THE ESTIMATION OF GENETIC PARAMETERS
FOR CATEGORICAL TRAITS

A thesis presented in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
in Animal Science
at Massey University.

ARTHUR RICHARD GILMOUR

1983

ABSTRACT

The estimation of heritabilities, genetic and environmental variances and covariances, and the prediction of breeding values, is a major concern among animal breeders. This study adapts existing statistical methods to provide a new method of estimating these parameters for categorical traits.

The problems associated with the analysis of categorical data arise because of the relationship between the mean and variance of the sampling distribution. The parameters of the sampling distribution are assumed to be a non-linear function of values on an underlying scale. It is further assumed that fixed and random effects are additive on the underlying scale. This scale cannot be observed and information about it must be deduced from the observed categorical trait.

A common practice has been to estimate parameters (fixed effects and variances) on the categorical variable itself and then to transform these estimates to values applicable to the underlying scale. This procedure is theoretically invalid
JUSTIFY? except for a fully random model in which the only fixed effect is the mean. The method developed in this thesis attempts to estimate parameters directly in the underlying scale by transforming the data before calculating estimates of the parameters. It is a synthesis of mixed model procedures (Henderson *et al.*, 1959) and generalized linear models (Nelder and Wedderburn, 1972) and is called the generalized linear mixed model. The general method is for analysing data presumed to arise from a two-stage sampling procedure when the second sampling has an error distribution
means belonging to the single parameter exponential family.

The detailed algebra for applying the new method to binomial and multinomial traits for the estimation of fixed effects is
JUSTIFY. presented. The logit transformation is used in this application and the resulting system of equations is called the logistic linear mixed model. A procedure for estimating

variances, covariances and the intraclass correlation on the underlying scale is also developed.

The logistic linear mixed model is evaluated by comparing parameter estimates from the method with true values used to generate the data being analysed. Biases appear to be small except for some extreme combinations of parameters when assumptions made while developing the algebra break down.

The logistic linear mixed model is applied to two real problems for which fixed and random effects and variance components are estimated and comparisons made with parameters estimated by other methods. The first problem is the analysis of data on the feet characteristics of 2513 lambs, the second is the analysis of 1396 lambing performance records.

The concluding discussion considers the general use of the logistic linear mixed model and its relationship with other models.

ACKNOWLEDGEMENTS

I gratefully acknowledge the guidance of my supervisors, Professor R.D.Anderson, Professor A.L. Rae and Dr.G.A. Wickham.

The Australian Wool Corporation provided the post-graduate scholarship which enabled this study to be undertaken. It is thanked, along with the New South Wales Department of Agriculture which provided study leave and financial assistance.

The Animal Science Department has provided a fine working environment and the opportunity to broaden my outlook.

My gratiude is extended to the friends we have made, particularly in Milson. Their willingness to include us in their community has made life both pleasant and rewarding for me and especially for my wife, Ellen, and our three daughters. In particular, I acknowledge the friendship of Graham and Lynette Toms and the members of Milson Christian Fellowship.

To Ellen and the three girls, thank you for being there when needed and bearing with me when times were busy.

To my God, thank you for providing all our needs - friends, a home, work and recreation, health, finance and the opportunity to serve - through Jesus Christ.

Table of Contents

Section Heading page

List of Tables

List of Figures

Chapter 1. Introduction

Animal breeders aim to improve the genetic merit of herds and flocks for productive performance by selecting, as parents for the next generation, those animals most likely to maximize genetic advance in the population. Genetic merit is defined by use of a function of the productive traits to be improved, the selection objective. Traits which might be included in a selection objective include the total weight of wool produced by sheep, the total amount of milk or milk fat produced by cows, the liveweight gain of meat animals, liability to disease and numbers of progeny.

The selection of animals to be used as parents is usually made before all the information required by the selection objective, is available. The assessment of merit must therefore be made on the basis of the observations that are available on the candidates for selection and/or their close relatives. The principal question for animal breeders is: What function of the observations is the best indicator of the potential merit of an animal if used as a parent? The various techniques proposed differ principally in what is assumed about the observations and the population under selection. Four situations are readily identifiable (Henderson, 1973):

1) When the form of the joint distribution of the observations and the variables to be predicted, together with the actual values of the parameters characterizing that distribution are known, selection can be on the conditional mean of the predicted values. This situation rarely occurs in practice.

2) When the form of the distribution is unknown but means, variances and covariances of the observations and variables to be predicted, are regarded as known, the selection index (Hazel, 1943) is used.

3) When the variances and covariances, but not the means, of the observations and variables to be predicted, are regarded as known, best linear unbiased prediction (BLUP) is used.
4) When variances and covariances are unknown, they must be estimated. BLUP procedures may then be used but will not be unbiased if the variances are estimated from the same observations.

Most of the literature on the estimation of breeding values and genetic parameters, namely genetic variances and covariances, relates to continuous (quantitative) traits. However, animal breeders are also interested in categorical (qualitative) traits such as liability to disease, reproductive performance measured as number of progeny, calving difficulty and 'stayability' (number of lactations).

Categorical variables arise when responses are classified into categories rather than being measured on a continuous scale. A continuous variable can be converted into a categorical variable but not vice-versa. For example, the height of trees can be recorded to the nearest millimeter (a continuous trait) or simply recorded as low, medium and high (a categorical trait). Sometimes the categories have a natural order and can be regarded as different levels of a single underlying variable as in the height example. This is the threshold model. When there is no natural order, the extremal model applies. It assumes each category represents a different underlying variable. Cause of death with categories 'accident', 'disease', 'exposure' and 'slaughter', is an example.

In this thesis, a modified BLUP procedure and a method of estimating variance and covariance components for categorical traits, are derived and discussed.

Chapter 2. Literature Review

> "Farm animals have on the whole been improved by the simple method of mating best to best. We do not understand at all clearly - in some cases hardly at all - the resulting changes in physiological and biochemical makeup. ... The scientist's contribution has been ... to clarify what is 'best', and ... at a statistical level, to indicate how the available information should be combined to make correct decisions. It is assumed that many genes are involved which can be manipulated individually only in rather special cases, and that genetic improvement is a gradual accumulation of small effects." (Robertson, 1982)

This review of literature will introduce the concept of heritability, describe categorical traits and discuss methods proposed for the analysis of categorical traits in an animal breeding context. The methods are all based on the assumption of multifactorial inheritance, i.e. that the trait is controlled by many pairs of genes, each with small additive effects. This assumption is unlikely to be strictly true for any particular trait but in the absence of specific genetic information, appears to be the most useful basis for generating general methods.

The ultimate purpose of statistical analysis in animal breeding is to rank the individuals available for selection. A secondary purpose is to estimate parameters needed for assessing the potential for genetic improvement by selection, and for forming a criterion upon which to rank the individuals available for selection.

The genetic parameter most widely quoted is heritability which, in the broad sense (Lush, 1949), is the proportion of variation in a trait which is genetically determined. As such, it depends on the magnitude of all the genetic variation relative to the total genetic and environmental variation. Heritability, in the narrow sense, is the ratio of additive genetic variance to the total variance.

The total variation must be partitioned into its components before heritability, and other genetic parameters, can be estimated. Searle (1971 a) reviewed methods of variance component estimation to that time. Harville (1977) presented a thorough review of more recent methods. The most widely used methods in animal breeding are those described by Henderson (1953) for unbalanced data. They estimate variance components by equating sums of squares from analyses of variance to their expectations. Other more recent methods, some based on maximum likelihood, are increasing in popularity. More details are given in section 3.1.3.

The principal method for predicting breeding values has been the selection index (Hazel, 1943). This is based on the principle of best linear prediction (Henderson, 1973), in which it is assumed that means and variances are known. This was extended to the case where means are not known by Henderson (1963, 1973, 1975) using the technique of best linear unbiased prediction (BLUP). These techniques are described in section 3.1.

2.1. Categorical data

Categorical data are data obtained by classification rather than measurement. Generally the categories are defined by descriptive names like 'dead' or 'alive', 'easy' or 'hard' and 'single' or 'multiple'. Some categorical characters arise from direct Mendelian genetic segregation. For these, a characteristic occurs only if a particular gene is present and expressed. Depending on the level of dominance, the gene combinations AA, Aa and aa might be distinguishable. This model will not be considered further.

Most categorical traits of interest to animal breeders are believed to be an expression of unobservable underlying continuous traits. Two models for multinomial traits (see, for instance, Bock, 1975, p513) are:

1) Threshold model A characteristic occurs only when the value of an underlying variable exceeds a threshold value. There may be several thresholds which generate a trait with observed states which have an obvious order.

2) Extremal model Each category arises from a different underlying threshold variable. The observed category depends on which variable reaches its threshold first.

The extremal and threshold models are equivalent when there are only two categories, i.e. for all-or-none traits.

Categorical data were commonly regarded as being the result of Mendelian segregation until Wright (1934 a, 1934 b) investigated the inheritance of digit (toe) numbers in guinea pigs. He realized that the number of digits did not follow a simple Mendelian segregation pattern. Rather, the number of digits appeared to be related to an underlying variable which was affected by both genetic and environmental variation. Wright assumed the underlying variable was normally distributed and used an inverse probability transformation to convert frequencies of digit classes to means on the underlying scale. A revised form of this transformation is now generally known as the probit transformation (Bliss, 1935). Wright used tetrachoric correlations (Pearson, 1901, see equation 2.7) to investigate the effect of mating-type, sex and age of dam on the relative frequencies of digit classes. The tetrachoric correlation is the correlation coefficient for an underlying bivariate normal distribution which would produce the specified 2 x 2 frequency table if truncated on both axes. By parent-offspring analysis using tetrachoric correlations, he showed that continued segregation of three- and four-digit types in some inbred lines was due to environmental and not genetic causes.

2.2. Analysis on the binary (0,1) scale

2.2.1. Binomial traits

The method of least squares is a convenient procedure for analysing binomial traits on the (0,1) scale. However, for these analyses to be valid, some assumptions must be satisfied. The principal assumption is that the error variances are constant (or known) and that errors are uncorrelated (or have known covariance). In the case of known variances and covariances, the method of weighted least squares is appropriate. Assumptions about the distribution of the data are required where tests of significance are performed. The method of least squares is known to be quite robust to departures from normality (Ghosh and Sinha, 1980, see also Bloomfield and Watson, 1975, and Knott, 1975). It is for this reason that many animal breeders, and others, use least squares procedures for analyses of binomial variables. By so doing, they often ignore differences in the error variance which they know exist. Another problem with the method of least squares, under a mixed model in particular, is that predicted probabilities are not constrained within the interval [0,1]. This problem is most likely to occur when continous covariates are included in the model. The problems with ordinary least squares have been noted by Gianola (1980 a) and his list is given in section 2.3.2.

Korkman (1948) used an unweighted analysis of variance of a binary (0,1) variable to investigate the genetic effects of twinning in cattle. Variance components for herds, sires-of-cows, cows and within-cows were obtained by equating mean squares to their expectations. No attempt was made to consider an underlying variable or to adjust for the dependence between mean and variance in binomial data.

Lush, Lamoreux and Hazel (1948) reported a study 'to measure the <u>extent</u> or <u>degree</u> to which resistance to disease is inherited in the fowl'. They analysed binary (0,1) data by least squares using a random model. The data from each of four years were analysed separately. Families with less than ten full sisters were omitted. Analyses of variance were obtained for three binary traits: mortality from all causes, mortality from leukosis only and mortality from other causes. Variance components were obtained for sires, dams-within-sires and full-sisters by equating mean squares to their expectations. Heritability was estimated from the correlation between full sisters and from the correlation between half sisters. Lush *et al.* (1948) observed that:

> "The actually observed variance in binomial data is correlated with the mean and hence becomes very small when the average percentage in either class approaches zero. This relationship makes the directly observed estimates of heritability depend in part upon the average incidence of mortality, unlike the situation in normally distributed characteristics. Hence correction for average incidence is required before comparisons can be made fairly between heritabilities of specific causes of mortality. This is also necessary for comparing estimates of heritability from different flocks in which the average incidence in mortality is not the same.
>
> "Transformation to Bliss' 'probit' scale provides a reasonable amount of correction for making such comparisons. ... The basic assumption for this transformation is that resistance to death among individual birds is a continuous and normally distributed variable with death being a threshold which divides the population into only two fractions. The probit value locates the population along that scale of real resistance. Heritability on the actually observed percentage scale can be transformed into the genetically more accurate heritability on the probit scale by multiplying it by $\frac{p(1-p)}{z^2}$ where p is the fraction which dies and z is the height of the ordinate which truncates p of the area of the normal curve."

Most of their discussion relates to heritabilities and genetic correlations estimated on the binary (0,1) scale. However, there was "apparent inconsistency" in heritabilities calculated on the binary scale. The heritability of total mortality (.22) was higher than for its two components, leukosis (.14) and other (.03). The problem "is resolved by the probit transformation. This results from removing the bias due to different average percentages of mortality in the three categories." The corresponding heritabilities on the probit scale were .33, .37 and .05 respectively. These results highlight the importance of using an underlying probit scale when comparing parameter estimates from (sub)populations with different mean incidence. Heritability on the probit scale is "genetically more accurate" when multifactorial inheritance applies and the central limit theorem might reasonably be invoked.

Robertson and Lerner (1949) analysed poultry mortality data on the (0,1) scale. They ignored half-sib families with less than 12 birds and full-sib families with less than 5 birds. They assumed each bird had a genetically-based probability of surviving to become a parent in the next generation with mean $\bar{p}$ and variance σ_p^2. They calculated heritability as

$$\text{heritability} = \frac{\text{genetic improvement}}{\text{phenotypic selection differential}}$$

$$h^2 = \frac{\sigma_p^2}{\bar{p}} \frac{1}{1-\bar{p}}$$

$$= \frac{\chi^2 - (N - 1)}{r\ k} \qquad (2.1)$$

where χ^2 is the heterogeneity χ^2 of the 2 x N data table,
σ_p^2 is estimated from an analysis of variance,
r is the difference in genetic relationship between and within classes and
k is the coefficient of $r\sigma_p^2$ in the expression for the expected value of the sum of squares.

Hill and Smith (1977) commented that this method was not preferable to the components of variance method.

Robertson and Lerner (1949) discussed the concept of "heritability on the probit scale" introduced by Lush <u>et al.</u> (1948). They presented an interpretation of the probit transformation which applies to mortality from a "single definite cause". They considered the probit transformation inappropriate for total mortality, an unspecified mixture of variously heritable traits, because "the dosage scale ceases to have any definite meaning". They showed that the heritability of a trait composed of ten equi-heritable traits is less than the heritability of just one of those traits at the same incidence.

Dempster and Lerner (1950) investigated "the conditions ... under which serious errors are likely to result from use of the convenient p scale" instead of the underlying scale, the X scale. They state

> "The probit transformation may be satisfactory for ... comparing heritability values but it is not apparent how it can be used for devising optimum selection indexes nor is it at all convenient in many situations for computing expected rates of gain."

They consider an underlying variable (X) whose value is the sum of two independent normal variables, g and e, with means zero and variances h^2 and $1-h^2$ respectively. Thus X is distributed N(0,1). The character is expressed in those individuals, and only those, in which X exceeds a threshold x'. Gene substitutions are presumed to have individually small and strictly additive effects on the underlying variate. The proportion, p, of individuals with the character is determined from the tail of the normal curve.

> "Genetic differences are ... thought of as shifting the mean of the environmental distribution, the threshold remaining fixed. Alternatively, they can equally well be thought of as shifting the threshold abscissa, the mean of the environmental distribution remaining fixed. Using either concept, the genotype of an individual on the p scale is the proportion of environments in which it would exhibit the character."

The relationship between heritability on the underlying scale and on the p scale, used by Lush *et al.* (1948), was proved by Robertson (in an appendix to Dempster and Lerner, 1950). The relationship, repeated below, assumes a normal distribution for the underlying variable.

$$h^2_{pa} = \frac{z^2 h^2}{p(1-p)} \qquad (2.2)$$

where z is the height of the ordinate at x',
h^2 is heritability on the X scale and
h^2_{pa} is heritability (additive) on the p scale.

The accuracy of this relationship between heritability on the p and X scales was studied empirically by Van Vleck (1972) and Olausson and Rönningen (1975) in simulation studies. Van Vleck (1972) concluded

> "Heritability on the normal scale generally will be only slightly over-estimated by multiplying the estimate from the paternal half-sib correlation from binomial data by $p(1-p)/z^2$ but may be over-estimated substantially if the estimate from parent-offspring correlation is similarly adjusted particularly when p is not between .25 and .75. The over-estimate will be greater when normal heritability is large than when it is small.
>
> "Genetic gain from mass selection as measured by the fraction of the population with the character will be under-estimated when only a small fraction of the population has the character and when heritability on the normal scale is high ...."

James and McGuirk (1982) recently considered the estimation of the heritability of all-or-none characters by offspring/parent regression. They allowed each generation to have a different incidence and obtained the relationship

$$h^2_{pa} = \frac{z\ z_o\ h^2}{p\ (1-p)} \qquad (2.3)$$

where p is the incidence in the parent generation (not the actual parents),

z is the ordinate corresponding to p and

z_o is the ordinate corresponding to the incidence in the offspring.

This function is dependent on both parent and offspring incidences. Plotting showed that the relationship is more sensitive to changes in offspring incidence than to changes in parent incidence. This is because it is proportional to z_o while changes in z are largely compensated by changes in p(1-p).

2.2.2. Multinomial traits

When a categorical trait is definitely ordered (as it often is in animal breeding), it has been common to analyse category codes as a continuous variable. In sheep research, fertility traits such as oestrous activity, ovulation rate, number of lambs born and lambs weaned have traditionally been analysed as continuous traits (e.g. Hohenboken and Clarke, 1981, Lamberson and Thomas, 1982, Dzakuma, Whiteman and NcNew 1982). In dairy cattle, calving difficulty (5 classes) has been analysed as if continuous (Pollak and Freeman, 1976, Thompson, Freeman and Berger, 1981).

An alternative approach is to replace the category codes with scores. Tong, Wilton and Schaeffer (1977) presented a method of computing scores based on work by Snell (1964) and using the logistic distribution. This was used by Schaeffer and Wilton (1977) and Hudson, Schaeffer and Wilton (1980) to give ease of calving scores of 100, 50, 30, 0 for the evaluation of beef sires. Scores could also be based on monetary value of the categories. Gianola (1979 a) has shown the relationship between heritability in the outward (scores) scale and in an underlying normal X scale to be

$$h_o^2 = h^2 \frac{\left[\sum_i^{m-1} z_i (a_{i+1} - a_i) \right]^2}{\mathrm{var}(\underset{\sim}{a}'\underset{\sim}{p})} \qquad (2.4)$$

where $\underset{\sim}{a} = \{a_i\}$ is a vector of m scores,

$\underset{\sim}{p} = \{p_i\}$ is a vector of m probabilities,

z_i is the ordinate of the normal distribution at the i th threshold,

h_o^2 is the heritability in the outward scale and

h^2 is the heritability in the underlying normal scale.

In general, the more categories in the trait, the closer the heritabilities will be. In the case of only two categories, equation 2.4 reduces to equation 2.2. The scores which maximize heritability on the outward scale are proportional to the means of truncated sections of the normal distribution

$$\text{i.e. } a_i = \frac{z_{i-1} - z_i}{p_i} \qquad (2.5)$$

where z_0 and z_m are both zero.

Landis and Koch (1977 c), in a study of the repeatability of psychological diagnoses, expressed a categorical trait of L categories as L binary variables. The categories did not have a natural order although one could possibly be imposed. They used multivariate analysis procedures and a one-way random model. They showed that the usual variance components procedure could be used to obtain an intraclass correlation coefficient for each variable. Quaas and Van Vleck (1980) extended this approach to animal breeding with the estimation of breeding values utilizing multiple trait mixed model equations (Henderson 1973). They considered the ordered trait stayability - the number of seasons a cow remains in a herd. Famula (1980) suggested this trait is better fitted by an exponential distribution than a multinomial but his method does not allow for genetic relationships. Van Vleck and Karner (1980) used multiple trait mixed model equations on

L-1 binary variables to evaluate 712 Brown Swiss bulls for twelve descriptive type traits. While this method makes allowance for the correlations among categories, it does not allow for variance differences resulting from different thresholds on the underlying scale i.e. variance differences arising from different mean incidences associated with fixed effects in the model.

2.2.3. Correlation between variables

Several authors have studied the relationship between genetic and phenotypic correlations estimated between two discrete variables, or one discrete and one continuous variable, with the correlation for two underlying continuous variables.

Tate (1955) considered the correlation of a normal variable and an assumed normal variable underlying a dichotomous trait. Cox (1974) extended this to a polychotomous trait. Olausson and Ronningen (1975) considered the genetic correlation between the underlying variable and another normal variable. Tallis (1962) considered the correlation on the underlying normal scale of two threshold characters. Vinson, White and Kliewer (1976) derived the general expression for correlation on an underlying scale of two polychotomous threshold characters.

Mao (1976) estimated genetic and phenotypic correlations between milk traits before and after recoding into discrete categories. He concluded

> "Phenotypic correlations on the continuous scale generally were under-estimated in absolute value when one or both traits were discrete. Under-estimation became greater as the number of discrete classes decreased and became substantial (50 %) when data on both traits were binomial-like. Genetic correlations on the continuous scale ... were closely estimated", using paternal half-sib analyses, when one or both traits were discrete. However, some slight over-estimation occurred and it

> increased when fewer categories were used.
> "Over-estimation was greater when the underlying continuous genetic correlation was negative or small positive than when the correlation was high positive. When the underlying correlation was high positive, neither over- nor under-estimation occurred. In all cases, the fewer the number of categories, the greater the sampling variation in correlation estimates."

Similar results were reported by Olausson and Rönningen (1975) who concluded

> "The genetic correlation between two traits, the one all-or-none and the other normally distributed, can be estimated by the half-sib correlation method in the ordinary way without transformation of the estimate except for low heritability (0.1) and for low incidence levels (5 and 10 %) of the all-or-none trait. ... The estimation of the phenotypic correlation ... by the same method as described for the genetic correlation, gives considerable bias in estimates."

They gave a correction factor which is a quadratic function of incidence to convert the estimate using binomial data to an estimate for the underlying scale.

Rutledge (1977) reported a Monte Carlo comparison of two estimators of repeatability of a binomial trait. He sampled from a bivariate normal distribution with equal variances and correlation parameter ranging 0.05, 0.15, ... 0.85. The sample correlation was computed before the sample was reduced to a 2 x 2 contingency table by truncating both axes.
Let the 2 X 2 contingency table be represented by

		Record 1		
		0	1	Total
Record 2	0	n_{00}	n_{10}	$n_{.0}$
	1	n_{01}	n_{11}	$n_{.1}$
Total		$n_{0.}$	$n_{1.}$	$n_{..}$

The estimators were:

$$\text{Lush} \qquad t_L = n_{11}/\, n_{1.} - n_{01}/\, n_{0.} \qquad (2.6)$$

$$\text{and Pearson } t_P = \sin \left[\frac{\pi}{2} \; \frac{\sqrt{n_{00}n_{11}} - \sqrt{n_{01}n_{10}}}{\sqrt{n_{00}n_{11}} + \sqrt{n_{01}n_{10}}} \right] \qquad (2.7)$$

which was rediscovered by Plackett (1965) in the form

$$t_P = - \cos \left[\psi \quad \pi/(1+\psi) \right] \qquad (2.8)$$

$$\text{where} \qquad \psi = \left[n_{00}n_{11}/n_{01}n_{10} \right]^{1/2}$$

The deviation from the true value of the correlation was regressed against the true value and the threshold values. A complex fourth power function was required to fit the deviations satisfactorily but the threshold was the principal source of bias.

2.2.4. Selection involving categorical traits

Lush et al. (1948) devised strategies for selection against mortality. Heritability on the probit scale was used as a basis for assessing the relative efficiency of individual and family selection at various levels of incidence. They concluded that family selection will generally be "many times more effective than mass selection" unless the incidence of desired animals is close to the proportion to be selected. However, the selection of families is not straightforward when they differ in size. They suggested correcting the average family deviation from the flock mean by $nrh^2/(1+t(n-1))$ where n is family size, t is the phenotypic correlation between family members, r is the genetic relationship between family members and h^2 is the heritability on the probit scale.

While genetic variance on the X scale is all additive (by definition above), the genetic variance on the p scale is not. Thus predictions based on heritability on the p scale will be in error. Dempster and Lerner (1950) showed this error was greater for highly heritable traits than for lowly heritable traits and when the incidence is near 0 or 1. The efficiency of mass selection is low unless the selection differential is close to the incidence. Efficiency can be increased by the use of family selection or a correlated trait.

The gains expected using mass selection have been discussed more recently by Rönningen (1976). Danell and Rönningen (1981) investigated the behaviour of a selection index based on a binomial and a normal trait at 180 combinations of genetic parameters, economic weights and incidences of the all-or-none trait. Genetic and phenotypic correlations between the normal trait and the normal trait underlying the binomial trait were always equal in the study. The efficiency of selection was measured as the ratio of the genetic merit of the group selected on the index, to the merit of the group which would have been selected using an index based on the normal trait and the underlying trait. They concluded

> "The varying efficiency of the index showed several patterns .... In situations with correlations between traits ranging from strongly favourable to moderately unfavourable it seemed possible to predict the efficiency pattern with the help of theoretically obtained correlations between aggregate genotype and index, plus what is known from selection on single binomial traits. For strongly unfavourable correlations the index was very unstable and the efficiency difficult to predict. ... Appropriate transformations of heritabilities, genetic and phenotypic correlations and of economic weights from the binomial scale to the underlying normal scale are needed for the predictions."

2.3. Analysis on the underlying scale

2.3.1. Falconer's liability method

The papers discussed in the preceding section all involved analyses of binary traits on the (0,1) scale. Three of them considered the use of an underlying continuous X scale and used a transformation to convert the heritability estimate on the p scale to one on the X scale. Ways of estimating heritability directly on the underlying scale were not considered.

Falconer (1965,1967) sought to estimate the heritability of disease in human populations. He called the underlying X scale 'liability', considering that susceptibility was too general in concept. He used the selection response method of Robertson and Lerner(1949) but applied to the X scale. The procedure requires estimates of the incidences in the general population and among relatives of affected individuals. Projecting these incidences on to two normal curves assumed to have the same variance and same threshold (hence different origins) enables the mean of the affected individuals, and the mean of their relatives, to be estimated on the X scale. The regression on the underlying scale of relatives on affected individuals is then estimated as

$$b = \frac{\text{response}}{\text{selection differential}}$$

$$= \frac{(\text{mean of relatives - general mean})}{(\text{mean of affected individuals - general mean})}$$

$$= \frac{r - 0}{a - 0} \qquad (2.9)$$

where 0 is mean of population,

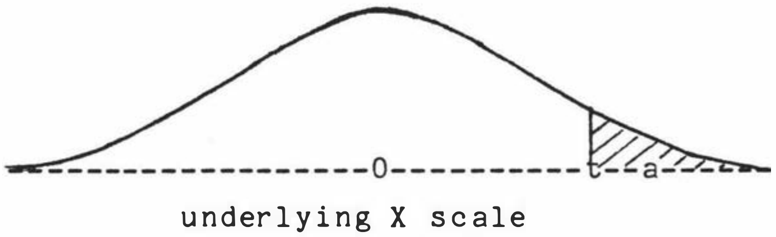
a is the mean of the affected individuals,

t is the threshold and

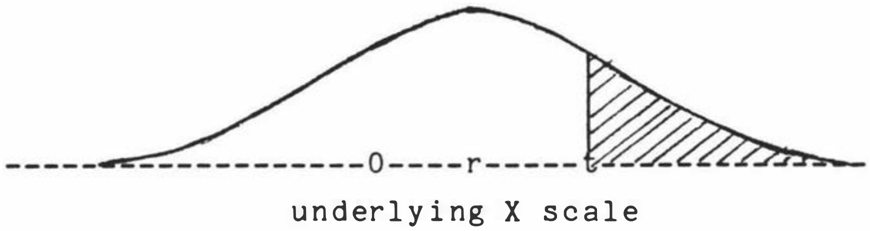
r is the mean of the relatives of affected individuals.

Heritability is then estimated as b/(degree of relationship).

Distribution for the total population



Distribution for the relatives of affected individuals



A difficulty with the model is that since the affected individuals form the upper tail of a normal distribution, the distribution of their relatives will be skewed and of smaller variance - the extent being greater for higher heritability and closer relationship. Falconer (1965) believed this error was only serious for monozygous twins. Edwards (1969) and Smith (1970), however, showed that this caused Falconer's method to under-estimate heritability by up to ten percent. Both presented graphs showing the correlation between relatives given the incidences in the population and in relatives of affected individuals. Smith (1970) also gives an approximate algebraic expression for his graph, namely;

$$\tan((0.25)\pi(1-r)(1+r^5)) = \log(p_R) - \log(p_P) \qquad (2.10)$$

where p_R is incidence in relatives of affected individuals,
p_P is incidence in the total population,
r is the required correlation and
r^5 can be ignored if r is less than .5 (i.e. except for a highly heritable trait in monozygous twins).

He also gives a correction formula for use when the relatives are from a different population to the affected individuals.

Reich, James and Morris (1972) adapted Falconer's (1965) method by allowing for the variance reduction in the distribution of relatives of affected individuals. They applied the method to multiple threshold data and attempted to distinguish single gene from multifactorial inheritance.

Both Edwards (1969) and Smith (1970) discussed the problem of bias introduced because of the common familial environment of affected individuals and their relatives. Smith suggested this bias can be assessed by comparison of heritabilities derived from relatives of different degree since more distant relatives will have different environments.

The advantage of assessing heritability on the liability scale is demonstrated in the above papers. Sub-populations, having different incidences, had quite different heritabilities on the p scale but very similar values on the X scale. This can, in fact, be a check that the heritability model being used is appropriate. This was done by Smith, Falconer and Duncan (1972) for a study of incidence of diabetes in humans. They used Falconer's original method for computational convenience, rather than Smith's revised form.

2.3.2. Generalized linear models and other proposals

Gianola (1980 a) summarized the "problems associated with genetic evaluation of categorical traits by linear models" as:

> "(1) Scores are arbitrarily assigned to response categories. (2) Mixed model solutions do not incorporate the restriction in the estimation space that the sum of response probabilities must total 1 across categories. (3) The variance in the observed scale is not constant and depends on the genotypic value of the candidates for selection. (4) The additive genetic variance in the observed scale depends on the mean incidence of the character in the sub-populations considered in the model.

> (5) Non-additive genetic variation is present on the observed scale. (6) Linear relationships fail outside a restricted range of the data. (7) Ranking optimality of best linear predictors is lacking when the conditional expectation of the predictand given the data is not linear."

He proposed a method of sire evaluation of dichotomies on the logit scale using empirical logits and weights. Gianola (1980 b) extended this to polychotomous traits treated as multiple traits.

Thompson (1979) suggested that generalized linear models (GLM, Nelder and Wedderburn, 1972) should be of use in animal breeding settings as they have been used successfully for dichotomous traits in other applications. He suggested a method of adapting the mixed model equations to the GLM form. This procedure offers a solution to the problem referred to by Dempster and Lerner (1950) of how to use the underlying scale for optimum selection index calculations. Gianola and Foulley (1982) have presented a derivation, using an empirical Bayes approach, of the method suggested by Thompson (1979).

Outside the area of animal breeding, Griffiths (1973) and Williams (1975, 1982) have used the beta-binomial for data with extra-binomial variation. This assumes that the data arises from binomial sampling where the binomial parameter is a random variable from the beta distribution. Kupper and Haseman (1977) discussed the beta-binomial and presented the correlated-binomial model, an alternative way of analysing data with extra-binomial variation which assumes the binomial sampling is not independent.

2.4. Scope of the thesis

In this thesis, the suggestion of Thompson (1979) mentioned above, is explored in detail. A general method for analysing data presumed to arise from a two-stage sampling process, when the second sampling has an error distribution belonging to the single parameter exponential family, is derived. This is a quasi-likelihood method (Wedderburn, 1974).

The general method is used to obtain equations for analysing binomial, threshold and extremal characters using the logit transformation. A procedure for obtaining the intraclass correlation is presented. The procedure is used for sire evaluation and estimation of heritability in several data sets and compared with other techniques in one data set.

Chapter 3. Review of statistical procedures required in later chapters

3.1. Some statistical procedures used in animal breeding

3.1.1. The selection index

The selection index (Smith, 1936 and Hazel, 1943) is an optimal linear method for ranking animals available for selection when information is available on the means and variances of the traits to be improved and the traits that have been recorded. It may be expressed as

$$I_i = \underset{\sim}{v}'\underset{\sim}{G}\ \underset{\sim}{P}^{-1}\underset{\sim}{Y}_i \qquad (3.1)$$

where $\underset{\sim}{v}$ is a vector of the relative economic weights of the traits to be improved (defining the selection objective),

$\underset{\sim}{G}$ is a matrix of genetic variances and covariances involving the traits in the selection objective and the traits for which data are available,

$\underset{\sim}{P}$ is a matrix of phenotypic variances and covariances among the traits for which data are available and

$\underset{\sim}{Y}_i$ is the vector of data available on the i th animal corrected for fixed effects.

The problems with the selection index approach are:

1) The economic weights that are relevant depend on the prices of future products which are unknown (Williams, 1962). However, the relative prices of future products may vary less than the actual prices so that this problem may not be serious.

2) The appropriate fixed effect correction factors are usually unknown. Sometimes estimates from independent data are available but this is unlikely, especially for seasonal effects which change continually.

3) Genetic and phenotypic variances and covariances are usually unknown and must be estimated. This problem is generally not as critical as the preceeding problem because the variances and covariances are less sensitive to seasonal

conditions and independent estimates are more likely available. Some techniques for this will be mentioned in the section 3.1.3.

3.1.2. Best linear unbiased prediction

The problem of obtaining appropriate fixed effect correction factors was overcome with the development of best linear unbiased prediction (BLUP) by Henderson (1963, 1973, 1975).

Define the linear mixed model

$$\underset{\sim}{y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{Z}\,\underset{\sim}{u} + \underset{\sim}{e} \qquad (3.2)$$

where $\underset{\sim}{\beta}$ is a vector of unknown fixed effects,

$\underset{\sim}{u}$ is a vector of random effects with zero mean and variance $\underset{\sim}{R}\sigma_u^2$,

$\underset{\sim}{R}$ is a matrix of additive genetic relationships and

$\underset{\sim}{e}$ is a vector of random effects with zero mean, variance $\underset{\sim}{Q}\sigma_e^2$ and independent of $\underset{\sim}{u}$.

$\underset{\sim}{y}$ is a vector with mean $\underset{\sim}{X}\underset{\sim}{\beta}$ and variance $(\underset{\sim}{Q} + \lambda\underset{\sim}{Z}\underset{\sim}{R}\underset{\sim}{Z}')\sigma_e^2$

and λ is σ_u^2/σ_e^2

When $\underset{\sim}{u}$, $\underset{\sim}{e}$ and hence $\underset{\sim}{y}$ are assumed to be multivariate normal, the maximum likelihood solution for $\underset{\sim}{\beta}$, say $\underset{\sim}{\beta}^o$, is given by

$$\underset{\sim}{X}'(\underset{\sim}{Q} + \lambda\,\underset{\sim}{Z}\,\underset{\sim}{R}\,\underset{\sim}{Z}')^{-1}\underset{\sim}{X}\underset{\sim}{\beta}^o = \underset{\sim}{X}'(\underset{\sim}{Q} + \lambda\,\underset{\sim}{Z}\,\underset{\sim}{R}\,\underset{\sim}{Z}')^{-1}\underset{\sim}{y} \qquad (3.3)$$

Henderson _et al._ (1959) derived the same solution for $\underset{\sim}{\beta}$ by maximizing the joint density function of $\underset{\sim}{y}$ and $\underset{\sim}{u}$. This is

$$f(\underset{\sim}{y},\underset{\sim}{u}) = g(\underset{\sim}{y}|\underset{\sim}{u})\,h(\underset{\sim}{u}) \qquad (3.4)$$

$$= C\exp[-(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta}-\underset{\sim}{Z}\underset{\sim}{u})'\underset{\sim}{Q}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta}-\underset{\sim}{Z}\underset{\sim}{u})/2\sigma_e^2 - \underset{\sim}{u}'\underset{\sim}{R}^{-1}\underset{\sim}{u}/2\sigma_u^2]$$

Differentiating with respect to $\underset{\sim}{\beta}$ and $\underset{\sim}{u}$ and equating to zero gives

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z} + \underset{\sim}{R}^{-1}/\lambda \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\beta}} \\ \hat{\underset{\sim}{u}} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} \end{bmatrix} \qquad (3.5)$$

These are known as Henderson's mixed model equations (MME).

That they give the same solution for $\underset{\sim}{\beta}$ is proved by eliminating $\underset{\sim}{\hat{u}}$ from 3.5 to get 3.3 using

$$\underset{\sim}{Q}^{-1} - \underset{\sim}{Q}^{-1}\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z} + \underset{\sim}{R}^{-1}/\lambda)^{-1}\underset{\sim}{Z}'\underset{\sim}{Q}^{-1} = (\underset{\sim}{Q} + \lambda\underset{\sim}{Z}\underset{\sim}{R}\underset{\sim}{Z}')^{-1} \qquad (3.6)$$

(see Henderson and Searle, 1981).

However, the interpretation of the solution for the random vector $\underset{\sim}{\hat{u}}$ in equations 3.5 was the major discovery leading to BLUP. Henderson (1963) derived the mixed model equations by imposing linearity, unbiasedness and minimum prediction error variance, but not normality, on the estimator of random effects. Thus $\underset{\sim}{\hat{u}}$ is the best linear unbiased predictor of the random effects $\underset{\sim}{u}$. He showed that this estimator of random effects is identical to the selection index estimator of breeding values when the generalized least squares estimates of fixed effects are used in place of the true values in the selection index equations. Henderson (1973) discussed the range of animal breeding problems which can be usefully solved using BLUP and Henderson (1975) showed how it can make allowance for bias arising from previous selection or culling.

3.1.3. Estimation of variance components

The matrices $\underset{\sim}{Q}$ and $\underset{\sim}{R}$ in equations 3.5 are usually known. $\underset{\sim}{Q}$ is often an identity matrix and $\underset{\sim}{R}$ a matrix of additive genetic relationships. But, the variances σ_e^2 and σ_u^2 are less likely to be known and must be estimated. Searle (1971) reviewed variance component estimation methods for balanced and unbalanced data available at that time. Anderson (1978) and Harville (1977) reviewed more recent methods.

Henderson (1953) described three methods of variance component estimation for unbalanced data which have been widely used in animal breeding. The methods are all based on equating sums of squares to their expectations. Method one applies to a fully random model. Sums of squares are formed by the direct methods applicable for balanced data. Their expectations are obtained and linear functions of them are used to obtain the estimates of variance components. In unbalanced data, 'sums of squares' relating to interactions derived when using this method are not necessarily positive and the resulting variance component estimates may be negative.

Method two is similar to Method one except that it allows for fixed effects in the model provided there are no interactions between fixed and random effects. The data is first adjusted for fixed effects and then Method one applied to the residuals. The adjustment for fixed effects should be done using the least squares effects obtained by fitting the full model including all fixed and random effects.

Method three overcomes the problem of negative 'sums of squares' while allowing for a mixed model. However, negative estimates of variance components may still arise. It equates a true sum of squares with its expectation. The sum of squares is obtained as 'variation explained by a model including both fixed and random effects' less 'variation explained by a sub-model obtained by dropping the random factor from the full model'.

i.e. For the model in equation 3.2

$$SS(\text{random}|\text{fixed}) = \underset{\sim}{y}'\underset{\sim}{W}(\underset{\sim}{W}'\underset{\sim}{W})^{-}\,\underset{\sim}{W}'\underset{\sim}{y} - \underset{\sim}{y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\,\underset{\sim}{X}'\underset{\sim}{y}$$

where $\underset{\sim}{W} = [\underset{\sim}{X}\ \underset{\sim}{Z}]$

$$EMS(\text{random}|\text{fixed}) = SS(\text{random}|\text{fixed}) / df$$

$$= \sigma_e^2 + k\sigma_r^2$$

where $k = \text{trace}(\ \underset{\sim}{Z}'\underset{\sim}{Z} - \underset{\sim}{Z}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\ \underset{\sim}{X}'\underset{\sim}{Z}\ ) / df$

df is the degrees of freedom associated with SS(random|fixed) and

$\underset{\sim}{Z}$ in the evaluation of the trace must be free of any constraints which might be applied when obtaining the sum of squares.

One problem with Method three is that for a model with several random effects, different orderings of the random effects will usually lead to different estimates of the variance components.

Another method of particular interest to this thesis is the restricted maximum likelihood (REML) procedure developed by Patterson and Thompson (1971). The method allows for several random factors in the model but for the purposes of this thesis will be described in relation to the simple model of equation 3.2. Let $\underset{\sim}{T}$ be that portion of the inverse of the left hand side of equation 3.5 corresponding to the random effects $\underset{\sim}{\hat{u}}$.

$$\begin{aligned}\underset{\sim}{T} &= (\underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z} + \underset{\sim}{R}^{-1}/\lambda - \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z})^{-1} \\ &= (\underset{\sim}{R}^{-1}/\lambda + \underset{\sim}{Z}'\underset{\sim}{S}\underset{\sim}{Z})^{-1} \qquad (3.7)\end{aligned}$$

where $\underset{\sim}{S} = \underset{\sim}{Q}^{-1} - \underset{\sim}{Q}^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Q}^{-1}$

and $(\underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{X})^{-}$ is a generalized inverse.

The REML estimators (c.f. Harville, 1977, eqns 6.3 & 6.4) are

$$\hat{\sigma}_e^2 = \underset{\sim}{y}'(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\hat{\beta}} - \underset{\sim}{Z}\underset{\sim}{\hat{u}}) \;/\; (N - p - q) \qquad (3.8)$$

and $$\hat{\sigma}_u^2 = (\;\underset{\sim}{\hat{u}}'\underset{\sim}{\hat{u}} + \mathrm{trace}(\underset{\sim}{T}\;))/q \qquad (3.9)$$

where N is the number of observations,
p is the rank of $\underset{\sim}{X}$ and
q is the length of $\underset{\sim}{\hat{u}}$.

Since $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$ are involved in both sides of 3.8 and 3.9, iterative solution is required.

3.2. Other statistical procedures

3.2.1. Logistic distribution

The probit transformation (Bliss, 1935) is often used to map probabilities [0,1] onto the real line. An alternative, the logit transformation has recently been promoted by Bock (1975, p515) because it is "closely similar" to the probit and easier to compute.

The logit of a probability, p_i, is

$$t_i = \ln \left(\frac{p_i}{1 - p_i} \right) \tag{3.10}$$

where t_i is the distance of a threshold from the mean of a logistic variable, x, with variance $\pi^2/3$ and

p_i is the probability that random values of x are below the threshold.

It is important to realize that $-t_i$ is the distance from the threshold to the mean of the variable. When comparing variables with different proportions, it is convenient to regard the threshold as the fixed reference point and $-t_i$ as the relative location of the i th mean. Therefore, comparison of t_i values is equivalent to comparison of means.

The standard logistic distribution has mean zero, variance $\pi^2/3$ and density function

$$\begin{aligned} f(x) &= \frac{1}{4} \operatorname{sech}^2\left(\frac{x}{2} \right) \\ &= \frac{1}{4} \left(\frac{2}{\exp(-\frac{x}{2})+\exp(\frac{x}{2})} \right)^2 \\ &= \frac{\exp(x)}{(1 + \exp(x))^2} \end{aligned} \tag{3.11}$$

The corresponding distribution function is

$$P(\ x<t\ ) = \frac{\exp(t)}{(1 + \exp(t))} \tag{3.12}$$

The logit transformation of 3.10 comes from inverting 3.12. Table 3.1 shows the relationship between 3.12 and the normal distribution on the same variance scale. The particular point of interest is that the logistic distribution is heavier tailed than the normal distribution.

Table 3.1 The relationship between the standard normal and standard logistic distributions when used as probability transformations. The threshold values are expressed in standard deviations relative to a mean of zero for both distributions.

threshold (t)	normal P(x<t)	logistic P(x<t)
0	.5	.5
1	.841	.860
2	.9772	.9741
3	.9986	.9957
4	.99997	.99929

The logit transformation has a simple extension for multinomial data (Bock, 1975, p521).

$$p_i = \frac{\exp(t_i)}{1 + \sum_j \exp(t_j)} \qquad \text{for } i = 1, \ldots c \qquad (3.13)$$

and c+1 categories,

which rearranges to

$$t_i = \ln \left(\frac{p_i}{1 + \sum_j p_j} \right) \qquad (3.14)$$

The marginal distributions are logistic with mean zero and variance $\pi^2/3$. The covariance (x_i, x_j) is $\pi^2/6$ and the corresponding correlation is 1/2. This provides a very close approximation to the multivariate normal with equal variances and equal correlations of 1/2.

3.2.2. Log likelihood expectations

The likelihood function for a random sample is the joint density of the random variables in the sample which is considered to be a function of the parameters of the joint density (see, for instance, Mood, Graybill and Boes, 3rd edition, 1974). The method of maximum likelihood is to find those parameter estimates which make this likelihood a maximum. It is usually more convenient to maximize the natural logarithm of the likelihood. The log likelihood is differentiated with respect to the parameters to be estimated and the resulting expressions equated to zero and solved.

Wedderburn (1974) has defined a quasi-likelihood for the case when the individual sample densities are not known but a relationship between the mean and variance is known. He showed that the quasi-likelihood has properties similar to log likelihoods. The likelihood method used in this thesis is, in effect, a quasi-likelihood method although the formulation is different. A two-stage sampling procedure is assumed in which the second stage is conditional on the first. Random variables, $\underset{\sim}{u}$, from the initial sampling are combined with parameters, $\underset{\sim}{\theta}$, to define the parameters of the distribution for the second sampling from which variables, $\underset{\sim}{z}$, are observed. The likelihood can be written

$$L = \prod_{k}^{K} \pi(z_k : \theta_k, \underset{\sim}{u})\ f(\underset{\sim}{u} : \phi) \qquad (3.15)$$

While this expression will be called a likelihood, it does not strictly satisfy the usual definition of a likelihood since the random variables, $\underset{\sim}{u}$, are not observed and the samples are not independent, being conditional on $\underset{\sim}{u}$. It is believed, however, to have most of the properties of a likelihood although this is not proved. Some expectation properties are shown below. These require that L satisfies the usual regularity conditions (see Mood <u>et al.</u>, 1974). When $f(\underset{\sim}{u}:\phi)$ is null, this reduces to the usual likelihood. Following Kendall and Stuart (1979, pp8,9)

$$\int \dots \int L\ dz_1 \dots dz_n = 1 \qquad (3.16)$$

Differentiating 3.16 inside the integral with respect to θ_i

$$\int \dots \int \frac{\partial L}{\partial \theta_i}\ dz_1 \dots dz_n = 0$$

$$\int \dots \int \frac{1}{L} \frac{\partial L}{\partial \theta_i}\ L\ dz_1 \dots dz_n = 0$$

$$\int \dots \int \frac{\partial \ln L}{\partial \theta_i}\ L\ dz_1 \dots dz_n = 0 \qquad (3.17)$$

$$\text{i.e. } E\left[\frac{\partial \ln L}{\partial \theta_i}\right] = 0 \qquad (3.18)$$

Differentiating 3.17 with respect to θ_j,

$$\int \ldots \int \left(\frac{\partial \ln L}{\partial \theta_i} \frac{\partial L}{\partial \theta_j} + \frac{\partial^2 \ln L}{\partial \theta_i \; \partial \theta_j} L \right) dz_1 \ldots dz_n = 0$$

$$\int \ldots \int \left(\frac{\partial \ln L}{\partial \theta_i} \frac{\partial \ln L}{\partial \theta_j} + \frac{\partial^2 \ln L}{\partial \theta_i \; \partial \theta_j} \right) L \, dz_1 \ldots dz_n = 0$$

$$\text{i.e. } E\left[\frac{\partial \ln L}{\partial \theta_i} \frac{\partial \ln L}{\partial \theta_j}\right] = - E\left[\frac{\partial^2 \ln L}{\partial \theta_i \; \partial \theta_j}\right] \qquad (3.19)$$

Equations 3.18 and 3.19 are the key results from this section.

3.2.3. Generalized linear models (GLM)

3.2.3.1. Derivation of the general expressions for members of the single parameter exponential family

Nelder and Wedderburn (1972) described a general technique for maximum likelihood estimation of parameters for distributions from the exponential family. The technique is presented for distributions from the single parameter exponential family such as binomial and poisson.

The density function for an observation, z, can be written as

$$\pi(z:\theta) = \exp(z\theta - g(\theta) + h(z)) \qquad (3.20)$$

where θ is the parameter of the distribution.

For K independent observations,

$$L = \prod_k^K \pi(z_k:\theta_k) = \prod_k^K \exp(z_k\theta_k - g(\theta_k) + h(z_k)) \qquad (3.21)$$

$$\ln L = \sum_k^K (z_k\theta_k - g(\theta_k) + h(z_k)) \qquad (3.22)$$

Differentiating 3.22 with respect to the parameters, θ_i,

$$\frac{\partial \ln L}{\partial \theta_i} = z_i - g'(\theta_i) \qquad (3.23)$$

$$\frac{\partial^2 \ln L}{\partial \theta_i^2} = - g''(\theta_i) \qquad (3.24)$$

$$\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} = 0 \qquad \text{for } i \neq j \qquad (3.25)$$

Using 3.18 on 3.23 gives an expression for the mean of z.

$$E [z_i - g'(\theta_i)] = 0$$

i.e. $$\mu_i \doteq E [z_i] = g'(\theta_i) \qquad (3.26)$$

Note also $$\frac{\delta \mu_i}{\delta \theta_i} = g''(\theta_i) \qquad (3.27)$$

Using 3.19 with 3.24, 3.25, 3.26 and 3.27,

$$\text{LHS of 3.19} = E \left[\frac{\partial \ln L}{\partial \theta_i} \; \frac{\partial \ln L}{\partial \theta_j} \right]$$

$$= E [(z_i - \mu_i) (z_j - \mu_j)]$$

$$= \operatorname{cov}(z_i, z_j)$$

$$\text{RHS of 3.19} = - E \left[\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right]$$

Therefore $$\operatorname{var}(z_i) = g''(\theta_i) = \frac{\delta \mu_i}{\delta \theta_i} \qquad (3.28)$$

and $$\operatorname{cov}(z_i, z_j) = 0 \qquad \text{for } i \neq j \qquad (3.29)$$

Equations 3.26, 3.28 and 3.29 are expressions for the mean, variance and covariance of the observed values, z_i, given as functions of the parameters to be estimated. This relationship must be inverted so that the parameters are a function of the observed values. The problem is generalized so that the parameter for an observation belongs to a family of parameters which are related through a linear model. Thus, each observation has concomitant information on m independent variables in $\underset{\sim}{X}$.

A linear predictor, $\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta}$ (3.30)
is related to the parameter $\underset{\sim}{\theta}$ by
a 'link function', $\theta_i = f(Y_i)$. (3.31)
The link function must be specified along with the distribution and is often the identity $\theta_i = Y_i$. If a solution can be obtained for $\underset{\sim}{\beta}$, say $\underset{\sim}{\beta}^o$, $\hat{\underset{\sim}{Y}}$ can be calculated as $\underset{\sim}{X}\underset{\sim}{\beta}^o$ and $\underset{\sim}{\theta}$ can be calculated using the link function. The problem is thus reformulated to finding the maximum likelikood solution for $\underset{\sim}{\beta}$. Equations to estimate $\underset{\sim}{\beta}^o$ are obtained by differentiating 3.22 with respect to $\underset{\sim}{\beta}$.
Since the observations are independent

$$\frac{\partial \ln L}{\partial \beta_i} = \sum_{k}^{K} \frac{\partial \ln L}{\partial \theta_k} \frac{d\,\theta_k}{d\,Y_k} \frac{\partial Y_k}{\partial \beta_i}$$

$$= \sum_{k}^{K} (z_k - \mu_k) \frac{d\,\theta_k}{d\,Y_k} x_{ik} \qquad (3.32)$$

and $$\frac{\partial^2 \ln L}{\partial \beta_i \, \partial \beta_j} = \sum_{k}^{K} \frac{\partial^2 \ln L}{\partial Y_k^2} \frac{\partial Y_k{}^2}{\partial \beta_i \, \partial \beta_j}$$

$$= \sum_{k}^{K} \frac{\partial^2 \ln L}{\partial Y_k^2} x_{ik} \, x_{jk}$$

where $$\frac{\partial^2 \ln L}{\partial Y_k^2} = \frac{\partial}{\partial Y_k} \left(\frac{\partial \ln L}{\partial \theta_k} \frac{d\,\theta_k}{d\,Y_k} \right)$$

$$= \frac{\partial \ln L}{\partial \theta_k} \frac{d^2\,\theta_k}{d\,Y_k^2} + \frac{d\,\theta_k}{d\,Y_k} \frac{\partial^2 \ln L}{\partial \theta_k^2} \frac{d\,\theta_k}{d\,Y_k}$$

$$= (z_k - \mu_k) \frac{d^2\,\theta_k}{d\,Y_k^2} - \left(\frac{d\,\theta_k}{d\,Y_k} \right)^2 \mathrm{var}(z_k)$$

$$= (z_k - \mu_k) \frac{d^2\,\theta_k}{d\,Y_k^2} - \left(\frac{d\,\theta_k}{d\,\mu_k} \frac{d\,\mu_k}{d\,Y_k} \right)^2 \mathrm{var}(z_k)$$

$$= (z_k - \mu_k) \frac{d^2\,\theta_k}{d\,Y_k^2} - \left(\frac{d\,\mu_k}{d\,Y_k} \right)^2 / \mathrm{var}(z_k)$$

after using 3.30.

$$E\left[\frac{\partial^2 \ln L}{\partial \beta_i \, \partial \beta_j} \right] = \sum_k^K E\left[(z_k - \mu_k) \frac{d^2 \theta_k}{d Y_k^2} - \left(\frac{d \mu_k}{d Y_k} \right)^2 / \mathrm{var}(z_k) \right] x_{ik} x_{jk}$$

$$= - \sum_k^K \left(\frac{d \mu_k}{d Y_k} \right)^2 x_{ik} \, x_{jk} \, / \, \mathrm{var}(z_k) \qquad (3.33)$$

after using 3.26 and 3.27.

