right\} \].  \tag{18}

We check the accuracy of the approximation numerically, divorced from stochastic aspects, by comparing it with the exact answer

\[ \int_{-\infty}^{\infty} \frac{1}{(2 \pi)^{\frac{1}{2}}} \sigma_B^2 \exp \left\{ -\frac{1}{2} \left( \frac{Y_{js} - e^{(\theta+A_j)x_j}}{2 \sigma_B^2} \right)^2 \right\} \exp \left( \frac{-A_j^2}{2 \sigma_A^2} \right) dA_j, \]  \tag{19}

based on simulated rather than empirical data for a particular choice of parameter values.

For illustration we take \( m = 10 \) samples with \( r = 5 \) observations in each sample, and values of the parameters \( \sigma_B = 0.25, \theta = 0.25, \{ x_j \} = \{ 1, 2, 3, 4, 5 \} \) and \( \sigma_A = 0.01 \). Figure 1
shows the log likelihood functions, over a reasonable range of $\theta$, corresponding to the exact likelihood, the approximation with only the leading term, and the approximation with the correction term, evaluated at the true parameter values. The plot demonstrates clearly that, even with a moderate standard deviation of $A$, the approximation performs well, the log likelihood functions being almost indistinguishable at this level. Note that for a single sample, the range of values of $\theta$ of interest will be wider than that depicted.

For the same $m$, $r$, $\sigma_B$, $\theta$ and $\{x_i\}$ as above, and with $\sigma_A$ increasing from 0.0025 to 0.02, Table 2 sets out the exact and approximate likelihoods evaluated at the true parameter values, together with the correction factor. For small to moderate $\sigma_A^2$ the agreement between the exact and approximate likelihoods is extremely good. It is worth noting that the variation in slope is detectable visually even at very small standard deviations of $A$. At these small levels, the correction term typically takes close to unit values.

### Table 2. Exact and approximate likelihoods for $\theta$, $\sigma_A$, $\sigma_B$ in exponential model (16), and correction factor, evaluated at the true parameter values, for increasing $\sigma_A$

<table>
<thead>
<tr>
<th>$\sigma_A$</th>
<th>Exact likelihood</th>
<th>Approximate likelihood (with correction)</th>
<th>Correction factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0025</td>
<td>2.70416</td>
<td>2.70419</td>
<td>0.99991</td>
</tr>
<tr>
<td>0.0075</td>
<td>0.16998</td>
<td>0.16807</td>
<td>0.94724</td>
</tr>
<tr>
<td>0.01</td>
<td>0.06164</td>
<td>0.06900</td>
<td>1.02115</td>
</tr>
<tr>
<td>0.0125</td>
<td>0.010198</td>
<td>0.09835</td>
<td>0.68910</td>
</tr>
<tr>
<td>0.015</td>
<td>0.01643</td>
<td>0.01015</td>
<td>0.32259</td>
</tr>
<tr>
<td>0.0175</td>
<td>$4 \times 10^{-9}$</td>
<td>$3 \times 10^{-9}$</td>
<td>0.38876</td>
</tr>
<tr>
<td>0.02</td>
<td>0.000201</td>
<td>0.00000852</td>
<td>142.868</td>
</tr>
</tbody>
</table>
Nonlinear component of variance models

However, Table 2 illustrates that as soon as the correction factor adjusts the leading term by 20 to 30%, the method may fail badly. For instance, for values of $\sigma_A$ nearer 0.02, the correction term can make much more than a minor difference. We recommend using the approximation provided the correction factor lies between 0.75 and 1.25. Although Table 2 shows that in particular cases the approximation may be reasonable for values of the correction factor lying outside this range, we caution that care is needed in application.

When the standard deviation of $A$ is known to be large relative to $\sigma_B$, an alternative approach is required.