The maximum likelihood solution for $\underset{\sim}{\beta}$, say $\underset{\sim}{\beta}^o$, satisfies the LHS of 3.32 set equal to zero. Since 3.32 may not be linear in $\underset{\sim}{\beta}^o$, it may not be easy to solve directly. Using an initial estimate of $\underset{\sim}{\beta}$, say $\underset{\sim}{t}$, a Taylor series expansion can be used to provide equations that can be solved. The problem of obtaining an initial estimate is discussed in section 3.2.5. It is convenient to write the expansion of 3.32 as matrices using the notation { } which defines the elements in the matrices. The first two terms of the Taylor series expansion about a previous estimate $\underset{\sim}{t}$ give

$$\left\{ \frac{\partial \ln L}{\partial \beta_i} \right\} \simeq \left\{ \frac{\partial \ln L}{\partial \beta_i} \right\}_{\underset{\sim}{t}} + (\underset{\sim}{\beta}^o - \underset{\sim}{t}) \left\{ \frac{\partial^2 \ln L}{\partial \beta_i \, \partial \beta_j} \right\} \qquad (3.34)$$

The Fisher-scoring method involves taking the expected value of 3.34 and solving the resulting equations for $\underset{\sim}{\beta}^o$, i.e. Solving

$$- E \left[\left\{ \frac{\partial^2 \ln L}{\partial \beta_i \, \partial \beta_j} \right\} \right]_{\underset{\sim}{t}} (\underset{\sim}{\beta}^o - \underset{\sim}{t}) = \left\{ \frac{\partial \ln L}{\partial \beta_i} \right\}_{\underset{\sim}{t}} \qquad (3.35)$$

Substituting from 3.32 and 3.33 gives

$$\left\{ \sum_k^K \left(\frac{\partial \mu_k}{\partial Y_k} \right)^2 x_{ik} x_{jk} / \mathrm{var}(z_i) \right\} (\underset{\sim}{\beta}^o - \underset{\sim}{t}) = \left\{ \sum_k^K (z_k - \mu_k) \frac{d \theta_k}{d Y_k} x_{ik} \right\} \qquad (3.36)$$

This is simplified by defining a weight matrix

$$\underset{\sim}{W} = \text{diag} \{ (\frac{d\ \mu_k}{d\ Y_k})^2 / \text{var}(z_k) \} \qquad (3.37)$$

where diag { } is a diagonal matrix with elements { }, and a vector of scaled residuals,

$$\underset{\sim}{y}^* = \{ (z_k - \mu_k) / (\frac{d\ \mu_k}{d\ Y_k}) \} \qquad (3.38)$$

Equations 3.36 may then be written as

$$\underset{\sim}{X}' \underset{\sim}{W} \underset{\sim}{X} (\underset{\sim}{\beta}^o - \underset{\sim}{t}) = \underset{\sim}{X}' \underset{\sim}{W} \underset{\sim}{y}^* \qquad (3.39)$$

From 3.30, the current estimate of $\underset{\sim}{Y}$ is $\hat{\underset{\sim}{Y}} = \underset{\sim}{X}\underset{\sim}{t}$.

When premultiplied by $\underset{\sim}{X}'\underset{\sim}{W}$, this gives

$$\underset{\sim}{X}' \underset{\sim}{W} \underset{\sim}{X} \underset{\sim}{t} = \underset{\sim}{X}' \underset{\sim}{W} \hat{\underset{\sim}{Y}} \qquad (3.40)$$

Adding 3.40 to 3.39 gives the general least squares equations

$$\underset{\sim}{X}' \underset{\sim}{W} \underset{\sim}{X} \underset{\sim}{\beta}^o = \underset{\sim}{X}' \underset{\sim}{W} \underset{\sim}{y} \qquad (3.41)$$

$$\text{where} \quad \underset{\sim}{y} = \hat{\underset{\sim}{Y}} + \underset{\sim}{y}^*$$

$$= \hat{\underset{\sim}{Y}} + \{ (z_k - \mu_k) / (\frac{d\ \mu_k}{d\ Y_k}) \} \qquad (3.42)$$

An intuitive argument for deriving the weighting matrix is helpful. Assume 3.42 defines a variable to be analysed where $\hat{\underset{\sim}{Y}}$ is fixed,

$\underset{\sim}{\mu} = E[\ \underset{\sim}{z}\ ]$ is a known function of $\hat{\underset{\sim}{Y}}$ and

$\frac{d\ \mu_k}{d\ Y_k}$ is a constant.

Since $\hat{Y}_k$ and μ_k have zero variance, $\text{var}(y_k)$ is $\text{var}(z_k)/(d\ \mu_k/d\ \hat{Y}_k)^2$. The weighted least squares analysis of $\underset{\sim}{y}$ requires a weight matrix, $\underset{\sim}{W}$, which is the inverse of this variance matrix. The argument extends to the more general case where μ_k is a function of several elements of $\hat{\underset{\sim}{Y}}$.

$\underset{\sim}{W} = \underset{\sim}{D}\ \underset{\sim}{V}^{-1}\ \underset{\sim}{D}$ where $\underset{\sim}{D}$ is $\{\partial\ \mu_k/\partial\ \hat{Y}_m\}$

and $\underset{\sim}{V}$ is $\text{var}(\underset{\sim}{z})$

3.2.3.2. Testing the 'goodness of fit' of a proposed model

There are two common measures of 'goodness of fit' used with categorical data. Both may be tested using the chi-square distribution. The first is the statistic

$$S = \Sigma \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$= (\underset{\sim}{z} - \underset{\sim}{\mu})' \, (\text{diag}(\underset{\sim}{\mu}))^{-1} \, (\underset{\sim}{z} - \underset{\sim}{\mu}) \qquad (3.43)$$

Another statistic, the deviance, is specifically for likelihood models.

$$D_i = -2 \, (\ln L_i - \ln L_c) \qquad (3.44)$$

where $\ln L_i$ is the value of the log likelihood under model i,

$\ln L_c$ is the value of the log likelihood under the complete model, i.e. when $\underset{\sim}{\mu} = \underset{\sim}{z}$.

The deviance can be used in a manner analagous to the use of sums of squares in an analysis of variance to create an analysis of deviance table. If an extra factor is added to model i to form model j, the significance of including that factor may be tested by comparing $D_j - D_i$ with a tabulated chi-square statistic. The statistics in 3.43 and 3.44 are both affected by class size and will be biased if sample size is 1 for every sample. If sample size is larger, they provide a test of the adequacy of the model as well as enabling the testing of individual factors in the model.

3.2.4. Maximum likelihood solution for parameters of the multinomial distribution by generalized linear models

The multinomial distibution does not belong to the single parameter exponential family except in its simplest form, the binomial. However, Cooney (*pers. comm.*) has shown how the GLM procedure described by Nelder and Wedderburn (1972) can be applied to the multinomial distribution.

The likelihood function for a multinomial sample of n observations in c+1 categories is

$$L = \text{constant} \prod_{i}^{c} p_i^{z_i} \, (1 - \sum_j p_j)^{(n - \sum_j z_j)} \qquad (3.45)$$

where p_i is the probability of the i th category and

z_i is the frequency of the i th category.

Taking logarithms,

$$\ln L = \text{constant} + (n - \sum_{i}^{c} z_i) \ln(1 - \sum_{i}^{c} p_i) + \sum_{i}^{c} z_i \ln(p_i)$$

$$= \text{constant} + n \ln(1 - \sum_{i}^{c} p_i) + \sum_{i}^{c} z_i \ln\left(\frac{p_i}{1 - \sum_j p_j} \right) \qquad (3.46)$$

This expression needs to be reparameterized into the form of the exponential family.

Using the multinomial logit from 3.13,

$$\theta_i = \ln\left(\frac{p_i}{1 - \sum_j p_j} \right) \qquad i = 1, \ldots c \qquad (3.47)$$

$$\text{implying } p_i = \frac{\exp(\theta_i)}{1 + \sum_j \exp(\theta_j)} \qquad (3.48)$$

$$\text{and} \qquad \ln(1 - \sum_j p_j) = - \ln(1 + \sum_j \exp(\theta_j)),$$

equation 3.46 becomes

$$\ln L = \text{constant} + \sum_i z_i \theta_i - n \ln(1 + \sum_j \exp(\theta_j)) \qquad (3.49)$$

Differentiating 3.49 and using 3.18 and 3.19 leads to expressions for the mean and variance of z.

$$\frac{\partial \ln L}{\partial \theta_i} = z_i - n \frac{\exp(\theta_i)}{1 + \sum_j \exp(\theta_j)} = z_i - np_i \qquad (3.50)$$

$$\frac{\partial^2 \ln L}{\partial \theta_i \, \partial \theta_j} = n \frac{\exp(\theta_i) \exp(\theta_j)}{(1 + \sum_k \exp(\theta_k))^2} = np_i p_j \quad \text{for } i \neq j \qquad (3.51)$$

$$\frac{\partial^2 \ln L}{\partial \theta_i^2} = -n \frac{\exp(\theta_i)}{1 + \sum_j \exp(\theta_j)} + n \frac{(\exp(\theta_i))^2}{(1 + \sum_j \exp(\theta_j))^2}$$

$$= -np_i(1-p_i) \qquad (3.52)$$

Using 3.18 on 3.50 implies $E[\, z_i \,] = np_i$ (3.53)

$$\text{LHS of 3.19} = E\left[\, \frac{\partial \ln L}{\partial \theta_i} \frac{\partial \ln L}{\partial \theta_j} \,\right]$$

$$= E[\, (z_i - np_i)\,(z_j - np_j) \,]$$

$$= \mathrm{cov}(z_i, z_j) \text{ using 3.53.} \qquad (3.54)$$

Using 3.54 on 3.51 and 3.52 gives

$$\mathrm{var}(z_i) = np_i(1 - p_i)$$

and $\mathrm{cov}(z_i, z_j) = -np_ip_j$

In matrix notation

$$\underset{\sim}{\mu} = E[\, \underset{\sim}{z} \,] = n\,\underset{\sim}{p} = \{\, np_i \,\} = \left\{\, \frac{n \exp(\theta_i)}{1 + \sum_j \exp(\theta_j)} \,\right\} \qquad (3.55)$$

$$\underset{\sim}{V} = \mathrm{var}(\underset{\sim}{z}) = n\,(\,\mathrm{diag}(\underset{\sim}{p}) - \underset{\sim}{p}\underset{\sim}{p}'\,) \qquad (3.56)$$

$$\left\{\, \frac{\partial \mu_i}{\partial \theta_j} \,\right\} = \underset{\sim}{V} \text{ (by differentiating 3.55)} \qquad (3.57)$$

The GLM method requires a link between $\underset{\sim}{\theta}$ and a linear model. Using the identity link,

$$\underset{\sim}{\theta} = \{\, \theta_i \,\} = \{\, Y_i \,\} = (\underset{\sim}{x}' \otimes \underset{\sim}{I})\,\underset{\sim}{\beta} \qquad (3.58)$$

where Y_i is the linear predictor for the i th category,
$\underset{\sim}{x}'$ is a vector of q concomitant variables,
$\underset{\sim}{\beta}$ is a qc vector of regression coefficients and
$\otimes$ represents the right direct product matrix operation.

Maximum likelihood solution for $\underset{\sim}{\beta}$ requires differentiation.

$$\frac{\partial \ln L}{\partial \beta_{ij}} = \frac{\partial \ln L}{\partial \theta_i} \frac{d\,\theta_i}{d\,Y_i} \frac{\partial Y_i}{\partial \beta_{ij}}$$

$$= (z_i - \mu_i)\; 1 \; x_j \text{ using 3.50 and 3.55}$$

In matrix notation,

$$\left\{ \frac{\partial \ln L}{\partial \beta_{ij}} \right\} = \underset{\sim}{x}' \otimes (\underset{\sim}{z} - \underset{\sim}{\mu}) \qquad (3.59)$$

The Taylor series expansion of the LHS of 3.59 about a previous solution for $\underset{\sim}{\beta}$, say $\underset{\sim}{t}$, is

$$\left\{ \frac{\partial \ln L}{\partial \beta_{ij}} \right\} \simeq \left\{ \frac{\partial \ln L}{\partial \beta_{ij}} \right\}_{\underset{\sim}{t}} + \left\{ \frac{\partial}{\partial \beta_{mn}} \left\{ \frac{\partial \ln L}{\partial \beta_{ij}} \right\} \right\} (\underset{\sim}{\beta}^{o} - \underset{\sim}{t})$$

Taking expected values and rearranging gives

$$E \left[\frac{\partial}{\partial \beta_{mn}} \left\{ \frac{\partial \ln L}{\partial \beta_{ij}} \right\} \right]_{\underset{\sim}{t}} (\underset{\sim}{\beta}^{o} - \underset{\sim}{t}) \quad \left\{ \frac{\partial \ln L}{\partial \beta_{ij}} \right\}_{\underset{\sim}{t}} \qquad (3.60)$$

$$\frac{\partial}{\partial \beta_{mn}} \frac{\partial \ln L}{\partial \beta_{ij}} = \frac{\partial^2 \ln L}{\partial \theta_i \, \partial \theta_m} \frac{d\,\theta_i}{d\,Y_i} \frac{d\,\theta_m}{d\,Y_m} \frac{\partial\,Y_i}{\partial \beta_{ij}} \frac{\partial\,Y_m}{\partial \beta_{mn}}$$

$$= \frac{\partial^2 \ln L}{\partial \theta_i \, \partial \theta_j} x_j \, x_n$$

Using 3.56, $E \left[\left\{ \frac{\partial}{\partial \beta_{mn}} \frac{\partial \ln L}{\partial \beta_{ij}} \right\} \right] = \underset{\sim}{x}\underset{\sim}{x}' \otimes \underset{\sim}{V}$ (3.61)

Substituting 3.59 and 3.61 in 3.60 and summing over K independent samples gives

$$\left(\sum_k \underset{\sim}{x}_k \underset{\sim}{x}'_k \otimes \underset{\sim}{V}_k \right) \quad (\underset{\sim}{\beta}^{o} - \underset{\sim}{t}) = \sum_k \underset{\sim}{x}'_k \underset{\sim}{x}_k (\underset{\sim}{z}_k - \underset{\sim}{\mu}_k)$$

which rearranged is

$$(\underset{\sim}{X} \otimes \underset{\sim}{I})' \underset{\sim}{V} (\underset{\sim}{X} \otimes \underset{\sim}{I}) (\underset{\sim}{\beta}^{o} - \underset{\sim}{t}) = (\underset{\sim}{X} \otimes \underset{\sim}{I})' \underset{\sim}{V} \underset{\sim}{y}^{*} \qquad (3.62)$$

where $\underset{\sim}{V} = \Sigma_k^{+} \underset{\sim}{V}_k$

and $\underset{\sim}{y}^{*} = \underset{\sim}{V}^{-1} (\underset{\sim}{z} - \underset{\sim}{\mu})$

Let $\underset{\sim}{y} = \hat{\underset{\sim}{Y}} + \underset{\sim}{y}^{*}$

where $\hat{\underset{\sim}{Y}} = (\underset{\sim}{X} \otimes \underset{\sim}{I}) \underset{\sim}{t}$

Adding $(\underset{\sim}{X} \otimes \underset{\sim}{I})' \underset{\sim}{V} (\underset{\sim}{X} \otimes \underset{\sim}{I}) \underset{\sim}{t} = (\underset{\sim}{X} \otimes \underset{\sim}{I})' \underset{\sim}{V} \hat{\underset{\sim}{Y}}$ to 3.62

gives $(\underset{\sim}{X} \otimes \underset{\sim}{I})' \underset{\sim}{V} (\underset{\sim}{X} \otimes \underset{\sim}{I}) \underset{\sim}{\beta}^{o} = (\underset{\sim}{X} \otimes \underset{\sim}{I})' \underset{\sim}{V} \underset{\sim}{y}$ (3.63)

These equations can be solved iteratively to produce maximum likelihood estimates of the effects in the linear model.

3.2.5. Solution of reweighted least squares equations

The weighted least squares equations to be solved may be expressed as

$$\underset{\sim}{X}' \underset{\sim}{W} \underset{\sim}{X} \underset{\sim}{b} = \underset{\sim}{X}' \underset{\sim}{W} \underset{\sim}{y} \qquad (3.64)$$

where $\underset{\sim}{X}$ is a known matrix,

$\underset{\sim}{W}$ is a weighting matrix, a known function of $\underset{\sim}{b}$, the parameters to be estimated and

$\underset{\sim}{y}$ is a working variable, a known function of $\underset{\sim}{b}$ and the obsevations $\underset{\sim}{z}$.

The problem of starting procedures has been discussed by several authors (e.g. Nelder and Wedderburn, 1972). For binomial and multinomial traits, the usual procedure is to estimate p_{ij} and y_{ij} as

$$p_{ij} = \frac{z_{ij} + 0.5}{n_{ij} + c} \qquad (3.65)$$

$$y_{ij} = \ln\left(\frac{p_{ij}}{1 - \sum_k p_{ik}} \right) \qquad (3.66)$$

where c is the number of categories less 1.

The weight matrix $\underset{\sim}{W}$ can be approximated using 3.65 and then equations 3.64 solved. Once the initial value of $\underset{\sim}{b}$ is obtained, it is used in the linear predictor to compute $\underset{\sim}{\hat{Y}} = \underset{\sim}{X}'\underset{\sim}{b}$. The weighting matrix, $\underset{\sim}{W}$, and the working variable, $\underset{\sim}{y}$, are reformed according to the appropriate equations. The process is repeated until the solution for $\underset{\sim}{b}$ is stable between cycles.

Chapter 4. Generalized linear mixed model

The generalized linear models (GLM) procedure (Nelder and Wedderburn, 1972) described in section 3.2.3, obtains the maximum likelihood solution by iterative re-weighted least squares. That procedure is for the analysis of data sampled from a distribution belonging to the exponential family when the parameter of the distribution is a function of a linear predictor. This chapter outlines two methods of extending GLM to mixed models. It serves to introduce the concepts which are applied to multinomial and binomial data in chapter 5 and as a reference for attempting to apply the methods to other distributions.

The data are assumed to have arisen by random sampling in a two-stage sampling process. The distribution for the second stage is assumed to belong to the exponential family. The parameter of the distribution is assumed to be a function of a linear predictor which includes the random variable obtained from the initial sampling (see section 3.2.2.).

Two alternative systems of equations are derived. The different approaches may be seen by reference to section 3.1.2 where it was shown that

$$\underset{\sim}{X}'(\underset{\sim}{Q} + \lambda\, \underset{\sim}{Z}\,\underset{\sim}{R}\,\underset{\sim}{Z}')^{-1}\underset{\sim}{X}\underset{\sim}{\beta}^{o} = \underset{\sim}{X}'(\underset{\sim}{Q} + \lambda\, \underset{\sim}{Z}\,\underset{\sim}{R}\,\underset{\sim}{Z}')^{-1}\underset{\sim}{y} \qquad (4.1)$$

gives the same solution for $\underset{\sim}{\beta}^{o}$ as

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z} + \underset{\sim}{R}^{-1}/\lambda \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta}^{o} \\ \underset{\sim}{u}^{o} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} \end{bmatrix} \qquad (4.2)$$

The first approach for non-normal distributions is to generalize the method used by Henderson _et al._ (1959) to derive equations 4.2. This will be referred to as the joint-maximization method since it involves maximization with respect to both the fixed and random effects. The resulting equations were suggested by Thompson (1979) for use in animal breeding and were obtained by Gianola and Foulley (1982) using a Bayes argument.

The second approach, the maximization-expectation method, is to attempt to obtain equations with a structured weight matrix as in equation 4.1 by taking expectations over the distribution of $\underset{\sim}{u}$. Equations like 4.2 may then provide a convenient computing strategy when solving for $\underset{\sim}{\beta}^{o}$.

4.1. The joint-maximization method

For this method, the distribution of both sampling stages must be known. The initial sampling is presumed to be from a normal $(0,\sigma_u^2)$ distribution, and the second sampling from a distribution belonging to the one-parameter exponential family.

The vector of random variables drawn from $N(0,\sigma_u^2)$ is combined with several fixed effects in a linear function to give a mixed model

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{Z}\underset{\sim}{u} \qquad (4.3)$$

where $\underset{\sim}{Y}$ is a linear predictor of length $N. = \sum_i N_i$,

$\underset{\sim}{X}$ is a matrix of concomitant variables of rank p,

$\underset{\sim}{\beta}$ is a vector of fixed effects,

$\underset{\sim}{Z}$ is a matrix mapping q random effects to $\underset{\sim}{Y}$,

$\underset{\sim}{u}$ is the vector of q random effects and

N_i is the number of samples associated with u_i.

The vector, $\underset{\sim}{\theta}$, of parameters for the second distribution, is related to the vector $\underset{\sim}{Y}$ by a 'link function'

$$\theta_{ij} = f(Y_{ij})$$

The data, $\underset{\sim}{z}$, are independent samples from a distribution whose density function can be written as

$$f(z_{ij}:\theta_{ij}) = \exp[z_{ij}\theta_{ij} - g(\theta_{ij}) + h(z_{ij})] \qquad (4.4)$$

The objective is to estimate the fixed effects, $\underset{\sim}{\beta}$, and the random effects, $\underset{\sim}{u}$. This is feasible because when the data are collected at the second sampling, the random effects, $\underset{\sim}{u}$, from the initial sampling, are in effect fixed.

Since the elements u_i may not have been independently sampled, their distribution will be expressed as a multivariate normal $(\underset{\sim}{0},\underset{\sim}{G})$. In the animal breeding situation, $\underset{\sim}{G}$ might be $\underset{\sim}{R}\sigma_u^2$ where $\underset{\sim}{R}$ is a matrix of additive genetic relationships. The density function of the total sample can be expressed as

$$L = [\prod_{i}^{q} \{ \prod_{j}^{N_i} \exp[z_{ij}\theta_{ij} - g(\theta_{ij}) + h(z_{ij})]\}]\ C \exp[-\underset{\sim}{u}'\underset{\sim}{G}^{-1}\underset{\sim}{u}/2]$$

and taking logarithms,

$$\ln L = \sum_{i}^{q} \{ \sum_{j}^{N_i} [z_{ij}\theta_{ij} - g(\theta_{ij}) + h(z_{ij})]\} + \ln C - \underset{\sim}{u}'\underset{\sim}{G}^{-1}\underset{\sim}{u}/2 \qquad (4.5)$$

C is a constant not involving $\underset{\sim}{u}$ or $\underset{\sim}{\theta}$.

In this method, both $\underset{\sim}{\theta}$ and $\underset{\sim}{u}$ are regarded as parameters to be estimated by maximum likelihood. The standard GLM procedure (section 3.2.3) is used directly on 4.5.

Define a mean vector, $\underset{\sim}{\mu} = \{ g'(\theta_{ij}) \}$ (4.6)

a variance matrix, $\underset{\sim}{V} = \text{diag}\{ g''(\theta_{ij}) \}$ (4.7)

a change of scale matrix, $\underset{\sim}{D} = \text{diag}\ \{ \frac{d\theta_{ij}}{\partial Y_{Ij}} \}$ (4.8)

The following results are obtained by differentiating 4.5.

$$\{ \frac{\partial \ln L}{\partial \theta_{ij}} \} = \{ z_{ij} - g'(\theta_{ij}) \} = \underset{\sim}{z} - \underset{\sim}{\mu} \qquad (4.9)$$

$$\{ \frac{\partial^2 \ln L}{\partial\theta_{ij}\partial\theta_{lm}} \} = \text{diag}\ \{ - g''(\theta_{ij}) \} = - \underset{\sim}{V} \qquad (4.10)$$

$$\{ \frac{\partial \ln L}{\partial \beta_k} \} = \underset{\sim}{X}'\underset{\sim}{D}\ (\underset{\sim}{z} - \underset{\sim}{\mu}) \qquad (4.11)$$

$$\{ \frac{\partial \ln L}{\partial u_i} \} = \underset{\sim}{Z}'\underset{\sim}{D}\ (\underset{\sim}{z} - \underset{\sim}{\mu}) - \underset{\sim}{G}^{-1}\underset{\sim}{u} \qquad (4.12)$$

$$\{ \frac{\partial^2 \ln L}{\partial \beta_k\ \partial\beta_l} \} = - \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{V}\underset{\sim}{D}\underset{\sim}{X} \qquad (4.13)$$

$$\{ \frac{\partial^2 \ln L}{\partial \beta_k\ \partial u_i} \} = - \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{V}\underset{\sim}{D}\underset{\sim}{Z} \qquad (4.14)$$

$$\{ \frac{\partial^2 \ln L}{\partial u_i\ \partial u_j} \} = - \underset{\sim}{Z}'\underset{\sim}{D}\underset{\sim}{V}\underset{\sim}{D}\underset{\sim}{Z} - \underset{\sim}{G}^{-1} \qquad (4.15)$$

Taking the Taylor series expansion of the first differential about prior estimates β^o and u^o results in the expression

$$\begin{bmatrix} X'D\,(z-\mu) \\ Z'D\,(z-\mu) - G^{-1}u \end{bmatrix} \approx \begin{bmatrix} X'D\,(z-\mu) \\ Z'D\,(z-\mu) - G^{-1}u \end{bmatrix}\Bigg|_{\substack{\beta^o \\ u^o}} + \begin{bmatrix} -X'DVDX & -X'DVDZ \\ -Z'DVDX & -Z'DVDZ - G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} - \beta^o \\ \hat{u} - u^o \end{bmatrix}$$

which, when set to zero, can be rearranged to give

$$\begin{bmatrix} X'DVDX & X'DVDZ \\ Z'DVDX & Z'DVDZ + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} - \beta^o \\ \hat{u} - u^o \end{bmatrix} = \begin{bmatrix} X'D\,(z-\mu) \\ Z'D\,(z-\mu) - G^{-1}u \end{bmatrix} \qquad (4.16)$$

Letting $y = \hat{Y} + (VD)^{-1}\,(z - \mu)$ (4.17)

where $\hat{Y} = X\beta^o + Zu^o$

the RHS of 4.16 becomes

$$\begin{bmatrix} X'DVDy \\ Z'DVDy \end{bmatrix} - \begin{bmatrix} X'DVDX & X'DVDZ \\ Z'DVDX & Z'DVDZ + G^{-1} \end{bmatrix} \begin{bmatrix} \beta^o \\ u^o \end{bmatrix}$$

Then rearranging 4.16 gives

$$\begin{bmatrix} X'DVDX & X'DVDZ \\ Z'DVDX & Z'DVDZ + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'DVDy \\ Z'DVDy \end{bmatrix} \qquad (4.18)$$

It is readily shown that if the distribution of z given u is $N(X\beta+Zu, Q)$, μ is Y, y is z, V is Q and D is Q^{-1} so that 4.18 becomes

$$\begin{bmatrix} X'Q^{-1}X & X'Q^{-1}Z \\ Z'Q^{-1}X & Z'Q^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'Q^{-1}y \\ Z'Q^{-1}y \end{bmatrix} \qquad (4.19)$$

which are Henderson's mixed model equations (equation 4.2).

4.2. The maximization-expectation method

The joint-maximization method has two potential problems in practice. First, the number of random effects is often large in animal breeding applications and repeated solution of a large set of equations may be impractical. Secondly, the amount of information on a particular random effect is often small and may lead to instability of the iteration.

The maximization-expectation method avoids these by estimating only the fixed effects by iteration. It uses weights based on the expected values of functions of the random effects rather than using functions of the current solution for the random effects. The estimation of the random effects, of special interest to animal breeders, by back solution, is discussed later in this section.

The specification of the problem is similar to that of the previous section except that the distribution used in the initial sampling need not be multivariate normal. The data, $\underset{\sim}{z}$, are independent samples from a distribution whose density function can be written as

$$f(z_{ij}|\dot{\theta}_{ij}) = \exp[z_{ij}\dot{\theta}_{ij} - g(\dot{\theta}_{ij}) + h(z_{ij})] \qquad (4.20)$$

The joint density function of the sample becomes

$$L = [\prod_{i}^{q} \{ \prod_{j}^{N_i} \exp[z_{ij}\dot{\theta}_{ij} - g(\dot{\theta}_{ij}) + h(z_{ij})] \}]\, U(\underset{\sim}{u}:\underset{\sim}{G}) \qquad (4.21)$$

where $\dot{\theta}_{ij} = f(Y_{ij} + u_i)$

$\underset{\sim}{Y}$ (= $\underset{\sim}{X}\underset{\sim}{\beta}$) is a linear predictor of length $N. = \sum_i N_i$,

$\underset{\sim}{X}$ is a matrix of concomitant variables of rank p,

$\underset{\sim}{\beta}$ is a vector of fixed effects,

$\underset{\sim}{Z}$ is a matrix mapping q random effects to $\underset{\sim}{Y}$,

$\underset{\sim}{u}$ is the vector of q random effects with mean zero, variance $\underset{\sim}{G}$ (= $\underset{\sim}{R}\sigma_u^2$) and density $U(\underset{\sim}{u}:\underset{\sim}{G})$,

$\underset{\sim}{R}$ is a matrix of additive genetic relationships and

N_i is the number of samples associated with u_i.

The log likelihood is

$$\ln L = [\prod_{i}^{q} \{ \prod_{j}^{N_i} [z_{ij}\dot{\theta}_{ij} - g(\dot{\theta}_{ij}) + h(z_{ij})] \}] + \ln (U(\underset{\sim}{u}:\underset{\sim}{G}))$$

An informal derivation of the method is attempted.

From 3.26 $E[\underset{\sim}{z} | \underset{\sim}{u}] = \{ g'(\dot{\theta}_{ij}) \}$

Therefore $\underset{\sim}{\mu} = E[\underset{\sim}{z}] = E[\{ g'(\dot{\theta}_{ij}) \}]$ (4.22)

From 3.28 $\mathrm{var}(\underset{\sim}{z} | \underset{\sim}{u}) = \mathrm{diag}\{ g''(\dot{\theta}_{ij}) \}$

Therefore $E[z_{ij}^2] = E[\quad \{g''(\dot{\theta}_{ij})\} + \{g'(\dot{\theta}_{ij})\ g'(\dot{\theta}_{ij})\}]$

and $\underset{\sim}{V} = \mathrm{var}(\underset{\sim}{z}) = E[\mathrm{diag}\{g''(\dot{\theta}_{ij})\} + \mathrm{var}(\{g'(\dot{\theta}_{ij}) \})$ (4.23)

The expectations in equations 4.22 and 4.23 are taken over the distribution of $\underset{\sim}{u}$.

By analogy with 3.42, the variable to be analysed at each iteration is $\underset{\sim}{y} = \underset{\sim}{\hat{Y}} + \underset{\sim}{D}^{-1}(\underset{\sim}{z}-\underset{\sim}{\hat{\mu}})$ (4.24)

where $\underset{\sim}{D} = \{ \dfrac{\partial\ \mu_{ij}}{\partial\ Y_{lm}} \}$

and $\underset{\sim}{\hat{Y}}$, $\underset{\sim}{D}$, $\underset{\sim}{V}$ and $\underset{\sim}{\hat{\mu}}$ are all estimated using the most recent solution for $\underset{\sim}{\beta}^{o}$.

By analogy with 3.41, the equations to be solved are

$$\underset{\sim}{X}'\ \underset{\sim}{D}\ \underset{\sim}{V}^{-1}\ \underset{\sim}{D}\ \underset{\sim}{X}\ \underset{\sim}{\hat{\beta}} = \underset{\sim}{X}'\ \underset{\sim}{D}\ \underset{\sim}{V}^{-1}\ \underset{\sim}{D}\ \underset{\sim}{y} \qquad (4.25)$$

Obviously $\underset{\sim}{V}$ has a complex structure and it is of interest to know under what conditions the structure is such that equations like 4.2 may be useful when solving 4.25 for $\underset{\sim}{\hat{\beta}}$.

In this regard, interest focuses on the second term in 4.23. By Taylor series approximation,

$$
\begin{aligned}
g'(\dot{\theta}_{ij}) &= g'(\ f(Y_{ij} + u_i)\ ) \\
&= g'(\ f(Y_{ij})\ ) + u_i\ g''(\ f(Y_{ij})\ )\ f'(Y_{ij})
\end{aligned}
$$

and $E[\ g'(\dot{\theta}_{ij})\ g'(\dot{\theta}_{lm}) - \mu_{ij}\mu_{lm}\ )\ ]$

$$
\begin{aligned}
&= (g'(\ f(Y_{ij})\ ) - \mu_{ij})(\ g'(\ f(Y_{lm})\ ) - \mu_{lm}) \\
&\quad + g''(\ f(Y_{ij})\ )\ f'(Y_{ij})\ r_{il}\sigma_u^2\ g''(\ f(Y_{lm})\ )\ f'(Y_{lm})
\end{aligned}
$$

where r_{il} is an element in $\underset{\sim}{R}$.

When $\mu_{ij} = g'(\ f(Y_{ij})\ )$,

$\underset{\sim}{A} = \text{diag}\{\ g''(\ f(Y_{ij})\ )\ f'(Y_{ij})\ \}$ and

$\underset{\sim}{E} = E[\ g''(\dot{\theta}_{ij})\ ]$,

then $\underset{\sim}{V} = \underset{\sim}{E} + \underset{\sim}{A}\ \underset{\sim}{Z}\ \underset{\sim}{G}\ \underset{\sim}{Z}'\ \underset{\sim}{A}$ (4.26)

and 4.25 has the same solution for $\hat{\underset{\sim}{\beta}}$ as

$$
\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{Z} + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\beta}} \\ \hat{\underset{\sim}{u}} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{y} \end{bmatrix} \qquad (4.27)
$$

Another approach is to argue in the reverse direction. The structure in 4.26 can be directly imputed to $\underset{\sim}{V}$ by arguing

that $\text{cov}(g'(\dot{\theta}_{ij}), g'(\dot{\theta}_{lm})) = \dfrac{d\ \mu_{ij}}{d\ Y_{ij}}\ \text{cov}(u_i, u_l)\ \dfrac{d\ \mu_{lm}}{d\ Y_{lm}}$ (4.28)

This leads to an expression for $E[\ g''(\dot{\theta}_{ij})\ ]$, at least for binomial characters. It follows that

$\underset{\sim}{V} = \underset{\sim}{E} + \underset{\sim}{D}\underset{\sim}{Z}\underset{\sim}{G}\underset{\sim}{Z}'\underset{\sim}{D}$ (4.29)

so that 4.27 becomes

$$
\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{Z} + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\beta}} \\ \hat{\underset{\sim}{u}} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{y} \end{bmatrix} \qquad (4.30)
$$

An advantage of this latter approach is that it immediately extends to multiple random factors in the mixed model. Since $\underset{\sim}{u}$ is not required during the iteration procedure, it can be absorbed out. The ease of absorption will depend on the structure of $\underset{\sim}{G}$.

It is readily shown that if the distribution of $\underset{\sim}{u}$ is $N(\underset{\sim}{0},\underset{\sim}{G})$ and of $\underset{\sim}{z}$ given $\underset{\sim}{u}$ is $N(\underset{\sim}{X}\underset{\sim}{\beta}+\underset{\sim}{Z}\underset{\sim}{u},\underset{\sim}{Q})$, $\underset{\sim}{\mu}$ is $\underset{\sim}{Y}$, $\underset{\sim}{D}$ is $\underset{\sim}{I}$, $\underset{\sim}{y}$ is $\underset{\sim}{z}$, $\underset{\sim}{E}$ is $\underset{\sim}{Q}$ and that 4.30 becomes

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{Z} + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\beta}} \\ \hat{\underset{\sim}{u}} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} \end{bmatrix}$$

which are Henderson's mixed model equations (equation 4.2). In this case the equations do not need to be iterated since $\underset{\sim}{y}$ is $\underset{\sim}{z}$ and is independent of the solution for $\underset{\sim}{\beta}$.

In this section, equations 4.27 and 4.30 are proposed as a computing strategy for solving equations 4.25. Their obvious similarity to the BLUP equations (3.5) suggests the solution $\hat{\underset{\sim}{u}}$ would be a predictor of the random effects on the underlying scale (see equation 4.21). Adapting equation 6.17 to the notation of equation 4.27, the solution

is $$\hat{\underset{\sim}{u}} = (\underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{Z} + \underset{\sim}{G}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu}) \qquad (4.31)$$

This expression would be used to solve for $\underset{\sim}{u}$ after a final solution for the fixed effects had been obtained. The properties of 4.31 as a predictor of the random effects have not been investigated.

Chapter 5. Deriving the logistic linear mixed model

Categorical data encountered in animal breeding research is often assumed to be sampled from either the binomial or the multinomial distributions. Multinomial characters may be of either the threshold type, where an order of categories is implicit, or the extremal type where the ordering of categories is arbitrary. This chapter derives the specific equations required for analysing binomial, extremal and multiple threshold data by the two methods outlined in chapter 4; the joint-maximization and the maximization-expectation methods. The derivations are based on the logit transformation (section 3.2.1). The name logistic linear mixed model (LLMM) will be used specifically for equations obtained by the maximization-expectation method for binomial and multinomial characters. Equations obtained by the joint-maximization method will not be utilized in subsequent chapters and are not included when the name LLMM is used.

5.1. Mixed model analysis of binomial data

The binomial distribution has density function

$$\begin{aligned} {}^{n}C_{z}\; p^{z}\,(1-p)^{n-z} &= \exp(\, z \ln(p/(1-p)) + n \ln(1-p) + \ln({}^{n}C_{z})) \\ &= \exp(z\,\theta - n \ln(1 + \exp(\theta)) + \ln({}^{n}C_{z})) \\ &= \exp(z\,\theta - g(\theta) + h(z)) \qquad (5.1) \end{aligned}$$

This is of the one-parameter exponential form of equation 4.4 where z is the binomial frequency of successes in n trials,

p is the binomial probability,

$\theta = \ln(p/(1-p))$ and

$g(\theta) = n \ln(1 + \exp(\theta))$ (5.2)

Differentiating $g(\theta)$ with respect to θ gives

$$g'(\theta) = n \,\frac{\exp(\theta)}{(1 + \exp(\theta))} = n\,p \qquad (5.3)$$

$$\text{and} \quad g''(\theta) = n \frac{\exp(\theta)}{(1 + \exp(\theta))^2} = np(1 - p) \qquad (5.4)$$

5.1.1. Solution by the joint-maximization method

From equation 4.3, the linear predictor is $\hat{\underset{\sim}{Y}} = \underset{\sim}{X}\hat{\underset{\sim}{\beta}} + \underset{\sim}{Z}\hat{\underset{\sim}{u}}$ (5.5)

Using the link function, $\hat{\theta}_{ij} = \hat{Y}_{ij}$, (5.6)

$\underset{\sim}{p}$ may be predicted by $\hat{\underset{\sim}{p}} = \left\{ \dfrac{\exp(\hat{\theta}_{ij})}{1 + \exp(\theta_{ij})} \right\}$ (5.7)

From 4.6 $\underset{\sim}{\mu} = \{n\ g'(\hat{\theta}_{ij})\} = n\ \hat{\underset{\sim}{p}}$ (5.8)

From 4.7 $\underset{\sim}{V} = \text{diag}\ \{ g''(\hat{\theta}_{ij}) \} = \text{diag}\ \{ n\ \hat{p}_{ij}(1-\hat{p}_{ij}) \}$ (5.9)

From 4.8 $\underset{\sim}{D} = \left\{ \dfrac{\delta\ \theta_{ij}}{\delta\ Y_{ij}} \right\} = \underset{\sim}{I}$ (5.10)

From 4.17 $\underset{\sim}{y} = \hat{\underset{\sim}{Y}} + \underset{\sim}{V}^{-1}\ (\underset{\sim}{z} - \underset{\sim}{\mu})$ (5.11)

Inserting these expressions into equation 4.18 gives

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{V}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{V}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{V}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{V}\underset{\sim}{Z} + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{pmatrix} \hat{\underset{\sim}{\beta}} \\ \hat{\underset{\sim}{u}} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{X}'\underset{\sim}{V}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{V}\underset{\sim}{y} \end{pmatrix} \qquad (5.12)$$

The iterative solution of these equations is quite straight-forward following the comments in section 3.2.5.

5.1.2. Solution by the maximization-expectation method

The difficult step with this method is to find satisfactory expressions for the expected values of $g'(\dot{\theta}_{ij})$, $g''(\dot{\theta}_{ij})$ and $(g'(\dot{\theta}_{ij})g'(\dot{\theta}_{lm}))$ to substitute in equations 4.22 and 4.23.

i.e. $\underset{\sim}{\mu} = \{\ E\ [\ g'(\dot{\theta}_{ij})\ ]\ \}$ (5.13)

$$\text{var}(z_{ij}) = E[g''(\dot{\theta}_{ij})] + E[g'(\dot{\theta}_{ij})^2] - \mu_{ij}^2 \qquad (5.14)$$

$$\text{cov}(z_{ij}, z_{lm}) = E\ [g'(\dot{\theta}_{ij})\ g'(\dot{\theta}_{lm})] - \mu_{ij}\mu_{lm} \qquad (5.15)$$

$$\text{where } g'(\dot{\theta}_{ij}) = n\dot{p}_{ij} = \frac{n\exp(\dot{\theta}_{ij})}{1+\exp(\dot{\theta}_{ij})} = \frac{n\exp(Y_{ij}+u_i)}{1+\exp(Y_{ij}+u_i)} \tag{5.16}$$

$$\text{and } g''(\dot{\theta}_{ij}) = n\dot{p}_{ij}(1-\dot{p}_{ij}) = \frac{n\exp(\dot{\theta}_{ij})}{(1+\exp(\dot{\theta}_{ij}))^2}$$

$$= \frac{n\exp(Y_{ij}+u_i)}{(1+\exp(Y_{ij}+u_i))^2} \tag{5.17}$$

Notation: A dot will be used over the symbol p when the probability is a function of a particular random element u_i, to distinguish it from a corresponding 'average' p which will not have a dot.

The three expectations in 5.13, 5.14 and 5.15 can be obtained as functions of $E[\dot{p}_{ij}]$ and $E[\dot{p}_{ij}\dot{p}_{lm}]$. The distribution of u_i, or at least its first moments, must be specified for these expectations to be formed. Three alternatives have been investigated.

1. u_i distibuted $N(0,\sigma_u^2)$.
2. u_i distributed with an unspecified symmetric distribution with mean zero and variance small (say less than 0.3).
3. u_i distributed with mean zero and variance σ_u^2. Expected values are obtained from the logistic distribution.

5.1.2.1. Using the Normal distribution

From 5.16

$$E[\dot{p}_{ij}] = E\left[\frac{\exp(Y_{ij}+u_i)}{1+\exp(Y_{ij}+u_i)}\right]$$

$$= \int \frac{\exp(Y_{ij}+u_i)}{1+\exp(Y_{ij}+u_i)} \frac{1}{\sqrt{2\pi}\,\sigma_u} \exp\left(\frac{-u_i^2}{2\sigma_u^2}\right) du_i \tag{5.18}$$

Using the standard series

$$\frac{a}{1+a} = a - a^2 + a^3 - a^4 + a^5 - a^6 \ldots \text{ for } a < 1. \tag{5.19}$$

and

$$\frac{a}{1+a} = 1 - a^{-1} + a^{-2} - a^{-3} + a^{-4} - a^{-5} \ldots \text{ for } a > 1.$$

but with $a = \exp(Y_{ij} + u_i)$

implying $a < 1$ if $u_i < -Y_{ij}$

and $a > 1$ if $u_i > -Y_{ij}$

allows 5.18 to be written as

$$E[\,\dot{p}_{ij}\,] = \int_{-\infty}^{-Y_{ij}} \Big(-\sum_k^{\infty} (-1)^k \exp(Y_{ij} + u_i)^k\Big) \frac{1}{\sqrt{2\pi}\,\sigma_u} \exp\Big(\frac{-u_i^2}{2\sigma_u^2}\Big)\, du_i$$

$$+ \int_{-Y_{ij}}^{\infty} \Big(1 - \sum_k^{\infty} (-1)^k \exp(Y_{ij} + u_i)^{-k}\Big) \frac{1}{\sqrt{2\pi}\,\sigma_u} \exp\Big(\frac{-u_i^2}{2\sigma_u^2}\Big)\, du_i$$

$$= \int_{-\infty}^{-Y_{ij}} \Big(-\sum_k^{\infty} (-1)^k \exp(Y_{ij})^k\Big) \frac{1}{\sqrt{2\pi}\,\sigma_u} \exp\Big(\frac{-u_i^2 + 2u_i k\sigma_u^2}{2\sigma_u^2}\Big)\, du_i$$

$$+ \int_{-Y_{ij}}^{\infty} \Big(1 - \sum_k^{\infty} (-1)^k \exp(Y_{ij})^{-k}\Big) \frac{1}{\sqrt{2\pi}\,\sigma_u} \exp\Big(\frac{-u_i^2 - 2u_i k\sigma_u^2}{2\sigma_u^2}\Big)\, du_i$$

$$= -\sum_k^{\infty} (-1)^k \exp(kY_{ij} + k^2\sigma_u^2/2) \int_{-\infty}^{\frac{-Y_{ij} - k\sigma_u^2}{\sigma_u}} \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-x_i^2}{2}\Big)\, dx_i$$

$$+ \int_{\frac{-Y_{ij}}{\sigma_u}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-x_i^2}{2}\Big)\, dx_i$$

$$+ \sum_k^{\infty} (-1)^k \exp(-kY_{ij} + k^2\sigma^2/2) \int_{\frac{-Y_{ij} + k\sigma_u^2}{\sigma_u}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-x_i^2}{2}\Big)\, dx_i$$

$$= 1 - \Phi\left(\frac{-Y_{ij}}{\sigma_u}\right) \tag{5.20}$$

$$+ \sum_{k}^{\infty} F(k) \left[\exp(-kY_{ij})\left(1-\Phi\left(\frac{-Y_{ij}+k\sigma_u^2}{\sigma_u}\right)\right) - \exp(kY_{ij})\Phi\left(\frac{-Y_{ij}-k\sigma_u^2}{\sigma_u}\right)\right]$$

where $F(k) = (-1)^k \exp(k^2\sigma_u^2/2)$

and after substituting

$x_i = \frac{u_i - k\sigma_u^2}{\sigma_u}$ in the first integral and

$x_i = \frac{u_i + k\sigma_u^2}{\sigma_u}$ in the second integral.

This expression is not practical computationally because F(k) quickly becomes large while the associated integrals become small but the product is not small. Numerical integration techniques may be feasible in small problems but are likely to be too demanding, in computing resources, for larger problems. The same comments apply to the expressions for $E[\dot{p}_{ij}\dot{p}_{lm}]$ obtained in Appendix B. Computing equations of the form of equation 4.25 are not derived in this section because their evaluation is impractical for the reasons just noted.

5.1.2.2. Using the moments of a symmetric distribution

This method requires the nomination of the variance, σ_u^2, but not the full distribution of u_i. It has the restriction that σ_u^2 may not be large because of the approximations in the derivation. The derivation follows the same principle as the first method and readily extends to multinomial extremal characters. The full algebraic development for multinomial characters is in Appendix C. The principle is outlined here but the expressions are obtained from the appendix.

From 5.16

$$E\left[\,\dot{p}_{ij}\,\right] = E\left[\,\frac{\exp(Y_{ij} + u_i)}{1 + \exp(Y_{ij} + u_i)}\,\right] \tag{5.21}$$

Assuming Y_{ij} is negative and $u_i < -Y_{ij}$ enables $\dot{p}_{ij}$ to be expanded using the series in 5.19

$$E\left[\,\dot{p}_{ij}\,\right] = E\left[\,-\sum_m \left(-\exp(Y_{ij} + u_i)\right)^m\,\right]$$

$$= E\left[\,-\sum_m \left(-\left(1 + \sum_n \frac{(Y_{ij} + u_i)^n}{n!}\right)\right)^m\,\right] \tag{5.22}$$

On taking expectations of the individual terms in 5.22, odd powers of u_i drop out because of the symmetry assumption. Further, terms with u_i raised to the power of 4 or more are neglected as being negligible when $\mathrm{var}(u_i)$ is small. From equation C.52,

$$E\left[\,\dot{p}_{ij}\right] = p_{ij} + p_{ij}(1-p_{ij})\,\sigma_u^2\,(1/2 - p_{ij}) \tag{5.23}$$

$$\text{where } p_{ij} = \frac{\exp(Y_{ij})}{1 + \exp(Y_{ij})} \tag{5.24}$$

In similar fashion, from equation C.80

$$E[\dot{p}_{ij}\dot{p}_{lm}] = p_{ij}p_{lm}\Big(1 + \sigma_u^2\big(\,(\tfrac{1}{2}-p_{ij})(1-p_{ij}) + (\tfrac{1}{2}-p_{lm})(1-p_{lm})\,\big)$$

$$+ \sigma_u^2 r_{il}(1-p_{ij})(1-p_{lm})\,\Big) \tag{5.25}$$

where r_{il} is the correlation between u_i and u_l.

The coefficient r_{il} is one when i equals l and otherwise it is usually zero. In the animal breeding context, i and l may refer to related sires and r_{il} would be their additive genetic relationship.

Substituting 5.23 and 5.25 in 5.13, 5.14 and 5.15 gives

$$\underset{\sim}{\mu} = \{\, n_{ij}p_{ij} + n_{ij}p_{ij}(1-p_{ij})\sigma_u^2(\tfrac{1}{2}-p_{ij})\,\} \tag{5.26}$$

$$\begin{aligned}\mathrm{Cov}(z_{ij},z_{lm}) = {} & n_{ij}p_{ij}n_{lm}p_{lm}(1 \\ & \quad + \sigma_u^2(\ (\tfrac{1}{2}-p_{ij})(1-p_{ij}) + (\tfrac{1}{2}-p_{lm})(1-p_{lm})\ ) \\ & \quad + \sigma_u^2 r_{il}(1-p_{ij})(1-p_{lm})\ ) \\ - {} & n_{ij}p_{ij}n_{lm}p_{lm}(1 \\ & \quad + \sigma_u^2(\ (\tfrac{1}{2}-p_{ij})(1-p_{ij}) + (\tfrac{1}{2}-p_{lm})(1-p_{lm})\ ))\end{aligned}$$

after dropping a term in σ_u^4.

$$= n_{ij}p_{ij}(1-p_{ij})\ \sigma_u^2 r_{il} n_{lm}p_{lm}(1-p_{lm}) \qquad (5.27)$$

for $ij \neq lm$.

$$\begin{aligned}\mathrm{var}(z_{ij}) &= \mu_{ij} - E[n_{ij}\dot{p}_{ij}^2] + E[(n_{ij}\dot{p}_{ij})^2] - \mu_{ij}^2 \\ &= n_{ij}p_{ij}(1 + \sigma_u^2(1-p_{ij})(\tfrac{1}{2} - p_{ij}) \\ &\quad - n_{ij}p_{ij}^2(1 + 2\sigma_u^2(1-p_{ij})(\tfrac{1}{2} - p_{ij}) + \sigma_u^2(1-p_{ij})^2) \\ &\quad + n_{ij}^2 p_{ij}^2(1-p_{ij})^2\sigma_u^2 \\ &= n_{ij}p_{ij}(1-p_{ij}) \\ &\quad + n_{ij}p_{ij}(1-p_{ij})\sigma_u^2(\tfrac{1}{2} - p_{ij} - p_{ij} + 2p_{ij}^2 - p_{ij}(1-p_{ij})) \\ &\quad + n_{ij}^2 p_{ij}^2(1-p_{ij})^2\sigma_u^2 \\ &= n_{ij}p_{ij}(1-p_{ij})\ (1 + \sigma_u^2(\tfrac{1}{2} - 3p_{ij}(1-p_{ij}))) \\ &\quad + n_{ij}^2 p_{ij}^2(1-p_{ij}^2)\ \sigma_u^2 \qquad (5.28)\end{aligned}$$

These expressions can be represented in matrix notation to conform to 4.26 as follows.

$$\underset{\sim}{E} = \mathrm{diag}\{\ n_{ij}p_{ij}(1-p_{ij})(1 + \sigma_u^2(\tfrac{1}{2} - 3p_{ij}(1-p_{ij})))\ \} \qquad (5.29)$$

$$\underset{\sim}{A} = \mathrm{diag}\ \{\ n_{ij}p_{ij}(1-p_{ij})\ \} \qquad (5.30)$$

$\underset{\sim}{G} = \sigma_u^2\ \underset{\sim}{R}$ where $\underset{\sim}{R}$ is a matrix of additive relationships among the random elements in $\underset{\sim}{u}$,

$$\underset{\sim}{V} = \mathrm{var}(\underset{\sim}{z}) = \underset{\sim}{E} + \underset{\sim}{AZGZ'A} \qquad (5.31)$$

Differentiating 5.26 with respect to Y_{ij},

$$\frac{d\ \mu_{ij}}{d\ Y_{ij}} = \frac{d\ \mu_{ij}}{d\ p_{ij}} \frac{d\ p_{ij}}{d\ Y_{ij}}$$

$$= \frac{d}{d\ p_{ij}} n_{ij}(p_{ij} + \sigma_u^2(\tfrac{1}{2}p_{ij} - \tfrac{1}{2}p_{ij}^2 - p_{ij}^2 + p_{ij}^3)) \frac{d\ p_{ij}}{d\ Y_{ij}}$$

$$= n_{ij}(1 + \sigma_u^2(\tfrac{1}{2} - 3p_{ij} + 3p_{ij}^2)\ )\ p_{ij}(1-p_{ij})$$

This is the same as the elements in $\underset{\sim}{E}$ so

$$\underset{\sim}{D} \text{ of } 4.24 = \underset{\sim}{E} \tag{5.32}$$

$$\underset{\sim}{y} \text{ of } 4.24 = \underset{\sim}{\hat{Y}} + \underset{\sim}{D}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu})$$

$$= \underset{\sim}{\hat{Y}} + \underset{\sim}{E}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu}) \tag{5.33}$$

Then 4.27 becomes

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{E}\underset{\sim}{E}^{-1}\underset{\sim}{E}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{E}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{E}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{Z} + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta}^{o} \\ \underset{\sim}{u}^{o} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{E}\underset{\sim}{E}^{-1}\underset{\sim}{E}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{E}\underset{\sim}{y} \end{bmatrix}$$

which simplifies to

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{E}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{Z} + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta}^{o} \\ \underset{\sim}{u}^{o} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{E}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{y} \end{bmatrix} \tag{5.34}$$

In equations 5.26, 5.29, 5.30, 5.32 and 5.33 is a set of simple expressions which can be easily used in 5.34. The starting rule suggested in section 3.2.5 may be used. Once equation 5.34 has been solved, the new solution for β^{o} is used in the linear predictor, $\underset{\sim}{\hat{Y}} = \underset{\sim}{X}\underset{\sim}{\beta}^{o}$ (from equation 4.21), to set up and solve equation 5.34 again. The process is repeated until changes in β^{o} between iterations are sufficiently small. As noted before, the equations in $\underset{\sim}{u}^{o}$ may be absorbed.

5.1.2.3. Using the logistic distribution

The logistic distribution is described in section 3.2.1. The simple relationship between probability and threshold forms a convenient basis for obtaining $E[\ \dot{p}_{ij}]$ and $E[\ \dot{p}_{ij}\dot{p}_{lm}\ ]$.

The expression from equation 5.16

$$\dot{p}_{ij} = \frac{\exp(Y_{ij} + u_i)}{1 + \exp(Y_{ij} + u_i)} \qquad (5.35)$$

is based on the standard logistic distribution. Under this distribution, $\dot{p}_{ij}$ is the probability that a random variable, x_{ij}, with mean zero and variance $\pi^2/3$ has a value less than $Y_{ij}+u_i$. The effect of assuming u_i is unknown with mean zero and variance σ_u^2 is to increase the variance on the underlying scale to $\pi^2/3 + \sigma_u^2$. The threshold point, Y_{ij}, does not change. It is convenient to assume the distribution of $x_{ij} + u_i$ is still logistic. $E[\dot{p}_{ij}]$ is then obtained as the probability that $x_{ij}+u_i$ is less than Y_{ij} and is, after a change of scale,

$$\bar{p}_{ij} = E[\dot{p}_{ij}] = \frac{\exp(s\,\hat{Y}_{ij})}{1 + \exp(s\,\hat{Y}_{ij})} \qquad (5.36)$$

where the scale factor is

$$s = \sqrt{\frac{\pi^2/3}{\pi^2/3 + \sigma_u^2}}$$

$$= \sqrt{1-\rho} \qquad (5.37)$$

and the intraclass correlation is

$$\rho = \frac{\sigma_u^2}{\pi^2/3 + \sigma_u^2} \qquad (5.38)$$

The intraclass correlation in this situation is the correlation between two observations on the underlying scale which have the same (unknown) random effect.

The second result that is required is an expression for

$$E[\, \dot{p}_{ij}\dot{p}_{lm} \,] = E\left[\frac{\exp(Y_{ij} + u_i)}{1 + \exp(Y_{ij} + u_i)} \; \frac{\exp(Y_{lm} + u_l)}{1 + \exp(Y_{lm} + u_l)} \right]$$

By definition, under the preceding assumptions,

$$\mathrm{Cov}(x_{ij}+u_i, x_{lm}+u_l) = r_{il}\,\sigma_u^2$$

$$= r_{il}\,\frac{\rho}{1-\rho}\,\pi^2/3 \qquad (5.39)$$

where ρ is the intraclass correlation from 5.37 and
r_{il} is the correlation between u_i and u_l.

On the basis of 4.28,

$$\mathrm{Cov}(\dot{p}_{ij}, \dot{p}_{lm}) = \frac{d\,\bar{p}_{ij}}{d\,\hat{Y}_{ij}}\,\mathrm{Cov}\,(x_{ij}+u_i, x_{lm}+u_l)\,\frac{d\,\bar{p}_{lm}}{d\,\hat{Y}_{lm}} \qquad (5.40)$$

Differentiating 5.36

$$\frac{d\,\bar{p}_{ij}}{d\,\hat{Y}_{ij}} = \frac{s\,\exp(s\,\hat{Y}_{ij})}{(1+\exp(s\,\hat{Y}_{ij}))^2} = s\,\bar{p}_{ij}(1-\bar{p}_{ij}) \qquad (5.41)$$

Substituting 5.39 and 5.41 in 5.40,

$$\mathrm{Cov}(\dot{p}_{ij}, \dot{p}_{lm}) = s\,\bar{p}_{ij}(1-\bar{p}_{ij})\,r_{il}\,\frac{\rho}{1-\rho}\,\pi^2/3\,s\,\bar{p}_{lm}(1-\bar{p}_{lm})$$

$$= \bar{p}_{ij}(1-\bar{p}_{ij})\,r_{il}\,\rho\,\pi^2/3\,\bar{p}_{lm}(1-\bar{p}_{lm}) \qquad (5.42)$$

since, from 5.37, $s^2 = 1-\rho$

By the definition of covariance,

$$E[\dot{p}_{ij}\dot{p}_{lm}] = \mathrm{Cov}(\dot{p}_{ij}, \dot{p}_{lm}) + E[\dot{p}_{ij}]\,E[\dot{p}_{lm}]$$

$$= \bar{p}_{ij}(1-\bar{p}_{ij})\,r_{il}\,\rho\,\pi^2/3\,\bar{p}_{lm}(1-\bar{p}_{lm}) + \bar{p}_{ij}\bar{p}_{lm} \qquad (5.43)$$

Substituting 5.36 and 5.43 in 5.13, 5.14 and 5.15 gives

$$\underset{\sim}{\mu} = \{\,n_{ij}\bar{p}_{ij}\,\} = \{\,\frac{n_{ij}\,\exp(s\,\hat{Y}_{ij})}{1+\exp(s\,\hat{Y}_{ij})}\,\} \qquad (5.44)$$

$$\mathrm{Cov}(z_{ij}, z_{lm}) = \mathrm{Cov}(n_{ij}\dot{p}_{ij}, n_{lm}\dot{p}_{lm})$$

$$= n_{ij}\bar{p}_{ij}(1-\bar{p}_{ij})\,r_{il}\,\rho\,\pi^2/3\,n_{lm}\bar{p}_{lm}(1-\bar{p}_{lm}) \qquad (5.45)$$

for $ij \neq lm$

$$\mathrm{var}(z_{ij}) = E[n_{ij}\dot{p}_{ij}] - E[n_{ij}\dot{p}_{ij}^2] + \mathrm{var}[n_{ij}\dot{p}_{ij}]$$

$$= n_{ij}\bar{p}_{ij} - n_{ij}(\mathrm{var}(\dot{p}_{ij})+\bar{p}_{ij}^2) + n_{ij}^2\mathrm{var}(\dot{p}_{ij})$$

$$= n_{ij}\bar{p}_{ij}(1-\bar{p}_{ij}) + n_{ij}(n_{ij}-1)\bar{p}_{ij}^2(1-\bar{p}_{ij})^2\,\rho\pi^2/3 \qquad (5.46)$$

These expressions can be represented in matrix notation to conform to 4.29 as follows.

$$\underset{\sim}{E} = \text{diag}\ \{\ n_{ij}\bar{p}_{ij}(1-\bar{p}_{ij})(1 - \bar{p}_{ij}(1-\bar{p}_{ij})\ \rho\ \pi^2/3\ )\ \} \qquad (5.47)$$

$$\underset{\sim}{A} = \text{diag}\ \{\ n_{ij}\bar{p}_{ij}(1-\bar{p}_{ij})\ \} \qquad (5.48)$$

$\underset{\sim}{G} = \rho\ \pi^2/3\ \underset{\sim}{R}$ where $\underset{\sim}{R}$ is a matrix of additive relationships among the random elements in $\underset{\sim}{u}$, (5.49)

$$\underset{\sim}{V} = \text{var}(\underset{\sim}{z}) = \underset{\sim}{E} + \underset{\sim}{A}\underset{\sim}{Z}\underset{\sim}{G}\underset{\sim}{Z}'\underset{\sim}{A}$$

$$\underset{\sim}{D} \text{ of } 4.24 = s\ \underset{\sim}{A} \qquad (5.50)$$

$$\begin{aligned}\underset{\sim}{y} \text{ of } 4.24 &= \hat{\underset{\sim}{Y}} + \underset{\sim}{D}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu}) \\ &= \hat{\underset{\sim}{Y}} + \underset{\sim}{A}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu})/s \qquad (5.51)\end{aligned}$$

Then 4.30 becomes, after dividing by s,

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{Z} + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta}^{o} \\ \underset{\sim}{u}^{o} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{y} \end{bmatrix} \qquad (5.52)$$

Writing $\underset{\sim}{W} = \underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}$

$$\begin{aligned} &= \text{diag}\ \left\{ \frac{n_{ij}\bar{p}_{ij}(1-\bar{p}_{ij})\ n_{ij}\bar{p}_{ij}(1-\bar{p}_{ij})}{n_{ij}\bar{p}_{ij}(1-\bar{p}_{ij})(1 - \bar{p}_{ij}(1-\bar{p}_{ij})\ \rho\ \pi^2/3)} \right\} \\ &= \text{diag}\ \left\{ \frac{n_{ij}\bar{p}_{ij}(1-\bar{p}_{ij})}{1 - \bar{p}_{ij}(1-\bar{p}_{ij})\ \rho\ \pi^2/3} \right\} \qquad (5.53)\end{aligned}$$

Equations 5.44, 5.48, 5.51 and 5.53 contain the expressions to be used in 5.52. A starting rule was suggested in section 3.2.5. Once 5.52 has been solved, the new solution for $\underset{\sim}{\beta}^{o}$ is used in the linear predictor, $\hat{\underset{\sim}{Y}} = \underset{\sim}{X}\underset{\sim}{\beta}^{o}$ (from 4.21), to set up and solve 5.52 again. The process is repeated until changes in $\underset{\sim}{\beta}^{o}$ between iterations are sufficiently small. As noted in section 4.2, the solution for $\underset{\sim}{u}^{o}$ is not required during iteration and it will often be convenient to absorb these equations when solving for $\underset{\sim}{\beta}^{o}$.

5.2. Analysis of multinomial data

The distinction between the two types of multinomial characters, extremal and threshold, has been discussed in section 2.1. A threshold character arises when there is a single underlying variable with several thresholds. An extremal character arises when there are several underlying variables, each having a single threshold and each category corresponding to a variable. Each observation adds a count to the category of the extremal character that corresponds to the first underlying variable to reach its threshold. When there is a single underlying variable with only one threshold, the extremal and threshold models both reduce to the binomial situation.

5.2.1. Analysis of extremal characters

The obvious choice of a probability distribution to map the probabilities associated with each category to an underlying continuous scale might be the multivariate normal distribution. A convenient and simple alternative is the multinomial logit introduced in section 3.2.1.

The development of the equations to analyse an extremal character by reweighted least squares under a fixed model was presented in section 3.2.4.

5.2.1.1. Solution by joint-maximization method

The extension of the results of section 3.2.4 to a mixed model is straight-forward. Following the format of section 5.1.1 with appropriate modification, from section 4.1, the linear predictor is $\underset{\sim}{Y} = (\underset{\sim}{X} \otimes \underset{\sim}{I})\ \underset{\sim}{B} + (\underset{\sim}{Z} \otimes \underset{\sim}{I})\ \underset{\sim}{u}$ (5.54)

and the link function is $\theta_{ijk} = Y_{ijk}$ (5.55)

where i refers to the i th random vector, u_i,
j refers to the j th sample within i,
k refers to the k th category,

From 3.48, $$\underset{\sim}{p}_{ij} = \{ p_{ijk} \} = \{ \frac{\exp(\theta_{ijk})}{1 + \sum\limits_{n} \exp(\theta_{ijn})} \} \qquad (5.56)$$

From 3.55 and 4.6, $$\mu_{ijk} = n_{ij} \frac{\exp(\theta_{ijk})}{1 + \underset{n}{\ } \exp(\theta_{ijn})} \qquad (5.57)$$

From 3.56 and 4.7, $$\underset{\sim}{V}_{ij} = \underset{\sim}{n}_{ij} (\text{diag } \{\underset{\sim}{p}_{ijk}\} + \underset{\sim}{p}_{ij}\underset{\sim}{p}'_{ij})$$

$$\text{and } \underset{\sim}{V} = \sum_i{}^{+} \sum_j{}^{+} \underset{\sim}{V}_{ij} \qquad (5.58)$$

From 4.8, $$\underset{\sim}{D} = \{ \frac{\delta\theta_{ijk}}{\delta Y_{ijk}} \} = \underset{\sim}{I} \qquad (5.59)$$

From 4.17, $$\underset{\sim}{y} = \underset{\sim}{\hat{Y}} + \underset{\sim}{V}^{-1} (\underset{\sim}{z} - \underset{\sim}{\mu}) \qquad (5.60)$$

Let $\underset{\sim}{G} = \underset{\sim}{R} \otimes \underset{\sim}{\Sigma}$ where $\underset{\sim}{R}$ is a matrix of additive relationships
among the vectors, $\underset{\sim}{u}_i$, and
$\underset{\sim}{\Sigma}$ is var($\underset{\sim}{u}_i$).

Inserting these expressions in 4.18 gives

$$\begin{vmatrix} (\underset{\sim}{X}\otimes\underset{\sim}{I})' \underset{\sim}{V} (\underset{\sim}{X}\otimes\underset{\sim}{I}) & (\underset{\sim}{X}\otimes\underset{\sim}{I})' \underset{\sim}{V} (\underset{\sim}{Z}\otimes\underset{\sim}{I}) \\ (\underset{\sim}{Z}\otimes\underset{\sim}{I})' \underset{\sim}{V} (\underset{\sim}{X}\otimes\underset{\sim}{I}) & (\underset{\sim}{Z}\otimes\underset{\sim}{I})' \underset{\sim}{V} (\underset{\sim}{Z}\otimes\underset{\sim}{I}) + \underset{\sim}{G}^{-1} \end{vmatrix} \begin{vmatrix} \underset{\sim}{B}^{o} \\ \underset{\sim}{u}^{o} \end{vmatrix} = \begin{vmatrix} (\underset{\sim}{X}\otimes\underset{\sim}{I})' \underset{\sim}{V}\underset{\sim}{y} \\ (\underset{\sim}{Z}\otimes\underset{\sim}{I})' \underset{\sim}{V}\underset{\sim}{y} \end{vmatrix}$$

(5.61)

These equations are readily solved (see section 3.2.5) but will often be of large order since the number of equations is

(number of categories less 1) times (rank of [$\underset{\sim}{X}$ $\underset{\sim}{Z}$])

5.2.1.2. Solution by the maximization-expectation method

The application of this method to extremal characters is algebraically cumbersome but straight-forward. The derivation will be based on section 4.2 but drawing from section 3.2.4.

Extending equation 3.49 in the fashion of equation 4.21 gives ln L = constant

$$+ \sum_{i}^{q} \sum_{j}^{N_i} \left(\left(\sum_{k}^{c} z_{ijk}\theta_{ijk} \right) - n_{ij} \ln\left(1 + \sum_{n}^{c} \exp(\theta_{ijn}) \right) \right)$$

$$+ \ln(U(\underset{\sim}{u}:\underset{\sim}{G})) \qquad (5.62)$$

where $$\theta_{ijk} = \ln\left(\frac{\dot{p}_{ijk}}{1 - \sum_n \dot{p}_{ijn}} \right) \qquad (5.63)$$

$$p_{ijk} = \frac{\exp(\theta_{ijk})}{1 + \sum_n \exp(\theta_{ijn})} \qquad (5.64)$$

and $$g(\theta_{ijk}) = n_{ij} \ln(1 + \sum \exp(\theta_{ijn})) \qquad (5.65)$$

i refers to the i th random vector, $\underset{\sim}{u}_i$,
j refers to the j th sample within i,
k refers to the k th of c+1 categories,
$\underset{\sim}{G} = \underset{\sim}{R} \times \underset{\sim}{\Sigma}$,
$\underset{\sim}{R}$ is a matrix of additive genetic relationships among the vectors $\underset{\sim}{u}_i$ and
$\underset{\sim}{\Sigma}$ is var($\underset{\sim}{u}_i$)

From equation 4.22 using equation 3.55

$$\begin{aligned} \underset{\sim}{\mu} &= E[\ \underset{\sim}{z}\ ] \\ &= E[\ \{\ g'(\theta_{ijk})\ \}\ ] \\ &= E[\ \{\ n_{ij}\dot{p}_{ijk}\ \}\ ] \end{aligned} \qquad (5.66)$$

From 4.23 using 3.56,

$$\begin{aligned} \underset{\sim}{V} &= \text{var}(\ \underset{\sim}{z}\ ) \\ &= \underset{\sim}{E} + \text{var}(\ \{\ g'(\theta_{ijk})\ \}\ ) \\ &= \sum_{i}^{+} \sum_{j}^{+} \underset{\sim}{E}_{ij} + \text{var}(\ \{\ n_{ij}\dot{p}_{ijk}\ \}\ ) \end{aligned} \qquad (5.67)$$

where $$\underset{\sim}{E}_{ij} = n_{ij}\ E[\ \text{diag}(\underset{\sim}{\dot{p}}_{ij}) - \underset{\sim}{\dot{p}}_{ij}\underset{\sim}{\dot{p}}'_{ij}\ ] \qquad (5.68)$$

From 4.21, the linear predictor is $\underset{\sim}{Y} = (\underset{\sim}{X} \otimes \underset{\sim}{I})\ \underset{\sim}{B}$ (5.69)

and the link function is $\theta_{ijk} = Y_{ijk}$. (5.70)

Equations 5.66 and 5.67 require expressions for $E(\dot{p}_{ijk})$ and $E(\dot{p}_{ijk}\dot{p}_{lmn})$. In section 5.1.2, three methods of obtaining these expectations were proposed. The first, a strict assumption of normality was found impractical and has not been attempted for this case. The third, a convenient use of the logistic distribution, has not been attempted but might lead to useful expressions. The second method, that of assuming only the moments are known, has been applied to this case. The algebra is in appendix C. From that appendix, Equation C.50 gives

$$E[\dot{p}_{ijk}] = p_{ijk}(1 - \tfrac{1}{2}\underset{\sim}{p}'_{ij}\underset{\sim}{d} + \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_k + \tfrac{1}{2}\sigma_k^2) \qquad (5.71)$$

and equation C.79 gives

$$\begin{aligned} E[\dot{p}_{ijk}\dot{p}_{lmn}] = p_{ijk}p_{lmn}(1 &+ \tfrac{1}{2}(\sigma_k^2 + 2r_{il}\sigma_{kn} + \sigma_n^2) \\ &+ \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} + r_{il}\underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{lm} + \underset{\sim}{p}'_{lm}\underset{\sim}{\Sigma}\underset{\sim}{p}_{lm} \\ &- \tfrac{1}{2}(\underset{\sim}{p}_{ij} + \underset{\sim}{p}_{lm})'\underset{\sim}{d} - \underset{\sim}{p}'_{ij}(\underset{\sim}{c}_k + r_{il}\underset{\sim}{c}_n) \\ &- \underset{\sim}{p}'_{lm}(\underset{\sim}{c}_n + r_{il}\underset{\sim}{c}_k) \qquad (5.72) \end{aligned}$$

where σ_k^2 is the k th diagonal element of $\underset{\sim}{\Sigma}$,

σ_{kn} is the kn th element of $\underset{\sim}{\Sigma}$,

$\underset{\sim}{d}$ is the vector with elements $\{\sigma_k^2\}$

c_k is the k th column of $\underset{\sim}{\Sigma}$,

$$p_{ijk} = \frac{\exp(Y_{ijk})}{1 + \Sigma\exp(Y_{ijn})}$$

$\underset{\sim}{p}_{ij}$ is $\{p_{ijk}\}$ and

r_{il} is an element from $\underset{\sim}{R}$.

For use in 5.67

$$\begin{aligned} \operatorname{cov}(n_{ij}\dot{p}_{ijk}n_{lm}\dot{p}_{lmn}) &= n_{ij}n_{lm}E[\dot{p}_{ijk}\dot{p}_{lmn}] \\ &\quad - n_{ij}n_{lm}E[\dot{p}_{ijk}]E[\dot{p}_{lmn}] \\ &= n_{ij}n_{lm}p_{ijk}p_{lmn}r_{il}(\sigma_{kn} + \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{lm} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_n - \underset{\sim}{p}'_{lm}\underset{\sim}{c}_k) \qquad (5.73) \end{aligned}$$

This can be conveniently expressed in matrix notation as

$$\mathrm{var}(\ \{\ n_{ij}\dot{p}_{ijk}\ \}\ ) = \underset{\sim}{A}\ (\underset{\sim}{Z} \otimes \underset{\sim}{I})\ \underset{\sim}{G}\ (\underset{\sim}{Z} \otimes \underset{\sim}{I})'\ \underset{\sim}{A} \qquad (5.74)$$

where $\underset{\sim}{A} = \sum_{i}^{+} \sum_{j}^{+} \underset{\sim}{A}_{ij}$

and $\underset{\sim}{A}_{ij} = n_{ij}\ (\ \mathrm{diag}\{\ p_{ijk}\ \} - \underset{\sim}{p}_{ij}\underset{\sim}{p}'_{ij}\ ) \qquad (5.75)$

Proof:

Let $\underset{\sim}{M} = r_{il}\underset{\sim}{A}_{ij}\underset{\sim}{\Sigma}\underset{\sim}{A}_{lm}$.

$$\begin{aligned} m_{kn} &= r_{il}n_{ij}(p_{ijk}\underset{\sim}{\delta}'_{k} - p_{ijk}\underset{\sim}{p}'_{ij})\ \underset{\sim}{\Sigma}\ (p_{lmn}\underset{\sim}{\delta}_{n} - p_{lmn}\underset{\sim}{p}_{lm})n_{lm} \\ &= r_{il}n_{ij}p_{ijk}n_{lm}p_{lmn}(\sigma_{kn} + \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{lm} - \underset{\sim}{c}'_{k}\underset{\sim}{p}_{lm} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_{n}) \\ &= \mathrm{cov}(n_{ij}\dot{p}_{ijk}n_{lm}\dot{p}_{lmn}) \text{ from } 5.73 \end{aligned}$$

with $\underset{\sim}{\delta}_{k}$ representing the k th column of an identity matrix.

QED

The kn th element in $\underset{\sim}{E}_{ij}$ of 5.68, from simplifying 5.72, is

$$\begin{aligned} \{e_{kn}\} = - n_{ij}\ p_{ijk}p_{ijn}(1 + \tfrac{1}{2}(\sigma_{k}^{2} + 2\sigma_{kn} + \sigma_{n}^{2}) + 3\underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} \\ - \underset{\sim}{p}'_{ij}\underset{\sim}{d} - 2\underset{\sim}{p}'_{ij}(\underset{\sim}{c}_{k} + \underset{\sim}{c}_{n})\ ) \qquad (5.76) \\ \text{for } k \neq n \end{aligned}$$

The k th diagonal element of $\underset{\sim}{E}_{ij}$, from 5.71 and 5.76, is

$$\begin{aligned} \{e_{kk}\} = n_{ij}p_{ijk}(1 - \tfrac{1}{2}\underset{\sim}{p}'_{ij}\underset{\sim}{d} + \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_{k} + \tfrac{1}{2}\sigma_{k}^{2}) \\ - n_{ij}\ p_{ijk}^{2}(1 + 2\sigma_{k}^{2} + 3\underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} - \underset{\sim}{p}'_{ij}\underset{\sim}{d} - 4\underset{\sim}{p}'_{ij}\underset{\sim}{c}_{k}) \ (5.77) \end{aligned}$$

Then $\underset{\sim}{V} = \mathrm{var}(\underset{\sim}{z}) = \underset{\sim}{E} + \underset{\sim}{A}\ (\underset{\sim}{Z} \otimes \underset{\sim}{I})\ \underset{\sim}{G}\ (\underset{\sim}{Z} \otimes \underset{\sim}{I})'\ \underset{\sim}{A} \qquad (5.78)$

From 5.66 using 5.71,

$$\begin{aligned} \underset{\sim}{\mu} &= \{\ E[\ n_{ij}\dot{p}_{ijk}\ ]\ \} \\ &= \{\ n_{ij}p_{ijk}(1 - \tfrac{1}{2}\underset{\sim}{p}'_{ij}\underset{\sim}{d} + \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_{k} + \tfrac{1}{2}\sigma_{k}^{2})\ \} \qquad (5.79) \end{aligned}$$

The matrix $\underset{\sim}{D}$ of 4.24 is obtained by differentiating 5.79.

The following differentials are required.

$$\frac{\partial p_{ijk}}{\partial Y_{ijk}} = p_{ijk}(1-p_{ijk})$$

and $$\frac{\partial p_{ijn}}{\partial Y_{ijk}} = -p_{ijk}p_{ijn} \qquad \text{for } k \neq n$$

Then $$\frac{\partial \underset{\sim}{p}_{ij}}{\partial Y_{ijk}} = p_{ijk}(\underset{\sim}{\delta}_k - \underset{\sim}{p}_{ij})$$

$$\frac{\partial \underset{\sim}{p}'_{ij}\underset{\sim}{d}}{\partial Y_{ijk}} = p_{ijk}(\underset{\sim}{\delta}_k - \underset{\sim}{p}_{ij})'\underset{\sim}{d} = p_{ijk}(\sigma_k^2 - \underset{\sim}{p}'_{ij}\underset{\sim}{d})$$

$$\frac{\partial \underset{\sim}{p}'_{ij}\underset{\sim}{c}_n}{\partial Y_{ijk}} = p_{ijk}(\underset{\sim}{\delta}_k - \underset{\sim}{p}_{ij})'\underset{\sim}{c}_n = p_{ijk}(\sigma_{kn} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_n)$$

and $$\frac{\partial \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij}}{\partial Y_{ijk}} = 2\underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}(\underset{\sim}{\delta}_k - \underset{\sim}{p}_{ij})p_{ijk} = 2p_{ijk}(\underset{\sim}{p}'_{ij}\underset{\sim}{c}_k - \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij})$$

Differentiating 5.79 using these expressions gives

$$\begin{aligned}\frac{\partial \mu_{ijn}}{\partial Y_{ijk}} = {} & n_{ij}p_{ijn}\big(0 - \tfrac{1}{2}p_{ijk}(\sigma_k^2 - \underset{\sim}{p}'_{ij}\underset{\sim}{d}) + 2p_{ijk}(\underset{\sim}{p}'_{ij}\underset{\sim}{c}_k - \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij}) \\ & \quad - p_{ijk}(\sigma_{kn} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_k)\ \big) \\ & - n_{ij}p_{ijk}p_{ijn}(1 - \tfrac{1}{2}\underset{\sim}{p}'_{ij}\underset{\sim}{d} + \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_n + \tfrac{1}{2}\sigma_n^2) \\ = {} & -\ n_{ij}p_{ijk}p_{ijn}\big(1 + \tfrac{1}{2}(\sigma_k^2 + 2\sigma_{kn} + \sigma_n^2) - 2\underset{\sim}{p}'_{ij}(\underset{\sim}{c}_n + \underset{\sim}{c}_k) \\ & \quad - \underset{\sim}{p}'_{ij}\underset{\sim}{d} + 3\underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij}\big) \qquad \text{for } k \neq n \qquad (5.80)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mu_{ijk}}{\partial Y_{ijk}} = {} & n_{ij}p_{ijk}\big(0 - \tfrac{1}{2}p_{ijk}(\sigma_k^2 - \underset{\sim}{p}'_{ij}\underset{\sim}{d}) + 2p_{ijk}(\underset{\sim}{p}'_{ij}\underset{\sim}{c}_k - \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij}) \\ & \quad - p_{ijk}(\sigma_k^2 - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_k)\ \big) \\ & + n_{ij}p_{ijk}(1-p_{ijk})(1 - \tfrac{1}{2}\underset{\sim}{p}'_{ij}\underset{\sim}{d} + \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_k + \tfrac{1}{2}\sigma_k^2)\end{aligned}$$

$$= n_{ij}p_{ijk}(1 - \tfrac{1}{2}\underset{\sim}{p}'_{ij}\underset{\sim}{d} + \underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij} - \underset{\sim}{p}'_{ij}\underset{\sim}{c}_k + \tfrac{1}{2}\sigma_k^2)$$

$$- n_{ij}p_{ijk}^2(1 + 2\sigma_k^2 - \underset{\sim}{p}'_{ij}\underset{\sim}{d} - 4\underset{\sim}{p}'_{ij}\underset{\sim}{c}_k + 3\underset{\sim}{p}'_{ij}\underset{\sim}{\Sigma}\underset{\sim}{p}_{ij}) \qquad (5.81)$$

$$\frac{\partial\ \mu_{ijk}}{\partial\ Y_{lmn}} = 0 \qquad \text{for } ij \neq lm$$

Comparison of 5.80 with 5.76, and 5.81 with 5.77 shows they are identical, i.e. $\underset{\sim}{D} = \underset{\sim}{E}$ (5.82)

$$\underset{\sim}{y} \text{ of } 4.24 = \hat{\underset{\sim}{Y}} + \underset{\sim}{D}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu})$$

$$= \hat{\underset{\sim}{Y}} + \underset{\sim}{E}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu}) \qquad (5.83)$$

Then 4.27 becomes

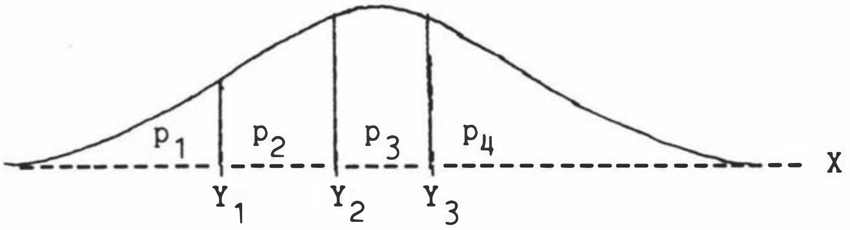
$$\begin{bmatrix} (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{E}(\underset{\sim}{X}\otimes\underset{\sim}{I}) & (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{A}(\underset{\sim}{Z}\otimes\underset{\sim}{I}) \\ (\underset{\sim}{Z}\otimes\underset{\sim}{I})'\underset{\sim}{A}(\underset{\sim}{X}\otimes\underset{\sim}{I}) & (\underset{\sim}{Z}\otimes\underset{\sim}{I})'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}(\underset{\sim}{Z}\otimes\underset{\sim}{I}) + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{pmatrix} \underset{\sim}{B}^{o} \\ \underset{\sim}{u}^{o} \end{pmatrix} = \left| \begin{matrix} (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{E}\underset{\sim}{y} \\ (\underset{\sim}{Z}\otimes\underset{\sim}{I})'\underset{\sim}{A}\underset{\sim}{y} \end{matrix} \right|$$

(5.84)

Equations 5.74, 5.76, 5.77, 5.79 and 5.83 contain the expressions required to solve 5.84. Once 5.84 has been solved, the new solution for $\underset{\sim}{B}$ is used in the linear predictor (equation 5.69) to set up and solve 5.84 again. The process is repeated until changes in $\underset{\sim}{B}$ between iterations are sufficiently small. As before, the equations in $\underset{\sim}{u}^{o}$ may be absorbed so that the order of the matrix to be inverted is (number of categories less 1) times (rank of X).

5.2.2. Analysis of multiple threshold characters

Many examples involving multinomial data are of the threshold type rather than extremal type. The threshold model proposes a single underlying variable, X, with c threshold values, Y_i, defining c+1 categories of response. The scale of X can be chosen so that the probability of response in a particular category relates to a known distribution, typically a normal or logistic distribution. The situation can be represented as



Such data are analysed in three ways:

1) Continuation ratios - the multiple categories are analysed as a series of independent binomial characters.
For example 1 versus 2 & 3 & 4
2 versus 3 & 4 omitting 1
3 versus 4 omitting 1 & 2

2) Repeated binomial analysis - the multiple categories are analysed as a series of binomial characters which are not independent. For example 1 versus 2 & 3 & 4
1 & 2 versus 3 & 4
1 & 2 & 3 versus 4

3) Simultaneous binomial analysis - the multiple categories are analysed by multivariate analysis of the binomial characters of method 2. This form of analysis is outlined by Bock (1975, p 541).

The first and second methods could be applied using the results of previous sections. However, neither provides a convenient basis for analysis of such data by a mixed model because the binomial characters are analysed separately. An additive random effect on the underlying scale would shift all thresholds and so must be estimated in an analysis that includes all thresholds. The third method does this and is presented in the remainder of this section.

For a single multiple-threshold sample of size n, let

p_k be the probability of response in category k,

$P_k = \sum_i^k p_i$ be the probability of response in the first k categories, (5.85)

z_k be the frequency of response in category k,

$S_k = \sum_i^k z_i$ be the frequency of response in the first k categories and let (5.86)

$Y_k = \ln\left(\frac{P_k}{1 - P_k} \right)$ be the k th threshold, implying

$$P_k = \frac{\exp(Y_k)}{1 + \exp(Y_k)} \qquad (5.87)$$

The natural parameters of the multinomial are the multinomial logits from section 3.2.4.

From 3.47 $\theta_i = \ln\left(\frac{p_i}{1 - \sum_j p_j} \right)$

From 3.48 $p_i = \frac{\exp(\theta_i)}{1 + \sum_j \exp(\theta_j)}$ (5.88)

The likelihood of the sample, from 3.45, is

$$L = \text{constant} \quad \left(\prod_k^c p_k^{z} \right) (1 - P_c)^{(n-S_c)}$$

and from 3.46,

$$\ln L = \text{constant} + \left(\sum_k^c z_k \ln(p_k) \right) + (n-S_c)\ln(1 - P_c) \qquad (5.89)$$

The analysis required, in the light of the introduction to this section, is linear on the χ scale. The natural parameters, $\underset{\sim}{\theta}$, are not linear functions of this scale but the Y parameters of equation 5.87 are. The link function (relating $\underset{\sim}{\theta}$ to $\underset{\sim}{Y}$) is not linear but this causes little difficulty if the analysis is performed using $\underset{\sim}{S}$ and $\underset{\sim}{P}$ rather than $\underset{\sim}{z}$ and $\underset{\sim}{p}$.

Equation 3.55 gives $E[\ z_i\ ] = np_i$ (5.90)

and equation 3.56 gives $\mathrm{var}(\underset{\sim}{z}) = n\ (\mathrm{diag}(\underset{\sim}{p}) - \underset{\sim}{p}\underset{\sim}{p}')$ (5.91)

Define $\underset{\sim}{F}$, a matrix of order c with zeroes above the diagonal and ones on and below the diagonal.

Then $\underset{\sim}{S} = \{\ S_k\ \} = \underset{\sim}{F}\underset{\sim}{z}$

and $E[\ \underset{\sim}{S}\ ] = E[\ \underset{\sim}{F}\underset{\sim}{z}\ ] = \underset{\sim}{F}\ E[\ \underset{\sim}{z}\ ] = n\ \underset{\sim}{F}\ \underset{\sim}{p} = n\underset{\sim}{P}$ (5.92)

using 5.85, 5.86 and 5.90.

Similarly, using 5.91,

$$\begin{aligned}\underset{\sim}{V} = \mathrm{var}(\ \underset{\sim}{S}\ ) &= \mathrm{var}(\ \underset{\sim}{F}\underset{\sim}{z}\ ) \\ &= \underset{\sim}{F}\ \mathrm{var}(\underset{\sim}{z})\ \underset{\sim}{F}' && (5.93) \\ &= n\ \underset{\sim}{F}\ \mathrm{diag}\{p_i\}\ \underset{\sim}{F}' - n\ \underset{\sim}{F}\underset{\sim}{p}\underset{\sim}{p}'\underset{\sim}{F}' \\ &= n \begin{bmatrix} P_1 & P_1 & P_1 & \cdots \\ P_1 & P_2 & P_2 & \cdots \\ P_1 & P_2 & P_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} - n\ \underset{\sim}{P}\underset{\sim}{P}' \\ &= n\ \{P_i(1-P_j)\} \quad \text{for } i <= j && (5.94)\end{aligned}$$

5.2.2.1. Fixed effects generalized linear model

To apply the GLM procedure, assume N independent samples. For a fixed effects model, the linear predictor is

$\underset{\sim}{Y} = (\underset{\sim}{X} \otimes \underset{\sim}{I})\ \underset{\sim}{B}$ (5.95)

From 5.92 $\underset{\sim}{\mu} = \{\ n_j\ P_{jk} \}$ (5.96)

From 5.94 $\underset{\sim}{V}_j = \{\ n_j\ P_{jk}(1-P_{jl})\ \}$ for $k <= l$ (5.97)

and $\underset{\sim}{V} = \sum_{j}^{N}{}^{+}\ \underset{\sim}{V}_j$ since the samples are independent. (5.98)

By analogy with 3.42, the working variable is

$$\underset{\sim}{y} = \underset{\sim}{\hat{Y}} + \underset{\sim}{D}^{-1} (\underset{\sim}{S} - \underset{\sim}{\mu}) \qquad (5.99)$$

where $\underset{\sim}{D} = \{ \frac{\partial \mu_{jk}}{\partial Y_{jl}} \}$

$$= \text{diag} \{ n_j P_{jk}(1-P_{jk}) \} \qquad (5.100)$$

Following 3.42, the equations to be solved are

$$(\underset{\sim}{X} \otimes \underset{\sim}{I})' \underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D} (\underset{\sim}{X} \otimes \underset{\sim}{I}) \underset{\sim}{B} = (\underset{\sim}{X} \otimes \underset{\sim}{I}) \underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D} \underset{\sim}{y} \qquad (5.101)$$

where $\underset{\sim}{V}^{-1} = \Sigma^{+}_{j} \underset{\sim}{V}_j^{-1}$

and $\underset{\sim}{V}_j^{-1}$ is a tri-diagonal matrix with

$$v^{ii} = \frac{1}{n} (\frac{1}{p_i} + \frac{1}{p_{i+1}}) \qquad (5.102)$$

$$\text{and } v^{i,i+1} = v^{i+1,i} = - \frac{1}{n} \frac{1}{p_{i+1}} \qquad (5.103)$$

Proof:

From 5.93 $\underset{\sim}{V} = n \underset{\sim}{F} (\text{diag}\{p_i\} - \underset{\sim}{p}\underset{\sim}{p}') \underset{\sim}{F}'$

$$\text{so } \underset{\sim}{V}^{-1} = \frac{1}{n} (\underset{\sim}{F}')^{-1} (\text{diag}\{p_i\} - \underset{\sim}{p}\underset{\sim}{p}')^{-1} \underset{\sim}{F}^{-1} \qquad (5.104)$$

Now $\underset{\sim}{F}^{-1}$ is zero except for ones on the diagonal and minus ones on the sub-diagonal. (5.105)

$$\text{and } (\text{diag}\{p_i\} - \underset{\sim}{p}\underset{\sim}{p}')^{-1} = (\text{diag}\{ \frac{1}{p_i} \}) + \frac{1}{1-P_i} \underset{\sim}{1}\underset{\sim}{1}' \qquad (5.106)$$

Substituting 5.105 and 5.106 in 5.104 gives the matrix in 5.102 and 5.103.

This result was given by Clayton(1974) but his paper neglected the minus signs in 5.103.

QED

It is not immediately obvious that 5.101 is the maximum likelihood solution since differentiating 5.89 does not appear to give ($\underset{\sim}{S} - \underset{\sim}{\mu}$). However,

$$\left\{ \frac{\partial \ln L}{\partial Y_k} \right\} = \left(\frac{z_k}{p_k} - \frac{z_{k+1}}{p_{k+1}} \right) P_k(1-P_k) \}$$

$$= \underset{\sim}{D}\, \underset{\sim}{V}^{-1} (\underset{\sim}{S} - \underset{\sim}{\mu}) \qquad (5.107)$$

and

$$E\left[\left\{ \frac{\partial^2 \ln L}{\partial Y_k \;\partial Y_l} \right\} \right] = - \underset{\sim}{D}\, \underset{\sim}{V}^{-1}\, \underset{\sim}{D} \qquad (5.108)$$

When the RHS of equation 5.107 is premultiplied by $(\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D})^{-1}$, the result is $\underset{\sim}{D}^{-1}(\underset{\sim}{S}-\underset{\sim}{\mu})$ which is the second term in the RHS of equation 5.99. $\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}$ is the weight matrix required and so equation 5.101 gives the maximum likelihood solution.

5.2.2.2. Mixed model threshold analysis by joint-maximization method

The results of the previous section extend directly to this model except that the linear predictor has the form

$$\underset{\sim}{Y} = (\underset{\sim}{X} \otimes \underset{\sim}{I})\, \underset{\sim}{B} + (\underset{\sim}{Z} \otimes \underset{\sim}{1})\, \underset{\sim}{u} \qquad (5.109)$$

The 1 vector associated with $\underset{\sim}{Z}$ and $\underset{\sim}{u}$ in 5.109 occurs because under the threshold model of a single underlying variable, the same random effect affects each of the thresholds associated with the particular observation vector.

The equations to be solved are, from 4.18

$$\begin{bmatrix} (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}(\underset{\sim}{X}\otimes\underset{\sim}{I}) & (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}(\underset{\sim}{Z}\otimes\underset{\sim}{1}) \\ (\underset{\sim}{Z}\otimes\underset{\sim}{1})'\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}(\underset{\sim}{X}\otimes\underset{\sim}{I}) & (\underset{\sim}{Z}\otimes\underset{\sim}{1})'\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}(\underset{\sim}{Z}\otimes\underset{\sim}{1})+\underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{B}} \\ \hat{\underset{\sim}{u}} \end{bmatrix} = \begin{bmatrix} (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}\underset{\sim}{y} \\ (\underset{\sim}{Z}\otimes\underset{\sim}{1})'\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}\underset{\sim}{y} \end{bmatrix} \qquad (5.110)$$

where $\underset{\sim}{V}$ of 4.18 is $\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}$ (defined in 5.98),

$\underset{\sim}{D}$ of 4.18 is $\underset{\sim}{I}$,

$\underset{\sim}{G}$ is $\sigma_u^2 \underset{\sim}{R}$ where $\underset{\sim}{R}$ is the matrix of additive genetic relationship and σ_u^2 is $\mathrm{var}(u_i)$ and

$\underset{\sim}{y}$ of 4.18 is 5.99.

Solution of 5.110 utilizes 5.87, 5.96, 5.97, 5.99 and 5.100.

5.2.2.3. Mixed model threshold analysis by maximization-expectation method

The requirement for this method is to find satisfactory expressions for $E[\, \underset{\sim}{S} \,]$ and $\mathrm{var}[\, \underset{\sim}{S} \,]$ when the log likelihood is

$$\ln L = K + \sum_{i}^{q} \sum_{j}^{N_i} \Big(\sum_{k}^{c} z_{ijk} \ln(\dot{p}_{ijk}) \Big) + (n_{ij} - S_{ijc}) \ln(1 - \dot{P}_{ijc}) + \ln(U(\underset{\sim}{u} : \underset{\sim}{G})) \qquad (5.111)$$

where $\dot{p}_{ijk}$ is $\dot{P}_{ijk} - \dot{P}_{ij(k-1)}$

$$\dot{P}_{ijk} = \frac{\exp(Y_{ijk} + u_i)}{1 + \exp(Y_{ijk} + u_i)} \qquad (5.112)$$

and $\mathrm{var}(u_i) = \sigma_u^2$

This implies the random effects are additive on the underlying scale (the scale of Y_{ijk}).

Extending the results from section 5.2.2 to take account of the extra term in the likelihood (compare 5.89 with 5.111),

From 5.92, $E[\, S_{ijk} \,] = E[\, n_{ij} \dot{P}_{ijk} \,] \qquad (5.113)$

From 5.94, taking expectations of both sides,

$$E[\, S_{ijk} S_{ijl} - n_{ij} \dot{P}_{ijk} n_{ij} \dot{P}_{ijl} \,] = E[\, n_{ij} \dot{P}_{ijk} (1 - \dot{P}_{ijl}) \,] \quad \text{for } k <= l \qquad (5.114)$$

These equations require expressions for $E[\, \dot{P}_{ijk}]$ and $E[\, \dot{P}_{ijk} \dot{P}_{lmn}]$. Either of the methods of sections 5.1.2.2 and 5.1.2.3 could be used. The latter of these is simpler and is used. It is based on the logistic distribution and the rationale is given in that section.

From 5.22,

$$\bar{P}_{ijk} = E[\dot{P}_{ijk}] = \frac{\exp(s\,Y_{ijk})}{1 + \exp(s\,Y_{ijk})} \qquad (5.115)$$

where the scale factor, from 5.37, is

$$s = \sqrt{1-\rho} \qquad (5.116)$$

and $\rho = \dfrac{\sigma_u^2}{\pi^2/3 + \sigma_u^2}$ is the intraclass correlation. (5.117)

The second result that is required is an expression for

$$E[\dot{P}_{ijk}\dot{P}_{lmn}] = E\left[\frac{\exp(Y_{ijk} + u_i)}{1 + \exp(Y_{ijk} + u_i)}\ \frac{\exp(Y_{lmn} + u_l)}{1 + \exp(Y_{lmn} + u_l)}\right]$$

By definition, under the preceding assumptions,

$$\begin{aligned} \mathrm{Cov}(x_{ijk}+u_i, x_{lmn}+u_l) &= r_{il}\,\sigma_u^2 \\ &= r_{il}\,\frac{\rho}{1-\rho}\,\pi^2/3 \qquad (5.118) \end{aligned}$$

where ρ is the intraclass correlation from 5.117 and
r_{il} is the correlation between u_i and u_l.

On the basis of 4.28,

$$\mathrm{Cov}(\dot{P}_{ijk}, \dot{P}_{lmn}) = \frac{d\,\bar{P}_{ijk}}{d\,Y_{ijk}}\ \mathrm{Cov}\ (x_{ijk}+u_i, x_{lmn}+u_l)\ \frac{d\,\bar{P}_{lmn}}{d\,Y_{lmn}} \qquad (5.119)$$

Differentiating 5.115

$$\frac{d\,\bar{P}_{ijk}}{d\,Y_{ijk}} = \frac{s\,\exp(s\,Y_{ijk})}{(1 + \exp(s\,Y_{ijk}))^2} = s\,\bar{P}_{ijk}(1-\bar{P}_{ijk}) \qquad (5.120)$$

Substituting 5.118 and 5.120 in 5.119,

$$\begin{aligned} \mathrm{Cov}(\dot{P}_{ijk}, \dot{P}_{lmn}) &= s\,\bar{P}_{ijk}(1-\bar{P}_{ijk})\ r_{il}\,\frac{\rho}{1-\rho}\,\pi^2/3\ s\,\bar{P}_{lmn}(1-\bar{P}_{lmn}) \\ &= P_{ijk}(1-\bar{P}_{ijk})\ r_{il}\,\rho\,\pi^2/3\ \bar{P}_{lmn}(1-\bar{P}_{lmn}) \qquad (5.121) \end{aligned}$$

since, from 5.116, $s^2 = 1-\rho$

By the definition of covariance,

$$E[\dot{P}_{ijk}\dot{P}_{lmn}] = \mathrm{Cov}(\dot{P}_{ijk},\dot{P}_{lmn}) + E[\dot{P}_{ijk}]\ E[\dot{P}_{lmn}]$$

$$= \bar{P}_{ijk}(1-\bar{P}_{ijk})\ r_{il}\ \rho\ \pi^2/3\ \bar{P}_{lmn}(1-\bar{P}_{lmn}) + \bar{P}_{ijk}\bar{P}_{lmn} \tag{5.122}$$

Substituting 5.115 and 5.122 in 5.113 and 5.114 gives

$$\underset{\sim}{\mu} = \{\ n_{ij}\bar{P}_{ijk}\ \} = \frac{n_{ij}\ \exp(s\ Y_{ijk})}{1 + \exp(s\ Y_{ijk})} \tag{5.123}$$

$$\mathrm{cov}(S_{ijk},S_{lmn}) = \mathrm{cov}(n_{ij}\dot{P}_{ijk},n_{lm}\dot{P}_{lmn})$$

$$= n_{ij}\bar{P}_{ijk}(1-\bar{P}_{ijk})\ r_{il}\ \rho\ \pi^2/3\ n_{lm}\bar{P}_{lmn}(1-\bar{P}_{lmn}) \tag{5.124}$$

$$\text{for } ij \neq lm$$

$$\mathrm{cov}(S_{ijk},\ S_{ijn}) = E[n_{ij}\dot{P}_{ijk}(1-\dot{P}_{ijn})] + \mathrm{cov}[n_{ij}\dot{P}_{ijk},\ n_{ij}\dot{P}_{ijn}]$$

$$= n_{ij}\bar{P}_{ijk} - n_{ij}(\mathrm{cov}(\dot{P}_{ijk},\ \dot{P}_{ijn}) + \bar{P}_{ijk}\bar{P}_{ijn})$$

$$+\ n_{ij}^2\mathrm{cov}(\dot{P}_{ijk},\ \dot{P}_{ijn})$$

$$= n_{ij}\bar{P}_{ijk}(1-\bar{P}_{ijn})$$

$$+\ n_{ij}(n_{ij}-1)\bar{P}_{ijk}(1-\bar{P}_{ijk})\bar{P}_{ijn}(1-\bar{P}_{ijn})\ \rho\ \pi^2/3 \tag{5.125}$$

$$\text{for } k <= n$$

These expressions can be represented in matrix notation to conform to 4.29 as follows.

$$\underset{\sim}{E}_{ij} = \{\ n_{ij}\bar{P}_{ijk}(1-\bar{P}_{ijn})(1 - \bar{P}_{ijn}(1-\bar{P}_{ijk})\ \rho\ \pi^2/3\ )\ \}$$

$$= \underset{\sim}{Q}_{ij} - \underset{\sim}{A}_{ij}\underset{\sim}{1}\ \lambda\ \underset{\sim}{1}'\underset{\sim}{A}_{ij} \tag{5.126}$$

where $\lambda = \rho\ \pi^2/3$

$$\underset{\sim}{Q}_{ij} = n_{ij}\bar{P}_{ijk}(1-\bar{P}_{ijn}) \text{ for } k <= n \tag{5.127}$$

$$\underset{\sim}{A}_{ij} = \mathrm{diag}\ \{\ n_{ij}\bar{P}_{ijk}(1-\bar{P}_{ijk})\ \} \tag{5.128}$$

Then $\underset{\sim}{V} = \mathrm{var}(\underset{\sim}{z}) = \underset{\sim}{E} + \underset{\sim}{A}\underset{\sim}{Z}\underset{\sim}{G}\underset{\sim}{Z}'\underset{\sim}{A}$ (5.129)

where $\underset{\sim}{E} = \Sigma^+_i\ \Sigma^+_j\ \underset{\sim}{E}_{ij}$

$\underset{\sim}{A} = \Sigma^+_i\ \Sigma^+_j\ \underset{\sim}{A}_{ij}$

$\underset{\sim}{G} = \lambda\ \underset{\sim}{R}$ where $\underset{\sim}{R}$ is a matrix of additive relationships among the random elements in $\underset{\sim}{u}$.

$$\underset{\sim}{D} \text{ of } 4.24 = s\,\underset{\sim}{A} \qquad (5.130)$$

$$\begin{aligned}\underset{\sim}{y} \text{ of } 4.24 &= \hat{\underset{\sim}{Y}} + \underset{\sim}{D}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu}) \\ &= \hat{\underset{\sim}{Y}} + \underset{\sim}{A}^{-1}(\underset{\sim}{S}-\underset{\sim}{\mu})/s \qquad (5.131)\end{aligned}$$

Then 4.30 becomes, after dividing by s,

$$\begin{bmatrix} (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}(\underset{\sim}{X}\otimes\underset{\sim}{I}) & (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}(\underset{\sim}{Z}\otimes\underset{\sim}{1}) \\ (\underset{\sim}{Z}\otimes\underset{\sim}{1})'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}(\underset{\sim}{X}\otimes\underset{\sim}{I}) & (\underset{\sim}{Z}\otimes\underset{\sim}{1})'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}(\underset{\sim}{Z}\otimes\underset{\sim}{1})+\underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \underset{\sim}{B}^{o} \\ \underset{\sim}{u}^{o} \end{bmatrix} = \begin{bmatrix} (\underset{\sim}{X}\otimes\underset{\sim}{I})'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{y} \\ (\underset{\sim}{Z}\otimes\underset{\sim}{1})'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{y} \end{bmatrix} \qquad (5.132)$$

The weight matrix is block diagonal because $\underset{\sim}{E}$ is block diagonal and $\underset{\sim}{A}$ is diagonal. Considering the ij th block, and omitting those subscipts,

$$\begin{aligned}\underset{\sim}{W} &= \underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A} \\ &= \underset{\sim}{A}\,(\,\underset{\sim}{Q} - \underset{\sim}{A}\,\underset{\sim}{1}\,\lambda\,\underset{\sim}{1}'\underset{\sim}{A}\,)^{-1}\,\underset{\sim}{A} && \text{using } 5.126 \\ &= \underset{\sim}{A}\,(\underset{\sim}{Q}^{-1} - \underset{\sim}{Q}^{-1}\underset{\sim}{A}\,\underset{\sim}{1}(\underset{\sim}{1}'\underset{\sim}{A}\underset{\sim}{Q}^{-1}\underset{\sim}{A}\underset{\sim}{1} - \tfrac{1}{\lambda})^{-1}\underset{\sim}{1}'\underset{\sim}{A}\,\underset{\sim}{Q}^{-1})\,\underset{\sim}{A} && \text{using } 3.6 \\ &= \underset{\sim}{A}\underset{\sim}{Q}^{-1}\underset{\sim}{A} - \underset{\sim}{A}\underset{\sim}{Q}^{-1}\underset{\sim}{A}\,\underset{\sim}{1}(\underset{\sim}{1}'\underset{\sim}{A}\underset{\sim}{Q}^{-1}\underset{\sim}{A}\underset{\sim}{1} - \tfrac{1}{\lambda})^{-1}\underset{\sim}{1}'\underset{\sim}{A}\underset{\sim}{Q}^{-1}\underset{\sim}{A} && (5.133)\end{aligned}$$

where $\underset{\sim}{Q}$ is of the form of $\underset{\sim}{V}$ in 5.97 and

$\underset{\sim}{Q}^{-1}$ is therefore available from 5.102 and 5.103.