4.2. Several 2×2 tables

We consider briefly the combination of information from several 2×2 contingency tables assumed to have a constant logistic difference $\Delta$ between the two ‘treatments’ under comparison via a binary response. Again, for simplicity, suppose that there are in the $j$th table $n$ individuals receiving treatment 0, $n$ receiving treatment 1, that the numbers of successes are $(Y_{j0}, Y_{j1})$ and that the logit probabilities of successes are $(\alpha_j, \alpha_j+\Delta)$, so that the contribution to the likelihood is

$$\exp \left\{ \alpha_j(y_{j0} + y_{j1}) + \Delta y_{j1} \right\} \over (1 + e^{\alpha_j})^n(1 + e^{\alpha_j+\Delta})^n$$

If the $\alpha_j$ are totally arbitrary nuisance parameters, inference about $\Delta$ is based on the likelihood for $\Delta$ conditional on $y_{j0} + y_{j1}$ ($j = 1, \ldots, m$), that is on a product of generalized hypergeometric forms. For testing $\Delta = 0$, the Mantel–Haenszel test results. If the $\alpha_j$ are assumed to be all the same, we pool the tables. An intermediate situation is to suppose the $\alpha_j$ to be random variables $A_j$ independently normally distributed with mean $\mu_A$ and variance $\sigma_A^2$.

The approximation (15) can now be used for the ensuing likelihood. In this outline investigation we concentrate on testing $\Delta = 0$ via the score statistic, i.e. the gradient of log likelihood with respect to $\Delta$ evaluated at $\Delta = 0$. Further we shall use a pre-specified value of $\sigma_A^2$, in particular because reasonably precise estimation of $\sigma_A^2$ is unlikely to be feasible unless the amount of data is very large. The unknown parameter $\mu_A$ will be replaced by the logit of the overall proportion of successes, which is the maximum likelihood estimate under $\Delta = 0$ and the values of $\sigma_A^2$ employed.

For our numerical work we have considered a one-sided test of $\Delta = 0$ with alternative $\Delta > 0$, taking $m = 10$ and studying power over the range $0 < \Delta \leq 1$. Reasonable values of $\sigma_A^2$ are often 0.05, 0.10; if the overall success rate is $\frac{1}{2}$, probabilities corresponding to $\alpha_j$ would have standard deviations approximately 0.056, 0.079, respectively.

The first derivative of the log likelihood ($\partial l / \partial \Delta$)$_{\Delta=0}$, depends on

$$t_2 = \sum (y_{j2} - n\pi), \quad t. = \sum (y_{j1} + y_{j2} - 2n\pi),$$

where $\pi = e^{\mu_A} / (1 + e^{\mu_A})$. Now $t_2$ is closely related to the Mantel–Haenszel statistic, in the derivation of which $t.$ is held constant as a conditioning variable. In detail

$$(\partial l / \partial \Delta)_{\Delta=0} = t_2 + \sigma_A^2(a_1 t. + b_1) + \sigma_A^2(a_2 t^2 + b_2 t. + c_2),$$
where

\[ a_1 = -n\pi(1 - \pi)\left(1 + 2\sigma^2n\pi(1 - \pi)^{-1}\right), \quad a_2 = n\pi(1 - \pi)(\pi - \frac{1}{2})\left(1 + 2\sigma^2n\pi(1 - \pi)^{-1}\right), \quad a_3 = n^2\sigma^2(1 - \pi)^2(\pi - \frac{1}{2}) - \frac{1}{2}n\pi(1 - \pi)(12\pi^2 - 10\pi + 1) + \frac{1}{2}n\pi^2(1 - \pi)(6\pi^2 - 6\pi + 1), \]

\[ b_1 = n\pi(1 - \pi)(\pi - \frac{1}{2})(12\pi^2 - 10\pi + 1), \quad b_2 = -\frac{1}{2}n\pi(1 - \pi)(6\pi^2 - 6\pi + 1), \]

The second derivative \(\frac{\partial^2}{\partial \Delta^2}\Delta=0\) has an elementary but lengthy form which will not be reproduced here; it is available from the authors on request.