The starting rule suggested with section 3.2.5 may be used. Once 5.132 has been solved, the new solution for $\underset{\sim}{B}^{o}$ is used in the linear predictor, $\hat{\underset{\sim}{Y}} = (\underset{\sim}{X} \otimes \underset{\sim}{I})\underset{\sim}{B}^{o}$ (from 4.21), to set up and solve 5.132 again. The process is repeated until changes in $\underset{\sim}{B}^{o}$ between iterations are sufficiently small. As noted previously, the solution for $\underset{\sim}{u}^{o}$ is not required during iteration and it will often be convenient to absorb these equations when solving for $\underset{\sim}{B}^{o}$.

It should be recognized that $(\underset{\sim}{X} \otimes \underset{\sim}{I})$ is only one of several possible structures for the fixed effects.

Chapter 6. Application of the logistic linear mixed model to animal breeding

The two topics covered in this chapter are the simultaneous analysis of categorical and continuous traits by the same mixed model and the estimation of breeding values and genetic variances. The discussion is only in relation to the logistic linear mixed model which was derived in chapter 5 by use of the maximization-expectation method of section 4.2.

6.1. The simultaneous analysis of categorical and continuous characters by the same mixed model

For categorical traits, the mixed model equations have the form, from equation 4.27,

$$\begin{bmatrix} \mathbf{X}'_d\mathbf{D}_d\mathbf{E}_d^{-1}\mathbf{D}_d\mathbf{X}_d & \mathbf{X}'_d\mathbf{D}_d\mathbf{E}_d^{-1}\mathbf{A}_d\mathbf{Z}_d \\ \mathbf{Z}'_d\mathbf{A}_d\mathbf{E}_d^{-1}\mathbf{D}_d\mathbf{X}_d & \mathbf{Z}'_d\mathbf{A}_d\mathbf{E}_d^{-1}\mathbf{A}_d\mathbf{Z}_d+\mathbf{G}_d^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}_d^o \\ \mathbf{u}_d^o \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_d\mathbf{D}_d\mathbf{E}_d^{-1}\mathbf{D}_d\mathbf{y}_d \\ \mathbf{Z}'_d\mathbf{A}_d\mathbf{E}_d^{-1}\mathbf{D}_d\mathbf{y}_d \end{bmatrix} \quad (6.1)$$

It is convenient to assume the normal distribution for continuous traits. The relevant mixed model equations are, from equation 4.31,

$$\begin{bmatrix} \mathbf{X}'_c\,\mathbf{Q}^{-1}\,\mathbf{X}_c & \mathbf{X}'_c\,\mathbf{Q}^{-1}\,\mathbf{Z}_c \\ \mathbf{Z}'_c\,\mathbf{Q}^{-1}\,\mathbf{X}_c & \mathbf{Z}'_c\,\mathbf{Q}^{-1}\,\mathbf{Z}_c+\mathbf{G}_c^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}_c^o \\ \mathbf{u}_c^o \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_c\,\mathbf{Q}^{-1}\,\mathbf{y}_c \\ \mathbf{Z}'_c\,\mathbf{Q}^{-1}\,\mathbf{y}_c \end{bmatrix} \quad (6.2)$$

The simultaneous solution of these equations is obtained by solving

$$\begin{bmatrix} \overset{''}{\mathbf{X}}{}'\mathbf{D}\mathbf{E}^{-1}\mathbf{D}\overset{''}{\mathbf{X}} & \overset{''}{\mathbf{X}}{}'\mathbf{D}\mathbf{E}^{-1}\mathbf{A}\overset{''}{\mathbf{Z}} \\ \overset{''}{\mathbf{Z}}{}'\mathbf{A}\mathbf{E}^{-1}\mathbf{D}\overset{''}{\mathbf{X}} & \overset{''}{\mathbf{Z}}{}'\mathbf{A}\mathbf{E}^{-1}\mathbf{A}\overset{''}{\mathbf{Z}}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}^o \\ \mathbf{u}^o \end{bmatrix} = \begin{bmatrix} \overset{''}{\mathbf{X}}{}'\mathbf{D}\mathbf{E}^{-1}\mathbf{D}\mathbf{y} \\ \overset{''}{\mathbf{Z}}{}'\mathbf{A}\mathbf{E}^{-1}\mathbf{D}\mathbf{y} \end{bmatrix} \quad (6.3)$$

where $\overset{''}{\mathbf{X}} = (\mathbf{X} \otimes \mathbf{I}_t)$ given $\mathbf{X}_d = (\mathbf{X} \otimes \mathbf{I}_d)$,

$\mathbf{X}_c = (\mathbf{X} \otimes \mathbf{I}_c)$,

$t = d + c$,

$$\overset{''}{\underset{\sim}{Z}} = (\underset{\sim}{Z} \otimes \underset{\sim}{H}_t) \text{ given } \underset{\sim}{Z}_d = (\underset{\sim}{Z} \otimes \underset{\sim}{H}_d),$$

$$\underset{\sim}{Z}_c = (\underset{\sim}{Z} \otimes \underset{\sim}{I}_c),$$

$$\underset{\sim}{H}_d = \underset{\sim}{1}_d \text{ or } \underset{\sim}{I}_d,$$

$$\underset{\sim}{H}_t = \begin{bmatrix} \underset{\sim}{H}_d & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{I}_c \end{bmatrix},$$

$$\underset{\sim}{D}_d = \underset{k}{\Sigma^{+}} \underset{\sim}{D}_k \text{ and } \underset{\sim}{D} = \underset{k}{\Sigma^{+}} \begin{bmatrix} \underset{\sim}{D}_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{I}_c \end{bmatrix}$$

$$\underset{\sim}{A}_d = \underset{k}{\Sigma^{+}} \underset{\sim}{A}_k \text{ and } \underset{\sim}{A} = \underset{k}{\Sigma^{+}} \begin{bmatrix} \underset{\sim}{A}_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{I}_c \end{bmatrix},$$

$$\underset{\sim}{E}_d = \underset{k}{\Sigma^{+}} \underset{\sim}{E}_k \text{ and } \underset{\sim}{E} = \underset{k}{\Sigma^{+}} \begin{bmatrix} \underset{\sim}{E}_k & \underset{\sim}{C}_k \\ \underset{\sim}{C}' & \underset{\sim}{Q}_k \end{bmatrix},$$

$$\underset{\sim}{Q} = \underset{k}{\Sigma^{+}} \underset{\sim}{Q}_k,$$

$$\underset{\sim}{G} = (\underset{\sim}{R} \otimes \underset{\sim}{\Sigma}) \text{ given } \underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Sigma}_d & \underset{\sim}{\Sigma}_{dc} \\ \underset{\sim}{\Sigma}'_{dc} & \underset{\sim}{\Sigma}_c \end{bmatrix},$$

$$\underset{\sim}{G}_d = (\underset{\sim}{R} \otimes \underset{\sim}{\Sigma}_d),$$

$$\underset{\sim}{G}_c = (\underset{\sim}{R} \otimes \underset{\sim}{\Sigma}_c),$$

$\underset{\sim}{R}$ is a matrix of additive genetic relationships,

$\underset{\sim}{B}^o$ contains $\underset{\sim}{B}^o_d$ and $\underset{\sim}{B}^o_c$ arranged as the columns of $\overset{''}{\underset{\sim}{X}}$,

$\underset{\sim}{u}^o$ contains $\underset{\sim}{u}^o_d$ and $\underset{\sim}{u}^o_c$ arranged as the columns of $\overset{''}{\underset{\sim}{Z}}$,

$\underset{\sim}{y}$ contains $\underset{\sim}{y}_d$ and $\underset{\sim}{y}_c$ arranged as the rows of $\overset{''}{\underset{\sim}{X}}$.

The two quantities included in 6.3 which do not occur in 6.1 or 6.2 are $\underset{\sim}{C}_k$ and $\underset{\sim}{\Sigma}_{dc}$. It is assumed that $\underset{\sim}{\Sigma}_{dc}$ is known as it refers to the covariances between the random effects in the categorical trait and the continuous trait. A method of estimating $\underset{\sim}{\Sigma}_{dc}$, indeed all of $\underset{\sim}{\Sigma}$, is considered in section 6.2.

6.1.1. Covariance between a binomial character and a normal character

Using a probit transformation to relate a binomial character, z, to an underlying standard normal character, x, gives the following.

Let z be 0 if $x<t$, and
z be 1 otherwise,
where $p = 1 - \Phi(t)$ and t is the threshold.

$$\begin{aligned}\mathrm{cov}(x,z) &= E[\ xz\ ] - E[\ x\ ]\,E[\ z\ ]\\ &= E[\ xz\ ] \text{ since } E[\ x\ ] \text{ is zero.}\\ &= \int_t^\infty x\,\phi(x)\,dx\\ &= \phi(t), \text{ the ordinate at } t. \qquad (6.4)\end{aligned}$$

The covariance of z with another normal variable v having variance σ_v^2 is

$$\mathrm{cov}(v,z) = c\,\sigma_v^2\,\phi(t) \qquad (6.5)$$

where $c\,\sigma_v^2$ is the cov(v,x). (6.6)

The procedures developed in this thesis have consistently used the logistic distribution as a convenient approximation to the normal distribution. If x is distributed as a standard logistic distribution, using equation 3.11,

$$\begin{aligned}\mathrm{cov}(z,x) &= \int_t^\infty x\,\frac{\exp(x)}{(1+\exp(x))^2}\,dx\\ &= \int_t^\infty x\left(\frac{d}{dx}\,\frac{\exp(x)}{1+\exp(x)}\right)dx\\ &= \left[\ x\,\frac{\exp(x)}{1+\exp(x)}\ \right]_t^\infty - \int_t^\infty \frac{\exp(x)}{1+\exp(x)}\,dx\\ &= \left[\ x\,\frac{\exp(x)}{1+\exp(x)} - \ln(1+\exp(x))\ \right]_t^\infty\end{aligned}$$

As $x \twoheadrightarrow \infty$ $\quad x\,\dfrac{\exp(x)}{1+\exp(x)} \dashrightarrow x$ and

$$\ln(1+\exp(x)) \dashrightarrow x.$$

$$\text{Therefore, } \operatorname{cov}(z,x) = \ln(1 + \exp(t)) - \frac{t\exp(t)}{1 + \exp(t)}$$

$$= -\ln(1-p) - p\ln\left(\frac{p}{1-p}\right)$$

$$= -p\ln(p) - (1-p)\ln(1-p) \qquad (6.7)$$

where, by 3.12, $p = \dfrac{\exp(t)}{1 + \exp(t)}$

$$\text{Thus, } \operatorname{cov}(v,z) = \frac{-\left(p\ln(p) + (1-p)\ln(1-p)\right) c\,\sigma_v^2}{\sqrt{\pi^2/3}} \qquad (6.8)$$

Equation 6.8 makes allowance for the change in variance between a standard normal and standard logistic but ignores the small other differences. It applies strictly to the case of a fixed effects model. In mixed models, it will be necessary to obtain the expected value of 6.8 with regard to the random factor. i.e. in a mixed model,

$$\operatorname{cov}(v,z) = E\left[\frac{-\left(\dot{p}\ln(\dot{p}) + (1-\dot{p})\ln(1-\dot{p})\right) c\,\sigma_v^2}{\sqrt{\pi^2/3}}\right] \qquad (6.9)$$

$$\text{where, from 5.16, } \dot{p} = \frac{\exp(Y_{ij} + u_i)}{1 + \exp(Y_{ij} + u_i)} \qquad (6.10)$$

The expectation in 6.9 is not easily obtained however and the following approximation is proposed. The diagonal elements in matrix $\underset{\sim}{E}_k$ (of 6.3) are terms like, from 5.47,

$$e_{ii} = E[\, n\dot{p}_{ij}(1-\dot{p}_{ij})\,] = n\bar{p}_{ij}(1-\bar{p}_{ij})(1 - f(\bar{p}_{ij},\rho)) \qquad (6.11)$$

The proposal is that cov(v,z) be evaluated by equation 6.8 using the same $\bar{p}_{ij}$ as used in 6.11, and then scaled by the square root of $(1 - f(\bar{p}_{ij},\rho))$. i.e.

$$\operatorname{cov}(v,z) = -\left(\bar{p}_{ij}\ln(\bar{p}_{ij}) + (1-\bar{p}_{ij})\ln(1-\bar{p}_{ij})\right) F \qquad (6.12)$$

$$\text{where } F = c\sqrt{\frac{\sigma_v^2(1 - f(\bar{p}_{ij},\rho)}{\pi^2/3}}.$$

Expression 6.12 applies directly to extremal and threshold characters and provides the equations to fill in the $\underset{\sim}{C}$ matrix of 6.3.

6.2. Estimation of variances

In the context of animal breeding, particular concern rests with the estimation of breeding values and variances. In terms of the general models developed in chapter 5, the breeding values are the random elements, $\utilde{u}_k$; the variances are var($\utilde{u}_k$). The purpose of this section is to show how to estimate the random effects given the fixed effects and how to estimate var($\utilde{u}_k$) = $\utilde{\Sigma}$. Equations 6.3 are solved iteratively, as discussed in section 4.2, to provide solutions for $\utilde{B}^o$. The solutions for $\utilde{u}^o$ are not required during the iteration and would usually not be computed until the final solution for $\utilde{B}^o$ was obtained. The random effects may be obtained by back-solution from the final fixed effects. Consider 6.3 rewritten as

$$\begin{bmatrix} \utilde{K} & \utilde{L} \\ \utilde{L}' & \utilde{M} \end{bmatrix} \begin{bmatrix} \utilde{B}^o \\ \utilde{u}^o \end{bmatrix} = \begin{bmatrix} \utilde{v} \\ \utilde{w} \end{bmatrix} \qquad (6.13)$$

The inverse of a partitioned matrix is

$$\begin{bmatrix} \utilde{K} & \utilde{L} \\ \utilde{L}' & \utilde{M} \end{bmatrix}^{-1} = \begin{bmatrix} \utilde{Q} & -\utilde{Q}\utilde{L}\utilde{M}^{-1} \\ -\utilde{M}^{-1}\utilde{L}'\utilde{Q} & \utilde{M}^{-1}+\utilde{M}^{-1}\utilde{L}'\utilde{Q}\utilde{L}\utilde{M}^{-1} \end{bmatrix} \qquad (6.14)$$

$$= \begin{bmatrix} \utilde{K}^{-1}+\utilde{K}^{-1}\utilde{L}\utilde{T}\utilde{L}'\utilde{K} & -\utilde{K}^{-1}\utilde{L}\utilde{T} \\ -\utilde{T}\utilde{L}'\utilde{K}^{-1} & \utilde{T} \end{bmatrix} \qquad (6.15)$$

where $\utilde{Q} = (\utilde{K} - \utilde{L}\utilde{M}^{-1}\utilde{L}')^{-1}$ and

$$\utilde{T} = (\utilde{M} - \utilde{L}'\utilde{K}^{-1}\utilde{L})^{-1}.$$

So $$\begin{bmatrix} \utilde{B}^o \\ \utilde{u}^o \end{bmatrix} = \begin{bmatrix} \utilde{Q}\utilde{v} & -\utilde{Q}\utilde{L}\utilde{M}^{-1}\utilde{w} \\ -\utilde{M}^{-1}\utilde{L}'\utilde{Q}\utilde{v} & +\utilde{M}^{-1}\utilde{w}+\utilde{M}^{-1}\utilde{L}'\utilde{Q}\utilde{L}\utilde{M}^{-1}\utilde{w} \end{bmatrix}$$

$$= \begin{pmatrix} \underset{\sim}{Q}(\underset{\sim}{v}-\underset{\sim}{L}\underset{\sim}{M}^{-1}\underset{\sim}{w}) \\ \underset{\sim}{M}^{-1}(\underset{\sim}{w}-\underset{\sim}{L}\underset{\sim}{B}^{o}) \end{pmatrix} \qquad (6.16)$$

The solution to 6.3 is therefore

$$\underset{\sim}{u}^{o} = \underset{\sim}{M}^{-1}(\underset{\sim}{\ddot{Z}}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{y} - \underset{\sim}{\ddot{Z}}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{\ddot{X}}\underset{\sim}{B}^{o})$$

$$= \underset{\sim}{M}^{-1}\underset{\sim}{\ddot{Z}}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{D}(\underset{\sim}{y} - \underset{\sim}{\ddot{X}}\underset{\sim}{B}^{o})$$

$$= \underset{\sim}{M}^{-1}\underset{\sim}{\ddot{Z}}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{r} \qquad (6.17)$$

where $\underset{\sim}{r} = \underset{\sim}{D}(\underset{\sim}{y} - \underset{\sim}{\ddot{X}}\underset{\sim}{B}^{o})$

$= \underset{\sim}{D}(\underset{\sim}{\ddot{X}}\underset{\sim}{B}^{o} + \underset{\sim}{D}^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu}) - \underset{\sim}{\ddot{X}}\underset{\sim}{B}^{o})$ using 4.24

$= \underset{\sim}{z}-\underset{\sim}{\mu}$

The exact nature of $\underset{\sim}{\ddot{Z}}$, $\underset{\sim}{A}$, $\underset{\sim}{E}$, $\underset{\sim}{\ddot{X}}$, $\underset{\sim}{D}$, $\underset{\sim}{y}$ and $\underset{\sim}{G}$ will depend on the particular problem being analysed.

One method of estimating the variance, $\underset{\sim}{\Sigma}$, is to equate $\underset{\sim}{\hat{u}}'\underset{\sim}{\hat{u}}$ to $E[\ \underset{\sim}{\hat{u}}'\underset{\sim}{\hat{u}}\ ]$. However, equations 6.3 are multiple trait and $\underset{\sim}{u}^{o}$ is made up of several subvectors. These subvectors can be written as columns of a matrix $\underset{\sim}{\hat{U}}$. The relationship between $\underset{\sim}{u}^{o}$ and $\underset{\sim}{\hat{U}}$ is given by $\underset{\sim}{u}^{o} = \mathrm{vec}(\underset{\sim}{\hat{U}}')$. The variance of a row of $\underset{\sim}{\hat{U}}$ was given in 6.3 as $\underset{\sim}{\Sigma}$. Let $\underset{\sim}{S}_i$ be a matrix which selects those elements in $\underset{\sim}{u}^{o}$ which comprise the i th column of $\underset{\sim}{\hat{U}}$. Then $\underset{\sim}{\hat{u}}_i = \underset{\sim}{S}_i\underset{\sim}{u}^{o}$ = i th column of $\underset{\sim}{\hat{U}}$.

$$= \underset{\sim}{S}_i\underset{\sim}{M}^{-1}\underset{\sim}{\ddot{Z}}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{r} \qquad (6.18)$$

Searle (1971 b) gives the expected value of a quadratic form as $E[\ \underset{\sim}{x}'\underset{\sim}{A}\underset{\sim}{x}\ ] = \mathrm{tr}(\underset{\sim}{A}\ \mathrm{var}(\underset{\sim}{x})) + \underset{\sim}{\mu}_x'\underset{\sim}{A}\underset{\sim}{\mu}_x$

where $\underset{\sim}{\mu}_x = E[\ \underset{\sim}{x}\ ]$

Using this result,

$$E[\ \underset{\sim}{\hat{u}}_i'\underset{\sim}{\hat{u}}_j\ ] = E[\ \underset{\sim}{r}'\underset{\sim}{E}^{-1}\underset{\sim}{A}\underset{\sim}{\ddot{Z}}\underset{\sim}{M}^{-1}\underset{\sim}{S}_i'\underset{\sim}{S}_j\underset{\sim}{M}^{-1}\underset{\sim}{\ddot{Z}}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{r}\ ]$$

$$= \mathrm{tr}(\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{M}^{-1}\underset{\sim}{S}'\underset{\sim}{S}\ \underset{\sim}{M}^{-1}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\ \mathrm{var}(\underset{\sim}{r})) \qquad (6.19)$$

since $E[\ \underset{\sim}{r}\ ] = 0$

Now $\mathrm{var}(\underset{\sim}{r}) = \mathrm{var}(\underset{\sim}{z}) + \mathrm{var}(\underset{\sim}{\mu}) - 2\ \mathrm{cov}(\underset{\sim}{z},\underset{\sim}{\mu})$

$$= \mathrm{var}(\underset{\sim}{z}) - \mathrm{var}(\underset{\sim}{\mu})$$

From 4.26, $\mathrm{var}(\underset{\sim}{z}) = \underset{\sim}{E} + \underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{G}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A} = \underset{\sim}{V}$ (6.20)

and from 3.6 $\underset{\sim}{V}^{-1} = \underset{\sim}{E}^{-1} - \underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}(\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}} + \underset{\sim}{G}^{-1})^{-1}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}$

$$= \underset{\sim}{E}^{-1} - \underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{M}^{-1}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1} \qquad (6.21)$$

Also $\mathrm{var}(\underset{\sim}{B}^{o}) = (\overset{\prime\prime}{\underset{\sim}{X}}{}'\underset{\sim}{D}\underset{\sim}{V}^{-1}\underset{\sim}{D}\overset{\prime\prime}{\underset{\sim}{X}})^{-1}$

$$= (\overset{\prime\prime}{\underset{\sim}{X}}{}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\overset{\prime\prime}{\underset{\sim}{X}} - \overset{\prime\prime}{\underset{\sim}{X}}{}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{M}^{-1}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{D}\overset{\prime\prime}{\underset{\sim}{X}})^{-1}$$

$$= (\underset{\sim}{K} - \underset{\sim}{L}\underset{\sim}{M}^{-1}\underset{\sim}{L}')^{-1} \text{ from 6.13} \qquad (6.22)$$

So $\mathrm{var}(\underset{\sim}{\mu}) = \underset{\sim}{D}\overset{\prime\prime}{\underset{\sim}{X}}(\underset{\sim}{K} - \underset{\sim}{L}\underset{\sim}{M}^{-1}\underset{\sim}{L}')^{-1}\overset{\prime\prime}{\underset{\sim}{X}}{}'\underset{\sim}{D}$ (6.23)

Combining equations 6.20 and 6.23,

$$\mathrm{var}(\underset{\sim}{r}) = \underset{\sim}{V} - \underset{\sim}{D}\overset{\prime\prime}{\underset{\sim}{X}}(\underset{\sim}{K} - \underset{\sim}{L}\underset{\sim}{M}^{-1}\underset{\sim}{L}')^{-1}\overset{\prime\prime}{\underset{\sim}{X}}{}'\underset{\sim}{D}$$

$$E[\underset{\sim}{u}_i'\underset{\sim}{u}_j] = \mathrm{tr}(\ \underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{M}^{-1}\underset{\sim}{S}_i'\underset{\sim}{S}_j\underset{\sim}{M}^{-1}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}(\underset{\sim}{V} - \underset{\sim}{D}\overset{\prime\prime}{\underset{\sim}{X}}(\underset{\sim}{K} - \underset{\sim}{L}\underset{\sim}{M}^{-1}\underset{\sim}{L}')^{-1}\overset{\prime\prime}{\underset{\sim}{X}}{}'\underset{\sim}{D}))$$

$$= \mathrm{tr}(\ \underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{M}^{-1}\underset{\sim}{S}_i'\underset{\sim}{S}_j\underset{\sim}{M}^{-1}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}(\underset{\sim}{E} + \underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{G}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}$$

$$- \underset{\sim}{D}\overset{\prime\prime}{\underset{\sim}{X}}(\underset{\sim}{K} - \underset{\sim}{L}\underset{\sim}{M}^{-1}\underset{\sim}{L}')^{-1}\overset{\prime\prime}{\underset{\sim}{X}}{}'\underset{\sim}{D}))$$

$$= \mathrm{tr}(\ \underset{\sim}{S}_i'\underset{\sim}{S}_j\underset{\sim}{M}^{-1}((\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}} + \overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{G}\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}})\underset{\sim}{M}^{-1}$$

$$- \overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{D}\overset{\prime\prime}{\underset{\sim}{X}}(\underset{\sim}{K} - \underset{\sim}{L}\underset{\sim}{M}^{-1}\underset{\sim}{L}')^{-1}\overset{\prime\prime}{\underset{\sim}{X}}{}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{M}^{-1})\ )$$

$$= \mathrm{tr}(\ \underset{\sim}{S}_i'\underset{\sim}{S}_j\underset{\sim}{M}^{-1}(\overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}}\underset{\sim}{G}(\underset{\sim}{G}^{-1} + \overset{\prime\prime}{\underset{\sim}{Z}}{}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{A}\overset{\prime\prime}{\underset{\sim}{Z}})\underset{\sim}{M}^{-1}$$

$$- \underset{\sim}{L}'(\underset{\sim}{K} - \underset{\sim}{L}\underset{\sim}{M}^{-1}\underset{\sim}{L}')^{-1}\underset{\sim}{L}\underset{\sim}{M}^{-1})\ )$$

$$\begin{aligned}
&= \mathrm{tr}(\ \mathbf{S}_i'\mathbf{S}_j\mathbf{M}^{-1}(\ddot{\mathbf{Z}}'\mathbf{A}\mathbf{E}^{-1}\mathbf{A}\ddot{\mathbf{Z}}\mathbf{G} - \mathbf{L}'(\mathbf{K} - \mathbf{L}\mathbf{M}^{-1}\mathbf{L}')^{-1}\mathbf{L}\mathbf{M}^{-1})\ ) \\
&= \mathrm{tr}(\ \mathbf{S}_i'\mathbf{S}_j\mathbf{M}^{-1}((\mathbf{M}-\mathbf{G}^{-1})\ \mathbf{G} + \mathbf{L}'(\mathbf{L}\mathbf{M}^{-1}\mathbf{L}' - \mathbf{K})^{-1}\mathbf{L}\mathbf{M}^{-1})\ ) \\
&= \mathrm{tr}(\ \mathbf{S}_i'\mathbf{S}_j\mathbf{G} - \mathbf{S}_i'\mathbf{S}_j(\mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{L}'(\mathbf{L}\mathbf{M}^{-1}\mathbf{L}' - \mathbf{K})^{-1}\mathbf{L}\mathbf{M}^{-1})\ ) \\
&= \mathrm{tr}(\ \mathbf{S}_i'\mathbf{S}_j\mathbf{G} - \mathbf{S}_i'\mathbf{S}_j(\mathbf{M} - \mathbf{L}'\mathbf{K}^{-1}\mathbf{L})^{-1}\ ) \\
&= \mathrm{tr}(\ \mathbf{S}_i'\mathbf{S}_j\mathbf{G} - \mathbf{S}_i'\mathbf{S}_j(\ddot{\mathbf{Z}}'\mathbf{A}\mathbf{E}^{-1}\mathbf{A}\ddot{\mathbf{Z}}+\mathbf{G}^{-1} \\
&\qquad - (\ddot{\mathbf{X}}'\mathbf{D}\mathbf{E}^{-1}\mathbf{A}\ddot{\mathbf{Z}})'(\ddot{\mathbf{X}}'\mathbf{D}\mathbf{E}^{-1}\mathbf{D}\ddot{\mathbf{X}})^{-1}(\ddot{\mathbf{X}}'\mathbf{D}\mathbf{E}^{-1}\mathbf{A}\ddot{\mathbf{Z}}))^{-1}\ ) \\
&= \mathrm{tr}(\ \mathbf{S}_j\mathbf{G}\mathbf{S}_i'\ ) - \mathrm{tr}(\ \mathbf{S}_j\mathbf{T}\mathbf{S}_i'\ ) \qquad (6.24)
\end{aligned}$$

where $\mathbf{T}$ is defined in 6.15 and is analagous to $\mathbf{T}$ of 3.7.

From 6.3, $\mathbf{G} = \mathbf{R} \times \boldsymbol{\Sigma}$.

Consideration of the nature of $\mathbf{S}_i$ and $\mathbf{S}_j$ reveals that

$\mathbf{S}_i\mathbf{G}\mathbf{S}_j' = \sigma_{ij}\ \mathbf{R}$ and

$\mathbf{S}_i\mathbf{T}\mathbf{S}_j' = \mathbf{T}_{ij}$ which contains only those rows of $\mathbf{T}$ relating to trait i and those columns of $\mathbf{T}$ relating to trait j.

Equating $\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j$ to its expectation gives

$$\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j = \sigma_{ij}\mathrm{tr}(\mathbf{R}) - \mathrm{tr}(\mathbf{T}_{ij}) \qquad (6.25)$$

which may be rearranged to give

$$\hat{\sigma}_{ij} = \frac{\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j + \mathrm{tr}(\mathbf{T}_{ij})}{\mathrm{tr}(\mathbf{R})} \qquad (6.26)$$

This estimator is analagous to the restricted maximum likelihood estimator if it is iterated. The best method of iterating for estimates of $\boldsymbol{\Sigma}$ has not been determined. The method used to produce results in subsequent chapters is discussed in section 6.3.2.

6.3. Implementation of the logistic linear mixed model in a generalized linear models programme

The analyses reported in subsequent chapters were performed using REG (Gilmour, 1983). REG is a generalized linear models programme which has been developed progressively since 1970 within the Biometrical Branch of the New South Wales Department of Agriculture. It has been expanded at Massey University to include some mixed model options and the logistic linear mixed model (LLMM) as defined in sections 5.1.2.3, 5.2.1.2 and 5.2.2.3. The implementation of these equations assumes an identity relationship matrix and absorbs the random effects during the formation of the weighted matrices. Several aspects of the programme including how the LLMM has been implemented are discussed in the following sections.

6.3.1. Measurement of 'lack of fit'

REG computes the deviance (equation 3.44) as a measure of lack of fit of the model. The deviance under the binomial distribution is

$$\text{dev} = -2\ \Sigma\ (\ z\ \ln(\ \frac{z}{np}\ ) + (n-z)\ \ln(\ \frac{n-z}{n(1-p)}\ )\ ) \qquad (6.27)$$

where p is the maximum likelihood solution.
Subscripts have been omitted but the summation is over the total sample.
The difference in deviances obtained under different models is a chi-square variable which may be used to test the difference between the models.

When extra-binomial variation is present, equation 6.27 is no longer the proper expression for the deviance since it ignores the extra variation. The likelihood expression given in equation 3.15 contains two parts. Equation 6.27 ignores the second part completely and by so doing creates an inflated estimate of deviance. REG, however, currently uses equation 6.27 to define the deviance whether or not extra-binomial variation is present. The proper expression for the deviance has not been worked out because it requires the full specification of the second part of the likelihood and this specification has not been required for the development of the logistic linear mixed model. Differences between deviances calculated by equation 6.27 should still be approximate χ^2 variables even though the deviances themselves are not when extra-binomial variation is present. An extra statistic provided by REG is the quantity labelled 'Deviance after random effects' in tables 8.9 and 8.10. This is the deviance obtained from 6.27 when the random effects are included in the linear model along with the fixed effects. The major difficulty with the statistic is that it is not obvious how many degrees of freedom are associated with it. The number of degrees of freedom is not simply the degrees of freedom remaining after fitting random and fixed effects because the random effects have been constrained by the use of the random variance in the model.

6.3.2. Estimation of variance components

REG can set up and solve mixed model equations of moderate order for continuous characters. Variance components may be estimated by Henderson's (1953) method 3 or by the REML equations 3.8 and 3.9. The solutions to these equations are not automatically iterated and obtaining a REML solution must be by trial and error.

For categorical traits, REG does not use the full REML-type equations given in 6.26 as the matrix $\underset{\sim}{T}$ is potentially large. Two approximations have been used. When there is only a single random factor (a threshold trait), the matrix $\underset{\sim}{X}$ has been replaced by the vector $\underset{\sim}{1}$ to give

$$\begin{aligned}\underset{\sim}{T}^{(1)} &= (\underset{\sim}{M} - \underset{\sim}{w}\underset{\sim}{w}'/d)^{-1} \\ &= \underset{\sim}{M}^{-1} + \underset{\sim}{M}^{-1}\underset{\sim}{w}(1/d - \underset{\sim}{w}'\underset{\sim}{M}^{-1}\underset{\sim}{w})^{-1}\underset{\sim}{w}'\underset{\sim}{M}^{-1} \qquad (6.28)\end{aligned}$$

where $\underset{\sim}{w} = \underset{\sim}{Z}'\underset{\sim}{A}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{1}$ and $d = \underset{\sim}{1}'\underset{\sim}{D}\underset{\sim}{E}^{-1}\underset{\sim}{D}\underset{\sim}{1}$.

So $\mathrm{tr}(\underset{\sim}{T}^{(1)}) = \mathrm{tr}(\underset{\sim}{M}^{-1}) + \mathrm{tr}(\ \underset{\sim}{w}'\underset{\sim}{M}^{-1}\underset{\sim}{M}^{-1}\underset{\sim}{w}\ )/(\ 1/d - \underset{\sim}{w}'\underset{\sim}{M}^{-1}\underset{\sim}{w}\ )$

For a single trait and an identity relationship matrix $\underset{\sim}{R}$, matrices $\underset{\sim}{M}$, $\underset{\sim}{A}$, $\underset{\sim}{D}$ and $\underset{\sim}{E}$ are diagonal and $\mathrm{tr}(\underset{\sim}{T}^{(1)})$ may be computed without actually forming $\underset{\sim}{T}^{(1)}$. For multiple trait situations and extremal characters, REG ignores the last term in equation 6.28 and computes $\mathrm{tr}(\underset{\sim}{T}^{(2)})$ where $\underset{\sim}{T}^{(2)} = \underset{\sim}{M}^{-1}$. This is analagous to ML-type estimators (Harville, 1977). The effect of both these approximations is to bias the estimates down, the bias being greater for the latter estimator.

Which ever form of equation 6.26 is used, some method of iteration is required. The simplest iterative procedure is the EM algorithm (Dempster, Laird and Rubin, 1977). This procedure requires cycling between an Expectation step represented by equation 6.26, and a Maximization step represented by equations 6.3. The EM algorithm is guaranteed to find a maximum, though not necessarily the global maximum, for an expression. It is however notoriously slow, especially in the case of several unknowns. Several methods of speeding up the algorithm are discussed by Dempster et al. (1977). The method used for this thesis has been to extrapolate and interpolate from previous iterations to obtain new trial values rather than just using the most recent solution as the new trial value. Alternatives to the EM algorithm would include the Newton-Raphson and Fisher scoring procedures which probably have forms analagous to those for normal characters (see Harville, 1977). These have not been considered in this study.

The application of the EM algorithm to an extremal trait is very slow to converge because of the number of parameters involved. The speeding up process used in chapter 9 is described here for a trait with three categories - dry-or-not (D), single-or-not (S) and twins-or-not (T). Specification of D and S implies T and specification of var(D), var(S) and cov(D,S) implies

$$\mathrm{var}(T) = \mathrm{var}(D) + \mathrm{var}(S) + 2\mathrm{cov}(D,S),$$
$$\mathrm{cov}(D,T) = (\mathrm{var}(S) - \mathrm{var}(D) - \mathrm{var}(T))/2 \text{ and}$$
$$\mathrm{cov}(S,T) = (\mathrm{var}(D) - \mathrm{var}(S) - \mathrm{var}(T))/2.$$

The estimation procedure used was again trial and error but required many more tries before reasonable estimates were obtained. Table 9.11 displays some of the solutions and demonstrates that the values included in table 9.9 are reasonable.

The problem of solving for several parameters simultaneously, encountered with the extremal model, is greater in the multiple trait situation because there is also an environmental covariance to consider. The procedure adopted to obtain the solutions presented in tables 9.12 and 9.13 involved three steps.
1) Set the environmental covariance to zero, use the variance estimates obtained from single trait equations and iterate towards an estimate of the genetic covariance.
2) Using the genetic correlation obtained from step one, iterate for revised estimates of the variances, if deemed necessary.
3) Repeat steps 1 and 2 but using alternative values for the environmental covariance.
A real danger exists of specifying error correlations which result in an error matrix which is not positive definite. The computing procedures used in REG include tests for this condition.

Chapter 7. Performance of the logistic linear mixed model in simulation studies.

This chapter investigates some characteristics of the logistic linear mixed model (LLMM) for binomial characters defined in sections 5.1.2.2 and 5.1.2.3. Topics considered are the validity of approximations used in the derivations and the precision of the estimates of breeding values and variances.

7.1. The behaviour of estimates of mean and variance under the logistic linear mixed model

In the derivation of $E[\ \dot{p}\ ]$ and $E[\ \dot{p}^2\ ]$ in section 5.1.2.2, an assumption was made that the random elements, u_k, were small enough that powers of u_k greater than 2 could be ignored. The validity of this assumption is examined by comparing the resulting expressions with those obtained when only powers of u_k greater than 4 are ignored and with expressions from section 5.1.2.3. The expressions retaining the fourth power are derived in Appendix C. From equation C.51,

$$E[\ \dot{p}\ ] = p + \sigma^2 p(1-p)(\tfrac{1}{2} - p)[1 + \sigma^2\{\tfrac{1}{4} - 3p(1-p)\}] \qquad (7.1)$$

where $\dot{p} = \dfrac{\exp(Y + u)}{1 + \exp(Y + u)}$,

$p = \dfrac{\exp(Y)}{1 + \exp(Y)}$,

$\sigma^2 = \mathrm{var}(u)$ and

$Y = \ln(\frac{p}{1-p})$.

Ignoring the σ^4 term, from C.52,

$$E[\ \dot{p}\ ] = p + \sigma^2 p(1-p)(\tfrac{1}{2} - p) \qquad (7.2)$$

Also, from equation C.81,

$$E[\ \dot{p}^2\ ] = p^2 + \sigma^2 p(1-p)(2p - 3p^2) +$$

$$\sigma^4 p(1-p)p[-\ \frac{p}{4} + (2 - 12p + 15p^2)(1-p)] \tag{7.3}$$

Therefore

$$E[\ \dot{p}(1-\dot{p})\ ] = E[\ \dot{p}\ ] - E[\ \dot{p}^2\ ]$$

$$= p - p^2 + \sigma^2 p(1-p)\{\frac{1}{2} - p - 2p + 3p^2 +$$

$$\sigma^2[(\frac{1}{2} - p)\{\frac{1}{4} - 3p(1-p)\} + p\{\frac{p}{4} - (2 - 12p + 15p^2)(1-p)\}]\}$$

$$= p(1-p) + \sigma^2 p(1-p)[\frac{1}{2} - 3p(1-p)] +$$

$$\sigma^4 p(1-p)[\frac{1}{8} - p\{\frac{1}{4} - \frac{1}{4}p + (\frac{3}{2} - 3p + 2 - 12p + 15p^2)(1-p)\}]$$

$$= p(1-p) + \sigma^2 p(1-p)[\frac{1}{2} - 3p(1-p)] +$$

$$\sigma^4 p(1-p)[\frac{1}{8} - 15p(1-p)\{\frac{1}{4} - p(1-p)\}] \tag{7.4}$$

On ignoring the σ^4 term,

$$E[\ \dot{p}(1-\dot{p})\ ] = p(1-p) + \sigma^2 p(1-p)[\frac{1}{2} - 3p(1-p)] \tag{7.5}$$

The equations from section 5.1.2.3 are,

$$\text{from 5.36,}\ \bar{p} = E[\ \dot{p}\ ] = \frac{\exp(sY)}{1 + \exp(sY)} \tag{7.6}$$

$$\text{where}\quad s = \sqrt{1-\rho}\quad \text{and}$$

$$\rho = \frac{\sigma^2}{\sigma^2 + \pi^2/3}$$

$$\text{From 5.43}\quad E[\ \dot{p}^2\ ] = \bar{p}^2(1-\bar{p})^2\rho\pi^2/3 + \bar{p}^2 \tag{7.7}$$

Combining equations 7.6 and 7.7 gives

$$E[\ \dot{p}(1-\dot{p})\ ] = \bar{p}(1-\bar{p}) - \bar{p}^2(1-\bar{p})^2\rho\pi^2/3 \tag{7.8}$$

It is important to understand what these expressions represent. The logistic distribution is being used to map random sub-populations, with the same variance but different means, onto the real line. The sub-populations have a population mean of Y, relative to a fixed threshold, and individual means of $Y + u_i$. The distance of the i th mean from the threshold determines the size of the probability $\dot{p}_i$ using the logistic distribution (see section 3.2.1). The non-linear nature of the logit transformation means that if the distribution of u_i is symmetric, the distribution of p will be skewed towards 0.5. Thus E[$\dot{p}_i$], the average of all $\dot{p}_i$, will be between 0.5 and p, the probability associated with a sub-population whose mean is Y. The three equations, 7.1, 7.2 and 7.6, all behave in the manner just described for small values of σ^2. However, equations 7.1 and 7.2 break-down for large σ^2. When $\sigma^2 p(1-p)$ is greater than 1, equation 7.2 estimates E[$\dot{p}_i$] beyond 0.5, which is clearly wrong.

The average value of the sub-population sampling variances, $E[\dot{p}_i(1-\dot{p}_i)]$, can also be deduced in a similar way. When the population mean is far from the threshold, all $\dot{p}_i$ will be close to 0 (or 1) and the average variance will be larger for the same reason that E[$\dot{p}_i$] was closer to 0.5. However, when the population mean is at the threshold, all sub-populations will be away from the threshold and hence have smaller variances than the 'average' sub-population. Equations 7.4, 7.5 and 7.8 all show this pattern for small values of σ^2. For large values, equations 7.4 and 7.5 break-down. For instance, equation 7.5 becomes negative for $p = 0.5$ and σ^2 greater than 4, but $\dot{p}_i(1-\dot{p}_i)$ is always positive.

Equations 7.6 and 7.8 are valid for any value of σ^2 and for this reason are preferred. Equations 7.1, 7.2, 7.4 and 7.5 break down for large σ^2 because the basic assumption in their derivation was that σ^2 was small. Figure 7.1 compares equations 7.1, 7.2 and 7.6, at a range of probabilities for variances up to 1, plotted in the form

$$\frac{E[\ \dot{p}\ ] - p}{p}$$

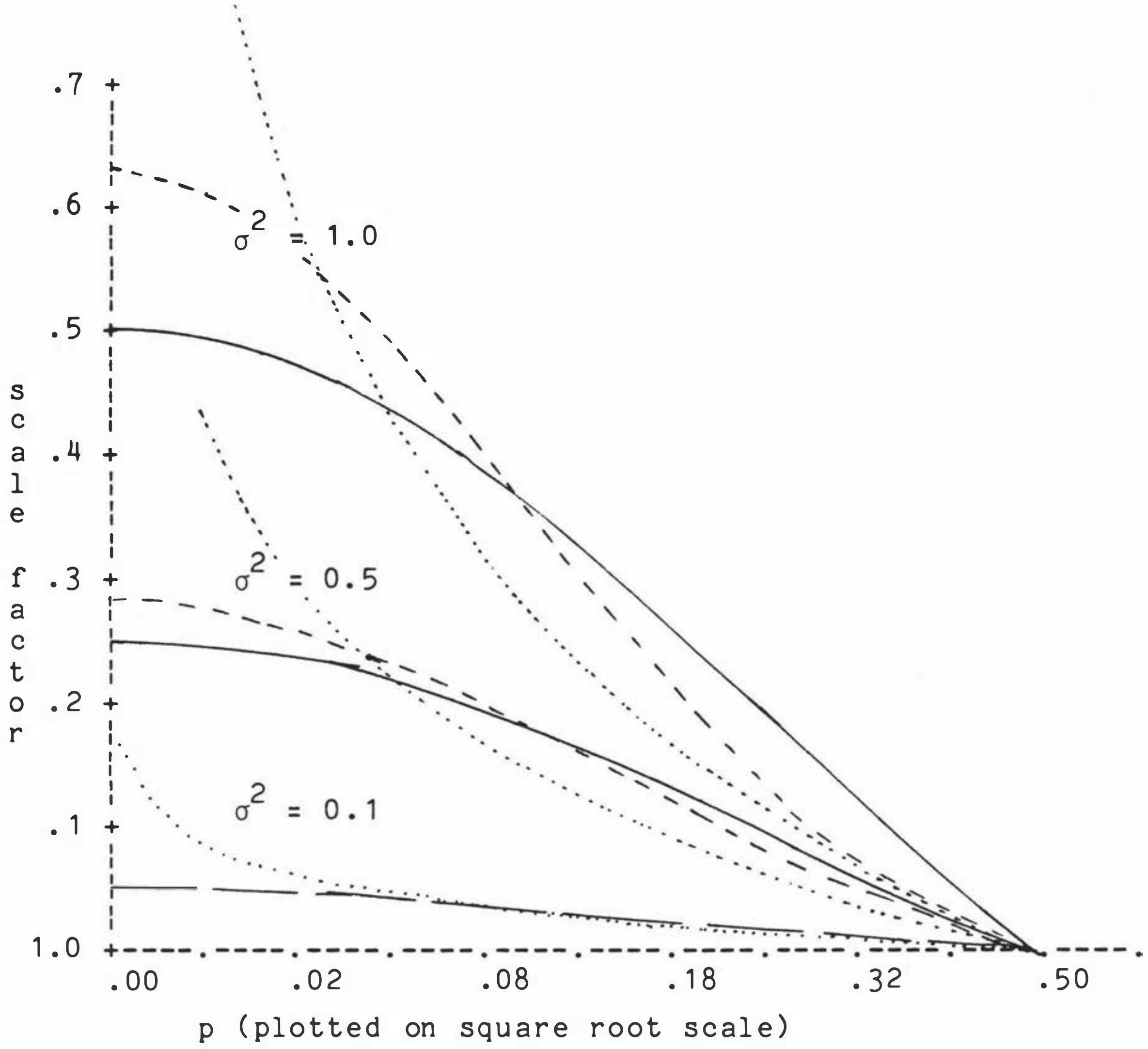


Figure 7.1 Comparison of the scale factors associated with the three expressions for $E[\dot{p}]$. The expressions are plotted in the form

$$\frac{E[\dot{p}] - p}{p}$$

from 7.1, - - - - $= \sigma^2(\frac{1}{2} - p)(1 + \sigma^2[\frac{1}{4} - 3p(1-p)])(1-p)$

from 7.2, ——— $= \sigma^2(\frac{1}{2} - p)(1-p)$

from 7.6, ········ $= \frac{\bar{p} - p}{p}$

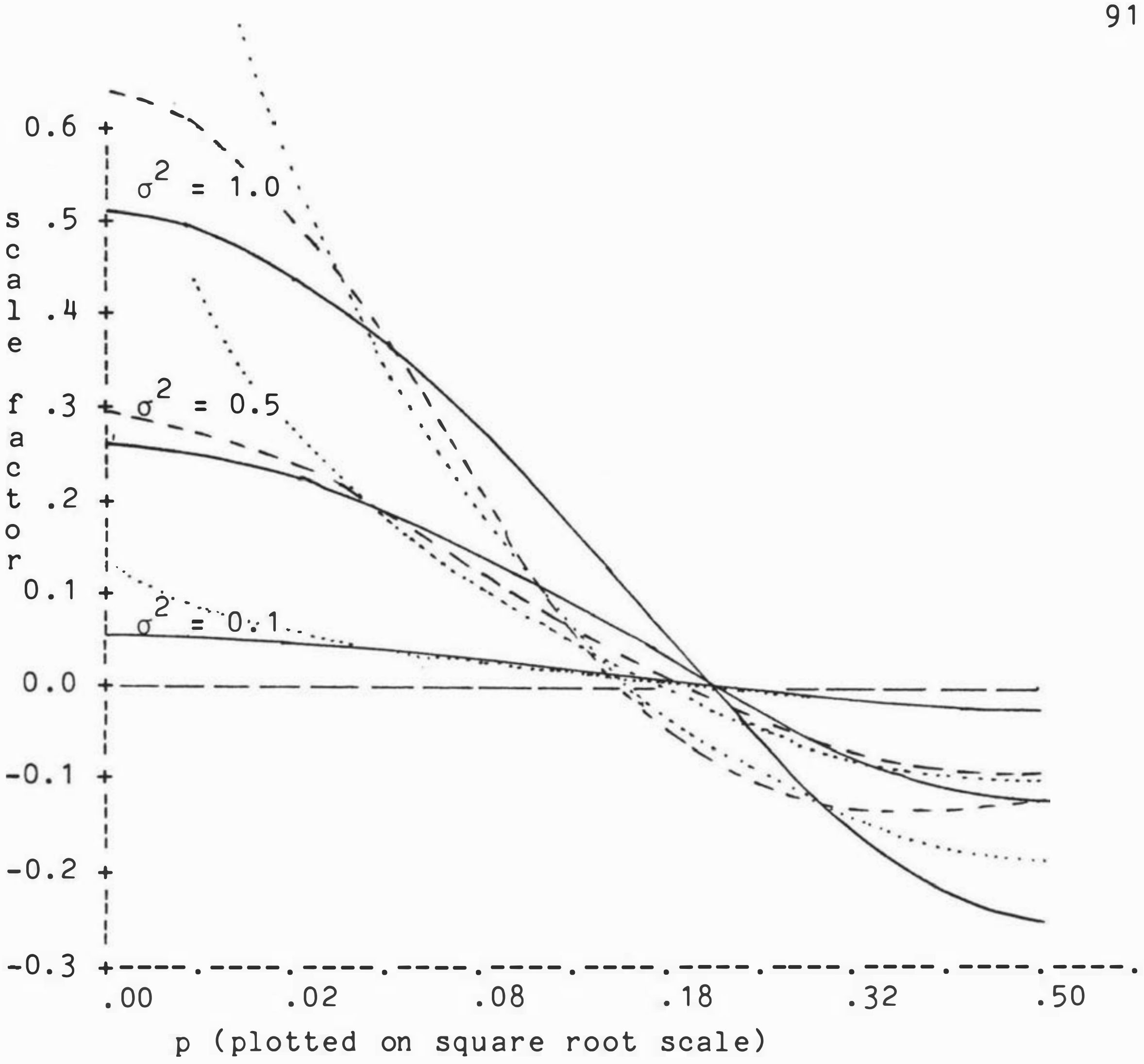


Figure 7.2 Comparison of the scale factors associated with the three expressions for $E[\ \dot{p}(1-\dot{p})\ ]$. The expressions are plotted in the form

$$\frac{E[\ \dot{p}(1-\dot{p})\ ] - p(1-p)}{p(1-p)}$$

from 7.5, -------- = $\sigma^2(\frac{1}{2} - 3p(1-p))$

from 7.4, - - - - = $\sigma^2[\frac{1}{2}-3p(1-p) + \sigma^2\{\frac{1}{8}-15p(1-p)(\frac{1}{4}-p(1-p))\}]$

from 7.8, ... = $\frac{\bar{p}(1-\bar{p}) - \bar{p}^2(1-\bar{p})^2\rho\pi^2/3}{p(1-p)} - 1$

Figure 7.2 compares equations 7.4, 7.5 and 7.8 in the form

$$\frac{E[\ \dot{p}(1-\dot{p})\ ] - p(1-p)}{p(1-p)}$$

Discussion of figures 7.1 and 7.2

Examination of figures 7.1 and 7.2 quickly establishes that for σ^2 less than 0.5, the expressions are nearly equivalent. Equation 7.1 under-estimates the adjustment by about 0.3 % when p is less than 0.1 and over-estimates it by about 3 % when p is between 0.4 and 0.5.

To evaluate the potential importance of these differences, the range of likely values of σ^2 must be anticipated. In half-sib analysis, a common animal breeding situation, the intraclass correlation should not exceed 0.25 and will often be less than 0.10. These represent values for σ^2 of 1.1 and 0.34 respectively. All of the expressions are satisfactory for problems of this type. In analyses where the intraclass correlation may exceed 0.25, as in analyses of repeatability, only equations 7.6 and 7.8 should be used for $E[\ \dot{p}\ ]$ and $E[\ \dot{p}(1-\dot{p})\ ]$ respectively.

7.2. The precision of estimates of the threshold and the intraclass correlation

The usefulness of the LLMM, while partly based on theoretical rationalization, also depends on precision. Robin Thompson (pers. comm.) found biases related to family size when the intraclass correlation was estimated after solving a set of equations for fixed and random effects. The equations he solved were similar to those derived by the joint-maximization method in section 5.1.1 and involved iterating both random and fixed effects in the model. This section reports a brief study of the influence of four factors on the estimates of the threshold and the intraclass correlation obtained using the LLMM as derived in section 5.1.2.3 by the maximization-expectation method.

The four factors considered were:

1) family sizes of 2, 3, 4, 6 and 8,
2) threshold values of 0, 0.85, 1.39, 2.20, 2.94, 3.48 and 4.60, corresponding to probabilities of 0.5, 0.7, 0.8, 0.90, 0.95, 0.97 and 0.99 respectively,
3) numbers of families of 100, 1000 and 10,000 and
4) intraclass correlations (ρ) of 0.8, 0.5, 0.3, 0.2, 0.1, 0.03, 0.01 and 0.003.

Method

A computer programme was written to generate binomial data with extra-binomial variation and then to obtain estimates of the parameters for comparison with the true values used to generate the data. The generation method was to form percentiles of a standard normal distribution and obtain the mean of each percentile. These 100 values were scaled by the square root of $\rho\pi^2/(3(1-\rho))$ and added to a threshold (from 2 above). Each modified threshold was converted to a probability by the inverse logit transformation. These probabilities were used to create binomial proportions for samples of sizes 2, 3, 4, 6 and 8. The 100 sets of proportions were added and multiplied by 1, 10 or 100 to give numbers of families of 100, 1000 and 10,000. The result of this process was a set of frequencies for each combination of the four factors. From each data set the threshold and intraclass correlations were estimated utilizing equations 7.6 and 7.8 for the expectations.

Results

The effect of an increase in the number of families was to reduce the variation in estimates of the intraclass correlation and of the fixed effects. Family size did not have any systematic influence on the estimates. This was true for all combinations of intraclass correlation and threshold. It is demonstrated for a particular combination of threshold and intraclass correlation in table 7.1. For the parameter values represented in table 7.1, both the thresholds and the intraclass correlations are under-estimated. The estimates of the threshold and correlation are positively correlated since the intraclass correlation governs the scale of measurement of the threshold.

Table 7.1 Effects of differing family size and numbers of families on estimates of the intraclass correlation and the threshold in binomial samples with extra-binomial variation.

Intraclass correlation estimates (true value 0.8).

Family sizes						Number of
2	3	4	6	8	average	families
.777	.775	.773	.782	.770	.775	100
.756	.755	.755	.756	.755	.755	1000
.753	.753	.753	.753	.753	.753	10,000

Estimates of the threshold (true value 0.85).

Family sizes						Number of
2	3	4	6	8	average	families
.690	.724	.736	.743	.702	.719	100
.678	.679	.681	.684	.686	.682	1,000
.676	.676	.676	.676	.677	.676	10,000

The value of the estimate of the intraclass correlation is affected by the location of the threshold (figure 7.3), and vice-versa, the estimate of the threshold is affected by the size of the intraclass correlation (figure 7.4). Figures 7.3 and 7.4 are based on 10,000 families of size 8 but are typical of other combinations. Several points arise from examination of figure 7.3.

1) For very low true intraclass correlations, the estimates will be low but resolution is not good. It was not possible to distinquish between intraclass correlations below 0.003.

2) For intraclass correlations below 0.1, the estimates were very close to the true values.

3) For intraclass correlations above 0.1, the estimates of the intraclass correlation only seem reasonable for threshold values less than 2.

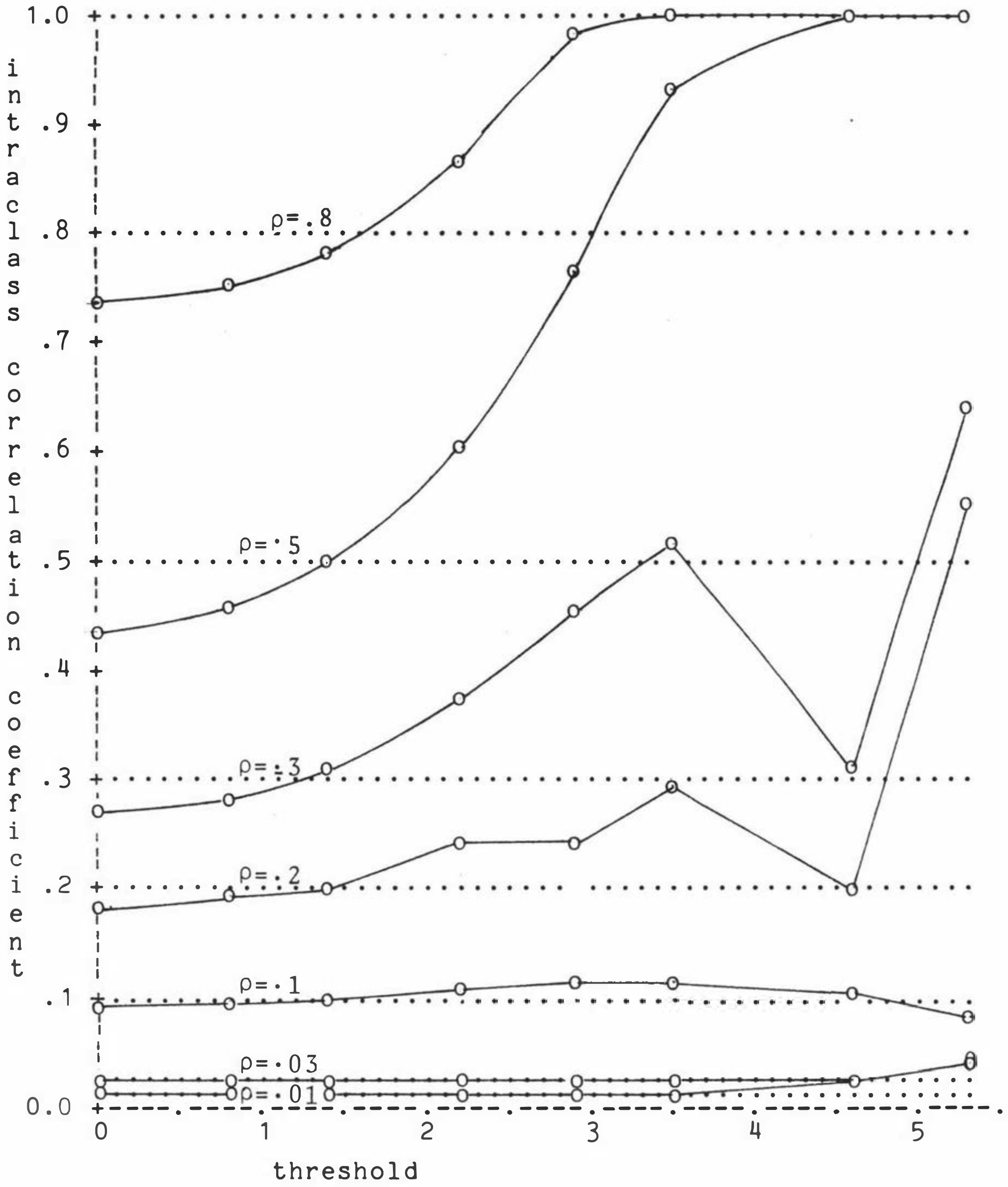


Figure 7.3 Relationship between actual and estimated values of the intraclass correlation coefficient for various threshold values.

o - estimates of the intraclass correlation

····· line of true intraclass correlation

—— line linking estimates of the intraclass correlation

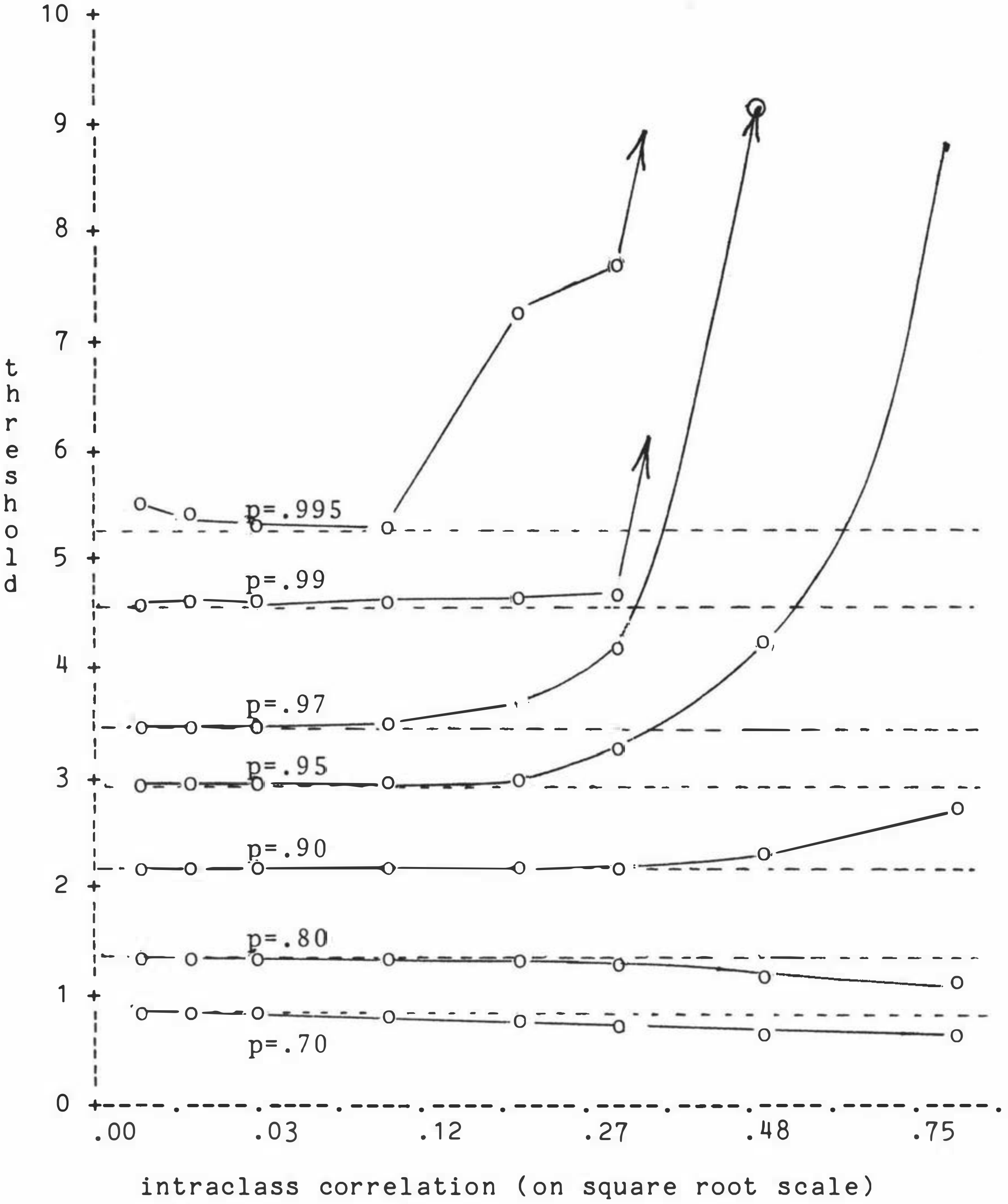


Figure 7.4 Relationship between actual and estimated values of the threshold for various intraclass correlations.

o - estimates of the threshold

---- line of true threshold values

—— line linking estimates of the threshold

Discussion

There are two possible partial explanations for the lack of perfect estimation shown in figures 7.3 and 7.4. First, weights based on average variances become less satisfactory as the range of true variances increases with increasing intraclass correlation. This problem is aggravated at extreme thresholds for then the distribution of true variances will be quite skewed. Second, the logistic is more heavily tailed than the normal distribution (see the comments on table 3.1 in section 3.2.1) and this will push extreme thresholds even further from the mean than would be the case for the normal distribution. This problem is aggravated at high intraclass correlations.

7.3. The estimation of breeding values under the logistic linear mixed model

The precision of the LLMM for estimating breeding values has been investigated by comparing it with an analagous BLUP analysis. The form of the LLMM used in the comparison was that of section 5.1.2.2; i.e. using equations 7.2 and 7.5 for the expectations. The estimates of random effects obtained from that model are compared with true values and with estimates obtained from the ordinary mixed model equations (equation 3.5).
Rewriting those equations,

$$\begin{bmatrix} \underset{\sim}{X}' \underset{\sim}{Q}^{-1} \underset{\sim}{X} & \underset{\sim}{X}' \underset{\sim}{Q}^{-1} \underset{\sim}{Z} \\ \underset{\sim}{Z}' \underset{\sim}{Q}^{-1} \underset{\sim}{X} & \underset{\sim}{Z}' \underset{\sim}{Q}^{-1} \underset{\sim}{Z} + \underset{\sim}{G}^{-1} \end{bmatrix} \begin{bmatrix} \underset{\sim}{B}_p \\ \underset{\sim}{u}_p \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}' \underset{\sim}{Q}^{-1} \underset{\sim}{z} \\ \underset{\sim}{Z}' \underset{\sim}{Q}^{-1} \underset{\sim}{z} \end{bmatrix} \qquad (7.9)$$

where $\underset{\sim}{Q} = \text{diag}\,\{n_{ij}\}$

Method

Binomial data were generated by computer for a simple model with known parameters and are discussed as if they relate to a trait measured on unrelated ewes mated to 100 unrelated rams in a flock. Names for factors in the model are incidental to the study and are used to make description easier.

A random sample of 100 normal (0, 0.1) random variables was drawn to represent the 100 random ram effects. To each of these was added five fixed effects giving a total of 500 threshold values. The fixed effects were taken in order from a recurring sequence of sixteen equidistant ordered values from -4.0 to 4.0. The thresholds were converted to probabilities using the inverse logit transformation (see equation 7.1). This whole procedure was repeated to form ten flocks. The actual means and variances of the ten samples of 100 random effects used to create the flocks, are displayed in table 7.2.

Table 7.2 Statistics for 10 samples of 100 N(0, .10) random variables used to define the 10 flocks.

	Mean	Variance (uncorrected)	Skewness (corrected)	Kurtosis (corrected)
Expected value	.0000	.1000	.0000	.0000
Flock 1	-.0173	.1152	-.0047	-.0014
Flock 2	.0050	.0783	-.0191	-.0211
Flock 3	-.0026	.1262	-.0180	-.0532
Flock 4	.0149	.1280	-.0001	-.0089
Flock 5	-.0020	.1062	.0053	-.0271
Flock 6	-.0047	.0989	.0098	.0029
Flock 7	-.0014	.0929	.0061	-.0492
Flock 8	.0096	.1016	.0258	.0250
Flock 9	.0107	.0860	.0065	-.0010
Flock 10	-.0097	.0833	-.0033	.0281
Standard Error	.0100		.0241	.0478

Binomial samples were then drawn for family sizes of n = 1, 2, 3, 4, 5, 25 and 100 ewes in each of the 500 combinations of fixed and random effects within each flock. In addition, n = 50 was used for five flocks and n=40, 60 was used for the other five flocks. Two sets of ten flocks with mixed family sizes were also formed. In these, family size was a uniform [0,5] and [0,25] random integer variable respectively. All sets of flocks with fixed family sizes used the same sequence of 500 probabilities because fixed and random effects were combined in the same order. However, the two sets with mixed family sizes used a different set of probabilities because of the way the programme dealt with null samples. Random effects associated with null samples were skipped but fixed effects were carried forward resulting in different combinations and hence different probabilities. The computer programme evaluated the binomial distribution fully for n less than 11. Otherwise it used a poisson or logistic approximation depending on the value of the mean.

A total of 105 sets of data representing 10 flocks and 12 family sizes were thus obtained. Each set was then analysed in four ways:

1 by the mixed model (equation 7.9) with $\underset{\sim}{G} = .1\ \underset{\sim}{I}$

2 by the logistic linear mixed model with $\underset{\sim}{G} = .5\ \underset{\sim}{I}$

3 by the logistic linear mixed model with $\underset{\sim}{G} = .1\ \underset{\sim}{I}$

4 by the logistic linear mixed model with $\underset{\sim}{G} = .02\ \underset{\sim}{I}$

Analysis 1 used an arbitrary coefficient in G. No attempt was made to obtain an optimal value. Analyses 2 and 4 were performed to see how sensitive the LLMM was to the use of a wrong coefficient in $\underset{\sim}{G}$. The fixed effects included in the model were the mean and a covariate which was the fixed effect used in obtaining the probabilities. i.e. The covariate had 16 possible values ranging from -4 to 4. The analyses were performed using REG (Gilmour, 1983, see section 6.3).

The analyses have been summarized by extracting the following ten statistics:

B(.50), B(.10) and B(.02), the regression coefficients for the covariate from analyses 2, 3 and 4, respectively,

G(.50), G(.10) and G(.02), the estimates of var($\underset{\sim}{u}$) from analyses 2, 3 and 4, respectively, Correlation of $\underset{\sim}{u}$ with analysis 3 solution for $\underset{\sim}{u}$, $\underset{\sim}{u}_x$,

Correlation of $\underset{\sim}{u}$ with analysis 1 solution for $\underset{\sim}{u}$, $\underset{\sim}{u}_p$,

Correlation of $\underset{\sim}{u}_x$ with $\underset{\sim}{u}_p$ and

Deviance from analysis 3.

The statistics G(.50), G(.10) and G(.02) are the estimates of var($\underset{\sim}{u}$) obtained after a single solution for fixed effects using prior estimates of .50, .10 and .02, respectively.

These ten statistics are plotted against family size in figures 7.5 to 7.14. They have also been analysed with a model testing linear, quadratic and cubic effects of log family size (LFS), flock and flock.LFS interaction. Where appropriate, the flock effects have been partitioned into 'due to ram variance differences between flocks' and 'due to other factors' by including as a covariate in the analysis, the variances from table 7.2.

Results

B(.50), B(.10) and B(.02) are plotted against family size in figures 7.5, 7.7 and 7.9 respectively. Family size has an obvious effect on variance of B in all three figures. The expected value of B is 1.00. The mean of B(.50) (analysis 2) was 1.08 showing a substantial positive bias resulting from using an inflated variance in the LLMM. Flock and LFS effects were not significant in predicting B(.50). The average value of B(.10) was 1.01. The flock.LFS interaction was nearly significant ($P<.10$) showing a tendency to distinguish the family characteristics as family size increases. The Mean of B(.02) was 1.00. The use of an underestimate of the variance of the random factor does not appear to bias the estimation of fixed effects. This may not hold if the fixed and random effects are not independent. The flock.LFS interaction had a probability of .1006 .

The variance estimate G(.50) in figure 7.6 was very dependent on the LFS. Linear, quadratic and cubic components are all significant ($P<.001$). The flock.LFS interaction was significant (P<.100) . Ninety-nine % of this flock effect was attributable to differences in the ram variances.

The variance estimate G(.10) in figure 7.8 had a nearly significant (P<.10) flock.LFS interaction of which 70 % was due to ram variance differences.

Family size had an obvious effect on the variance estimate G(.02) in figure 7.10. The flock.LFS effect was nearly significant (P<.10) and 44 % of this variation was associated with the differences in ram variances.

Figure 7.11 shows that LFS had a highly significant effect (P<.001) on the correlation between the random N(0,0.10) variables used to generate the data for the flock and the estimate of those variables obtained under the LLMM. Flock effects were also significant (P<.05) and 71 % of the flock variation was attributable to differences in ram variances.

The LFS effect is strongly quadratic in figure 7.12 highlighting the inefficiency of BLUP when applied to data for which the LLMM is more appropriate. The flock effect was significant (P<.05) but only 9 %, a non significant portion of it, was attributable to the differences in ram variances. In figure 7.13, the correlation between the BLUP and LLMM predictors declines with increasing LFS. The quadratic and cubic components were significant (P<.001) although only accounting for 3 % of the variation explained by LFS. Flock effects were not significant.

The deviance plotted in figure 7.14, includes all variation attributable to rams. It is only when family size exceeded 25 ewes per ram that this deviance exceeded its expected value of 499 (assuming the Chi-square distribution). The deviance values were not related to flock (P>.05) but were quadratically related to LFS (P<.001).

Discussion and Conclusions

For family sizes of less than 25 per random effect, the LLMM did little better than the ordinary mixed model equations at predicting the random effects - possibly because there was not enough precision in the data. The LLMM was better at estimating fixed effects than the ordinary mixed model equations. With the larger families, the advantage of using the LLMM increased, especially when there was a range of levels of fixed effects. The estimation of fixed effects will be biased if the wrong variance is used. This appears not to be critical for the ranking of random effects.

If the analyst is not confident of the variance estimate and has a substantial amount of data, a revised estimate may be obtained from the data using the iterative procedure described in section 6.3.2. There appears to be little advantage in accuracy beyond the first significant digit.

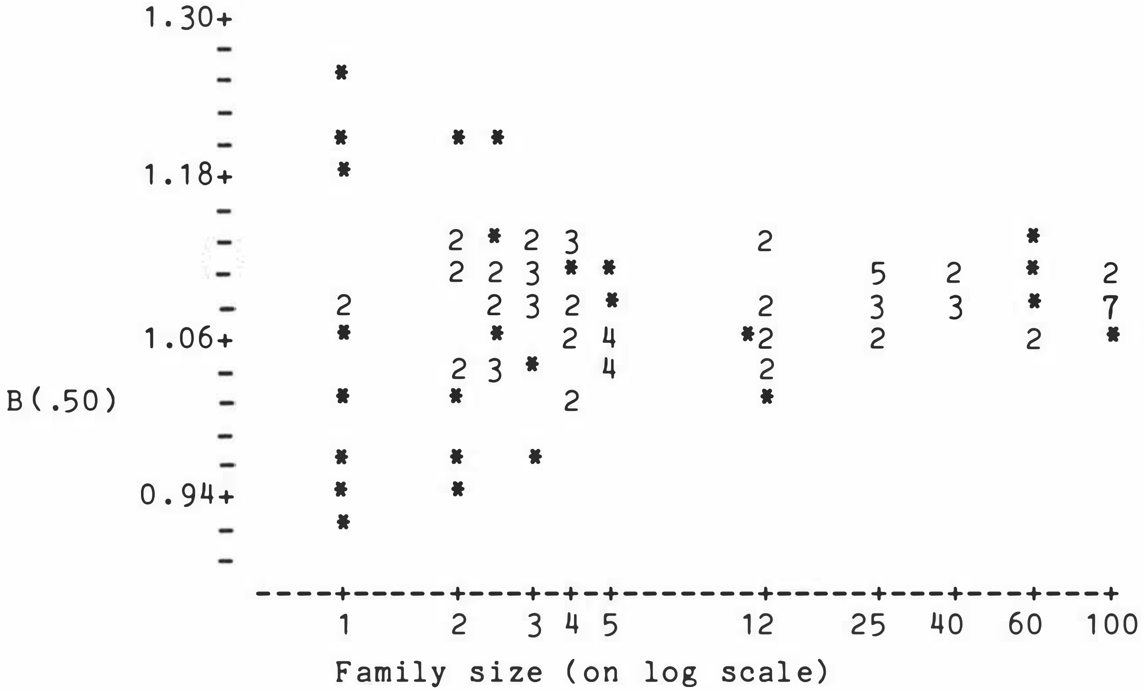


Figure 7.5 Relationship between B(.50) and family size.

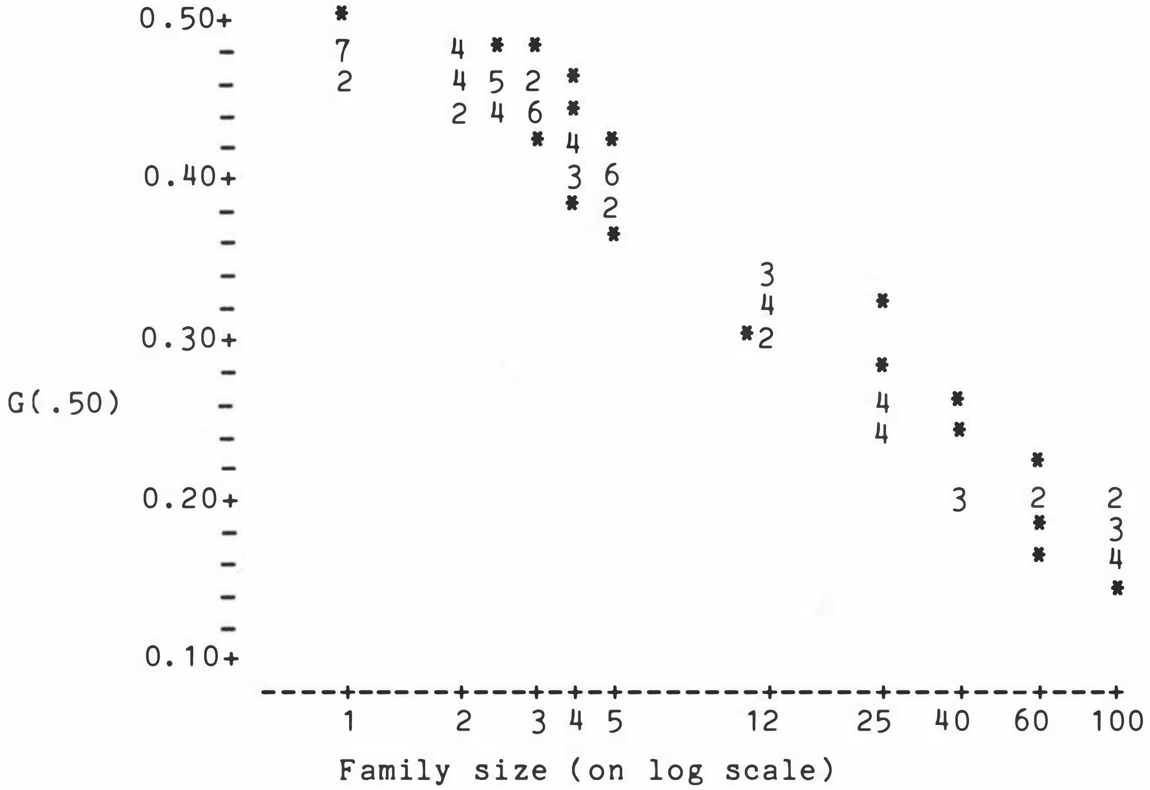


Figure 7.6 Relationship between G(.50) and family size.

```
       1.18+     *
           -             *
           -     *     *
           -     *
           -                           *
       1.06+           2 2 2 3         *     *   *   *
           -     2     2 2 3 * *       2     2   * 2 *   4
B(.10)     -             * 3 3 2      **     5   2 3 2   5
           -     *     * 3 * * 5       3     2   *   *   *
           -     *     * *   2 2
       0.94+           *               *
           -               *
           -     *     *
           -     2     *
           -
       0.82+
             ----+-----+---+-+-+-------+-----+---+---+---+
                 1     2   3 4 5      12    25  40  60  100
                 Family size (on log scale)
```

Figure 7.7 Relationship between B(0.10) and family size.

```
      0.130+
           -
           -                                       *
           -
           -                                             *
      0.120+
           -                                     *       *
           -
           -
           -                                 *       *   *
      0.110+                           2         * *     *
           -                           *     2
           -                 *         *     *           *
G(.10)     -             3 2 *         2     *     2 *
           -     *     4 2 3 3 3       *     *       *
      0.100+     9     5 4 5 * 4             2           2
           -           * *   3 2       *
           -                 *         *         2 * *   *
           -                   *      *                  *
           -                                 *   *   *
      0.090+                                 *
           -
           -                                             *
           -
           -
      0.080+
             ----+-----+---+-+-+-------+-----+---+---+---+
                 1     2   3 4 5      12    25  40  60  100
                 Family size (on log scale)
```

Figure 7.8 Relationship between G(.10) and family size.

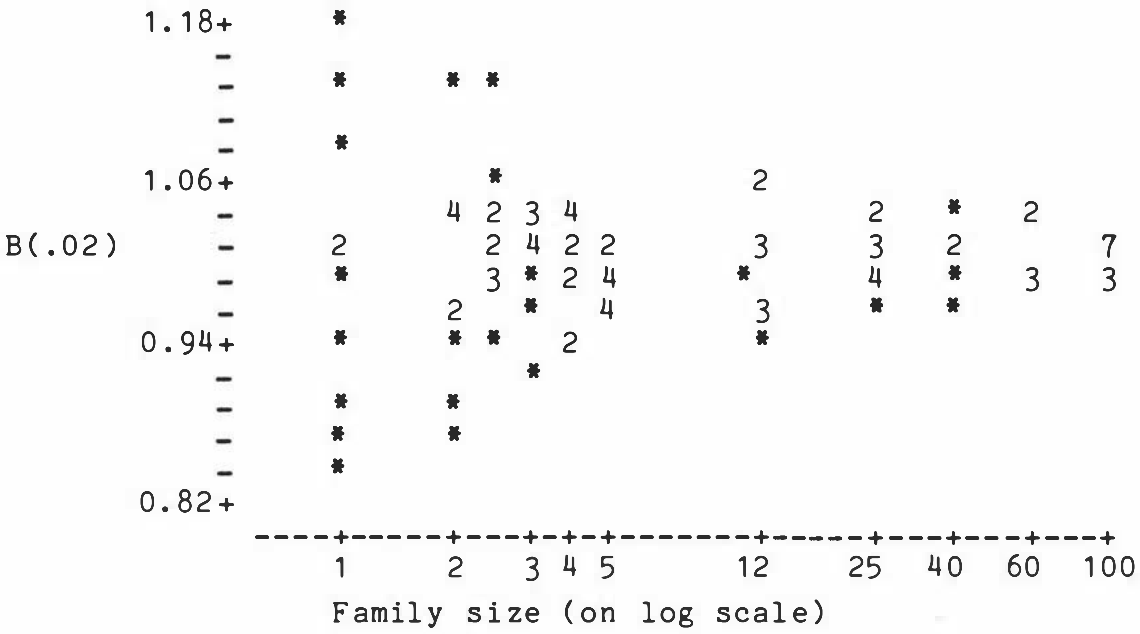


Figure 7.9 Relationship between B(.02) and family size.

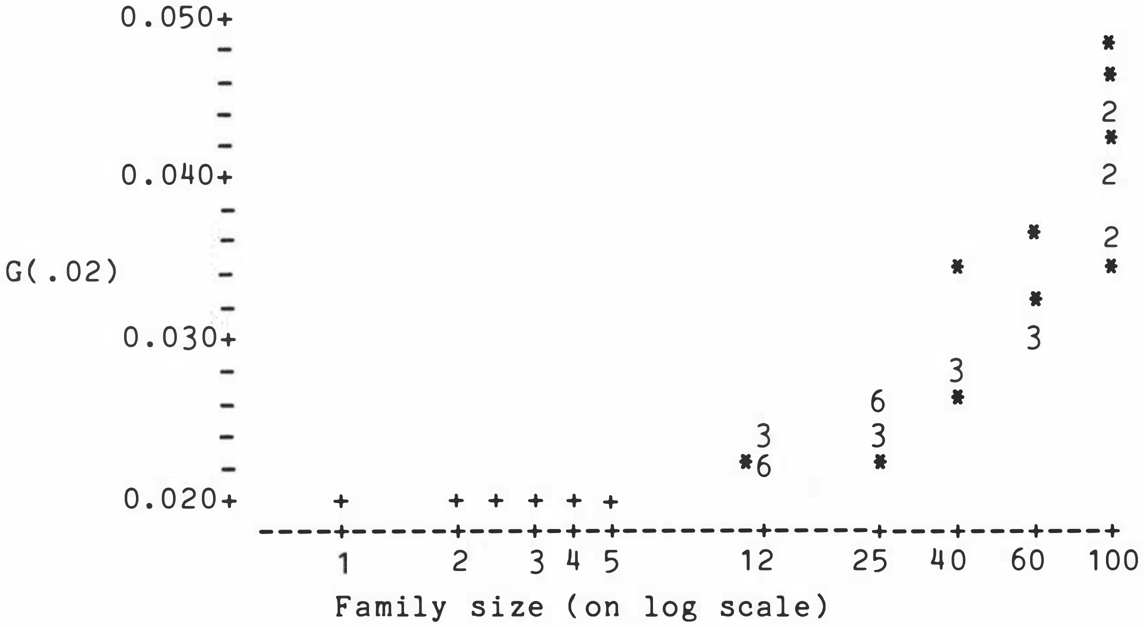


Figure 7.10 Relationship between G(.02) and family size.

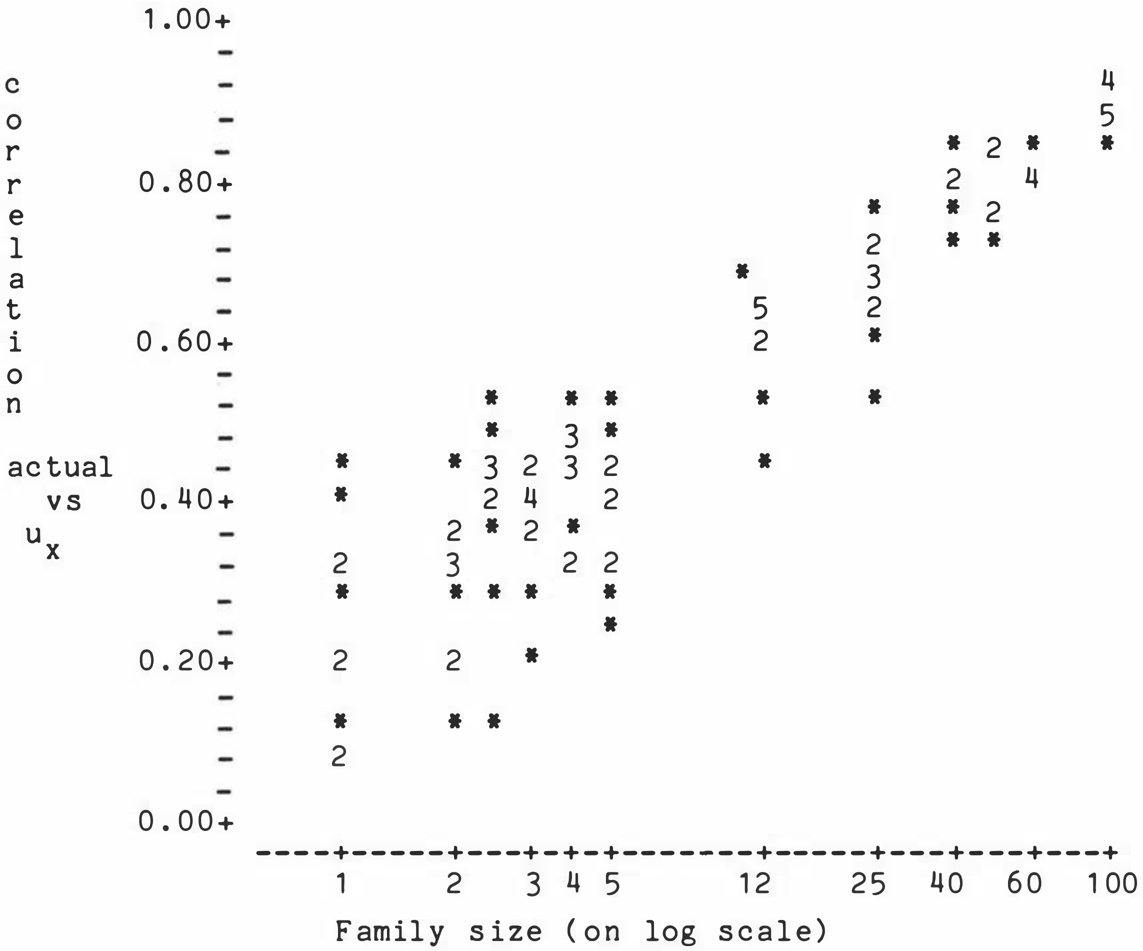


Figure 7.11 Relationship between correlation (u,u_x) and family size.

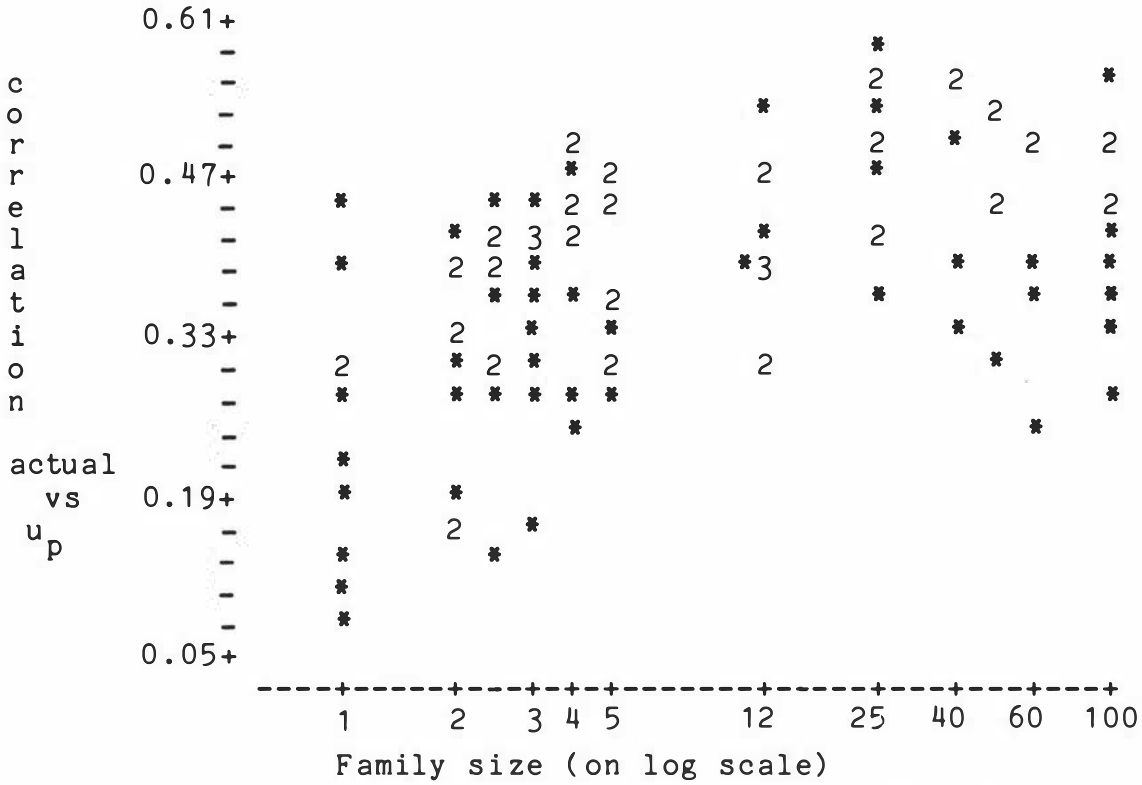


Figure 7.12 Relationship between correlation (u,u_p) and family size.

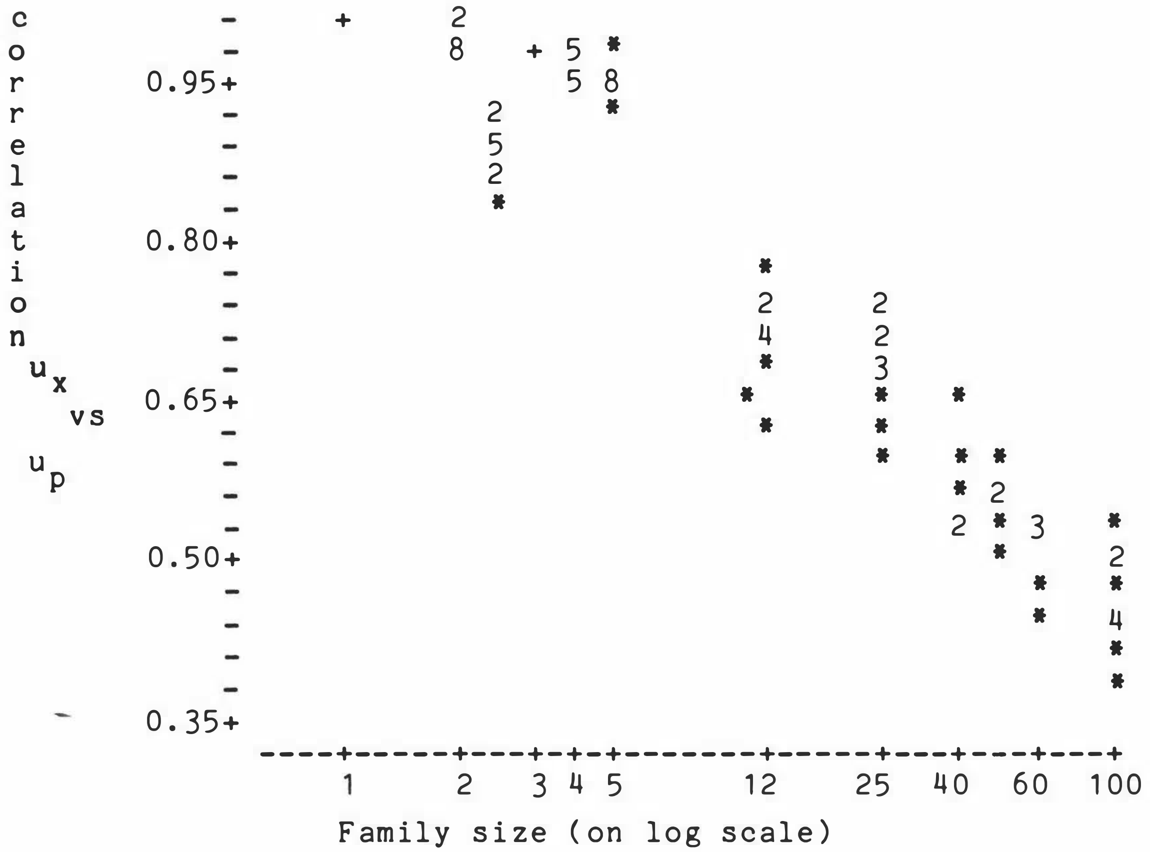


Figure 7.13 Relationship between correlation (u_x,u_p) and family size.

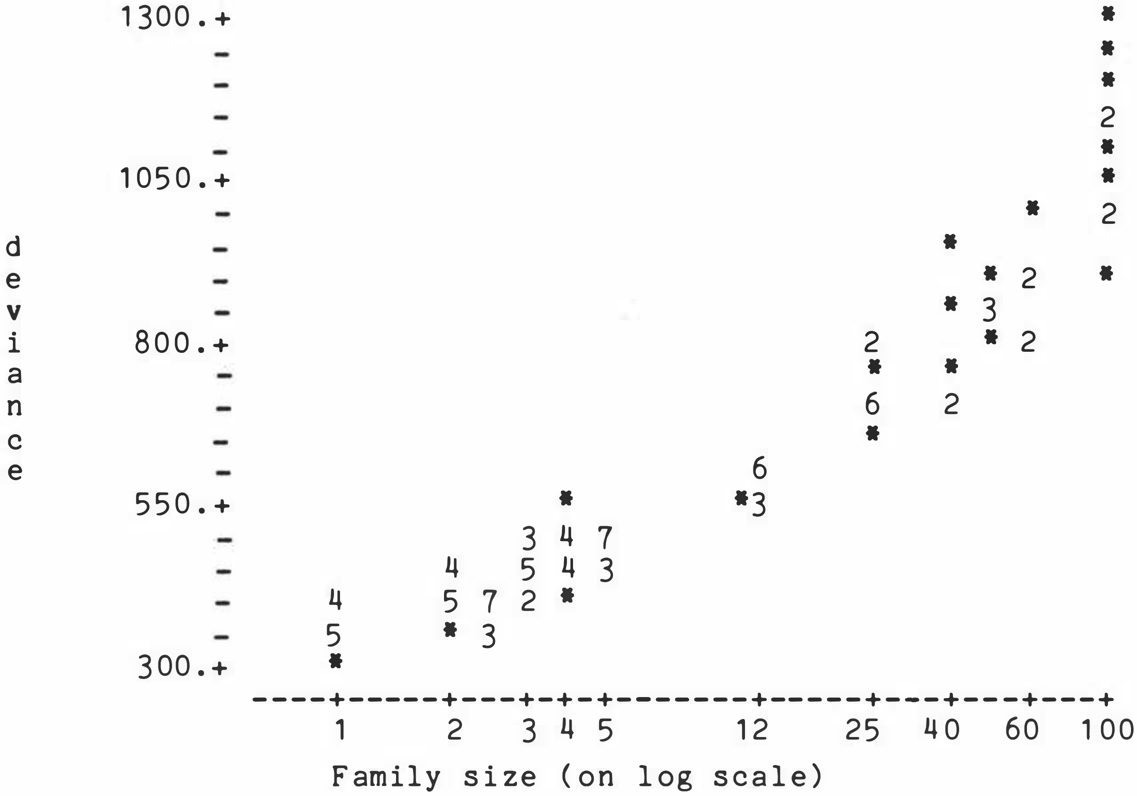


Figure 7.14 Relationship between deviance (499 df) and family size.

Chapter 8. A study of foot ailments associated with Merino-cross sheep grazing damp conditions.

The data analysed in this chapter were collected by M. Alwan for a research project reported in his thesis submitted for the degree of M. Agr. Sc. in the Department of Animal Science at Massey University in 1983. The objective of the work was to investigate the incidence and inheritance of foot ailments in Merino-cross sheep and included information on Perendale sheep as a control. Some comments about the breeds are in order. The highly prolific Booroola strain of Merino was used. Current opinion is that the high prolificacy is due to a major gene complex. The Merino is a breed known to be susceptible to foot ailments which may lead to management problems. The Perendale is a New Zealand synthetic breed based on a Cheviot x Romney cross and is reputed to be less susceptible to the foot ailments than Merino strains.

The traits studied were foot-shape, foot-scald and foot-rot. All feet of the sheep were examined. The foot-shape score was 5 if all feet were of acceptable shape, 4 if there was a small problem with the shape of at least one digit, and 3 if there was a more severe distortion. This classification is obviously subjective. The foot-scald score was one if any feet showed inflamation between the digits and zero otherwise. The foot-rot score was one if any digits showed signs of breakdown and zero otherwise.

The data are used here to demonstrate how the logistic linear mixed model (LLMM) defined in sections 5.1.2.3 and 5.2.2.3 might be used. All analyses reported in this chapter were performed using REG (Gilmour, 1983, see section 6.3). Interest centres in these analyses on the use of the intraclass correlation to estimate repeatability of assessment in section 8.1 and to estimate heritability in section 8.2.

8.1. Repeatability of assessment of foot-shape

Alwan (1983) estimated the repeatability of assessment for foot traits. The feet of 97 (Booroola x Romney) x Perendale ram hoggets were assessed in the morning and again in the afternoon by two observers in different random orders. The data for the three traits are in Appendix tables D.1, D.2 and D.3. Repeatability was estimated for only one trait, foot-shape score; there was insufficient variation in the other two traits for analysis to be meaningful. The data for foot-shape (table D.1) were analysed both as a continuous variable (table 8.1) and as a double-threshold variable (table 8.2).

Table 8.1 Analysis of variance of foot-shape scores.

Source of Variation	df	mean square	E[mean square]
Times	1	3.528	
Observers	1	.312	
Times.Observer	1	.023	
Hoggets	96	1.359	$\sigma_e^2 + 4\sigma_a^2$
Residual	288	.171	σ_e^2
Total	387		

Table 8.2 Analysis of deviance of foot-shape scores when analysed as a double-threshold trait with an intraclass correlation of 0.731.

Source of Variation	df	deviance
Times	2	7.23
Observers	2	3.80
Times.Observers	2	.74
Residual	768	690.16

From the analysis of variance (table 8.1), the intraclass correlation was calculated as follows. Let

$\hat{\sigma}_e^2$ = Residual mean square

$$\hat{\sigma}_a^2 = \frac{\text{Between animals mean square} - \text{Residual mean square}}{\text{number of obsevations per animal}} \quad (8.1)$$

$$= \frac{1.359 - .171}{4} = \frac{1.188}{4} = .297$$

$$\text{Then } \rho = \text{Intraclass correlation} = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_e^2 + \hat{\sigma}_a^2} \quad (8.2)$$

$$= \frac{.297}{.171 + .297} = .635$$

This was compared with the intraclass correlation estimate, under the LLMM, of 0.731. The latter value is expected to be larger than the former (0.635) because information is lost by converting a continuous character (on an underlying scale) to a threshold character. The ratio between these is 0.869 (=0.635/0.731). The expected magnitude of this ratio is not readily obtained. Considering the simplified trait, classified in category 5, with an average incidence of 0.554 (=215/388), the relationship between repeatability on the underlying scale and on the observed scale, using equation 2.2, is given by

$z^2/p(1-p) = 0.632.$ (8.3)

This is an under-estimate of the expected value of the ratio (0.869) since less information is lost by forming 3 categories than by forming 2. Therefore the ratio 0.869 is a reasonable value.

8.2. Estimation of heritability of incidence of foot ailments

Alwan (1983) scored 2513 lambs at 5 months of age, for foot-shape (1, 2, ... 5), presence of foot-scald (0, 1) and presence of foot-rot (0, 1). The lambs were born in August (late winter), 1980 and 1981. The ewe lambs from the 1980 lambing were retained and reassessed at 11 months and at 17 months of age. By contrast, ram lambs from the 1981 lambing were reassessed at 11 months of age. The lambs were from five mating-groups.

1) 7 Perendale rams mated to Perendale ewes in 1980,
2) 6 Booroola x Romney rams mated to Perendale ewes in 1980,
3) 3 Booroola rams mated to Romney ewes in 1980,
4) 6 Perendale rams mated to Perendale ewes in 1981, and
5) 12 Booroola x Romney rams (from mating-group 3) mated to Perendale ewes.

The Booroola x Romney rams in mating-group 2 came from a different research centre to the Booroola rams used in mating-group 3. The progeny from mating-groups 2 and 5 were to form the basis of a new selection flock. The data are in table D.4.

The three traits recorded at three ages are labelled

LC L54 L5 LS LR
HC H54 H5 HS HR
TC T54 T5 TS TR

where L- is a trait recorded on lambs (5 months),
H- is a trait recorded on hoggets (11 months),
T- is a trait recorded on two-tooth ewes (17 months),
-C is the foot-shape score (1 to 5)
-54 is two binary variables indicating the incidence of foot-shape categories 5 and 4, respectively,
-5 is a binary variable indicating the incidence of of foot-shape category 5,
-S is a binary variable indicating the presence of foot-scald and
-R is a binary variable indicating the presence of foot-rot.