In the first instance we examine the distribution of the score test statistic under the null hypothesis of no logistic difference; the mean and variance of the statistic should be 0 and 1 respectively. Based on 200 simulations with \(n = 10\) individuals receiving each treatment, and assuming the true values \(\mu_A = 0\) and \(\sigma_A^2 = 0.05\), the estimated mean of the statistic is 0.097 with standard deviation 1.012. Similarly, assuming a true value of \(\sigma_A^2 = 0.10\), the corresponding estimates are -0.021 and 0.953. We found that the distribution of the score test statistic is not much changed when the wrong \(\sigma_A^2\) is substituted.

Table 3 sets out the results from analogous simulations, but here we substitute the maximum likelihood estimate of \(\mu_A\) for its true value \(\mu_A = 0\) in calculating the mean and variance of the test statistic. We also take \(n = 25\) individuals, \(m = 25\) and assume the true value of \(\sigma_A^2\) over a range. Again, the aim is to see whether the estimated mean and variance of the score test statistic are near 0 and 1. The observed negative values of the means are all very small, although the standard deviations are consistently less than one. This may suggest that using the wrong value of the mean \(\mu_A\) is potentially more serious than using the wrong variance. Increasing both \(m\) and \(n\) would be expected to improve the asymptotic behaviour of the test statistic, which is nonetheless reasonable even under the present assumptions.

We were interested to compare the power of the score test to that of the Mantel–Haenszel test with continuity correction. Whichever test has the larger mean is likely to be more

**Table 3. Mean and standard deviation of score test statistic under the null hypothesis \(\Delta = 0\) with true value of \(\sigma_A^2\) and maximum likelihood estimate of \(\mu_A\). Results from 200 simulations with \(m = 25\), \(n = 25\), \(\mu_A = 0\)**

<table>
<thead>
<tr>
<th>(\sigma_A^2)</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.029</td>
<td>0.929</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.009</td>
<td>0.923</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.076</td>
<td>0.829</td>
</tr>
</tbody>
</table>

**Table 4. Mean and standard deviation of score test statistic and Mantel–Haenszel test statistic, with continuity correction, for ranges of true values of \(\Delta\) and \(\sigma_A^2\); \(n = m = 10\), 30 simulations. Standard deviation in parentheses**

<table>
<thead>
<tr>
<th>(\Delta)</th>
<th>Score</th>
<th>Mantel–Haenszel</th>
<th>(\sigma_A^2 = 0.05)</th>
<th>Score</th>
<th>Mantel–Haenszel</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.132 (0.972)</td>
<td>0.184 (0.919)</td>
<td>0.049 (0.977)</td>
<td>0.234 (0.961)</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.232 (0.933)</td>
<td>0.009 (1.081)</td>
<td>0.467 (0.805)</td>
<td>0.231 (1.113)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>2.241 (0.818)</td>
<td>1.421 (0.830)</td>
<td>2.288 (0.823)</td>
<td>1.841 (0.984)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.721 (1.053)</td>
<td>2.747 (1.065)</td>
<td>3.754 (0.847)</td>
<td>3.139 (0.956)</td>
<td></td>
</tr>
</tbody>
</table>
sensitive to departures from the null hypothesis, as we are looking to recognize values beyond the 5% tail. Table 4 presents selected distributions for the test statistics evaluated over a range of true parameter values. In each case the mean and standard deviation are based on 30 paired simulations, the standard errors of the differences being approximately 0.12. The results from simulations illustrate that under departures from the null hypothesis, the score test is the more sensitive.

The score test thus has increased power under certain conditions. Whether it is wise to use the score test in practice is doubtful. As compared with the Mantel–Haenszel procedure, the new test assumes more, is more difficult to apply and lacks the simple direct appeal of the Mantel–Haenszel test. Nevertheless the possibility of improved power in problems with many nuisance parameters of similar interpretation illustrates a general phenomenon of broad theoretical interest. We stress that there are a number of aspects of this problem for which detailed exploration is desirable but will not be attempted here.

ACKNOWLEDGEMENTS

P. J. Solomon thanks SERC and the 1990/91 Australian Bicentennial Fellowship for support. The work was completed while P. J. Solomon was at the National Centre for Epidemiology and Population Health, Canberra.

REFERENCES


[Received December 1990. Revised May 1991]