Method

The data were analysed by ordinary least squares (OLS) and by the LLMM. The typical full model compared the five mating-groups, sexes and sex.mating-group interaction (9 degrees of freedom). This reduced to just the mating-group effects when analysing the hogget and two-tooth data. The OLS analyses were used to estimate the variance component for sires using Henderson's (1953) method 3 (see section 3.1.3).

The breeding values reported for this method are the sire effects adjusted to a mating-group mean of zero. Only lamb traits were analysed by OLS. For the LLMM analysis, relationships between sires were ignored although they were known to exist. The intraclass correlation under the LLMM was estimated by the iterative procedure outlined in section 6.2 using the full model, even if some factors were not significant. Heritability was estimated as four times this intraclass correlation.

Results and Discussion

A typical sequence of trial values and solutions for estimating the intraclass correlation under the LLMM using the EM algoritm (see section 6.3.2) is shown in table 8.3 for the double-threshold trait L54. The point to note from the table is that interpolation provides a satisfactory method of locating the solution once an upper and lower bound have been identified. The procedure for identifying an upper and lower bound was *ad hoc.*. If the initial estimate was too low, a higher trial value was used. This was 1.5 times the previous solution subject to the limit of an intraclass correlation of 1. If the initial estimate was too high, the new solution was halved. By this method, traits with zero intraclass correlation will have an estimate close to zero - subject to the precision in the data.

Two analyses of deviance for the same trait, L54, are shown in table 8.4. The analyses are for an arbitrary (.03) and an estimated (.043) value of the intraclass correlation. The differences are minimal. In general, analyses of deviance were not altered much by changes in the intraclass correlation because of the way the deviance was computed (see the discussion in section 6.3). Table 8.5 contains estimates of the intraclass correlations and heritabilities for all of the traits based on the assumption that the sire variance is the same for the Perendale, Booroola and Booroola cross breeds. The estimates of heritability range from 0.002 to 0.524.

The heritability estimates for foot-shape were reasonably consistent. The use of a double threshold instead of a single threshold had little effect on the estimate of heritability under the LLMM. It had a greater effect in the OLS analysis as expected. Nevertheless, the difference between 0.165 and 0.154 is not as great as was expected using the theoretical change of scale factor discussed in section 8.1 (see equation 8.3).

The overall average incidence of foot-scald was 0.068 (=172/2513) which produces a scale factor of 0.245. The actual factor was 0.202 (=0.106/0.524). Similarly for incidence of foot foot-rot, the overall average was 0.027 (=67/2513) giving a scale factor of 0.147 compared to an actual factor of 0.175 (=0.040/0.229). The estimates of heritability for foot-scald and foot-rot were not very consistent across ages of assessment. The later assessments were on fewer animals and were also affected by seasonal conditions. The high incidence of foot-scald at the hogget stage in 1980 seems to have obliterated the sire differences while the low incidence in 1981 prevented its expression. The estimate for heritability of foot-rot in two-tooth ewes was poor because of the low numbers of ewes assessed. There were 442 ewes assessed, representing 13 sires and with an average incidence of .09.

Tables D.5 and D.6 contain estimates of breeding value for the traits assessed on the lambs, and for foot-shape score assessed on hoggets. The correlations between the breeding values are in tables 8.5 and 8.6. The sire rankings were nearly identical and the correlations between breeding values obtained by LLMM and by OLS were very high. The principal reason for this concordance was the relatively small differences in fixed effects. Such agreement would not be expected if there was an important covariate present which affected the individuals within mating-groups. In this data, the only factor other than mating-group was sex and the sex.mating-group interactions were generally not significant.

This demonstrates one of the essential features of the LLMM: in the absence of fixed effect differences within mating-groups, the sire rankings are basically the same as OLS (or BLUP) and the LLMM since all observations in the mating-groups are given the same weights. This topic is discussed furthur in section 10.2.

Tables 8.8 to 8.10 contain analyses of deviance for testing the fixed effects in the models while tables 8.11 to 8.14 contain analyses of variance for comparison. There were some differences between the analyses of deviance and analyses of variance in what was significant. The sex effect for L54 in table 8.8 is not quite significant; in table 8.11, sex is significant. However, sex in table 8.8, is tested on two degrees of freedom because it allows for different sex effects on each threshold. Examining the regression coefficients (-0.1310, -0.2868) shows that the sex effects were similar for both thresholds and if the sex effect was partitioned into two separate single degrees of freedom, an average effect and the difference between the thresholds, only the former would be significant. This discrepancy is therefore a consequence of the different formulations of the model and not a real difference between the analyses.

Another small discrepancy occured for the significance of the sex.B1 interaction on L5 in tables 8.9 and 8.12. In the analysis of deviance, this interaction was not quite significant but in the analysis of variance, it was just significant.

The one case of a substantial difference was in the analysis of foot-scald. In table 8.10, sex.B1 was not significant while in table 8.13, it was highly significant. This difference is a consequence of the non-linear change of scale. This is apparent in table 8.15 which displays the incidences of foot-scald for the various sex.mating-group classes. What to the OLS analysis was a large sex.B1 interaction was part of the sex main effect in the LLMM analysis.

Table 8.3 Iterative sequence for estimating the intraclass correlation for the double threshold trait, L54 (lamb foot shape score 5, 4 and less than 4).

Trial $\rho\pi^2/3$	New estimate	intraclass correlation
0.10000	0.11553	0.03512
0.17330	0.15583	0.04737
0.13450	0.13713	0.04168
0.14146	0.14084	0.04281
0.14014	0.14016	0.04260
0.14019	0.14018	0.04261
0.14017	0.14017	0.04261

Table 8.4 Analysis of deviance for the double-threshold trait L54 (lamb foot-shape score 5, 4 and less than 4).

Source of Variation		df	ρ=.04261 deviance	ρ=.03000 deviance	
Year	(YR = 1980 v 1981)	2	41.6	41.8	*
P v BR, B	1980 (B1 = 1 v 2, 3)	2	63.5	63.8	*
BR v B	1980 (B2 = 2 v 3)	2	30.3	29.8	*
P v BR	1981 (B3 = 4 v 5)	2	18.9	18.9	*
Sex	(SX)	2	5.3	5.3	ns
Sex.Year	(SX.YR)	2	0.1	0.1	ns
Sex.B1		2	3.7	3.8	ns
Sex.B2		2	0.2	0.2	ns
Sex.B3		2	0.6	0.6	ns
Residual		116	203.8	203.8	

where * means the deviance exceeds $\chi^2_{2,.05}$ (= 6.0), and ns means not significant (P>0.05).

Table 8.5 Estimates of intraclass correlation and heritability (4 times intraclass correlation) for 12 sheep foot traits.

Trait	Sires in 1980	Sires in 1981	Intraclass correlation	heritability MLM	heritability OLS
L54	16	18	.0426	.170	.171
L5	16	18	.0413	.165	.154
H54	16	18	.0870	.348	
H5	16	18	.0826	.331	
T54	13	-	.0512	.205	
T5	13	-	.0643	.257	
LS	16	12	.1309	.524	.106
HS	16	12	.0009	.004	
TS	13	-	.1129	.451	
LR	16	12	.0572	.229	.040
HR	16	12	.0852	.341	
TR	13	-	.0005	.002	

Table 8.6 Correlation between estimates of breeding value for foot-shape score, obtained by various methods.

	MLM.L54	OLS.LC	MLM.L5	OLS.L5	MLM.H54	
MLM.L54	1.					
OLS.LC	.9728	1.				
MLM.L5	.9912	.9486	1.			
OLS.L5	.9792	.9592	.9904	1.		
MLM.H54	.4995	.5429	.4741	.4647	1.	
MLM.H5	.4901	.5345	.4652	.4557	.9942	1

Table 8.7 Correlation between estimates of breeding value of 28 sires in four mating-groups for presence of foot-scald and presence of foot-rot.

	MLM.LS	OLS.LS	MLM.LR	OLS.LR
MLM.LS	1.			
OLS.LS	.9827	1.		
MLM.LR	.5509	.5884	1.	
OLS.LR	.6234	.6643	.9770	1.

Table 8.8 Analysis of deviance for foot-shape traits L54 and H54, using 34 sires.

Trait		L54		H54
Intraclass correlation		.0426		.0870
Source of Variation	df	Deviance	df	Deviance
1980 VS 1981	2	41.65 *	2	62.90 *
P v BR, B 1980 (B1)	2	63.52 *	2	2.23 ns
BR v B 1980 (B2)	2	30.27 *	2	12.53 *
P v BR 1981 (B3)	2	18.95 *	2	30.40 *
Sex	2	5.27 ns		
Sex.Year	2	0.07 ns		
Sex.B1	2	3.73 ns		
Sex.B2	2	0.20 ns		
Sex.B3	2	0.61 ns		
Residual	116	203.82	58	148.05

where * means the deviance exceeds $X^2_{2,.05}$ (= 6.0), and ns means not significant (P>0.05).

Table 8.9 Analysis of deviance for foot-shape traits L5 and H5, using 34 sires.

Trait		L5			H5	
Intraclass correlation		0.04134			0.08266	
Source of Variation	df	Deviance		df	Deviance	
1980 VS 1981	1	22.25	*	1	62.81	*
P v BR, B 1980 (B1)	1	50.06	*	1	1.24	ns
BR v B 1980 (B2)	1	22.79	*	1	12.28	*
P v BR 1981 (B3)	1	2.37	ns	1	30.18	*
Sex	1	4.07	*			
Sex.Year	1	-0.02	ns			
Sex.B1	1	3.75	ns			
Sex.B2	1	0.17	ns			
Sex.B3	1	0.58	ns			
Residual	58	138.98		29	95.86	
Deviance after random effects		37.87			13.18	

where * means the deviance exceeds $X^2_{1,.05}$ (= 3.84), and ns means not significant (P>0.05).

Table 8.10 Analysis of deviance for presence of foot-scald (LS) and presence of foot-rot (LR) in lambs, using 28 sires. Mating-group 4 (Perendale sires in 1981) was omitted from the analysis because no progeny had either foot-scald or foot-rot.

Trait		LS			HR	
Intraclass correlation		0.13093			0.05723	
Source of Variation	df	Deviance		df	Deviance	
1980 VS 1981	1	3.20	ns	1	0.07	ns
P v BR, B 1980 (B1)	1	32.02	*	1	7.49	*
BR v B 1980 (B2)	1	0.79	ns	1	1.58	ns
Sex	1	23.17	*	1	0.01	ns
Sex.Year	1	0.07	ns	1	0.99	ns
Sex.B1	1	1.74	ns	1	0.00	ns
Sex.B2	1	0.00	ns	1	0.01	ns
Residual	48	104.26		48	59.79	
Deviance after random effects		37.28			41.12	

where * means the deviance exceeds $X^2_{1,.05}$ (= 3.84), and ns means not significant (P>0.05).

Table 8.11 Analysis of Variance for foot-shape score (LC), using 34 sires. K coefficient is 73.4388.

Source of Variation	df	mean square	E[mean square]
1980 VS 1981	1	12.871506 *	
P v BR, B 1980 (B1)	1	16.478943 *	
BR v B 1980 (B2)	1	10.688408 *	
P v BR 1981 (B3)	1	3.785004 *	
Sex	1	1.899033 *	
Sex.Year	1	0.094002 ns	
Sex.B1	1	1.222366 ns	
Sex.B2	1	0.005821 ns	
Sex.B3	1	0.323463 ns	
Sires	29	1.402212	$\sigma_e^2 + K\sigma_u^2$
Error	2474	0.327616	σ_e^2
Total	2512		

where * means the F statistic is significant ($P<0.05$), and ns means not significant ($P>0.05$).

Table 8.12 Analysis of Variance for foot-shape trait L5, using 34 sires. K coefficient is 73.4388.

Source of Variation	df	mean square	E[mean square]
1980 VS 1981	1	5.121124 *	
P v BR, B 1980 (B1)	1	10.448410 *	
BR v B 1980 (B2)	1	5.190765 *	
P v BR 1981 (B3)	1	0.545158 ns	
Sex	1	0.890709 *	
Sex.Year	1	0.033005 ns	
Sex.B1	1	0.856270 *	
Sex.B2	1	0.032177 ns	
Sex.B3	1	0.158112 ns	
Sires	29	0.832369	$\sigma_e^2 + K\sigma_u^2$
Error	2474	0.211528	σ_e^2
Total	2512		

where * means the F statistic is significant ($P<0.05$), and ns means not significant ($P>0.05$).

Table 8.13 Analysis of Variance for presence of foot-scald (LS) using 28 sires. K coefficient is 69.7024. Mating-group 4 (Perendale sires in 1981) was omitted from the analysis because no progeny had foot-scald.

Source of Variation	df	mean square		E[mean square]
1980 VS 1981	1	0.241678	ns	
P v BR, B 1980 (B1)	1	2.523417	*	
BR v B 1980 (B2)	1	0.097591	ns	
Sex	1	1.778001	*	
Sex.Year	1	0.006284	ns	
Sex.B1	1	0.616025	*	
Sex.B2	1	0.011092	ns	
Sires	24	0.219928		$\sigma_e^2 + K\sigma_u^2$
Error	1928	0.075914		σ_e^2
Total	1959			

where * means the F statistic is significant ($P<0.05$), and ns means not significant ($P>0.05$).

Table 8.14 Analysis of Variance for presence of foot-rot (LR) using 28 sires. K coefficient is 69.7024. Mating-group 4 (Perendale sires in 1981) was omitted from the analysis because no progeny had foot-rot.

Source of Variation	df	mean square		E[mean square]
1980 VS 1981	1	0.002551	ns	
P v BR, B 1980 (B1)	1	0.224025	*	
BR v B 1980 (B2)	1	0.073370	ns	
Sex	1	0.000492	ns	
Sex.Year	1	0.033212	ns	
Sex.B1	1	0.001080	ns	
Sex.B2	1	0.002336	ns	
Sires	24	0.055429		$\sigma_e^2 + K\sigma_u^2$
Error	1928	0.032698		σ_e^2
Total	1959			

where * means the F statistic is significant (P<0.05), and ns means not significant (P>0.05).

Table 8.15 Incidence of foot-scald in lambs (5 months) on percentage and logit scales.

	Percentage scale			Logit scale		
Mating-group	Male	Female	diff	Male	Female	diff
1 (7 Perendales 1980)	4.8	3.4	1.4	-2.99	-3.35	.36
2 (6 Boor.xRom. 1980)	16.9	8.2	8.7	-1.59	-2.41	.82
3 (3 Booroola 1980)	21.2	9.4	11.8	-1.31	-2.27	.95
4 (6 Perendales 1981)	.0	.0		omitted from analysis		
5 (12 Boor.xRom 1981)	9.5	4.5	5.0	-2.25	-3.05	.80

Chapter 9. Reproductive performance of Perendale ewes

The Perendale is a New Zealand synthetic sheep breed derived by inter-breeding Cheviot x Romney cross animals. The following information was available on 1396 two-tooth ewes born in Massey University's registered Perendale flock from 1961 to 1972:

63 Sire identification	See table 9.1
Rearing rank of sire (RRSIRE)	(1003 single, 393 multiple).
Dam identification	
Dam age (DA)	See table 9.2.
Rearing rank of dam (RRDAM)	(816 single, 580 multiple).
Year born (YB) (61 to 72)	See table 9.3.
Rearing rank of ewe (RREWE)	(869 single, 527 multiple).
Weaning weight of ewe (WW)	(mean 23.7 kg, s.d. 3.8 kg).
Hogget weight of ewe (HW)	(mean 39.1 kg, s.d. 4.9 kg).
Number of lambs the ewe weaned as a two-year-old (NLW).	(308 nil, 961 single, 127 multiple).

These data were studied by Elliot(1975) and Lewer(1978) who have discussed the management of the flock. Concerning the selection of replacements within the flock, Lewer(1978) states,

> "Selection of candidates for entry into the flock is based on maintaining the Perendale 'type' and involves culling abnormally coarse woolled sheep or those which fail to exhibit wool-free points. The selection of fine wool sheep was given particular emphasis in the early 1960s but recently fertility has been given attention ... . Poor conformation or physical unsoundness would also exclude a ewe for consideration as a replacement. About half of the hoggets available enter the flock as two-tooths. Towards the end of the period under consideration, this proportion represented about 230 sheep, but previously fewer replacements were available, or required, when the flock was smaller."

Ram usage (table 9.1) was very uneven. Twenty-seven rams had 10 or less daughters, 30 rams had 11 to 50 daughters and 6 rams had more than 50 daughters. Twenty-three rams were used in more than one year. Each consecutive pair of years had at least two rams in common. Eight of the 63 sires, represented by 178 daughters, were obtained from outside the flock.

Table 9.1 Summary of the number of daughters sired by each of the 63 rams used in the flock from 1961 to 1972. RR is the rearing rank of the sire (1=single, 2=multiple) and NoD is the number of daughters. A * marks those sires born outside the flock.

matrix may show this better.

Ram	RR	NoD	years	Ram	RR	NoD	years	Ram	RR	NoD	years
1	1	14	61	22	2	12	66,67	43	2	102	70,71
2*	1	37	61-63	23	1	35	66-68	44	1	49	70,71
3	1	13	62	24	1	10	66	45	1	9	70
4	2	15	65,66	25	1	75	66-70	46	2	1	70
5	1	10	61	26	1	4	67	47	1	71	71,72
6	2	26	61,62	27	2	14	67	48	1	5	71
7	1	32	61-63	28*	1	9	67	49	2	9	71
8	1	4	61	29*	1	16	67	50	1	10	71
9	1	25	62,63	30	1	22	68	51	2	5	71
10*	1	59	63-65	31	2	11	68	52	2	32	71,72
11	1	12	64,65	32	2	45	68-70	53*	2	1	72
12*	1	22	64	33	1	9	68	54*	1	2	72
13	1	10	63	34	1	18	68	55	2	2	71
14	1	42	63-66	35*	1	30	68,69	56	2	2	72
15	1	32	64,65	36	2	10	69	57	2	6	72
16	1	1	64	37	1	18	69,71	58	2	10	72
17	1	24	64-67	38	1	9	69	59	2	23	72
18	1	4	64	39	1	5	69	60	1	5	71
19	1	85	64-68	40	2	18	69	61	2	13	72
20	2	15	65	41	1	21	69,71	62	2	5	72
21	1	10	68	42	2	16	69	63	1	135	68-70

Table 9.2 Dam age means for weaning weight, hogget weight and number of lambs weaned.

dam age	no. of ewes	weaning weight	hogget weight	no. of lambs weaned
2	363	23.3	38.5	0.865
3	355	23.8	39.6	0.862
4	265	24.1	39.2	0.879
5	204	23.2	38.7	0.902
6+	209	24.2	39.4	0.852

Table 9.3 Year means for weaning weight, hogget weight and number of lambs weaned.

year born	no. of ewes	weaning weight	hogget weight	no. of lambs weaned
61	52	24.8	35.5	0.904
62	76	23.1	39.3	1.013
63	77	21.6	37.6	0.948
64	69	22.1	36.2	0.957
65	115	20.6	33.6	0.930
66	87	22.7	35.6	0.770
67	98	25.3	35.9	0.918
68	148	24.6	42.8	0.959
69	183	25.1	40.9	0.803
70	149	24.5	41.2	0.799
71	212	23.3	39.6	0.792
72	130	24.7	42.1	0.862

Methods for estimating the heritability of some aspects of reproductive performance were compared. The trait of primary importance, NLW, was analysed as a multiple-threshold trait and as a combination of two traits. The multiple-threshold model implies that the genetic and environmental factors which control whether or not a ewe lambs, also control the number of lambs she actually has. The multiple-threshold model allows for different factors to control these two aspects of the trait NLW. The results indicate which of these models is more satisfactory.

The following traits were derived from NLW.

Dry-or-not (D),	a binary trait with 1 indicating the ewe did not rear any lambs.
Single-or-not (S),	a binary trait with 1 indicating the ewe reared a single lamb.
Twins-or-not (T),	a binary trait with 1 indicating the ewe reared twins.
Lambed-or-not (L),	a binary trait with 1 indicating the ewe reared one or more lambs.

Genetic analyses of these data may utilize both sire and dam information. The three factors RREWE, RRDAM and RRSIRE indicate whether the ewe's dam, maternal grandam and paternal grandam, respectively, raised a single or twin lambs. These factors were used to perform parent-offspring regression analyses on the trait twins-or-not given lambed but with two qualifications. The factors are conditional on lambed-or-not since an animal cannot leave progeny if it is not reared. Secondly, they refer to various (unspecified) dam ages whereas twins-or-not refers specifically to the two-year-old lambing. These qualifications would be removed if the two-tooth reproductive performance records were available on the dams and grandams.

The sire identification enabled paternal half-sib estimates of genetic parameters to be obtained. Sires were assumed unrelated in these analyses since genetic relationships were not known. RREWE and RRDAM were included as covariates since the genetic information in them comes from the dam side and is independent of the sire. RREWE contains a maternal environment component as well as a maternal genetic component. This is particularly important for the continuous traits WW and HW.

No attempt was made to estimate the repeatabilities of any of these traits over increasing dam ages. Weaning and hogget weights were analysed as continuous traits which may be genetically and/or environmentally correlated with reproductive performance. Dam age (DA) and year born (YB) were included in the models as fixed effects.

All analyses were performed using REG (Gilmour, 1983, see section 6.3).

9.1. Analysis of Perendale data by ordinary multivariate least squares

The traits WW, HW and NLW were analysed by ordinary multivariate least squares fitting the factors dam age (DA), year of birth (YB), rearing rank of ewe (RREWE), rearing rank of dam (RRDAM) and sires. Sires were included either as RRSIRE or as 63 individual sire effects. In the latter case, Henderson's (1953) Method 3 was used to estimate variance and covariance components (table 9.4) for the factor sires under the full model. Heritabilities derived from these variance components are in table 9.5.

Table 9.4 Variance and covariance components estimated by Henderson's (1953) Method 3. Covariance components are below and correlations are above the diagonal.

Error components - on 1316 degrees of freedom

Trait	weaning weight	hogget weight	number of lambs	lambed-or-not	twins-or-not
Weaning weight	8.60	.516	.027	.036	-.001
hogget weight	5.64	13.9	.020	-.003	.042
no. of lambs	.042	.041	.288	-	-
lambed-or-not	.0431	-.0049	-	.167	.168
twins-or-not	-.0004	.0454	-	.020	.082

Sires components - on 62 degrees of freedom
- using k coefficient 17.74

Trait	weaning weight	hogget weight	no. of lambs	lambed-or-not	twins-or-not
weaning weight	.321	.257	.115	-.991	.210
hogget weight	.156	1.15	.386	.065	.551
no. of lambs	.0042	.0266	.0041	-	-
lambed-or-not	-.0031	.0038	-	.0030	.089
twins-or-not	.0073	.0228	-	.00014	.00082

Table 9.5 Heritabilities of five traits by Henderson's (1953) Method 3.

Trait	weaning weight	hogget weight	no. of lambs	lambed-or-not	twins-or-not
Heritability	.14	.30	.056	.071	.040

The conclusions from these analyses are:

1) The weaning and hogget weights of ewe lambs were lower ($P<0.05$) if they were reared as a twin than if they were reared as a single lamb and differed ($P<0.05$) between years (year born is fully confounded with year measured).

2) The weaning weight of ewe lambs from two-tooth dams was lower ($P<0.05$) than for ewe lambs from older dams.

3) The hogget weights were lower (-0.9 ± 0.2 kg, $P<0.05$) for those ewe lambs whose sire was raised as a twin than for those ewe lambs whose sire was raised a single lamb. This suggests a negative genetic relationship between twinning and hogget weight.

4) None of the factors in the model affected NLW ($P>0.05$), the number of lambs weaned, analysed as a continuous trait. When analysed as two traits, lambed-or-not and twins-or-not, lambed-or-not was lower ($P<.05$) for ewes whose dams were reared as twins. YB differed ($P<.05$) when fitted before sires but not when fitted after.

9.2. Analysis of binary traits on the probit scale under a fixed effects model

Iterative re-weighted least squares with the probit transformation was used to analyse the binomial traits lambed-or-not, twins-or-not and twins-or-not given lambed.

The model included the factors YB, DA, RREWE, RRDAM and RRSIRE. Sires were not included as separate effects in the model to avoid instability in the iterative process resulting from few observations per sire and to keep the model small. The three rearing rank factors, RREWE, RRDAM and RRSIRE, were coded 0 for reared as a single and 1 for reared as a twin. Thus, the corresponding regression coefficients represent the differences (on the probit scale) between rearing rank sub-population means.

None of the factors in the probit analyses were significant except for YB and RRDAM on lambed-or-not where RRDAM had a negative coefficient (consistent with the analysis on the 0,1 scale). The frequencies of weaning types for the two RRDAM classes (unadjusted for year or other effects) were

number of lambs	0	1	2
RRDAM=single	164 (20%)	582 (71%)	70 (9%)
RRDAM=twin	144 (25%)	379 (65%)	57 (10%)

There is no immediately obvious explanation as to why ewes whose mothers were reared as twins were more likely to be dry than ewes whose mothers were raised as single lambs. This suggests that NLW is not a double-threshold trait because the negative sign indicates a negative genetic correlation and contradicts the concept that the environmental and genetic factors which control whether a ewe lambs also control the number of lambs.

Heritabilities on the underlying scale were estimated from the regression coefficients for RREWE, RRDAM and RRSIRE using Falconer's (1965) method descibed in section 2.3.1. The estimates are in table 9.6. Adapting equation 2.9, heritability is

$$h^2 = \frac{b}{r\ d} \qquad (9.1)$$

where b is the mean difference between the relatives of the 'affected' and the relatives of the 'unaffected' groups, the response,

d is the difference in mean between 'affected' and 'unaffected' groups, the selection differential, and

r is the coefficient of relationship between the classified subjects and their relatives.

In this case, the classified subjects were the dams (for RREWE), the maternal grandams (for RRDAM) and the paternal grandams (for RRSIRE). The relatives were the two-tooth ewes with the reproductive performance record. A difficulty with applying this method to this data arises with the definition of d, the difference between the means of the 'affected' and 'unaffected' groups. As emphasized by James and McGuirk (1983), the relevant proportions to use in calculating these means are the population proportions, not just the proportions in the sample. The proportion of twinning two-tooth ewes was taken as 12% [127/(127+961)]. The difficulty is that the classification was not made on a two-tooth lambing reproductive performance record but on a record from an unspecified age. The incidence of twins in the mixed age ewes appears to be 23% [263.5/(869+263.5)] provided there is no differential in survival between lambs of different sex. Using the population incidence of .121 gives a threshold value of 1.17 and ordinate of .2013 from the standard normal curve. The difference in means is then

$$d = \frac{.2013}{.121} - \frac{-.2013}{.879}$$

$$= 1.664 + .229 = 1.893$$

The heritability estimates for lambed-or-not and twinned-or-not in table 9.6 were calculated assuming a genetic correlation of 1 between these traits and the trait twinned-or-not given lambed.

Table 9.6 Heritability on the probit scale by Falconer's (1965) method.
(r) is the coefficient of relationship.
(b) is the regression coefficient.
(h^2) is the heritability estimate using eqn. 9.1.

factor		lambed-or-not		twins-or-not		twins-or-not given lambed	
	(r)	(b)	(h^2)	(b)	(h^2)	(b)	(h^2)
RREWE	.5	.122	.129	.071	.075	.059	.062
RRDAM	.25	-.192	-.406	.075	.159	.125	.264
RRSIRE	.25	.135	.286	.090	.190	.074	.157

The estimates of heritability in table 9.6 refer to the underlying normal scale. They were compared with the former estimates (table 9.5) using the relationship (Dempster and Lerner, 1950, see equation 2.2)

$$h_p^2 = \frac{z^2 h^2}{p(1-p)} \qquad (9.2)$$

where h_p^2 is heritability on the (0,1) scale,

h^2 is heritability on the underlying normal scale,

p is the mean proportion with the trait and

z is the ordinate of the standard normal curve corresponding to an intregral of p.

The estimates of heritability on the underlying scale, using this relationship on the values from table 9.5, were 0.140 for lambed-or-not and 0.122 for twins-or-not.

Falconer's method of analysis has the advantage of providing some insight into the applicability of the model in that the estimates from various relatives should concur if the model is appropriate. It also indicates the direction of an effect; estimation by a variance components approach assumes a positive intraclass correlation.

The model fitted on the probit scale was revised by the inclusion of hogget weight as a covariate to adjust for the maternal effect. The biggest change in the revised estimates (table 9.7) was that the heritability estimate for twins-or-not based on RREWE had doubled.

<u>Table 9.7</u> Heritability on the probit scale using Falconer's method after including hogget weight as a covariate and assuming genetic correlations of 1.
(r) is the coefficient of relationship.
(b) is the regression coefficient.
(h^2) is the heritability estimate using eqn. 9.1.

factor		lambed-or-not		twins-or-not		twins-or-not given lambed	
	(r)	(b)	(h^2)	(b)	(h^2)	(b)	(h^2)
RREWE	.5	.125	.132	.123	.130	.122	.129
RRDAM	.25	-.193	-.408	.069	.146	.121	.256
RRSIRE	.25	.136	.288	.112	.237	.103	.218

Sire evaluation

Sires were ranked using the residuals (on the 0,1 scale) from the probit analysis (residuals on the probit scale will all be plus or minus infinity). These residuals were reanalysed in a mixed model analysis using binomial weights calculated from the expected values obtained from the probit analysis.

Table 9.8 Sires in order of average rank for lambed-or-not and twins-or-not. Rankings are based on weighted residuals from the probit analyses.

Sire details			lambed-or-not	twins-or-not
id	NoD	RR	rank	rank
40	18	2	2	1
57	6	2	4	8
28	9	1	11	4
15	32	1	18	3
50	10	1	10	12
35	30	1	1	21
48	5	1	13	10
27	14	2	12	17
61	13	2	14	18
33	9	1	28	6
7	32	1	33	5
22	12	2	23	15
47	71	1	20	20
3	13	1	24	19
17	24	1	15	31
2	37	1	37	11
55	2	2	47	2
63	135	1	29	22
23	35	1	25	27
6	26	2	9	43
31	11	2	46	7
19	85	1	16	39
44	49	1	35	23
43	102	2	30	28
59	23	2	43	16
46	1	2	22	38
24	10	1	52	9
60	5	1	17	44
58	10	2	3	58
16	1	1	26	36
1	14	1	8	54
36	10	2	6	57
13	10	1	50	14
25	75	1	41	24
32	45	2	36	29
11	12	1	53	13
45	9	1	5	62
18	4	1	7	63
9	25	1	39	32
29	16	1	38	34
10	59	1	21	51
52	32	2	48	25
12	22	1	34	41
42	16	2	19	56
34	18	1	27	52
14	42	1	32	48
21	10	1	55	26
20	15	2	51	35
30	22	1	49	40
8	4	1	40	49
4	15	2	57	33
38	9	1	44	46
62	5	2	31	59
41	21	1	62	30
53	1	2	60	37
56	2	2	54	47
49	9	2	42	60
54	2	1	61	42
39	5	1	59	45
51	5	2	45	61
5	10	1	58	53
37	18	1	56	55
26	4	1	63	50

That is, the equations

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{W}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{W}\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{W}\underset{\sim}{X} & \underset{\sim}{Z}'\underset{\sim}{W}\underset{\sim}{Z} + k\underset{\sim}{I} \end{bmatrix} \begin{bmatrix} \underset{\sim}{b} \\ \underset{\sim}{s} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{W}\underset{\sim}{y} \\ \underset{\sim}{Z}'\underset{\sim}{W}\underset{\sim}{y} \end{bmatrix} \qquad (9.3)$$

were solved where $\underset{\sim}{X}$ was the $\underset{\sim}{1}$ vector, $\underset{\sim}{Z}$ was the sires incidence matrix, $\underset{\sim}{b}$ was a mean, $\underset{\sim}{s}$ was a vector of sire breeding values, $\underset{\sim}{y}$ was the vector of residuals, $\underset{\sim}{I}$ was an identity matrix and k was $(4-h^2)/h^2$.

This analysis was performed on lambed-or-not and twins-or-not using k = 25.67 corresponding to a heritability of 0.15. The results are in table 9.8.

9.3. Analysis of the Perendale data using the logistic linear mixed model

The data were described in the preceding sections. The model used in the analyses reported in this section included RREWE, DA (dam age) and YB (year of birth) as fixed effects, and sires as a random effects factor. The sires were regarded as unrelated.

The traits analysed were, in terms of implicit binary variables:

1) lambed-or-not estimating the probability of lambing,
2) twins-or-not estimating the probability of twinning,
3) lamb-t, a threshold trait estimating the probability of (1) being dry and (2) being dry or having a single lamb.
4) lamb-e, an extremal trait estimating the probability of (1) being dry and (2) having a single lamb.
5) WW, the weaning weight of the ewe.
6) HW, the hogget weight of the ewe.

These traits were analysed separately using the appropriate logistic or usual mixed model equations. The results are summarized in tables 9.9 and 9.10. The four categorical traits were then analysed with hogget weight in a multiple trait analysis using the genetic variances obtained from the single trait analyses.

Table 9.11 displays some of the values obtained when attempting to locate a solution for the variances and covariances of the extremal trait lamb-e using the method described in section 6.3.2. It demonstrates that the values included in table 9.9 are reasonable.

Tables 9.12 and 9.13 contain some results from attempting to estimate genetic covariances between the four categorical traits and hogget weight. The method used is described in section 6.3.2. Environmental error correlations of -0.3, 0.0 and 0.3 were used for the threshold traits and error correlations of of 0.0 used for the extremal trait.

Discussion

The effect of using different models for a categorical trait may be seen in table 9.9. The threshold trait lamb-t, is equivalent to the two binary traits lambed-or-not and twins-or-not after swapping zeroes and ones. Thus, except for the change in sign, the regression coefficients are obviously similar. The deviance used for testing the significance of fixed effects affecting a threshold trait has not been partitioned into separate components for each trait or into average and difference components. This could be done with more programming. The regression coefficients for lamb-e appear quite different but this is principally because of the form of the multinomial logit transformation. The first variable in lamb-e is dry-or-not which is the opposite of lambed-or-not. A relationship can be seen by comparing the coefficients for these traits across factors which are deviations (RREWE and year of birth).

Table 9.10 contains the estimates of heritabilities obtained under the logistic linear mixed model (LLMM). These were compared with those in tables 9.6 and 9.7, based on equation 9.1, and those obtained from table 9.5 using equation 9.2.

The major difference is that the heritability of lambed-or-not is very small under the LLMM but is 0.14 using equation 9.2. It could be argued that the lower value is more likely to be true because this is a trait under continuous natural selection.

Another interesting aspect of the heritability estimates is the low heritability of lamb-t despite the value of 0.143 for twins-or-not. This strongly suggests that the genetic predisposition to lambing is quite different to the predisposition to twinning, an observation also evident in table 9.8.

The heritability estimates under the extremal model were slightly higher for dry-or-not than they were for lambed-or-not (its complement). This suggests it is important to treat number of lambs as an extremal character, the genetic relationship between the categories does not follow the threshold model. The other point to note is that the implicit heritability for twins-or-not (= 1 - dry-or-not - single-or-not) obtained under the extremal estimates is (0.02 + 0.03 + 0.03 + 0.073) = 0.153 which is close to the direct estimate of 0.143.

The conclusion from this discussion is that there is little scope for genetically increasing the proportion of ewes lambing as two-tooths but there is scope for increasing the level of twinning, which has a heritability on the underlying scale of about 0.15.

The heritability estimates for hogget weight were around 0.3 which is reasonably high. It is of interest to know whether there is a genetic correlation between hogget weight and reproductive performance as such a relationship may enable increases in selection pressure. The only strong genetic correlations in table 9.12 are for twins-or-not. The correlation is positive ranging from 0.406 assuming an environmental correlation of 0.3, to 0.885 assuming an environmental correlation of -0.3. The intermediate value of 0.6275 is supported by the correlations under the extremal model being -0.6 (implying 0.6 for twins-or-not). This is consistent with the estimate of 0.55 obtained by Henderson's (1953) Method 3 (table 9.4). The agreement between these estimates was expected on the basis of the simulation studies by Olausson and Rönningen (1975) quoted in section 2.2.3. This correlation is positive however, contrary to the sire-offspring regression result mentioned as point 3 towards the end of section 9.1. There is an apparent contradiction between the positive estimates of genetic correlation and the negative sire-offspring regression coefficient mentioned under point 3 towards the end of section 9.1. The unadjusted hogget weight means for rearing rank of sire were

RRSIRE = single 39.037
RRSIRE = twin 39.139

This difference of 0.102 kg is of opposite sign to the adjusted mean difference. Table 9.14 shows the breeding values for twins-or-not and hogget weight obtained under the assumption of a genetic correlation of 0.625. Table 9.15 shows the correlations between the breeding values and the rearing rank of the sire. The negative relationship between rearing rank and hogget weight persists even though the breeding values for twins-or-not are positively correlated with both rearing rank and hogget weight. A large component of the negative correlation is due to sire 58 which had ten daughters in the data. One of these daughters had a hogget weight of 25kg. The next lowest was 32kg. It would appear that the least squares procedures have given special dominance to this low value. This does not fully explain the contradiction but does point up the need for using substantial sets of data when analysing categorical traits.

Table 9.9 Fixed effects for all traits estimated by maximum likelihood single trait methods. Statistical significance (sig.) is indicated by n.s. (P>.05), * (P<.05) and ** (P<.01).

Dam age coefficients

Age	lambed-or-not	twins-or-not	lamb-threshold		lamb-extremal		weaning weight	hogget weight
2	1.20	-2.54	-1.20	2.46	1.10	2.19	25.3	42.2
3	1.18	-2.70	-1.17	2.62	1.25	2.35	26.4	43.0
4	1.18	-2.52	-1.18	2.44	1.10	2.18	26.8	43.2
5	1.44	-2.86	-1.44	2.78	1.21	2.58	26.7	42.8
6+	1.10	-2.65	-1.08	2.59	1.27	2.27	27.3	43.4
sig.	n.s.	n.s.	n.s.		n.s.		**	**

Effect of ewe being raised as a twin rather than as single

	lambed-or-not	twins-or-not	lamb-threshold		lamb-extremal		weaning weight	hogget weight
RREWE	.22	.17	-.23	-.17	-.32	-.12	-4.1	-1.9
sig.	n.s.	n.s.	n.s.		n.s.		**	**

Effect of years relative to year 1 (1972).

year	lambed-or-not	twins-or-not	lamb-threshold		lamb-extremal		weaning weight	hogget weight
1961	.42	-.31	-.42	.33	-.04	.41	.1	-6.4
1962	.75	.61	-.75	-.62	-1.17	-.48	-2.0	-2.9
1963	.07	.91	-.08	-.82	-.80	-.88	-3.6	-4.6
1964	.22	.67	-.20	-.61	-.74	-.62	-2.7	-6.1
1965	.31	.33	-.31	-.23	-.49	-.24	-3.8	-8.5
1966	-.21	-1.07	.20	1.27	1.30	1.09	-2.0	-6.5
1967	.36	.07	-.36	-.03	-.32	.02	.5	-6.6
1968	.36	.61	-.38	-.54	-.79	-.51	-.6	.4
1969	-.23	-.27	.23	.27	.44	.23	.5	-1.5
1970	-.42	.32	.41	-.31	.03	-.43	-1.1	-1.9
1971	-.41	.17	.41	-.11	.20	-.24	-1.6	-2.9
sig.	**	*	**		**		**	**

Table 9.10 Genetic and environmental variances, heritabilities and breeding values estimated by maximum likelihood single trait methods. Sires marked an asterisk were raised as twins.

Estimates of genetic and environmental variances

	lambed-or-not	twins-or-not	lamb-threshold	lamb-extremal		weaning weight	hogget weight
σ^2_u	.0033	.1182	.0033	.020	.030	.1917	1.139
				.030	.073		
σ^2_e	n.a.	n.a.	n.a.	n.a.	n.a.	9.144	14.67

Estimates of heritability

lambed-or-not	twins-or-not	lamb-threshold	lamb-extremal		weaning weight	hogget weight
.0040	.1433	.0040	.048	.088	.082	.288

Estimates of breeding values of sires

sire	lambed-or-not	twins-or-not	lamb-threshold	lamb-extremal		weaning weight	hogget weight
1	.004	-.092	-.001	.035	.103	-.101	-.076
2	-.004	.115	.000	-.039	-.108	.036	.236
3	.002	.042	-.003	-.004	.000	-.569	.331
4*	-.006	-.006	.005	-.015	-.063	-.206	.578
5	-.005	-.064	.006	.003	-.017	.273	.789
6*	.008	-.069	-.004	.038	.126	.084	-.660
7	.001	.279	-.009	-.076	-.172	.345	-.891
8	-.001	-.031	.002	.004	.001	-.079	-.296
9	-.001	-.042	.002	.010	.017	.245	.376
10	.006	-.179	.001	.063	.172	-.471	.495
11	-.007	.078	.004	-.039	-.125	-.244	-.470
12	-.001	-.074	.003	.019	.042	.143	-.556
13	-.003	.035	.002	-.018	-.056	.081	-1.243
14	.002	-.356	.009	.104	.250	-.112	-.508
15	.005	.554	-.021	-.142	-.303	.183	1.486
16	.001	-.017	-.000	.006	.019	-.143	-.518
17	.005	-.087	-.002	.036	.110	.019	-.238
18	.002	-.066	-.001	.025	.071	.298	1.026
19	.008	-.327	.004	.107	.275	-.059	-.366
20*	-.005	-.055	.005	.002	-.018	-.197	-.834
21	-.005	-.014	.004	-.010	-.046	-.019	-.177
22*	.002	.050	-.003	-.006	-.002	.151	-.816
23	.004	.018	-.005	.006	.037	.397	1.809

continued

Table 9.10 continued Breeding values for all traits estimated by maximum likelihood single trait methods. Sires marked with an asterisk were raised as twins.

Estimates of breeding values of sires

sire	lambed-or-not	twins-or-not	lamb-threshold	lamb-extremal		weaning weight	hogget weight
24	-.004	.087	.003	-.034	-.100	.137	.342
25	-.010	-.056	.010	.001	-.041	.044	1.751
26	-.008	-.038	.008	-.014	-.072	.009	-.538
27*	.004	.084	-.006	-.009	.000	-.088	-.463
28	.001	.142	-.005	-.032	-.068	-.032	.673
29	-.002	-.039	.003	.006	.004	-.119	-.324
30	-.005	-.077	.006	.010	-.000	.373	-.972
31*	-.001	.170	-.004	-.047	-.114	.032	.128
32*	.003	-.010	-.002	.009	.033	-.187	-.926
33	.002	.201	-.007	-.048	-.105	.036	1.068
34	.000	-.125	.003	.035	.084	.024	-.736
35	.018	.044	-.017	.030	.157	-.089	.570
36*	.005	-.072	-.003	.033	.104	-.047	-.152
37	-.010	-.129	.012	.006	-.038	-.148	-.108
38	-.002	-.058	.003	.008	.009	-.071	-.209
39	-.005	-.035	.006	-.008	-.046	.047	-.572
40*	.012	.420	-.021	-.075	-.115	.033	-.276
41	-.021	-.055	.021	-.041	-.199	.034	-.045
42*	.003	-.108	-.001	.037	.102	-.518	-.569
43*	.004	-.050	-.002	.018	.054	.372	.444
44	-.005	-.030	.005	-.003	-.030	.125	1.064
45	.006	-.106	-.003	.044	.131	.044	-.200
46*	.001	-.011	-.001	.006	.021	-.006	.287
47	.005	.039	-.006	-.001	.021	.083	.450
48	.002	.064	-.003	-.011	-.017	-.084	.047
49*	-.001	-.091	.002	.021	.045	-.277	-.126
50	.003	.009	-.003	.007	.034	-.123	-.206
51*	-.002	-.055	.003	.008	.011	.241	.692
52*	-.008	.030	.007	-.026	-.096	-.082	.136
53*	-.003	-.011	.003	-.006	-.028	-.063	.012
54	-.005	-.017	.005	-.012	-.055	-.130	-.418
55*	-.001	.099	-.001	-.029	-.076	-.088	.373
56*	-.002	-.018	.002	-.001	-.011	.142	.410
57*	.005	.063	-.006	-.002	.020	.024	.704
58*	.007	-.092	-.004	.045	.141	-.517	-2.259
59*	-.004	.132	.000	-.043	-.118	.036	-.640
60	.001	-.046	-.000	.015	.041	.016	-.612
61*	.003	.002	-.003	.009	.038	.282	.249
62*	.001	-.043	.000	.013	.034	.195	.473
63	-.001	.095	-.003	-.030	-.069	.285	1.002

Table 9.11 Some results from using various trial values for the variances and covariances associated with an extremal trait. The composite binary traits are dry-or-not (D), single-or-not (S) and twins-or-not (T). All values are multiplied by 10,000.

	var(D)	cov(D,S)	var(S)	cov(D,T)	cov(S,T)	var(T)
Trial	200	306	730	-506	-1036	1541
Result	201	307	729	-507	-1035	1543
Difference	1	1	-1	-1	0	1
Trial	700	299	200	-999	-499	1499
Result	689	297	200	-986	-497	1483
Difference	-10	-2	-0	13	3	-16
Trial	500	400	500	-900	-900	1800
Result	500	403	505	-904	-908	1812
Difference	0	3	5	-4	-8	12
Trial	200	337	700	-537	-1037	1573
Result	201	338	701	-539	-1039	1578
Difference	1	1	1	-2	-3	5
Trial	700	337	200	-1037	-537	1573
Result	691	334	199	-1025	-533	1557
Difference	-9	-3	-1	12	4	-16
Trial	500	450	500	-950	-950	1900
Result	502	453	504	-955	-958	1913
Difference	2	3	4	-5	-8	13
Trial	200	262	700	-462	-962	1424
Result	201	263	698	-464	-961	1425
Difference	1	1	-2	-2	1	1
Trial	700	262	200	-962	-462	1424
Result	688	260	200	-948	-461	1408
Difference	-12	-2	0	14	1	-15
Trial	500	350	500	-850	-850	1700
Result	499	353	504	-852	-858	1710
Difference	-1	3	4	-2	-8	10

Table 9.12 Genetic variances and covariances between two binary traits, lambed-or-not and twins-or-not and hogget weight estimated for three environmental correlations, 0.0, 0.3 and -0.3.
All values are multiplied by 10,000.

Lambed-or-not (L) and Hogget weight (H)

correlations - environmental 0.0, genetic 0.0 and -0.045

	var(L)	cov(L,H)	var(H)	var(L)	cov(L,H)	var(H)
Trial	50	0	11390	50	-34	11390
Result	50	-1	10826	50	-33	10825
Diff.	-0	-1	-564	-0	1	-565

correlations - environmental 0.3, genetic 0.0 and -0.464

	var(L)	cov(L,H)	var(H)	var(L)	cov(L,H)	var(H)
Trial	50	0	11390	50	-350	11390
Result	50	-10	11028	50	-344	11013
Diff.	-0	-10	-362	-0	6	-377

correlations - environmental -0.3, genetic 0.0 and 0.44

	var(L)	cov(L,H)	var(H)	var(L)	cov(L,H)	var(H)
Trial	50	0	11390	50	332	11390
Result	50	8	11091	50	327	11089
Diff.	-0	8	-299	-0	-5	-301

Twins-or-not (T) and Hogget weight (H)

correlations - environmental 0.0, genetic 0.0 and 0.6275

	var(T)	cov(T,H)	var(H)	var(T)	cov(T,H)	var(H)
Trial	1182	0	11390	1182	2303	11390
Result	1178	236	10826	1179	2254	10943
Diff.	-5	236	-564	-4	-49	-447

correlations - environmental 0.3, genetic 0.0 and 0.406

	var(T)	cov(T,H)	var(H)	var(T)	cov(T,H)	var(H)
Trial	1182	0	11390	1182	1490	11390
Result	1174	152	10690	1178	1450	10829
Diff.	-9	152	-700	-5	-40	-561

correlations - environmental -0.3, genetic 0.0 and 0.885

	var(T)	cov(T,H)	var(H)	var(T)	cov(T,H)	var(H)
Trial	1182	0	11390	1182	3248	11390
Result	1188	334	11218	1182	3225	11239
Diff.	6	334	-172	-1	-23	-151

Table 9.13 Genetic variances and covariances between two categorical traits, lamb-t and lamb-e, and hogget weight, estimated for three environmental correlations, 0.0, 0.3 and -0.3. All values are multiplied by 10,000.

Lamb-t (t) and Hogget weight (H)

correlations - environmental 0.0, genetic 0.0 and -0.36

	var(t)	cov(t,H)	var(H)	var(t)	cov(t,H)	var(H)
Trial	50	0	11390	50	-272	11390
Result	50	-9	10826	50	-265	10860
Diff.	-0	-9	-564	-0	7	-530

correlations - environmental 0.3, genetic 0.0 and -0.98

	var(t)	cov(t,H)	var(H)	var(t)	cov(t,H)	var(H)
Trial	50	0	11390	50	-740	11390
Result	50	-23	11478	51	-749	11528
Diff.	-0	-23	88	1	-10	138

correlations - environmental 0.3, genetic -0.9

	var(t)	cov(t,H)	var(H)	var(t)	cov(t,H)	var(H)
Trial	200	-1394	12000	180	-1312	11800
Result	199	-1385	11877	180	-1311	11748
Diff.	-1	9	-123	0	0	-52

correlations - environmental -0.3, genetic 0.0 and -0.0985

	var(t)	cov(t,H)	var(H)	var(t)	cov(t,H)	var(H)
Trial	50	0	11390	50	74	11390
Result	50	2	10875	50	73	10863
Diff.	-0	2	-515	-0	-2	-527

Lamb-e (D,S) and Hogget weight (H)

correlations - environmental 0.0, genetic 0.0 and 0.0

	var(D)	cov(D,S)	var(S)	cov(D,H)	cov(S,H)	var(H)
Trial	200	299	700	0	0	11390
Result	201	301	700	-64	-147	10826
Diff.	1	1	-0	-64	-147	-564

correlations - environmental 0.0, genetic -0.6 and -0.6

	var(D)	cov(D,S)	var(S)	cov(D,H)	cov(S,H)	var(H)
Trial	200	299	700	-906	-1694	11390
Result	200	299	698	-886	-1657	10926
Diff.	0	-0	-2	19	37	-464

Table 9.14 Comparison of breeding values for lambed-or-not and hogget weight obtained assuming environmental correlation of 0.0 and genetic correlations of 0.0 and 0.6275. Sires raised as twins are identified by an asterisk beside the sire number.

sire	T0.0	H0.0	T0.6	H0.6	sire	T0.0	H0.0	T0.6	H0.6
1	-.092	-.08	-.091	-.16	33	.201	1.07	.374	1.27
2	.115	.24	.121	.30	34	-.125	-.74	-.225	-.81
3	.042	.33	.095	.38	35	.044	.57	.139	.58
4*	-.006	.58	.104	.55	36*	-.072	-.15	-.088	-.23
5	-.064	.79	.100	.71	37	-.129	-.11	-.120	-.21
6*	-.069	-.66	-.167	-.67	38	-.058	-.21	-.088	-.27
7	.279	-.89	.069	-.69	39	-.035	-.57	-.145	-.61
8	-.031	-.30	-.088	-.34	40*	.420	-.28	.283	.08
9	-.042	.38	.024	.34	41	-.055	-.05	-.046	-.09
10	-.179	.50	-.072	.40	42*	-.108	-.57	-.193	-.65
11	.078	-.47	-.021	-.38	43*	-.050	.44	.049	.40
12	-.074	-.56	-.151	-.60	44	-.030	1.06	.163	1.01
13	.035	-1.24	-.200	-1.17	45	-.106	-.20	-.125	-.32
14	-.356	-.51	-.357	-.69	46*	-.011	.29	.048	.27
15	.554	1.49	.679	1.76	47	.039	.45	.100	.46
16	-.017	-.52	-.121	-.55	48	.064	.05	.068	.13
17	-.087	-.24	-.106	-.30	49*	-.091	-.13	-.100	-.23
18	-.066	1.03	.141	.91	50	.009	-.21	-.030	-.20
19	-.327	-.37	-.299	-.48	51*	-.055	.69	.088	.60
20*	-.055	-.83	-.194	-.87	52*	.030	.14	.051	.15
21	-.014	-.18	-.043	-.19	53*	-.011	.01	-.008	-.01
22*	.050	-.82	-.121	-.78	54	-.017	-.42	-.100	-.44
23	.018	1.81	.329	1.77	55*	.099	.37	.170	.53
24	.087	.34	.136	.40	56*	-.018	.41	.066	.38
25	-.056	1.75	.233	1.69	57*	.063	.70	.194	.78
26	-.038	-.54	-.141	-.59	58*	-.092	-2.26	-.510	-2.31
27*	.084	-.46	-.014	-.37	59*	.132	-.64	-.011	-.52
28	.142	.67	.254	.83	60	-.046	-.61	-.162	-.68
29	-.039	-.32	-.086	-.35	61*	.002	.25	.049	.26
30	-.077	-.97	-.224	-.99	62*	-.043	.47	.056	.41
31*	.170	.13	.169	.31	63	.095	1.00	.235	.99
32*	-.010	-.93	-.144	-.90					

Table 9.15 Correlations between breeding values in table 9.14 and the rearing rank of the sire.

	RRSIRE	T0.0	H0.0	T0.6	H0.6
RRSIRE	1.0000	.0871	-.1479	-.0392	-.1259
T0.0		1.0000	.2463	.7378	.3650
H0.0			1.0000	.8324	.9910
T0.6				1.0000	.8956

Chapter 10. General Discussion

This thesis has presented a new method for analysing categorical data under a mixed model. It has been applied to binomial and multinomial data. Several questions relating to this method are now raised and discussed.

10.1. How relevant is an estimate of heritability on the underlying scale?

The whole basis of the method is that there is an underlying continuous trait, often called liability. It is assumed that when measured on a convenient scale, this underlying trait has an approximately normal distribution and that genetic effects on this scale are largely additive, at least more additive than on the categorical scale. If there is no reasonable biological base for making these assumptions it would be misleading to use the method. The biology of the trait may point to different assumptions which require a completely different kind of solution. This issue has been discussed by Kidd and Cavalli-Sforza (1973), Elston (1973), Elston (1977) and Hill and Smith (1977).

10.2. What relationship exists between the logistic linear mixed model and the ordinary mixed model analyses?

For a fully random model, the ordinary mixed model equations and logistic linear mixed model equations are essentially equivalent except for the change of scale. Quaas and Van Vleck (1980) described the mixed model analysis of a multinomial trait under a fully random model. The variances were assumed known on 'traditional advice'. For comparative purposes it is sufficient to consider the binomial situation only. The equations can then be written as

$$\begin{bmatrix} n & \underset{\sim}{1}'\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{1} & \underset{\sim}{Z}'\underset{\sim}{Z} + k\underset{\sim}{R}^{-1} \end{bmatrix} \begin{bmatrix} \beta_p \\ \underset{\sim}{u}_p \end{bmatrix} = \begin{bmatrix} \underset{\sim}{1}'\underset{\sim}{z} \\ \underset{\sim}{Z}'\underset{\sim}{z} \end{bmatrix} \qquad (10.1)$$

where $k = \frac{\sigma_e^2}{\sigma_u^2} = \frac{(4 - h_p^2)}{h_p^2}$

σ_e^2 is the environmental variation on the p scale,

σ_u^2 is the genetic variation on the p scale and

h_p^2 is the heritability of the trait on the p scale.

The comparative logistic linear mixed model equations are

$$\begin{bmatrix} nd^2/e & ad/e\ \underset{\sim}{1}'\underset{\sim}{Z} \\ ad/e\ \underset{\sim}{Z}'\underset{\sim}{1} & a^2/e\ \underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{R}^{-1}/g \end{bmatrix} \begin{bmatrix} \beta_x \\ \underset{\sim}{u}_x \end{bmatrix} = \begin{bmatrix} d^2/e\ \underset{\sim}{1}'\underset{\sim}{y} \\ ad/e\ \underset{\sim}{Z}'\underset{\sim}{y} \end{bmatrix} \quad (10.2)$$

where ρ, the intraclass correlation on the underlying scale,

$= \frac{h_x^2}{4} = \frac{\bar{p}\bar{q}\ h_p^2}{4z^2}$, using equation 2.2,

h_x^2 is the heritability on the underlying scale,

z is the ordinate of the standard normal curve at the point which corresponds to a probability of p,

$s = \sqrt{1 - \rho}$ from 5.37,

$\bar{p} = \frac{\exp(s\beta_x)}{1 + \exp(s\beta_x)}$ using 5.35,

$\bar{q} = 1 - \bar{p}$,

$e = \bar{p}\bar{q}\ (1 - \bar{p}\bar{q}\ \rho\ \pi^2/3)$ using 5.47,

$a = \bar{p}\bar{q}$ using 5.48,

$g = \rho\ \pi^2/3$ using 5.49,

$d = sa$ using 5.50 and

$\underset{\sim}{y} = \beta_x \underset{\sim}{1} + (\underset{\sim}{z} - \bar{p}\underset{\sim}{1})/d$ using 5.51.

Equations 10.2 may be rearranged into the form of equations 10.1 as follows:

1) multiply throughout by e,

2) divide the first equation by d,

3) divide the remaining equations by a,

4) replace $\underset{\sim}{y}$ by $\beta_x \underset{\sim}{1} + (\underset{\sim}{z} - \bar{p}\underset{\sim}{1})/d$ and
5) rearrange the terms.
The result is

$$\begin{bmatrix} n & \underset{\sim}{1}'\underset{\sim}{Z} \\ \underset{\sim}{Z}'\underset{\sim}{1} & \underset{\sim}{Z}'\underset{\sim}{Z} + c\underset{\sim}{R}^{-1} \end{bmatrix} \begin{bmatrix} d\beta_x - d\beta_x + \bar{p} \\ a\underset{\sim}{u}_x \end{bmatrix} = \begin{bmatrix} \underset{\sim}{1}'\underset{\sim}{z} \\ \underset{\sim}{Z}'\underset{\sim}{z} \end{bmatrix}$$

$$\text{where } c = \frac{e}{a^2 g} = \frac{\bar{p}\bar{q}(1 - \bar{p}\bar{q}\ \rho\ \pi^2/3)}{(\bar{p}\bar{q})^2 \rho\ \pi^2/3}$$

$$= \frac{4 - (\bar{p}\bar{q})^2\ h_p^2\ \pi^2/3z^2}{(\bar{p}\bar{q})^2\ h_p^2\ \pi^2/3z^2}. \tag{10.3}$$

It is obvious that equations 10.1 and 10.2 are the same when $c = k$. This occurs when the coefficient of h_p^2 in 10.3 is 1, i.e. when $x = 1.003$ implying $p = 0.158$ and $z = 0.2413$.

A consequence of the preceeding discussion is the fact that ordinary BLUP will be nearly as efficient as the LLMM equations for estimating breeding values whenever there is little variation associated with fixed effects. However, when incidences within fixed effect groups are outside the range of 0.3 to 0.7, the fluctuation in weights associated with small changes in incidence can be quite large and should not be ignored.

10.3. What can be done in very large problems where the iterative procedures would be prohibitively expensive?

When the procedure as defined cannot be used because of computing technology limitations, it will be necessary to consider simpler solutions. In the analysis of binomial data under a fixed model, four strategies are possible. These are
1) Analyse the binary data as if it were a normal continuous character. This approach ignores the relationship between mean and variance for binomial characters.
2) Apply a 'variance stabilizing' transformation to the binary data and then analyse it by ordinary least squares.

3) Calculate empirical weights from the data and analyse it by weighted least squares on the binomial scale.
4) Used the GLM methodology and analyse the character by iteratively reweighted least squares on a logit (or other) scale.

In the mixed model situation, an ordinary BLUP analysis corresponds to approach 1 while the use of the GLMM corresponds to approach 4. Approaches 2 and 3 generally require grouped data. (The use of grouped data for approaches 1 and 4 is not necessary theoretically but is advantageous for computing efficiency.) Approach 3 seems to provide a possibility for simplifying the GLMM procedures in large data sets. When applying BLUP to large data sets, it is common to absorb the fixed effects such as herd-year-season. Provided the herd-year-season groups are large enough to provide reasonable measures of incidence of the binary character within the group, the same approach could be used for categorical data. The incidence in the group would be calculated and then used to generate the weights and other quantities required by the GLMM equations. The fixed effects would then be absorbed and breeding values predicted from a single pass through the data. The efficiency of this method has not been investigated but will obviously depend on the incidences within each herd-year-season.

10.4. When might it be more efficient to use equations based on the joint-maximization rather than those based on the maximization-expectation method?.

Questions of the general efficiency of the methods have not been investigated. Some points can be made however. In data with substantial information on both fixed and random effects, the joint-maximization method may be better at estimating the random effects. In the case of a single random factor, it is essentially the number of observations per group and the mean incidence of the group that determines the probability of all individuals in the group being classified into the same response category. If some effects in the model have observed proportions of 0 or 100 per cent, the group mean on the logit scale will rapidly iterate towards plus or minus infinity and result in failure within computing routines. This problem is more likely to occur with the joint-maximization procedure, which iterates both fixed and random effects, than the maximization-expectation procedure which iterates only the fixed effects. The force of this argument is reduced by the fact that the size of the random effects is restricted, under both procedures, by the assumption of the variance of the random effects. Table 10.1 shows the probability of all observations being in the same category for a range of sample sizes and incidences.

Table 10.1 The probability that a group of nominated size and nominated incidence has observations all in the same category.

	incidence in the population sampled.		
size of the group	50%	10%	1%
1	1.0	1.0	1.0
2	0.5	0.82	0.98
3	0.25	0.73	0.97
4	0.125	0.656	0.961
8	0.008	0.430	0.92
16	<0.001	0.185	0.85
32	<0.001	0.034	0.72

It is immediately obvious that if the incidence of the trait is low, the sample sizes must be quite high to avoid many groups having extreme observations.

The essence of the matter is that the joint solution will only be better when the variance estimate based on the observations within the group is likely to be closer to the true variance than an average variance for that group. By using expected values for the variance, the maximization-expectation method is likely to give much more numerically stable results.

Section 7.2 provided some information on the robustness of the LLMM equations to small family sizes. Those investigations were based on having a large number of families. In practical applications, it is possible that family sizes may be only one. The reliability of any estimate of breeding value for a categorical trait based on small family size will always be poor - especially for traits with incidences outside to range of 10 to 90 percent.

10.5. What further work is required?

The method of solving for variance components used in this study has been a modified EM algorithm. Newton-Raphson and Fisher-scoring techniques probably follow directly from their counterparts in REML estimation for the normal case. These techniques were preferred by Thompson in his discussion of the EM technique (Dempster, Laird and Rubins, 1977). No consideration has been given in this thesis to estimating confidence bounds for estimates of variance components. The use of Newton-Raphson or Fisher-scoring techniques for estimating the variance components would lead to estimates of the variances of the estimates.

Computational feasability will often control what use is made of a technique. The additional computing required for using the LLMM equations rather than BLUP equations has two components. The iteration component is related to the number of iterations and may be minimized by absorbing the random effects since the solutions for these are not required during the iteration phase. The other component involves an exponentiation in the retransformation of the logit and the multiplication of various weight functions. This component is proportional to the number of records and can be reduced by grouping the data.

Some theoretical aspects of the method remain to be investigated.

Henderson (1963) has shown that BLUP provides a predictor which, in the class of linear predictors, minimizes the mean squared error of prediction. Can an analagous result be shown for the GLMM equations when used for predicting random variables?

The question of testing 'goodness of fit' of a mixed model has not been adequately covered in this thesis.

The GLMM method should be able to be used for Poisson data providing a suitable interpretation of an underlying scale is developed.

The issue of when to use the threshold as opposed to the extremal model for multinomial traits has not been discussed. The difficulty lies with traits which are coded is if threshold traits but which have an inadequate biological basis for that coding. The trait number-of-lambs-weaned (NLW) is a possible example. It seems likely, in the light of the results in chapter 9, that whether or not a ewe bears any lambs is under the control of different genes to those which control the number of lambs she actually bears, given that she has at least one lamb. Similarly, the trait calving-difficulty is coded in categories 'none', 'slight', 'easy pull', 'hard pull' and 'ceasarian'. These categories may misrepresent the problem which from a biological point of view may need to consider size of cow, size of calf, whether the calf is full term and other such matters.

10.6. Conclusion

The generalized linear mixed model, developed in this thesis, provides a computationally feasible method of analysis for categorical data under a mixed model. The specific computing expressions provided for binomial and multinomial characters have worked satisfactorily on the examples studied.

Appendix A
Some matrix symbols and operations

(1) Matrix operators

$\underset{\sim}{I}_n$	Identity matrix of order n.
$\underset{\sim}{1}_n$	a vector of n 1s.
$\underset{\sim}{J}_{mn}$	an m x n matrix of 1s (= $\underset{\sim}{1}_m\underset{\sim}{1}_n'$).
$\underset{\sim}{A}^{-1}$	inverse of arbitrary positive definite matrix $\underset{\sim}{A}$.
$\underset{\sim}{A}^{-}$	generalized inverse of $\underset{\sim}{A}$ (such that $\underset{\sim}{A}\ \underset{\sim}{A}^{-}\ \underset{\sim}{A} = \underset{\sim}{A}$).
$\underset{\sim}{A}'$	transpose of arbitrary matrix $\underset{\sim}{A}$.
$\{a_{ij}\}$	a matrix with elements a_{ij}
$\Sigma^{+}_{k}\underset{\sim}{D}_k$	a block diagonal matrix with matrices $\underset{\sim}{D}_k$ on diagonal
$\text{diag}\{a_k\}$	a diagonal matrix with diagonal elements a_k
$\text{diag}(\underset{\sim}{a})$	a diagonal matrix with the elements of vector $\underset{\sim}{a}$ on the diagonal.
$\underset{\sim}{A} \otimes \underset{\sim}{B}$	Right direct product of arbitrary matrices $\underset{\sim}{A}$ and $\underset{\sim}{B}$
$\underset{\sim}{A}*\underset{\sim}{B}$	Hadamard (element by element) product of matrices $\underset{\sim}{A}$ and $\underset{\sim}{B}$ (both of the same order)
$\text{vec}\underset{\sim}{A}$	vector formed by stacking columns of $\underset{\sim}{A}$ one beneath the other.

$(\underset{\sim}{A} \otimes \underset{\sim}{B})(\underset{\sim}{X} \otimes \underset{\sim}{Y}) = \underset{\sim}{A}\underset{\sim}{X} \otimes \underset{\sim}{B}\underset{\sim}{Y}$

$(\underset{\sim}{A} \otimes \underset{\sim}{B})^{-} = \underset{\sim}{A}^{-} \otimes \underset{\sim}{B}^{-1}$

$(\underset{\sim}{A} \otimes \underset{\sim}{B})^{-1} = \underset{\sim}{A}^{-1} \otimes \underset{\sim}{B}^{-1}$

$|\underset{\sim}{A}_{pp} \otimes \underset{\sim}{B}_{rr}| = |\underset{\sim}{A}|^{r}\ |\underset{\sim}{B}|^{p}$

$(\underset{\sim}{A} \otimes \underset{\sim}{B})' = \underset{\sim}{A}' \otimes \underset{\sim}{B}'$

$\text{vec}(\underset{\sim}{A}\underset{\sim}{B}\underset{\sim}{C}) = (\underset{\sim}{C}' \otimes \underset{\sim}{A})\ \text{vec}\ \underset{\sim}{B}$

$\text{trace}(\underset{\sim}{A}'\underset{\sim}{B}) = (\text{vec}\ \underset{\sim}{A})'\ \text{vec}\ \underset{\sim}{B}$

$\text{trace}\ (\underset{\sim}{A} \otimes \underset{\sim}{B}) = \text{trace}(\underset{\sim}{A}).\text{trace}(\underset{\sim}{B})$

(2) Matrix differentiation.

a) Derivative of a function with respect to a vector.

For $f(\underset{\sim}{x}) = \underset{\sim}{a}'\underset{\sim}{x} = \underset{\sim}{x}'\underset{\sim}{a}$ ($\underset{\sim}{a}$ and $\underset{\sim}{x}$ being vectors)

$$\partial f(\underset{\sim}{x})/\partial \underset{\sim}{x} = \underset{\sim}{a}$$

Hence $\partial \underset{\sim}{A}\underset{\sim}{x}/\partial \underset{\sim}{x} = \underset{\sim}{A}'$

For A symmetric, $\partial \underset{\sim}{x}'\underset{\sim}{A}\underset{\sim}{x}/\partial \underset{\sim}{x} = 2\underset{\sim}{A}\underset{\sim}{x}$

b) Derivative of a function with respect to a matrix.

For $\underset{\sim}{X}_{mn}$ a real matrix of mn independent variables,

$\partial f/\partial \underset{\sim}{X} = \{ \partial f/\partial x_{ij} \}$ is an mn-matrix of partial derivatives

If $f(\underset{\sim}{X}) = \underset{\sim}{a}'\underset{\sim}{X}\underset{\sim}{b}$, then $\partial f/\partial \underset{\sim}{X} = \underset{\sim}{a}\,\underset{\sim}{b}'$

and for $\underset{\sim}{X}$ symmetric, $\partial a'\underset{\sim}{X}a/\partial \underset{\sim}{X} = 2\underset{\sim}{a}\underset{\sim}{a}' - dg(\underset{\sim}{a}\underset{\sim}{a}')$

where $dg(\underset{\sim}{a}\underset{\sim}{a}')$ is a diagonal matrix with the same diagonal elements as $\underset{\sim}{a}\underset{\sim}{a}'$.

c) Derivative of a matrix with respect to a scalar.

$\partial \underset{\sim}{Y}/\partial \underset{\sim}{x}_{ij} = \{ \partial(f_{pq}(\underset{\sim}{X})) / \partial x_{ij} \}$, an mn-matrix of partial derivatives with respect to x_{ij}.

d) Derivative of a Determinant.

$\partial|\underset{\sim}{X}|/\partial \underset{\sim}{X} = \{ x^{ij} \}$ where x^{ij} is the cofactor of x_{ij}.

i.e. $\underset{\sim}{X}^{-1} = 1./|\underset{\sim}{X}| \{ x^{ij} \}$

Appendix B.

More expectations for section 5.1.2.1 using the normal distribution

In similar fashion to section 5.1.2.1, u_i and u_l are $N(\underset{\sim}{0},\underset{\sim}{\Sigma})$

where $\underset{\sim}{\Sigma} = \sigma^2 \begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix}$

$$E\,[\,\dot{p}_{ij}\,\dot{p}_{lm}\,] = E\,\left[\,\frac{\exp(Y_{ij}+u_i)}{1+\exp(Y_{ij}+u_i)}\;\frac{\exp(Y_{lm}+u_l)}{1+\exp(Y_{lm}+u_l)}\,\right]$$

$$= \int_{-\infty}^{-Y_{ij}}\int_{-\infty}^{-Y_{lm}} \Big(\sum_{k}^{\infty}(-1)^k \exp(Y_{ij}+u_i)^k\Big)\Big(\sum_{n}^{\infty}(-1)^n \exp(Y_{lm}+u_l)^n\Big) f$$

$$- \int_{-\infty}^{-Y_{ij}}\int_{-Y_{lm}}^{\infty} \Big(\sum_{k}^{\infty}(-1)^k \exp(Y_{ij}+u_i)^k\Big)\Big(1+\sum_{n}^{\infty}(-1)^n \exp(Y_{lm}+u_l)^{-n}\Big) f$$

$$- \int_{-Y_{ij}}^{\infty}\int_{-\infty}^{-Y_{lm}} \Big(1+\sum_{k}^{\infty}(-1)^k \exp(Y_{ij}+u_i)^{-k}\Big)\Big(\sum_{n}^{\infty}(-1)^n \exp(Y_{lm}+u_l)^n\Big) f$$

$$+ \int_{-Y_{ij}}^{\infty}\int_{-Y_{lm}}^{\infty} \Big(1+\sum_{k}^{\infty}(-1)^k \exp(Y_{ij}+u_i)^{-k}\Big)\Big(1+\sum_{n}^{\infty}(-1)^n \exp(Y_{lm}+u_l)^{-n}\Big) f \qquad (B.1)$$

where $f = \dfrac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(\dfrac{-(u_i^2-2\rho u_i u_l+u_l^2)}{2\sigma^2(1-\rho^2)}\right) du_i\; du_l$

For arbitrary ρ, B.1 would require special numerical procedures. For $\rho = 1$, i.e. when $u_i = u_1$, and $Y_{ij} < Y_{im}$

$$E[\dot{p}_{ij}\,\dot{p}_{im}]$$

$$= \int_{-\infty}^{-Y_{ij}} \Big(\sum_k^{\infty} (-1)^k \exp(Y_{ij}+u_i)^k\Big)\Big(\sum_n^{\infty} (-1)^n \exp(Y_{im}+u_i)^n\Big) f$$

$$- \int_{-Y_{ij}}^{-Y_{im}} \Big(1+\sum_k^{\infty} (-1)^k \exp(Y_{ij}+u_i)^{-k}\Big)\Big(\sum_n^{\infty} (-1)^n \exp(Y_{im}+u_i)^n\Big) f$$

$$+ \int_{-Y_{im}}^{\infty} \Big(1+\sum_k^{\infty} (-1)^k \exp(Y_{ij}+u_i)^{-k}\Big)\Big(1+\sum_n^{\infty} (-1)^n \exp(Y_{im}+u_i)^{-n}\Big) f$$

$$\text{where } f = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-u_i^2}{2\sigma^2}\right) du_i$$

$$= 1 - \Phi\left(\frac{-Y_{im}}{\sigma}\right)$$

$$+ \sum_k (-1)^k \exp(k^2\sigma^2/2)\ \Big\{ \exp(-kY_{ij})\left[1 - \Phi\left(\frac{-Y_{im}}{\sigma} + k\sigma\right)\right]$$

$$+ \exp(-kY_{im})\left[1 - \Phi\left(\frac{-Y_{im}}{\sigma} + k\sigma\right)\right]$$

$$- \exp(kY_{im})\left[\Phi\left(\frac{-Y_{im}}{\sigma} - k\sigma\right) - \Phi\left(\frac{-Y_{ij}}{\sigma} - k\sigma\right)\right]\Big\}$$

$$+ \sum_k \sum_n (-1)^{k+n} \Big[\exp((k+n)^2\sigma^2/2)\Big\{ \exp(kY_{ij}+nY_{im})\,\Phi\left[\frac{-Y_{ij}}{\sigma} - (k+n)\sigma\right]$$

$$+ \exp(-kY_{ij}-nY_{im})\left(1-\Phi\left[\frac{-Y_{im}}{\sigma} + (k+n)\sigma\right]\right)\Big\}$$

$$-\exp((n-k)^2\sigma^2/2)\ \exp(nY_{im}-kY_{ij})\Big\{\Phi\left[\frac{-Y_{im}}{\sigma} - (n-k)\sigma\right]$$

$$-\Phi\left[\frac{-Y_{ij}}{\sigma} - (n-k)\sigma\right]\Big\}\Big]$$

When $\dot{p}_{ij} = \dot{p}_{im}$, B.1 reduces to

$$
\begin{aligned}
E\ [\ \dot{p}_{ij}^{2}\ ] = 1 - \Phi\left(\frac{-Y_{ij}}{\sigma} \right) \\
+ \sum_{k}^{\infty} (-1)^{k}\ [\ 2\exp(k^{2}\sigma^{2}/2 + kY_{ij})\ \{1 - \Phi[\frac{-Y_{ij}}{} + k\sigma]\} \\
-k\ \exp((k+1)^{2}\sigma^{2}/2)\ \{\exp((k+1)Y_{ij})\Phi[\frac{-Y_{ij}}{\sigma} - (k+1)\sigma] \\
+ \exp(-(k+1)Y_{ij})(1-\Phi[\frac{-Y_{ij}}{\sigma} + (k+1)\sigma]\}]
\end{aligned}
$$

Appendix C

Derivation of the basic expressions required to apply the Generalised Linear Mixed Model to Multinomial data using the Multinomial Logit.

The multinomial logit (Bock 1975,p522) can be expressed as

$$p_i = \frac{\exp(\theta_i)}{1 + \sum_j \exp(\theta_j)} \quad \text{for } i = 1,\ldots c \text{ and } c + 1 \text{ categories.} \tag{C.1}$$

Extended to the mixed logit, this becomes

$$\dot{p}_i = \frac{\exp(\dot{\theta}_i)}{1 + \sum_j \exp(\dot{\theta}_j)}$$

$$= \frac{\exp(\theta_i + u_i)}{1 + \sum_j \exp(\theta_j + u_j)} \tag{C.2}$$

Now $\exp(\theta_k + u_k) = \exp(\theta_k) \exp(u_k)$

$$= t_k \; v_k \tag{C.3}$$

And $v_k = \exp(u_k)$

$$= 1 + u_k + u_k^2/2 + u_k^3/6 + u_k^4/24 + \ldots\ldots \tag{C.4}$$

Define $\underset{\sim}{u} = \{v_k\}$

$$= \underset{\sim}{1} + \underset{\sim}{v} + \underset{\sim}{w} + \underset{\sim}{x} + \underset{\sim}{y} \tag{C.5}$$

Where $\underset{\sim}{v} = \{u_k\}$ (C.6)

$\underset{\sim}{w} = \{u_k^2/2\}$ (C.7)

$\underset{\sim}{x} = \{u_k^3/6\}$ (C.8)

$\underset{\sim}{y} = \{u_k^4\} /24\}$ (C.9)

Also define

$$\underset{\sim}{t} \quad = \quad \{t_k\} \qquad \text{(C.10)}$$

$$\text{and } p \quad = \quad \{t_k\} \quad = \quad \frac{t_i}{1 + s} \qquad \text{(C.11)}$$

$$\text{where } s \quad = \quad \sum_i t_i \quad = \underset{\sim}{t}' \underset{\sim}{1}$$

$$\text{So} \quad p_i \quad = \quad \frac{\exp(\Theta_i + u_i)}{1 + \sum_j \exp(\Theta_j + u_j)}$$

$$= \frac{t_i \; v_i}{1 + \sum_j t_j \; v_j}$$

$$= \frac{t_i \; v_i}{1 + \underset{\sim}{t}' \; \underset{\sim}{u}} \qquad \text{(C.12)}$$

The distribution of u_i is unspecified but is assumed to be symmetric with the following expectations.

$$E\,(u_i) \quad = \quad 0 \qquad \text{(C.13)}$$

$$E\,(u_i{}^2) \quad = \quad \sigma_i^2 \qquad \text{(C.14)}$$

$$E\,(u_i u_j) = \sigma_{ij} \quad \text{if } u_i \text{ and } u_j \text{ are in the same group,}$$

$$= 0 \quad \text{if } u_i \text{ and } u_j \text{ are in different groups.}$$

(C.15)

$$\text{Define} \quad \underset{\sim}{\Sigma} \quad = \quad \text{Var}\,(\underset{\sim}{u}) \quad = \quad E\,(\underset{\sim}{u}\underset{\sim}{u}') \quad = \quad \{\sigma_{ij}\} \qquad \text{(C.16)}$$

$$\underset{\sim}{c}_j \quad = \quad \{\sigma_{ij}\} \quad = \quad j\text{th column of } \underset{\sim}{\Sigma} \qquad \text{(C.17)}$$

$$\underset{\sim}{d} \quad = \quad \{\sigma_i^2\} \quad = \text{vector containing diagonal of } \underset{\sim}{\Sigma}$$

(C.18)

$$\underset{\sim}{e} \quad = \quad \underset{\sim}{\Sigma} \; \underset{\sim}{p} \qquad \text{(C.19)}$$

The symbol * is used to denote the Hadamard product (Rao and Mitra, 1971).

ie $\underset{\sim}{A}_{a \times b} * \underset{\sim}{B}_{a \times b} = \{a_{ij}\, b_{ij}\} = \underset{\sim}{C}_{a \times b}$

In expressions where it occurs, this product must be evaluated before all other matrix products.

Note that for zero kurtosis and zero mean

$$E\,(u_i\, u_j\, u_k\, u_l) = \sigma_{ij}\,\sigma_{kl} + \sigma_{ik}\,\sigma_{jl} + \sigma_{il}\,\sigma_{jk} \quad (C.20)$$

$$E\,(u_i\, u_j\, u_k^2) = \sigma_{ij}\,\sigma_k^2 + 2\sigma_{ik}\,\sigma_{jk} \quad (C.21)$$

$$E\,(u_i^2\, u_j^2) = \sigma_i^2\,\sigma_j^2 + 2\sigma_{ij}^{\,2} \quad (C.22)$$

$$E\,(u_i\, u_j^3) = 3\sigma_{ij}\,\sigma_j^2 \quad (C.23)$$

$$E\,(u_i^{\,4}) = 3\sigma_i^{\,4} \quad (C.24)$$

The following series are required,

$$(1+s)^{-1} = 1 - s + s^2 - s^3 + s^4 - s^5 + \ldots.. \quad (C.25)$$

$$(1+s)^{-2} = 1 - 2s + 3s^2 - 4s^3 + 5s^4 - 6s^5 + \ldots. \quad (C.26)$$

$$(1+s)^{-3} = 1 - 3s + 6s^2 - 10s^3 + 15s^4 - \ldots. \quad (C.27)$$

$$(1+s)^{-4} = 1 - 4s + 10s^2 - 20s^3 + \ldots. \quad (C.28)$$

$$(1+s)^{-5} = 1 - 5s + 15s^2 - \ldots. \quad (C.29)$$

These series are valid only for $|s| < 1$.

Equation C.25 is now used with C.5 and C.10 to obtain an expression for the denominator of C.12.

Now $\utilde{t}'\utilde{u} = \utilde{t}'(\utilde{1}+\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y})$

$$= s+\utilde{t}'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y}) \qquad (C.30)$$

So, ignoring high powers of u,

$$(\utilde{t}'\utilde{u})^2 = s^2 + 2s\utilde{t}'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y}) + \utilde{t}'(\utilde{v}\utilde{v}'+2\utilde{v}\utilde{w}'+\utilde{w}\utilde{w}'+2\utilde{v}\utilde{x}')\utilde{t} \qquad (C.31)$$

$$(\utilde{t}'\utilde{u})^3 = s^3 + 3s^2\utilde{t}'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y}) + 3s\utilde{t}'(\utilde{v}\utilde{v}'+2\utilde{v}\utilde{w}'+\utilde{w}\utilde{w}'+2\utilde{v}\utilde{x}')\utilde{t}$$
$$+ \utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'(\utilde{v} + 3\utilde{w}) \qquad (C.32)$$

$$(\utilde{t}'\utilde{u})^4 = s^4 + 4s^3\utilde{t}'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y}) + 6s^2\utilde{t}'(\utilde{v}\utilde{v}'+2\utilde{v}\utilde{w}'+\utilde{w}\utilde{w}'+2\utilde{v}\utilde{x}')\utilde{t}$$
$$+4s\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'(\utilde{v}+3\utilde{w}) + (\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}) \qquad (C.33)$$

$$(\utilde{t}'\utilde{u})^5 = s^5+5s^4\utilde{t}'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y}) + 10s^3\utilde{t}'(\utilde{v}\utilde{v}'+2\utilde{v}\utilde{w}'+\utilde{w}\utilde{w}'+2\utilde{v}\utilde{x}')\utilde{t}$$
$$+10s^2\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'(\utilde{v}+3\utilde{w}) + 5s\,\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'\utilde{v}\utilde{v}'\utilde{t} \qquad (C.34)$$

And so, from C.25,

$$\frac{1}{(1+\utilde{t}'\utilde{u})} = 1 - (\utilde{t}'\utilde{u}) + (\utilde{t}'\utilde{u})^2 - (\utilde{t}'\utilde{u})^3 + (\utilde{t}'\utilde{u})^4 - (\utilde{t}'\utilde{u})^5 \ldots$$

$$= 1 - s + s^2 - s^3 + s^4 - s^5 + \ldots.$$
$$-\utilde{t}'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y})(1-2s+3s^2-4s^3+5s^4-\ldots..)$$
$$+\utilde{t}'(\utilde{v}\utilde{v}'+2\utilde{v}\utilde{w}'+\utilde{w}\utilde{w}'+2\utilde{v}\utilde{x}')\utilde{t}\,(1-3s+6s^2-10s^3+\ldots.)$$
$$-\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'(\utilde{v}+3\utilde{w})(1-4s+10s^2-\ldots.)$$
$$+\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\,(1-5s + \ldots..$$

$$= \frac{1}{1+s} - \frac{\utilde{t}'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y})}{(1+s)^2} + \frac{\utilde{t}'(\utilde{v}\utilde{v}'+2\utilde{v}\utilde{w}'+\utilde{w}\utilde{w}'+2\utilde{v}\utilde{x}')\utilde{t}}{(1+s)^3}$$
$$- \frac{\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'(\utilde{v}+3\utilde{w})}{(1+s)^4} + \frac{\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}\utilde{t}'\utilde{v}\utilde{v}'\utilde{t}}{(1+s)^5} \qquad (C.35)$$

But from C.11,

$$\frac{\utilde{t}}{(1+s)} = \frac{\{\ t_i\ \}}{(1+s)} = \{\ p_i\ \} = \utilde{p} \qquad (C.36)$$

From C.12,

$$\begin{aligned}\dot{p}_i &= \frac{t_i v_i}{(1 + \underset{\sim}{t}'\underset{\sim}{u})} \\ &= p_i \; (1 + u_i + u_i^2/2 + u_i^3/6 + u_i^4/24) \\ &\qquad (1 - \underset{\sim}{p}'(\underset{\sim}{v}+\underset{\sim}{w}+\underset{\sim}{x}+\underset{\sim}{y}) + \underset{\sim}{p}'(\underset{\sim}{v}\underset{\sim}{v}'+2\underset{\sim}{v}\underset{\sim}{w}'+\underset{\sim}{w}\underset{\sim}{w}'+2\underset{\sim}{v}\underset{\sim}{x}')\underset{\sim}{p} \\ &\qquad\quad - \underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}\underset{\sim}{p}'(\underset{\sim}{v}+3\underset{\sim}{w}) + \underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p} \;) \qquad (C.37)\end{aligned}$$

Expectation of $\underline{\dot{p}_i}$

The expected value of $\dot{p}_i$ is obtained as the expected value of C.37. In the expansion of C.37, terms with odd powers (which have expectation zero), and terms with powers of u_i greater than 4, have been ignored.

From C.37,

$$\begin{aligned}E\,(\dot{p}_i) = p_i \; E(1 &- \underset{\sim}{p}'(\underset{\sim}{w}+\underset{\sim}{y}) + \underset{\sim}{p}'(\underset{\sim}{v}\underset{\sim}{v}'+\underset{\sim}{w}\underset{\sim}{w}'+2\underset{\sim}{v}\underset{\sim}{x}')\underset{\sim}{p} \\ &- 3\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}\underset{\sim}{p}'\underset{\sim}{w} + \underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p} \\ &- u_i(\underset{\sim}{p}'(\underset{\sim}{v}+\underset{\sim}{x}) - 2\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{w}'\underset{\sim}{p} + \underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}\underset{\sim}{p}'\underset{\sim}{v}) \\ &+ u_i^2/2 \; (1 - \underset{\sim}{p}'\underset{\sim}{w} + \underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}) \\ &- u_i^3/6 \; \underset{\sim}{p}'\underset{\sim}{v} + u_i^4/24 \;) \qquad (C.38)\end{aligned}$$

The expected values of each of the terms in C.38 are derived in equations C.39 to C.48 using equations C.6 to C.24.

$$E\,(\underset{\sim}{p}'(\underset{\sim}{w}+\underset{\sim}{y})) = \sum_i p_i \; E\left(\frac{1}{2}u_i^2 + \frac{1}{24}u_i^4\right)$$

$$= \sum_i p_i \left(\frac{1}{2}\sigma_i^2 + \frac{3}{24}\sigma_i^4\right)$$

$$= \frac{1}{2}\underset{\sim}{p}'\underset{\sim}{d} + \frac{1}{8}\underset{\sim}{p}'\underset{\sim}{d}*\underset{\sim}{d} \qquad (C.39)$$

$$E\,(\underset{\sim}{p}'(\underset{\sim}{v}\underset{\sim}{v}' + \underset{\sim}{w}\underset{\sim}{w}' + 2\underset{\sim}{v}\underset{\sim}{x}')\underset{\sim}{p})$$

$$= \sum_i \sum_j p_i p_j \; E\left(u_i u_j + \frac{1}{4}u_i^2 u_j^2 + \frac{2}{6}u_i u_j^3\right)$$

$$= \sum_i \sum_j p_i p_j \left(\sigma_{ij} + \frac{1}{4}(\sigma_i^2\sigma_j^2 + 2\sigma_{ij}^2) + \frac{1}{3}(3\sigma_{ij}\sigma_j^2)\right)$$

$$= \underset{\sim}{p}'\Sigma\underset{\sim}{p} + \frac{1}{4}\underset{\sim}{p}'\underset{\sim}{d}\underset{\sim}{p}'\underset{\sim}{d} + \frac{1}{2}\underset{\sim}{p}'\Sigma_*\Sigma\underset{\sim}{p} + \underset{\sim}{p}'\underset{\sim}{d}*\underset{\sim}{e} \qquad (C.40)$$

$$E\,(3\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}\underset{\sim}{p}'\underset{\sim}{w}) = \frac{3}{2}\sum_i \sum_j \sum_k p_i\, p_j\, p_k \; E\,(u_i\, u_j\, u_k^2)$$

$$= \frac{3}{2}\sum_i \sum_j \sum_k p_i\, p_j\, p_k \; (\sigma_{ij}\,\sigma_k^2 + 2\sigma_{ik}\,\sigma_{jk})$$

$$= \frac{3}{2}\underset{\sim}{p}'\,\underset{\sim}{e}\,\underset{\sim}{p}'\underset{\sim}{d} + 3\,\underset{\sim}{p}'\,\underset{\sim}{e}*\underset{\sim}{e} \qquad (C.41)$$

$$E\,(\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}) = \sum_i \sum_j \sum_k \sum_l p_i\, p_j\, p_k\, p_l \; E\,(u_i\, u_j\, u_k\, u_l)$$

$$= \sum_i \sum_j \sum_k \sum_l p_i\, p_j\, p_k\, p_l \; (\sigma_{ij}\,\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$$

$$= 3\;\underset{\sim}{p}'\underset{\sim}{e}\;\underset{\sim}{p}'\underset{\sim}{e} \qquad (C.42)$$

$$E\,(u_i\underset{\sim}{p}'(\underset{\sim}{v}+\underset{\sim}{x})) = \sum_j p_j \; E\left(u_i u_j + \frac{1}{6}u_i u_j^3\right)$$

$$= \sum_j p_j \left(\sigma_{ij} + \frac{1}{2}\sigma_{ij}\sigma_j^2\right)$$

$$= \underset{\sim}{p}'\underset{\sim}{c}_i + \frac{1}{2}\underset{\sim}{p}'\underset{\sim}{d}*\underset{\sim}{c}_i \qquad (C.43)$$

$$E\,(2u_i\,\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{w}\underset{\sim}{p}) = 2\sum_j\sum_k p_j\,p_k\,E\,(u_i\,u_j\,\tfrac{1}{2}\,u_k^2)$$

$$= \sum_j\sum_k p_j\,p_k\,(\sigma_{ij}\,\sigma_k^2 + 2\sigma_{ik}\,\sigma_{jk})$$

$$= \underset{\sim}{p}'\underset{\sim}{c}_i\,\underset{\sim}{p}'\underset{\sim}{d} + 2\,\underset{\sim}{p}'\underset{\sim}{c}_i{*}\underset{\sim}{e} \qquad \text{(C.44)}$$

$$E\,(u_i\,\underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}\underset{\sim}{p}'\underset{\sim}{v}) = \sum_j\sum_k\sum_l p_j\,p_k\,p_l\;E\,(u_i\,u_j\,u_k\,u_l)$$

$$= \sum_j\sum_k\sum_l p_j\,p_k\,p_l\;(\sigma_{ij}\,\sigma_{kl} + \sigma_{ik}\,\sigma_{jl} + \sigma_{il}\,\sigma_{jk})$$

$$= 3\,\underset{\sim}{p}'\underset{\sim}{e}\,\underset{\sim}{p}'\underset{\sim}{c}_i \qquad \text{(C.45)}$$

$$E\;(u_i^2/2\;\;(1 - \underset{\sim}{p}'\underset{\sim}{w} + \underset{\sim}{p}'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}))$$

$$= \tfrac{1}{2}\{E\,(u_i^2) - \sum_j p_j\;E\,(\tfrac{1}{2}u_i^2\,u_j^2) + \sum_j\sum_k p_j\,p_k E(u_i^2\,u_j\,u_k)\,\}$$

$$= \tfrac{1}{2}\,\sigma_i^2 - \tfrac{1}{4}\sum_j p_j\,(\sigma_i^2\sigma_j^2 + 2\sigma_{ij}^{\,2}) + \tfrac{1}{2}\sum_j\sum_k p_j p_k\,(\sigma_i^2\sigma_{jk} + 2\sigma_{ij}\sigma_{ik})$$

$$= \tfrac{1}{2}\,\sigma_i^2 - \tfrac{1}{4}\,\sigma_i^2\,\underset{\sim}{p}'\underset{\sim}{d} - \tfrac{1}{2}\,\underset{\sim}{p}'\underset{\sim}{c}^{*}_i\underset{\sim}{c}_i + \tfrac{1}{2}\,\sigma_i^2\,\underset{\sim}{p}'\underset{\sim}{\Sigma}\underset{\sim}{p} + \underset{\sim}{p}'\underset{\sim}{c}_i\,\underset{\sim}{p}'\underset{\sim}{c}_i$$

$$= \sigma_i^2(\tfrac{1}{2} - \tfrac{1}{4}\underset{\sim}{p}'\underset{\sim}{d} + \tfrac{1}{2}\underset{\sim}{p}'\underset{\sim}{e}) - \tfrac{1}{2}\underset{\sim}{p}'\underset{\sim}{c}_i{*}\underset{\sim}{c}_i + \underset{\sim}{p}'\underset{\sim}{c}_i\underset{\sim}{p}'\underset{\sim}{c}_i \qquad \text{(C.46)}$$

$$E\,(u_i^3/6\;\underset{\sim}{p}'\underset{\sim}{v}) = \tfrac{1}{6}\sum_j p_j\;E\,(u_i^3\,u_j)$$

$$= \tfrac{1}{6}\sum_j p_j\,(3\sigma_i^2\,\sigma_{ij})$$

$$= \tfrac{1}{2}\;\sigma_i^2\,\underset{\sim}{p}'\underset{\sim}{c}_i \qquad \text{(C.47)}$$

$$E\,(u_i^4/24\,) = \tfrac{1}{8}\,\sigma_i^4 \qquad \text{(C.48)}$$

Substituting C.39 to C.48 in C.38 gives

$$\begin{aligned}
E\,(\dot{p}_i) = p_i\ \Big(1 &- \tfrac{1}{2}\underset{\sim}{p}'\underset{\sim}{d} - \tfrac{1}{8}\underset{\sim}{p}'\underset{\sim}{d}*\underset{\sim}{d} + \underset{\sim}{p}'\underset{\sim}{e} + \tfrac{1}{4}\underset{\sim}{p}'\underset{\sim}{d}\ \underset{\sim}{p}'\underset{\sim}{d} + \tfrac{1}{2}\underset{\sim}{p}'\underset{\sim}{\Sigma}*\underset{\sim}{\Sigma}\underset{\sim}{p} + \underset{\sim}{p}'\underset{\sim}{d}*\underset{\sim}{e} \\
&- \tfrac{3}{2}\,\underset{\sim}{p}'\underset{\sim}{e}\ \underset{\sim}{p}'\underset{\sim}{d} - 3\underset{\sim}{p}'\underset{\sim}{e}*\underset{\sim}{e} + 3\underset{\sim}{p}'\underset{\sim}{e}\underset{\sim}{p}'\underset{\sim}{e} \\
&- \underset{\sim}{p}'\underset{\sim}{c}_i - \tfrac{1}{2}\underset{\sim}{p}'\underset{\sim}{d}*\underset{\sim}{c}_i + \underset{\sim}{p}'\underset{\sim}{c}_i\ \underset{\sim}{p}'\underset{\sim}{d} + 2\underset{\sim}{p}'\underset{\sim}{c}_i*\underset{\sim}{e} - 3\underset{\sim}{p}'\underset{\sim}{e}\underset{\sim}{p}'\underset{\sim}{c}_i \\
&+ \tfrac{1}{2}\sigma_i^2\ (1-\tfrac{1}{2}\underset{\sim}{p}'\underset{\sim}{d} + \underset{\sim}{p}'\underset{\sim}{e} - \underset{\sim}{p}'\underset{\sim}{c}_i) \\
&- \tfrac{1}{2}\underset{\sim}{p}'\underset{\sim}{c}_i*\underset{\sim}{c}_i + \underset{\sim}{p}'\underset{\sim}{c}_i\underset{\sim}{p}'\underset{\sim}{c}_i + \tfrac{1}{8}\sigma_i^4\Big) \qquad (C.49)
\end{aligned}$$

$$= p_i\ (1 - \tfrac{1}{2}\ \underset{\sim}{p}'\underset{\sim}{d} + \underset{\sim}{p}'\underset{\sim}{e} - \underset{\sim}{p}'\underset{\sim}{c}_i + \tfrac{1}{2}\sigma_i^2\ ) \tag{C.50}$$

if fourth power terms are ignored.

Simplifying to the binomial logit,

$$= \ \sigma^2 = \sigma_i^2 = d = \Sigma = c_i$$

$$\begin{aligned}
E\,(\dot{p}) = p\ \big(1 &+ \sigma^2(\tfrac{1}{2}(1-p) - p(1-p)) \\
&+ \sigma^4(-\tfrac{1}{8}\,p + (\tfrac{1}{4}+\tfrac{1}{2}+1)\ p^2 - \tfrac{9}{2}\ p^3 + 3p^4 - \tfrac{1}{2}p \\
&\qquad + 3p^2 - 3p^3 - \tfrac{1}{4}p + \tfrac{1}{2}p^2 - \tfrac{1}{2}p + p^2 + \tfrac{1}{8}\ )\big)
\end{aligned}$$

$$= p\ \ (1+\sigma^2(1-p)(\tfrac{1}{2}-p) + \sigma^4(1-p)(\ \tfrac{1}{8} - \tfrac{7}{4}\ p + \tfrac{9}{2}\ p^2 - 3p^3))$$

$$= p\ \ + \sigma^2 p(1-p)(\tfrac{1}{2}-p)(1+\sigma^2(\tfrac{1}{4}-3p(1-p))$$

$$= p\ \ + \sigma^2 pq(\tfrac{1}{2}-p)(1+\sigma^2(\tfrac{1}{4}-3pq)) \tag{C.51}$$

where $q = 1-p$

When the fourth power term is omitted, C.51 becomes

$$E\ (\dot{p}) = p + \sigma^2 pq(\tfrac{1}{2}-p) \tag{C.52}$$

Expectation of $\dot{p}_i \dot{p}_j$

The expectation of $\dot{p}_i \dot{p}_j$ is obtained for the most general situation where $\dot{p}_i$ and $\dot{p}_j$ are assumed to be from different data records. It is further assumed that the correlation between u_i and u_j is $r\,\sigma_{ij}$ where r is a constant which is specific to the pair of data records.

For convenience, subscripts 1 and 2 are used to denote different records. Subscripts have been omitted from $\utilde{v},\utilde{w},\utilde{x}$, and $\utilde{y}$ as the subscript of the adjacent $\utilde{p}$ applies.

From C.12 and C.37,

$$
\begin{aligned}
\dot{p}_i \dot{p}_j &= \frac{t_{i1}\; t_{j2}\; \exp(u_{i1}+u_{j2})}{(1+\utilde{t}_1'\utilde{u}_1)(1+\utilde{t}_2'\utilde{u}_2)} \\
&= p_{i1}p_{j2}(1+u_{i1}+u_{j2}+(u_{i1}+u_{j2})^2/2+(u_{i1}+u_{j2})^3/6 \\
&\qquad\qquad +(u_{i1}+u_{j2})^4/24) \\
&\quad (1-\utilde{p}_1'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y})+\utilde{p}_1'(\utilde{v}\utilde{v}'+2\utilde{v}\utilde{w}'+\utilde{w}\utilde{w}'+2\utilde{v}\utilde{x}')\utilde{p}_1 \\
&\qquad -\utilde{p}_1'\utilde{v}\utilde{v}'\utilde{p}_1\utilde{p}_1'(\utilde{v}+3\utilde{w}) + \utilde{p}_1'\utilde{v}\utilde{v}'\utilde{p}\utilde{p}_1'\utilde{v}\utilde{v}'\utilde{p}_1) \\
&\quad (1-\utilde{p}_2'(\utilde{v}+\utilde{w}+\utilde{x}+\utilde{y}) + \utilde{p}_2'(\utilde{v}\utilde{v}'+2\utilde{v}\utilde{w}'+\utilde{w}\utilde{w}'+2\utilde{v}\utilde{x}')\utilde{p}_2 \\
&\qquad -\utilde{p}_2'\utilde{v}\utilde{v}'\utilde{p}_2\utilde{p}_2'(\utilde{v}+3\utilde{w}) + \utilde{p}_2'\utilde{v}\utilde{v}'\utilde{p}_2\utilde{p}_2'\utilde{v}\utilde{v}'\utilde{p}_2) \qquad (C.53)
\end{aligned}
$$

In expanding this expression, terms with odd powers of u are omitted because they have expectation zero. Terms with powers of u greater than four are also omitted.

Expanding C.53,

$$
\begin{aligned}
E\,(\dot{p}_{i1}\ \dot{p}_{j2}) = p_{i1}\ p_{j2}\ & E\,(1 - \underset{\sim}{p}_1'(\underset{\sim}{w}+\underset{\sim}{y}) - \underset{\sim}{p}_2'(\underset{\sim}{w}+\underset{\sim}{y}) + \underset{\sim}{p}_1'(\underset{\sim}{v}\underset{\sim}{v}'+\underset{\sim}{w}\underset{\sim}{w}'+\underset{\sim}{v}\underset{\sim}{x}'+\underset{\sim}{x}\underset{\sim}{v}')\underset{\sim}{p}_2 \\
& + \underset{\sim}{p}_1'(\underset{\sim}{v}\underset{\sim}{v}'+\underset{\sim}{w}\underset{\sim}{w}'+2\underset{\sim}{v}\underset{\sim}{x}')\underset{\sim}{p}_1+\underset{\sim}{p}_2'(\underset{\sim}{v}\underset{\sim}{v}'+\underset{\sim}{w}\underset{\sim}{w}'+2\underset{\sim}{v}\underset{\sim}{x}')\underset{\sim}{p}_2 \\
& - 3\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_1'\underset{\sim}{w} - \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{w} - 2\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_1'\underset{\sim}{w} \\
& - 3\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{w} - \underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_1'\underset{\sim}{w} - 2\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{w} \\
& + \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1 + \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2 \\
& + \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2 + \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2+\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2 \\
& +(u_{i1}+u_{j2})\,(-\underset{\sim}{p}_1'(\underset{\sim}{v}+\underset{\sim}{x})-\underset{\sim}{p}_2'(\underset{\sim}{v}+\underset{\sim}{x})+\underset{\sim}{p}_1'(\underset{\sim}{v}\underset{\sim}{w}'+\underset{\sim}{w}\underset{\sim}{v}')\underset{\sim}{p}_2 \\
& \qquad +2\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{w}'\underset{\sim}{p}_1+2\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{w}'\underset{\sim}{p}_2-\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_1'\underset{\sim}{v}) \\
& \qquad -\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{v}-\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{v}-\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{v}) \\
& + \frac{(u_{i1}+u_{j2})^2}{2}\ (1 - \underset{\sim}{p}_1'\underset{\sim}{w} - \underset{\sim}{p}_2'\underset{\sim}{w} \\
& \qquad + \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1 + \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2 + \underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2) \\
& + \frac{(u_{i1}+u_{j2})^3}{6}\ (-\underset{\sim}{p}_1'\underset{\sim}{v}-\underset{\sim}{p}_2'\underset{\sim}{v}) \\
& + \frac{(u_{i1}+u_{j2})^4}{24}\ ) \qquad (C.54)
\end{aligned}
$$

The expectation of many of these terms has already been obtained in equations C.39 to C.48. The remaining expectations required are:

$$E\,(\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2) = \sum_i \sum_j p_i\, p_j\, E\,(u_{i1}\, u_{j2})$$

$$= \sum_i \sum_j p_i\, p_j\;\, r\, \sigma_{ij}$$

$$= r\, \underset{\sim}{p}_1'\, \underset{\sim}{\Sigma}\, \underset{\sim}{p}_2 \qquad \text{(C.55)}$$

$$E\,(\underset{\sim}{p}_1'\underset{\sim}{w}\underset{\sim}{w}'\underset{\sim}{p}_2) = \sum_i \sum_j p_i\, p_j\, \tfrac{1}{4}\, E(u_{i1}^2 u_{j2}^2)$$

$$= \sum_i \sum_j p_i\, p_j\, \tfrac{1}{4}(\sigma_i^2\, \sigma_j^2 + 2r^2\sigma_{ij}{}^2)$$

$$= \tfrac{1}{4}\underset{\sim}{p}_1'\underset{\sim}{d}\; \underset{\sim}{p}_2'\underset{\sim}{d} + \frac{r^2}{2}\, \underset{\sim}{p}_1'\underset{\sim}{\Sigma}*\underset{\sim}{\Sigma}\; \underset{\sim}{p}_2 \qquad \text{(C.56)}$$

$$E\,(\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{x}'\underset{\sim}{p}_2) = \sum_i \sum_j\; p_i\, p_j\, \frac{1}{6}\, E\,(u_{i1}\, u_{j2}^3)$$

$$\sum_i \sum_j p_i\, p_j\, \frac{1}{6}(3r\, \sigma_{ij}\, \sigma_j{}^2)$$

$$= \frac{r}{2}\; \underset{\sim}{p}_2'\underset{\sim}{d}*\underset{\sim}{d}\underset{\sim}{e}_1 \qquad \text{(C.57)}$$

In similar fashion,

$$E\,(\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{w}) = \tfrac{1}{2}\, \underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{d} + r^2\underset{\sim}{p}_2'\; \underset{\sim}{e}_1*\, \underset{\sim}{e}_2 \qquad \text{(C.58)}$$

$$E\,(\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_1'\underset{\sim}{w}) = \frac{r}{2}\, \underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_2\underset{\sim}{p}_1'\underset{\sim}{d} + r\, \underset{\sim}{p}_1'\; \underset{\sim}{e}_1*\underset{\sim}{e}_2 \qquad \text{(C.59)}$$

$$E\,(\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2) = 3r\; \underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{\Sigma}\underset{\sim}{p}_2 \qquad \text{(C.60)}$$

$$E\,(\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2) = \underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{\Sigma}\underset{\sim}{p}_2 + 2r^2\underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_2\underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_2 \qquad \text{(C.61)}$$

$$E\ (u_{i1}\ \underset{\sim}{p}_2'(\underset{\sim}{v}+\underset{\sim}{x})) = r\ \underset{\sim}{p}_2'\underset{\sim}{c}_i + \tfrac{1}{2}r\ \underset{\sim}{p}_2'\underset{\sim}{d}*\underset{\sim}{c}_i \tag{C.62}$$

$$E\ (u_{i1}\ \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{w}'\underset{\sim}{p}_2) = \tfrac{1}{2}\sum_j\sum_k p_j p_k\ \ E\ (u_{i1}u_{j1}u_k^{2}) \tag{C.63}$$

$$= \tfrac{1}{2}\ \underset{\sim}{p}_1'\ \underset{\sim}{c}_i\underset{\sim}{p}_2'\underset{\sim}{d} + r^2\underset{\sim}{p}_2'\ \underset{\sim}{c}_i*\ \underset{\sim}{e}_1 \tag{C.63}$$

$$E\ (u_{i1}\ \underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{w}'\underset{\sim}{p}_2) = \frac{r}{2}\ \underset{\sim}{p}_2'\underset{\sim}{c}_i\underset{\sim}{p}_2'\underset{\sim}{d} + r\ \underset{\sim}{p}_2'\ \underset{\sim}{c}_i*\underset{\sim}{e}_2 \tag{C.64}$$

$$E\ (u_{i1}\ \underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{v}) = 3r\underset{\sim}{p}_2'\underset{\sim}{\Sigma}\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{c}_i \tag{C.65}$$

$$E\ (u_{i1}\ \underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2\underset{\sim}{p}_1'\underset{\sim}{v}) = \underset{\sim}{p}_2'\underset{\sim}{\Sigma}\underset{\sim}{p}_2\underset{\sim}{p}_1'\underset{\sim}{c}_i + 2r^2\underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_2\underset{\sim}{p}_2'\underset{\sim}{c}_i \tag{C.66}$$

$$E\ (u_{i1}\ \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{v}) = r\ \underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_1\underset{\sim}{p}_2'\underset{\sim}{c}_i + 2r\ \underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_2\underset{\sim}{p}_1'\underset{\sim}{c}_i \tag{C.67}$$

$$E\ (\tfrac{1}{2}u_{i1}{}^2\underset{\sim}{p}_2'\underset{\sim}{w}) = \tfrac{1}{4}\sigma_i^{2}\ \ \underset{\sim}{p}_2'\underset{\sim}{d} + \frac{r}{2}{}^{2}\ \underset{\sim}{p}_2'\underset{\sim}{c}_i*\underset{\sim}{c}_i \tag{C.68}$$

$$E\ (\tfrac{1}{2}u_{i1}{}^2\underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2) = \frac{r}{2}\sigma_i{}^2\ \underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_2 + r\ \underset{\sim}{p}_i'\underset{\sim}{c}_i\ \underset{\sim}{p}_2'\ \underset{\sim}{c}_i \tag{C.69}$$

$$E\ (\tfrac{1}{2}u_{i1}{}^2\underset{\sim}{p}_2'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2) = \tfrac{1}{2}\sigma_i{}^2\ \underset{\sim}{p}_2'\underset{\sim}{\Sigma}\underset{\sim}{p}_2 + r^{2}\ \underset{\sim}{p}_2'\underset{\sim}{c}_i\ \underset{\sim}{p}_2'\ \underset{\sim}{c}_i \tag{C.70}$$

$$E\ (u_{i1}\ u_{j2}) = r\sigma_{ij} \tag{C.71}$$

$$E\ (u_{i1}\ u_{j2}\ \underset{\sim}{p}_1'\underset{\sim}{w}) = \frac{r}{2}\sigma_{ij}\ \underset{\sim}{p}_1'\underset{\sim}{d} + r\ \underset{\sim}{p}_1'\ \underset{\sim}{c}_i*\underset{\sim}{c}_j \tag{C.72}$$

$$E\ (u_{i1}\ u_{j2}\ \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_1) = r\sigma_{ij}\ \underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_1 + 2r\ \underset{\sim}{p}_1'\ \underset{\sim}{c}_i\ \underset{\sim}{p}_1'\underset{\sim}{c}_j \tag{C.73}$$

$$E\ (u_{i1}\ u_{j2}\ \underset{\sim}{p}_1'\underset{\sim}{v}\underset{\sim}{v}'\underset{\sim}{p}_2) = r^2\sigma_{ij}\underset{\sim}{p}_1'\underset{\sim}{\Sigma}\underset{\sim}{p}_2 + \underset{\sim}{p}_1'\underset{\sim}{c}_i\underset{\sim}{p}_2'\underset{\sim}{c}_j + r^2\underset{\sim}{p}_1'\underset{\sim}{c}_j\underset{\sim}{p}_2'\underset{\sim}{c}_i \tag{C.74}$$

$$E\ (\tfrac{1}{6}(u_{i1}+u_{j2})^3\underset{\sim}{p}_1'\underset{\sim}{v}) = \tfrac{1}{2}(\sigma_i{}^2+2r\sigma_{ij}+\sigma_j{}^2)(p_1'c_i+rp_1'c_j) \tag{C.75}$$

$$E\ (\tfrac{1}{24}\ (u_{i1} + u_{j2})^4) = \frac{1}{8}\ (\sigma_i{}^2 + 2r\sigma_{ij} + \sigma_j{}^2)^2 \tag{C.76}$$

Substituting C.39 to C.48 and C.55 to C.76 into C.54 gives

$$E\ (\dot{p}_{i1}\ \dot{p}_{j2}) = p_{i1}\ p_{j2}$$

$$1 - (\utilde{p}_1 + \utilde{p}_2)'\ (\tfrac{1}{2}\utilde{d} + \tfrac{1}{8}\ \utilde{d}{*}\utilde{d})$$

$$+\ \utilde{p}_1'\ \utilde{e}_1 + r\utilde{p}_1'\utilde{e}_2 + \utilde{p}_2'\utilde{e}_2$$

$$+\ \tfrac{1}{4}(\utilde{p}_1'\utilde{d}\utilde{d}'\utilde{p}_1 + \utilde{p}_1'\utilde{d}\utilde{d}'\utilde{p}_2 + \utilde{p}_2'\utilde{d}\utilde{d}'\utilde{p}_2)$$

$$+\ \tfrac{1}{2}(\utilde{p}_1'\ \utilde{\Sigma}*\utilde{\Sigma}\utilde{p}_1 + r^2\utilde{p}_1'\utilde{\Sigma}*\utilde{\Sigma}\ \utilde{p}_2 + \utilde{p}_2'\utilde{\Sigma}*\utilde{\Sigma}\utilde{p}_2)$$

$$+\ \utilde{p}_1'\utilde{d}*\utilde{e}_1 + \tfrac{1}{2}r(\utilde{p}_1'\utilde{d}*\utilde{e}_2 + \utilde{p}_2'\utilde{d}*\utilde{e}_1) + \utilde{p}_2'\utilde{d}*\utilde{e}_2$$

$$-\ 3(\tfrac{1}{2}\utilde{p}_1'\utilde{e}_1\utilde{p}_1'\utilde{d} + \utilde{p}_1'\utilde{e}_1*\utilde{e}_1 + \tfrac{1}{2}\utilde{p}_2'\utilde{e}_2\utilde{p}_2'\utilde{d} + \utilde{p}_2'\utilde{e}_2*\utilde{e}_2)$$

$$-\ (\tfrac{1}{2}\ \utilde{p}_1'\utilde{e}_1\utilde{p}_2'\utilde{d} + r^2\utilde{p}_2\utilde{e}_1*\utilde{e}_1 + \tfrac{1}{2}\utilde{p}_2'\utilde{e}_2\utilde{p}_1'\utilde{d} + r^2\utilde{p}_1'\utilde{e}_2*\utilde{e}_2)$$

$$-\ r(\utilde{p}_1'\utilde{e}_2\utilde{p}_1'\utilde{d} + 2\utilde{p}_1'\utilde{e}_1*\utilde{e}_2 + \utilde{p}_1'\utilde{e}_2\utilde{p}_2'\utilde{d} + 2\utilde{p}_2'\utilde{e}_1*\utilde{e}_2)$$

$$+\ 3(\utilde{p}_1'\utilde{e}_1\utilde{p}_1'\utilde{e}_1 + r\utilde{p}_1'\utilde{e}_1\utilde{p}_1'\utilde{e}_2 + r\utilde{p}_1'\utilde{e}_2\utilde{p}_2'\utilde{e}_2 + \utilde{p}_2'\utilde{e}_2\utilde{p}_2'\utilde{e}_2)$$

$$+\ \utilde{p}_1'\utilde{e}_1'\utilde{p}_2\utilde{e}_2 + 2r^2\utilde{p}_1'\utilde{e}_2\utilde{p}_1'\utilde{e}_2$$

$$-\ \utilde{p}_1'\ (\utilde{c}_i + r\utilde{c}_j + \tfrac{1}{2}\utilde{d}*\utilde{c}_i + \tfrac{1}{2}r\utilde{d}*\utilde{c}_j)$$

$$-\ \utilde{p}_2'\ (\utilde{c}_j + r\utilde{c}_i + \tfrac{1}{2}\utilde{d}*\utilde{c}_j + \tfrac{1}{2}r\utilde{d}*\utilde{c}_i)$$

(continued)

$$+ \tfrac{1}{2}\underset{\sim}{p}_1{}'(\underset{\sim}{c}_i+r\underset{\sim}{c}_j)\underset{\sim}{p}_2{}'\underset{\sim}{d} + r\underset{\sim}{p}_2{}'\underset{\sim}{e}_1*(\underset{\sim}{c}_j+r\underset{\sim}{c}_i)$$

$$+ \tfrac{1}{2}\underset{\sim}{p}_2{}'(\underset{\sim}{c}_j+r\underset{\sim}{c}_i)\underset{\sim}{p}_1{}'\underset{\sim}{d} + r\underset{\sim}{p}_1{}'\underset{\sim}{e}_2*(\underset{\sim}{c}_i+r\underset{\sim}{c}_j)$$

$$+ \underset{\sim}{p}_1{}'(\underset{\sim}{c}_i+r\underset{\sim}{c}_j)\underset{\sim}{p}_1{}'\underset{\sim}{d} + 2\underset{\sim}{p}_1{}'\underset{\sim}{e}_1*(\underset{\sim}{c}_i+r\underset{\sim}{c}_j)$$

$$+ \underset{\sim}{p}_2{}'(\underset{\sim}{c}_j+r\underset{\sim}{c}_i)\underset{\sim}{p}_2{}'\underset{\sim}{d} + 2\underset{\sim}{p}_2{}'\underset{\sim}{e}_2*(\underset{\sim}{c}_j+r\underset{\sim}{c}_i)$$

$$- 3\underset{\sim}{p}_1{}'\underset{\sim}{e}_1\underset{\sim}{p}_1{}'(\underset{\sim}{c}_i+r\underset{\sim}{c}_j)-3\underset{\sim}{p}_2{}'\underset{\sim}{e}_2\underset{\sim}{p}_2{}'(\underset{\sim}{c}_j+r\underset{\sim}{c}_i)$$

$$- \underset{\sim}{p}_1{}'\underset{\sim}{e}_1\underset{\sim}{p}_2{}'(\underset{\sim}{c}_j+r\underset{\sim}{c}_i)-\underset{\sim}{p}_2{}'\underset{\sim}{e}_2\underset{\sim}{p}_1{}'(\underset{\sim}{c}_i+r\underset{\sim}{c}_j)$$

$$- 2r\underset{\sim}{p}_1{}'\underset{\sim}{e}_2(\underset{\sim}{p}_1{}'(\underset{\sim}{c}_i+r\underset{\sim}{c}_j)+\underset{\sim}{p}_2{}'(\underset{\sim}{c}_j+r\underset{\sim}{c}_i))$$

$$+ \tfrac{1}{2}\ (\sigma_i{}^2+2r\sigma_{ij}+\sigma_j{}^2)$$

$$- \tfrac{1}{4}\sigma_i{}^2(\underset{\sim}{p}_1+\underset{\sim}{p}_2)'d- \tfrac{1}{2}(\underset{\sim}{p}_1+r^2\underset{\sim}{p}_2)'\underset{\sim}{c}_i*\underset{\sim}{c}_i$$

$$- \tfrac{1}{4}\sigma_j{}^2(\underset{\sim}{p}_1+\underset{\sim}{p}_2)'d- \tfrac{1}{2}(\underset{\sim}{p}_2+r^2\underset{\sim}{p}_1)'\underset{\sim}{c}_j*\underset{\sim}{c}_j$$

$$- \tfrac{1}{2}r\sigma_{ij}(\underset{\sim}{p}_1+\underset{\sim}{p}_2)'d- r(\underset{\sim}{p}_1+\underset{\sim}{p}_2)'\underset{\sim}{c}_i*\underset{\sim}{c}_j$$

$$+ \tfrac{1}{2}r(\sigma_i{}^2+2r\sigma_{ij}+\sigma_j{}^2)\ \underset{\sim}{p}_1{}'\underset{\sim}{e}_2 + \underset{\sim}{p}_1{}'\underset{\sim}{c}_i\ \underset{\sim}{p}_2{}'\underset{\sim}{c}_j$$

$$+ r\underset{\sim}{p}_1{}'\underset{\sim}{c}_i\underset{\sim}{p}_2{}'\underset{\sim}{c}_i + r\underset{\sim}{p}_1{}'\underset{\sim}{c}_j\ \underset{\sim}{p}_2{}'\underset{\sim}{c}_j + r^2\underset{\sim}{p}_1{}'\underset{\sim}{c}_j\ \underset{\sim}{p}_2{}'\underset{\sim}{c}_i$$

(continued)

$$+ \tfrac{1}{2}(\sigma_i{}^2+2r\sigma_{ij}+\sigma_j{}^2)(\underset{\sim}{p}_1{}'\underset{\sim}{e}_1+\underset{\sim}{p}_2{}'\underset{\sim}{e}_2)$$

$$+ \underset{\sim}{p}_1{}'\underset{\sim}{c}_i\underset{\sim}{p}_1{}'(\underset{\sim}{c}_i+2r\underset{\sim}{c}_j) + \underset{\sim}{p}_2{}'\underset{\sim}{c}_j\ \underset{\sim}{p}_2{}'(\underset{\sim}{c}_j+2r\underset{\sim}{c}_i)$$

$$+ r^2\underset{\sim}{p}_2{}'\underset{\sim}{c}_i\underset{\sim}{p}_2{}'\underset{\sim}{c}_i + r^2\underset{\sim}{p}_1{}'\underset{\sim}{c}_j\underset{\sim}{p}_1{}'\underset{\sim}{c}_j$$

$$- \tfrac{1}{2}(\sigma_i{}^2+2r\sigma_{ij}+\sigma_j{}^2)(\underset{\sim}{p}_1{}'\underset{\sim}{c}_i + \underset{\sim}{p}_2{}'\underset{\sim}{c}_j)$$

$$- \tfrac{1}{2}r(\sigma_i{}^2+2r\sigma_{ij} + \sigma_j{}^2)(\underset{\sim}{p}_1{}'\underset{\sim}{c}_j + \underset{\sim}{p}_2{}'\underset{\sim}{c}_i)$$

$$+ \frac{1}{8}(\sigma_i{}^2 + 2r\sigma_{ij} + \sigma_j{}^2)(\sigma_i{}^2+ 2r\sigma_{ij} + \sigma_j{}^2)) \qquad (C.77)$$

which upon rearrangement equals

$$p_{i_1}p_{j_2}\{(1+\tfrac{1}{2}(\sigma_i{}^2 + 2r\sigma_{ij} + \sigma_j{}^2))\ \ (1 + r\underset{\sim}{p}_1{}'\underset{\sim}{e}_2$$

$$+ \underset{\sim}{p}_1{}'(\underset{\sim}{e}_1 - \tfrac{1}{2}\underset{\sim}{d} - \underset{\sim}{c}_i - r\underset{\sim}{c}_j)$$

$$+ \underset{\sim}{p}_2{}'(\underset{\sim}{e}_2 - \tfrac{1}{2}\underset{\sim}{d} - \underset{\sim}{c}_j - r\underset{\sim}{c}_i))$$

$$+ (\underset{\sim}{p}_1{}'(\underset{\sim}{e}_1 - \tfrac{1}{2}\underset{\sim}{d} - \underset{\sim}{c}_i - r\underset{\sim}{c}_j))^2$$

$$+ (\underset{\sim}{p}_2{}'(\underset{\sim}{e}_2 - \tfrac{1}{2}\underset{\sim}{d} -\underset{\sim}{c}_j - r\underset{\sim}{c}_i))^2$$

$$+ \underset{\sim}{p}_1{}'(\underset{\sim}{e}_1-\tfrac{1}{2}\underset{\sim}{d}-\underset{\sim}{c}_i-r\underset{\sim}{c}_j)\ \underset{\sim}{p}_2{}'(\underset{\sim}{e}_2-\tfrac{1}{2}\underset{\sim}{d}-\underset{\sim}{c}_j-r\underset{\sim}{c}_i)$$

$$+ \frac{1}{8}\ ((\sigma_i{}^2+2r\sigma_{ij}+\sigma_j{}^2)^2 - (\underset{\sim}{p}_1+\underset{\sim}{p}_2)'\underset{\sim}{d}*\underset{\sim}{d})$$

$$+ \tfrac{1}{2}\ (\underset{\sim}{p}_1{}'\underset{\sim}{\Sigma}*\underset{\sim}{\Sigma}\underset{\sim}{p}_1 + \underset{\sim}{p}_2{}'\underset{\sim}{\Sigma}*\underset{\sim}{\Sigma}\underset{\sim}{p}_2 + r^2\underset{\sim}{p}_1{}'\underset{\sim}{\Sigma}*\underset{\sim}{\Sigma}\underset{\sim}{p}_2)$$

(continued)

$$
\begin{aligned}
&+ \tfrac{1}{2}r(\underset{\sim}{p}_1{}'\underset{\sim}{d}*\underset{\sim}{e}_2 + \underset{\sim}{p}_2{}'\underset{\sim}{d}*\underset{\sim}{e}_1) + \underset{\sim}{p}_1{}'\underset{\sim}{d}*\underset{\sim}{e}_1 + \underset{\sim}{p}_2{}'\underset{\sim}{d}*\underset{\sim}{e}_2 \\
&- (3\underset{\sim}{p}_1 + r^2\underset{\sim}{p}_2)'\underset{\sim}{e}_1*\underset{\sim}{e}_1 - (3\underset{\sim}{p}_2 + r^2\underset{\sim}{p}_1)'\underset{\sim}{e}_2*\underset{\sim}{e}_2 \\
&- 2r(\underset{\sim}{p}_1 + \underset{\sim}{p}_2)'\underset{\sim}{e}_1*\underset{\sim}{e}_2 - r(\underset{\sim}{p}_1 + \underset{\sim}{p}_2)'\underset{\sim}{d}\ \underset{\sim}{p}_1{}'\underset{\sim}{e}_2 \\
&+ 2(\underset{\sim}{p}_1{}'\underset{\sim}{e}_1\underset{\sim}{p}_1{}'\underset{\sim}{e}_1 + \underset{\sim}{p}_2{}'\underset{\sim}{e}_2\underset{\sim}{p}_2{}'\underset{\sim}{e}_2 + r^2\underset{\sim}{p}_1{}'\underset{\sim}{e}_2\underset{\sim}{p}_1{}'\underset{\sim}{e}_2) \\
&+ 3r\underset{\sim}{p}_1{}'\underset{\sim}{e}_2(\underset{\sim}{p}_1{}'\underset{\sim}{e}_1 + \underset{\sim}{p}_2{}'\underset{\sim}{e}_2) \\
&- \tfrac{1}{2}\ (\underset{\sim}{p}_1{}'\underset{\sim}{e}_1\underset{\sim}{p}_1{}'\underset{\sim}{d} + \underset{\sim}{p}_2{}'\underset{\sim}{e}_2\underset{\sim}{p}_2{}'\underset{\sim}{d}) \\
&- \tfrac{1}{2}\ (\underset{\sim}{p}_1{}'\underset{\sim}{d} * (\underset{\sim}{c}_i + r\underset{\sim}{c}_j) + \underset{\sim}{p}_2{}'\underset{\sim}{d} * (\underset{\sim}{c}_j + r\underset{\sim}{c}_i)) \\
&+ \underset{\sim}{p}_1{}'(r\underset{\sim}{e}_2+2\underset{\sim}{e}_1) * (\underset{\sim}{c}_i+r\underset{\sim}{c}_j) + \underset{\sim}{p}_2{}'(r\underset{\sim}{e}_1+2\underset{\sim}{e}_2) * (\underset{\sim}{c}_j+r\underset{\sim}{c}_i) \\
&- \underset{\sim}{p}_1{}'(\underset{\sim}{e}_1+2r\underset{\sim}{e}_2)\ \underset{\sim}{p}_1{}'(\underset{\sim}{c}_i+r\underset{\sim}{c}_j) - \underset{\sim}{p}_2{}'(\underset{\sim}{e}_2+2r\underset{\sim}{e}_1)\ \underset{\sim}{p}_2{}'(\underset{\sim}{c}_j+r\underset{\sim}{c}_i) \\
&- \tfrac{1}{2}(\underset{\sim}{p}_1+r^2\underset{\sim}{p}_2)'\underset{\sim}{c}_i*\underset{\sim}{c}_i - \tfrac{1}{2}(\underset{\sim}{p}_2+r^2\underset{\sim}{p}_1)'\underset{\sim}{c}_j*\underset{\sim}{c}_j \\
&- r(\underset{\sim}{p}_1 + \underset{\sim}{p}_2)'\ \underset{\sim}{c}_i*\underset{\sim}{c}_j \qquad (C.78)
\end{aligned}
$$

Three simplified expressions, derived from C.78, are:

1) General expression neglecting 4th power terms

$$\begin{aligned} E\,(\dot{p}_{i_1}\,\dot{p}_{j_2}) = \; & p_{i1}\,p_{j2}(1 + \tfrac{1}{2}\,(\sigma^2{}_i + 2r\sigma_{ij} + \sigma_j{}^2) \\ & + p_1'e_1 + p_2'e_2 + rp_1'e_2 \\ & - \tfrac{1}{2}(p_1 + p_2)'d \\ & - p_1'(c_i + rc_j) - p_2'(c_j + rc_i)) \end{aligned} \qquad (C.79)$$

2) Binomial logit expression neglecting 4th power terms.

$$\begin{aligned} E\,(\dot{p}_1\,\dot{p}_2) = \; & p_1\,p_2\,(1 + \sigma^2\,((\tfrac{1}{2}-p_1)(1+r-p_1) + rp_1p_2 \\ & + (\tfrac{1}{2}-p_2)(1+r-p_2)) \end{aligned} \qquad (C.80)$$

3) Binomial logit expression including 4th power but for

$r = 1$, $p = p_1 = p_2$, and $q = 1-p$

$$\begin{aligned} E\,(\dot{p}\dot{p}) = \; & pp\;\;(1 + \sigma^2 q(2-3p) \\ & +\sigma^4 q(2-14\tfrac{1}{4}p + 27p^2 - 15p^3\,) \\ = \; & pp + \sigma^2 pq(2p-3p^2 + p\sigma^2(-\tfrac{1}{4}p+2q - 12pq + 15p^2q)) \end{aligned} \qquad (C.81)$$

Appendix D.

Some additional tables relating to chapter 8.

Table D.1 Foot-shape scores recorded at two times by two observers on 97 lambs.

Score pattern				
observer 1		observer 2		
am	pm	am	pm	frequency
5	5	5	5	32
5	5	5	4	3
5	5	4	5	7
5	5	4	4	7
5	4	4	4	2
4	5	5	5	5
4	5	5	4	1
4	5	4	5	6
4	5	4	4	4
4	4	5	5	1
4	4	4	5	2
4	4	4	4	11
4	4	3	4	4
4	4	3	3	1
4	3	5	5	1
3	4	3	4	3
3	4	3	3	1
3	3	5	5	1
3	3	3	4	2
3	3	3	3	3

Table D.2 Foot-scald scores recorded at two times by two observers on 97 lambs.

Score pattern				
observer 1		observer 2		
am	pm	am	pm	frequency
0	0	0	0	80
0	1	0	0	1
1	0	0	0	4
1	1	0	0	3
1	1	0	1	2
1	1	1	0	1
1	1	1	1	6

Table D.3 Foot-rot scores recorded at two times by two observers on 97 lambs.

Score pattern				
observer 1		observer 2		
am	pm	am	pm	frequency
0	0	0	0	90
1	0	0	0	3
1	1	0	0	1
1	1	0	1	1
1	1	1	0	2

Table D.4a Experiment data for 1980 lambing

No. No. No.

YR	B1	B2	3	SX	SR	LN	L5	L4	LS	LR	HN	H5	H4	HS	HR	TN	T5	T4	TS	TR
1	1	0		1	1	39	33	6	6	1										
1	1	0		0	1	50	41	9	2	0	47	38	9	2	1	36	28	8	0	1
1	1	0		1	2	35	30	4	1	0										
1	1	0		0	2	29	20	9	0	0	30	26	4	1	0	10	8	2	0	0
1	1	0		1	3	29	21	8	1	0										
1	1	0		0	3	48	31	17	4	1	47	33	12	6	6	21	13	7	3	1
1	1	0		1	4	35	23	12	3	0										
1	1	0		0	4	32	26	5	1	0	32	30	1	0	1	18	17	1	0	0
1	1	0		1	5	26	23	3	0	0										
1	1	0		0	5	25	19	6	1	1	26	21	5	5	0	8	6	1	1	1
1	1	0		1	6	34	31	3	0	1										
1	1	0		0	6	28	23	5	1	1	28	24	3	1	2	25	18	7	1	1
1	1	0		1	7	54	46	8	1	2										
1	1	0		0	7	54	50	4	1	2	52	52	0	2	0	41	40	1	0	0
1	-1	1		1	1	43	33	10	3	1										
1	-1	1		0	1	61	49	11	4	3	57	44	11	15	12	53	21	26	3	6
1	-1	1		1	2	57	35	19	13	1										
1	-1	1		0	2	53	35	17	5	2	54	47	7	16	4	46	25	16	3	8
1	-1	1		1	3	44	33	11	5	1										
1	-1	1		0	3	50	42	8	4	2	51	46	5	16	11	44	30	9	0	4
1	-1	1		1	4	49	32	12	7	2										
1	-1	1		0	4	38	23	15	1	3	42	35	7	13	3	35	17	12	2	3
1	-1	1		1	5	64	37	26	12	3										
1	-1	1		0	5	46	33	11	4	2	49	41	7	17	10	38	26	8	0	4
1	-1	1		1	6	62	27	30	16	4										
1	-1	1		0	6	56	30	22	6	1	62	53	4	25	8	57	33	15	2	10
1	-1	-1		1	1	22	10	10	7	1										
1	-1	-1		0	1	19	11	7	1	2	19	15	2	8	3					
1	-1	-1		1	2	23	10	6	5	2										
1	-1	-1		0	2	16	7	6	2	2	16	8	7	7	6					
1	-1	-1		1	3	21	7	13	2	1										
1	-1	-1		0	3	18	6	10	2	0	17	10	6	3	1					

(1260.

SEX SIRE

Table D.4b Experiment data for 1981 lambing

YR	B1	B2	B3	SX	SR	LN	L5	L4	LS	LR	HN	H5	H4	HS	HR
-1			1	1	8	67	40	21	0	0	69	58	11	0	1
-1			1	0	8	55	39	11	0	0					
-1			1	1	9	47	27	16	0	0	46	36	8	0	0
-1			1	0	9	31	23	7	0	0					
-1			1	1	10	52	38	8	0	0	49	44	4	0	0
-1			1	0	10	37	25	10	0	0					
-1			1	1	11	40	20	12	0	0	36	28	7	0	0
-1			1	0	11	43	27	12	0	0					
-1			1	1	12	50	17	21	0	0	49	19	24	0	0
-1			1	0	12	51	20	20	0	0					
-1			1	1	13	32	16	12	0	0	28	24	4	0	0
-1			1	0	13	48	30	13	0	0					
-1			-1	1	7	30	17	12	0	0	28	21	7	0	0
-1			-1	0	7	37	14	21	1	0					
-1			-1	1	8	31	23	8	2	1	30	21	9	0	0
-1			-1	0	8	20	13	6	0	1					
-1			-1	1	9	19	12	6	0	0	18	12	6	0	1
-1			-1	0	9	22	14	7	0	0					
-1			-1	1	10	24	11	10	5	2	25	11	8	3	3
-1			-1	0	10	26	17	8	5	2					
-1			-1	1	11	39	15	17	8	4	40	16	19	3	4
-1			-1	0	11	20	15	3	1	0					
-1			-1	1	12	31	20	9	7	2	28	12	12	0	5
-1			-1	0	12	23	17	6	2	2					
-1			-1	1	13	23	20	3	0	0	23	16	7	1	2
-1			-1	0	13	27	21	5	0	1					
-1			-1	1	14	33	17	14	1	0	36	18	16	1	2
-1			-1	0	14	40	25	13	0	0					
-1			-1	1	15	33	20	13	1	0	33	20	12	0	2
-1			-1	0	15	26	15	9	0	0					
-1			-1	1	16	46	33	11	6	1	44	20	21	2	5
-1			-1	0	16	40	30	9	3	0					
-1			-1	1	17	22	14	5	6	4	20	12	4	3	6
-1			-1	0	17	34	21	12	3	3					
-1			-1	1	18	30	19	10	0	1	29	14	13	0	4
-1			-1	0	18	24	14	8	0	1					

Table D.5 Estimates of breeding values for the trait foot-shape score - obtained by various methods.

	Method and Trait					
	MLM	OLS	MLM	OLS	MLM	MLM
Sire	L54	LS	L5	L5	H54	H5
1	.401	.093	.394	.088	.8041	.7887
2	.258	.074	.252	.069	-.0010	.0114
3	.138	.037	.132	.032	-.2286	-.2396
4	.077	.027	.072	.023	-.1799	-.1868
5	-.092	-.032	-.083	-.020	.0633	.0527
6	-.285	-.080	-.273	-.070	.3618	.3685
7	-.496	-.118	-.495	-.122	-.8198	-.7949
8	.481	.166	.458	.133	.3149	.2977
9	.438	.143	.425	.120	-.3998	-.3932
10	-.055	-.013	-.071	-.021	-.0383	-.0411
11	-.087	-.027	-.087	-.026	.1546	.1359
12	-.113	-.049	-.085	-.027	-.0370	-.0553
13	-.663	-.220	-.639	-.178	.0055	.0559
14	.249	.146	.204	.084	.3463	.4021
15	-.110	-.036	-.227	-.094	-.0769	-.0997
16	-.139	-.110	.023	.010	-.2694	-.3024
17	.387	.164	.413	.129	.5822	.5819
18	.245	.125	.198	.064	.0995	.1097
19	.240	.099	.228	.066	.4064	.3786
20	-.040	-.015	-.057	-.019	.3676	.3438
21	-.086	-.045	-.063	-.021	.0893	.0836
22	-.745	-.327	-.718	-.219c	-1.5452	-1.4977
23	.528	.226	.525	.192	.4287	.3694
24	.348	.127	.347	.107	-.2825	-.3245
25	.249	.118	.226	.082	.4785	.4175
26	.179	.079	.171	.061	-.4109	-.3588
27	.022	.012	.019	.007	.3266	.2652
28	-.030	-.023	-.009	-.004	-.0345	.1109
29	-.034	-.015	-.036	-.014	-.1808	-.2080
30	-.059	-.011	-.088	-.032	.1973	.1503
31	-.149	-.053	-.158	-.052	-.1262	-.1694
32	-.195	-.093	-.179	-.067	-.4844	-.3140
33	-.405	-.210	-.324	-.114	-.5097	-.4881
34	-.454	-.156	-.493	-.165	.5980	.5496

Table D.6 Estimates of breeding values for the traits presence of foot-scald and presence of foot-rot, obtained by two methods.

	Method and Trait			
Sire	MLM Scald	OLS Scald	MLM Rot	OLS Rot
1	-.410	-.022	.314	.021
2	-.366	-.025	.156	.016
3	.804	.050	-.081	-.005
4	-.284	-.021	.023	.003
5	-.378	-.026	-.176	-.016
6	.252	.019	-.183	-.016
7	.383	.026	-.053	-.004
8	-.215	-.025	-.093	-.009
9	-.433	-.049	-.027	-.003
10	.113	.014	.066	.005
11	.320	.038	-.154	-.013
12	-.294	-.038	.181	.017
13	.509	.060	.026	.002
14	.231	.038	.030	.006
15	-.322	-.055	-.218	-.042
16	.090	.016	.187	.036
17 : 22	not estimated because no lambs in this mating-group had either foot-rot or scald.			
23	-.685	-.069	-.116	-.017
24	.357	.031	-.267	-.026
25	-.356	-.038	.011	.000
26	.883	.091	.284	.035
27	-.629	-.070	-.226	-.037
28	1.011	.095	.724	.089
29	-.729	-.075	.000	-.001
30	-.582	-.058	-.305	-.038
31	-.629	-.055	-.344	-.037
32	1.258	.129	.322	.042
33	.700	.072	.241	.028
34	-.599	-.054	-.324	-.037

Bibliography

ALWAN, M.T. (1983).
Studies on the mating performance of the Booroola Merino crossbred ram lambs and the foot conditions in Booroola Merino crossbreds and Perendales grazed on hill country. Thesis, Master in Animal Science, Massey University, Palmerston North, New Zealand.

ANDERSON, R.D. (1978).
Studies on the estimation of variance components. Thesis, PhD, Cornell University, U.S.A., 151pp.

BLISS, C.I. (1934 a).
The method of probits. Science, 79:38

BLISS, C.I. (1934 b).
The method of probits - a correction. Science, 79:409.

BLISS, C.I. (1935).
The calculation of the dosage mortality curve. Annals of Applied Biology, 22: 134-167.

BLOOMFIELD, P. and WATSON, G.S. (1975)
The inefficiency of least squares. Biometrika, 62: 121-128.

BOCK, R.D. (1975).
Multivariate Statistical Methods in Behavioural Research. McGraw-Hill Book Co., New York.

COCHRAN, W.G. (1940).
The analysis of variance when experimental errors follow the poisson or binomial laws. Annals Math. Stat., 11: 335-347.

COONEY, G.H.
Macquarie University, New South Wales, Australia.

COX, Neil R. (1974).
Estimation of the correlation between a continuous and a discrete variable. *Biometrics*, 30: 171-178.

CLAYTON, D.G. (1974).
Some odds ratio statistics for the analysis of ordered categorical data. *Biometrika*, 61: 525-531.

CROWDER, M.J. (1979).
Inference about the intra-class correlation coefficient in the beta binomial ANOVA for proportions. *Journal of the Royal Statistical Society, Series B*, 41: 230-234.

CURNOW, R.N. and SMITH, Charles (1975).
Multifactorial models for familial diseases in man. *Journal of the Royal Statistical Society, Series A*, 138: 131-169.

DANELL, O. and RÖNNINGEN, K. (1981).
All-or-none traits in index selection. *Z. Tierzuchtg. Zuchtgsbiol.*, 98: 265-284.

DEMPSTER, A.P., Laird, N.M. and RUBIN, D.B. (1977).
Maximum likelihood from incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society, Series B*, 39: 1-38.

DEMPSTER, E.R. and LERNER, I.M. (1950).
Heritability of threshold characters. *Genetics*, 35: 212-236.

DUN, R.B. (1958).
Selection for an invariant character 'Vibrissa Number' in the house mouse. *Nature*, 181: 1018-1019.

DUN, R.B. and FRASER, A.S. (1959).
Selection for an invariant character 'Vibrissa Number' in the house mouse. Australian Journal of Biological Science, 12: 506-523.

DZAKUMA, J.M., WHITEMAN, J.V. and McNEW, R.W. (1982).
Repeatability of lambing rate. Journal of Animal Science, 54: 540-543.

EDWARDS, J.H. (1969).
Familial predisposition in man. British Medical Bulletin, 25: 58-63.

ELLIOT, K.H. (1975).
A study of the inheritance of some productive traits in Perendale sheep. Thesis, Master in Animal Science, Massey University, Palmerston North, New Zealand, 136pp.

ELSTON, R.C. (1977).
Estimating 'heritability' of a dichotomous trait. Biometrics, 33: 231-236.

FALCONER, D.S. (1965).
The inheritance of liability to certain diseases, estimated from the incidence among relatives. Annals of Human Genetics, London, 29: 51-76.

FALCONER, D.S. (1967).
The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitis. Annals of Human Genetics, London, 31: 1-20.

FAMULA, T.R. (1981).
Exponential stayability model with censoring and covariates. Journal of Dairy Science, 64: 538-545.

FRASER, A.S., NAY, T. and KINDRED, B.M. (1959).
Variation of vibrissa Number in the mouse. Australian Journal of Biological Science, 12: 331-339.

FRASER, A.S. and KINDRED, B.M. (1960).
Selection for an invariant character, vibrissa number, in the house mouse. II Limits to Variability. Australian Journal of Biological Science, 13: 48-58.

GABRIEL, K.R. (1963).
Analysis of Variance of proportions Journal of the American Statistical Association, 58: 1133-1157.

GIANOLA, D. (1979 a).
Heritability of polychotomous characters. Genetics, 93: 1051-1055.

GIANOLA, D. (1979 b).
Analysis of discrete variables in animal breeding contexts. Journal of Dairy Science, 62: 1471-1478.

GIANOLA, D. (1980 a).
A method of sire evaluation for dichotomies. Journal of Animal Science, 51: 1266-1271.

GIANOLA, D. (1980 b).
Genetic evaluation of animals for traits with categorical responses. Journal of Animal Science, 51: 1272-1276.

GIANOLA, D. (1980 c).
Theory and analysis of threshold characters. Symposium on Statistical Analysis of Categorical Data in animal research, 72 Annual Meeting of the American Society of Animal Science, Cornell University, July 28, 1980

GIANOLA, D. and FOULLEY, J.L. (1982).
Non-linear prediction of latent genetic liability with binary expression: an empirical Bayes approach. 2nd World Congress on Genetics applied to Livestock Production, 4th to 8th October, 1982, Madrid, Spain. 7: 293-303.

GILMOUR, A.R. (1983).
REG - A generalized linear models programme. 1983 edition in preparation, to be published by Biometrical Branch, New South Wales Department of Agriculture, Sydney, Australia.

GHOSH, M. and SINHA, B.K. (1980).
On the robustness of least squares procedures in regression models. Journal of Multivariate Analysis, 10: 332-342.

GRIFFITHS, D.A. (1973).
Maximum likelihood estimation for the beta binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics, 29: 637-648.

GRIZZLE, J.E., STARMER, C.F. and KOCH, G.G. (1969).
Analysis of categorical data by linear models. Biometrics, 25: 489-504.

HARVILLE, D.A. (1977).
Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72: 320-340.

HAZEL, L.N. (1943).
The genetic basis for constructing selection indexes. Genetics, 28: 476-490.

HENDERSON, C.R. (1953).
Estimation of variance and covariance componenets. Biometrics, 9: 226-252.

HENDERSON, C.R., KEMPTHORNE, O., SEARLE, S.R. and VON KROSIGK, C.M. (1959).
The estimation of environmental and genetic trends from records subject to culling. Biometrics, 15: 192-218.

HENDERSON, C.R. (1963).
Selection index and expected genetic advance. Statistical Genetics and Plant Breeding (W.D. Hanson and H.F. Robinson, Eds.). Washington: National Academy of Sciences - National Research Council, 141-163.

HENDERSON, C.R. (1973).
Sire evaluation and genetic trends. Proc. of the Anim. Breeding and Genet. Symp. in Honor of Dr. Jay L. Lush. July 29, 1972, Blacksburg, Va. 10-41.

HENDERSON, C.R. (1975).
Best linear unbiased estimation and prediction under a selection model. Biometrics, 31: 423-447.

HENDERSON, C.R. (1977).
Prediction of future records. Proceedings of the International Conference on Quantitative Genetics, Iowa State University Press, Ames, Iowa. 615-638.

HENDERSON, H.V. and SEARLE, S.R. (1981)
On deriving the inverse of a sum of matrices. Siam Review, 23: 53-60.

HILL, W.G. (1977).
Comments on statistical efficiency in the bull progeny testing for calving difficulty. Livestock Production Science, 4: 203-207.

HILL, W.G. and SMITH, Charles (1977).
Estimating 'heritability' of a dichotomous trait. Biometrics, 33: 231-236.

HOHENBOKEN, W.D. and CLARKE, S.E. (1981).
Genetic, environmental and interaction effects on lamb survival, cumulative lamb production and longevity of crossbred ewes. Journal of Animal Science, 53: 966-976.

HUDSON, G.F.S., SCHAEFFER, L.R. and WILTON, J.W. (1980).
Breed and age of dam influences on beef sire evaluations for ease of calving. Canadian Journal of Animal Science, 60: 825-831.

JAMES, J.W. and McGUIRK, B.J. (1982).
Regression of offspring on parent for an all-or-none trait. Tierzuchtung Kapitel, 41: in press.

KENDALL, M.G. and STUART, A.L. (1979).
The advanced theory of statistics. Vol. II, 4th edition. London: Griffin.

KIDD, K.K. and CAVALLI-SFORZA, L.L. (1973).
An analysis of the genetics of schizophrenia. Social Biology, 20: 254-265.

KINDRED, Berenice (1961).
A maternal effect on vibrissa score due to the tabby gene. Australian Journal of Biological Science, 14: 627-636.

KINDRED, Berenice (1962).
A correlated response mediated through a maternal effect in the house mouse. Australian Journal of Biological Science, 15: 352-361.

KINDRED, B.M. (1963).
Selection for an invariant character, 'vibrissa number', in the house mouse. IV Probit analysis Genetics, 48: 621-632.

KNOTT, M. (1975).
On the minimum efficiency of least squares. Biometrika, 62: 129-132.

KOCH, G.G., GRIZZLE, J.E., SEMENYA, Kofi and SEN, P.K. (1978)
Statistical methods for evaluation of mastitis treatment data. Journal of Dairy Science, 61: 829-847.

KOCH, G.G., LANDIS, J.R., FREEMAN, J.L., FREEMAN, D.H.Jr and LEHNEN, R.G. (1977).
A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics, 33: 133-158.

KOCH, G.G. and REINFURT, D.W. (1971).
The analysis of categorical data from mixed models. Biometrics, 27: 157-173.

KORKMAN, Nils (1948).
Genetic variation in the frequency of multiple births in cattle. Hereditas, 34: 23-34.

KUPPER, L.L. and HASEMAN, J.K. (1978).
The use of a correlated binomial model for the analysis of certain toxicological experiments. Biometrics, 34: 69-76.

LAMBERSON, W.R. and THOMAS, D.L. (1982).
Effects of season and breed of sire on incidence of estrous and ovulation rate in sheep. Journal of Animal Science, 54: 533-539.

LANDIS, J.R. and KOCH, G.G. (1977 a).
The measurement of observer agreement for categorical data. Biometrics, 33: 159-174.

LANDIS, J.R. and KOCH, G.G. (1977 b).
An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics, 33: 363-374.

LANDIS, J.R. and KOCH, G.G. (1977 c).
A one-way components of variance model for categorical data. Biometrics, 33: 671-679.

LEWER R.P. (1978).
A study of genetic and environmental variation and covariation in productive traits of a flock of Perendale sheep. Thesis, PhD, Massey University, Palmerston North, New Zealand, 178pp.

LUSH, J.L. (1949).
Heritability of quantitative characters in farm animals. In Proceedings of the eighth international congress of genetics issued as a supplementary volume of Hereditas, edited by Bonnier, G. and Larsson, L., Lund, Berlingska Boktryckeriet.

LUSH, J.L., LAMOREUX, W.F. and HAZEL, L.N. (1948).
The heritability of resistance to death in the fowl. Poultry Science, 27: 375-388.

MAO, I.L. (1976).
Correlation estimates from discontinuous data. Journal of Dairy Science, 59: 2115-2121.

MOOD, A.M., GRAYBILL, F.A. and BOES, D.C. (1974).
Introduction to the theory of statistics, 3rd edition, McGraw-Hill series in Probability and Statistics, McGraw-Hill Kogakusha, Ltd,. Tokyo.

MORENO, A.M., WIGGANS, G.R. and VAN VLECK, L.D. (1979). Genetic and herd-year variation in type traits of Brown Swiss cows. Journal of Dairy Science, 62: 486-492.

McCULLAGH, P (1977). A logistic model for paired comparisons with ordered categorical data. Biometrika, 64: 449-453.

NELDER, J.A. and WEDDERBURN, R.W.M. (1972). Generalized linear models. Journal of the Royal Statistical Society, Series A, 135: 370-384.

OLAUSSON, A. and RÖNNINGEN, K. (1975). Estimation of genetic parameters for threshold characters. Acta. Agriculturae Scandinavica, 25: 201-208.

PATTERSON, H.D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58: 545-554.

PEARSON, Karl (1901). I. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. Philosophical Transactions of the Royal Society of London, Series A, 195: 1-47.

PLACKETT, R.L. (1965). A class of bivariate distributions. Journal of the American Statistical Association, 60: 516-522.

POLLAK, E.J. and FREEMAN, A.E. (1976). Parameter estimation and sire evaluation for dystocia and calf size in Holsteins. Journal of Dairy Science, 59: 1817-1824.

QUAAS, R.L. and VAN VLECK, L.D. (1980).
Categorical trait sire evaluation by best linear unbiased prediction of future progeny category frequencies. Biometrics, 36: 117-123.

REICH, T, JAMES, J.W. and MORRIS, C.A. (1972).
The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. Annals of Human Genetics, London, 36: 163-184.

RENDAL, J.M. (1959).
Variation and dominance at the scute locus in Drosophila melanogaster. Australian Journal of Biological Science, 12: 524-533.

RENDAL, J.M. (1959).
Canalization of the scute phenotype of drosophila. Evolution, 13: 425-439.

ROBERTSON, A. (1982).
Genetics engineering in animal production. 2nd World Congress on Genetics applied to Livestock Production, 4th to 8th October, 1982, Madrid, Spain. 6: 139-145.

ROBERTSON, A. and LERNER, I.M. (1949).
The heritability of all-or-none traits: viability of poultry. Genetics, 34: 395-411.

RÖNNINGEN, K. (1976).
The estimation of genetic parameters for all-or-none traits. Z. Tierzuchtg. Zuchtgsbiol., 93: 226-236.

RUTLEDGE, J.J. (1977).
Repeatability of threshold traits. Biometrics, 33: 395-399.

SCHAEFFER, L.R. and WILTON, J.W. (1977).
Evaluation of beef sires across breeds for calving ease. *Canadian Journal of Animal Science*, 57: 635-645.

SCHAEFFER, L.R., WILTON, J.W. and THOMPSON, R. (1978).
Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics*, 34: 199-208.

SEARLE, S.R. (1971 a).
Topics in variance component estimation. *Biometrics*, 27: 1-76.

SEARLE, S.R. (1971 b).
Linear Models. New York: John Wiley and Sons.

SMITH, Charles (1970).
Heritability of liability and concordance in monozygous twins. *Annals of Human Genetics, London*, 34: 85-91.

SMITH, Charles (1971).
Recurrence risks for multifactorial inheritance. *American Journal of Human Genetics*, 23: 578-589.

SMITH, Charles, FALCONER, D.S. and DUNCAN, L.J.P. (1972).
A statistical and genetical study of diabetes. II Heritability of liability. *Annals of Human Genetics, London*, 35: 281-299.

SMITH, Fairfield (1937).
A discriminant function for plant selection. *Annals of Eugenics*, 7: 240-250.

SNELL, E.J. (1964).
A scaling procedure for ordered categorical data. *Biometrics*, 20: 592.

TALLIS, G.M. (1962).
The maximum likelihood estimation of correlation from contingency tables. Biometrics, 18: 342-353.

TATE, R.F. (1955).
The theory of correlation between two continuous variables when one is dichotomized. Biometrika, 42: 205-216.

THOMPSON, J.R., FREEMAN, A.E. and BERGER, P.J. (1980).
Age of dam and maternal considerations in Holstein dystocia. Dairy cattle Improvement Section, American Dairy Science Association 75th Annual Meeting. June 15-18, 1980, Virginia Tech, Blocksburg, Virginia

THOMPSON, J.R., FREEMAN, A.E. and BERGER, P.J. (1981).
Age of dam and maternal effects for dystocia in Holsteins. Journal of Dairy Science, 64: 1603-1609.

THOMPSON, Robin.
A.R.C. Unit of Statistics, University of Edinburgh,
Mayfield Rd., Edinburgh, EH9 3JZ, Scotland.

THOMPSON, R. (1979).
Sire evaluation. Biometrics, 35: 339-353.

TONG, A.K.W., WILTON, J.W. and SCHAEFFER, L.R. (1976).
Evaluation of ease of calving for Charolais sires. Canadian Journal of Animal Science, 56: 17-26.

TONG, A.K.W., WILTON, J.W. and SCHAEFFER, L.R. (1977).
Application of a scoring procedure and transformations to dairy type classification and beef ease of calving categorical data. Canadian Journal of Animal Science, 57: 1-5.

TONG, A.K.W., WILTON, J.W. and SCHAEFFER, L.R. (1977).
Application of a scoring procedure and transformations to dairy type classification and beef ease of calving categorical data. _Canadian Journal of Animal Science_, 57: 1-5.

VAN VLECK, L.D. (1972).
Estimation of heritability of threshold characters. _Journal of Dairy Science_, 55: 218-225.

VAN VLECK, L.D. (1980).
Stayability evaluation as a categorical trait and by considering other traits. _Journal of Dairy Science_, 63: 1172-1180.

VAN VLECK, L.D. and KARNER, P.J. (1980).
Sire evaluation by best linear unbiased prediction for categorically scored type traits. _Journal of Dairy Science_, 63: 1328-1333.

VINSON, W.E., WHITE, J.M. and KLIEWER, R.H. (1976).
Overall classification as a selection criterion for improving categorically scored components of type in Holsteins. _Journal of Dairy Science_, 59: 2104-2114.

WADDINGTON, C.H. (1956).
Genetic assimilation of the bithorax phenotype. _Evolution_, 10: 1-13.

WEDDERBURN, R.W.M. (1974).
Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. _Biometrika_, 61: 439-447.

WHITE, J.M., VINSON, W.E. and PEARSON, R.E. (1981).
Dairy cattle improvement and Genetics. _Journal of Dairy Science_, 64: 1305-1317.

WILLIAMS, D.A. (1975).
The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. _Biometrics_, 31: 949-952.

WILLIAMS, D.A. (1982).
Extra-binomial variation in logistic linear models. _The Journal of the Royal Statistical Society, Series C (Applied Statistics)_, 31: 144-148.

WILLIAMS, J.S. (1962).
The evaluation of a selection index. _Biometrics_, 18: 375-393.

WOOD, C.L. (1978).
Comparison of linear trends in binomial proportions. _Biometrics_, 34: 496-504.

WRIGHT, S (1934 a).
An analysis of variability in number of digits in an inbred strain of guinea pigs. _Genetics_, 19: 506-536.

WRIGHT, S (1934 b).
The results of crosses between inbred strains of guinea pigs, differing in number of digits. _Genetics_, 19: 537-551.

YAO, S.C., WHITE, J.M., VINSON, W.E. and KLIEWER, R.H.(1979).
Environmental correlations among paternal half sisters for type. _Journal of Dairy Science_, 62: 493-498.