THE ANALYSIS OF

DEPENDENT COUNT DATA

Promotor: dr. P. van der Laan, hoogleraar in de wiskundige statistiek

J. Engel

THE ANALYSIS OF DEPENDENT COUNT DATA

Proefschrift

ter verkrijging van de graad van
doctor in de landbouwwetenschappen,
op gezag van de rector magnificus,
dr. C.C. Oosterlee,
in het openbaar te verdedigen
op vrijdag 29 mei 1987
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen.

STELLINGEN

1. Laat $F(x)$ de verdelingsfunktie zijn van een stochastische variabele X; laat het interval $(-\infty, \infty)$ de drager zijn van $F(x)$. Laat voorts de relatie

   $F(x)(1 - F(x - \theta))/[F(x - \theta)(1 - F(x))] = \phi(\theta)$

   geldig zijn voor alle reële x en θ, waarbij $\phi(\theta)$ een niet-constante funktie is van θ die niet van x afhangt.

   Dan is

   $F(x) = [1 + \exp(-b(x - a))]^{-1}$

   voor reële a en $b > 0$, zodat de kansverdeling van X de logistische verdeling is.

   Engel, J. (1985), Some characterizations of distributions by regression models for ordinal response data, Metrika 32, 65-72.

2. De keuze van de link function in toepassingen van gegeneraliseerde lineaire modellen wordt veelal uitsluitend gemaakt op grond van argumenten van wiskundige aard, hierbij gemakshalve voorbijgaand aan de interpreteerbaarheid van het resultaat.

3. De resultaten die zijn verkregen door Brier ten aanzien van de kansverdeling van de G^2-toets in het geval van een groot aantal onafhankelijke trekkingen uit de Dirichlet multinomiale verdeling volgen direct uit de theorie van quasi-likelihood toetsen zoals die is afgeleid door McCullagh.

   Brier, S.S. (1980), Analysis of contingency tables under cluster sampling, Biometrika 67, 591-596.
   McCullagh, P. (1983), Quasi-likelihood functions, Ann. Statist. 11, 59-67.

4. De in de regressieanalyse gangbare benaming "verklarende variabelen" ter aanduiding van de regressoren dient niet de suggestie te wekken dat het hier een "wetenschappelijk verklaren" betreft. In dit verband is de aard van de statistiek niet verklarend doch beschrijvend en derhalve dient een verklaring dan ook anderszins te worden gegeven.

5. Beschouw een homogeen Poisson proces met intensiteitsparameter $\lambda = 1$ en laat N(Y) het aantal punten zijn in het tijdsinterval [0, Y), waarbij Y een niet-negatieve stochastische variabele is. Dan wordt de kansverdeling van Y een-eenduidig bepaald door de kansverdeling van N(Y). Zo geldt dat de kansverdeling van N(Y) de negatief binomiale verdeling is met parameters α en $\theta(1 + \theta)^{-1}$ dan en slechts dan als Y gamma verdeeld is met parameters α en θ.

   Engel, J. and Zijlstra, M. (1980), A characterization of the gamma distribution by the negative binomial distribution, J. Appl. Prob. 17, 1138-1144.

6. Het *optreden van extra-Poisson variatie in een produktieproces* kan wijzen op statistische onbeheerstheid van dat proces.

7. Bij de statistische modelbouw kan men soms met vrucht gebruik maken van de resultaten die zijn verkregen ten aanzien van de karakterisering van kansverdelingen.

8. Indien met behulp van de methode van kwantielanalyse de parameters van de logistische verdelingsfunktie dienen te worden geschat en indien daarbij een kleine kans op succes wordt verwacht bij een zekere toegediende stimulus, kan het zinvol zijn niet van te voren het aantal trials te fixeren, doch het aantal successen.

   Engel, J. (1984), Kwantielanalyse en de negatief binomiale verdeling, Kwantitatieve Methoden 13, 42-62.

9. Gezien het belang van experimentele resultaten voor de ontwikkeling van de natuurwetenschappen dient de door de statistiek verworven kennis ten aanzien van het opzetten van experimenten in ruime mate te worden uitgedragen bij het opleiden van natuurwetenschappers.

10. De informatie in de bijsluiter van een geneesmiddel omtrent de bijwerkingen van dat geneesmiddel is vaak onvoldoende om zich een oordeel te kunnen vormen omtrent het risico dat de patiënt loopt bij gebruik van het middel. Dit hangt samen met het feit dat veelal gegevens ontbreken over risicoverhogende factoren en over de frequentie waarmee bijwerkingen optreden.

11. Het creëren van samenwerkingsverbanden tussen afdelingen voor statistische consulatie en afdelingen voor statistisch onderzoek is stimulerend voor de ontwikkeling van de toegepaste statistiek.

Stellingen
behorend bij het proefschrift
THE ANALYSIS OF DEPENDENT COUNT DATA
J. Engel, Wageningen, 29 mei 1987

Aan mijn moeder

Aan Saapke

TEN GELEIDE

Bij het tot stand komen van dit proefschrift wil ik graag diegenen bedanken die hieraan een bijdrage hebben geleverd. Dit betreft in de eerste plaats mijn promotor prof. dr. P. van der Laan, wiens kritische en inspirerende opmerkingen erg stimulerend waren tijdens het onderzoek dat tot dit proefschrift leidde. In een eerder stadium van dit onderzoek waren suggesties van prof. ir. J.W. Sieben mede richtingbepalend.

De Nederlandse Philips Bedrijven B.V. dank ik voor het gebruik van de computer, voor correcties van de Engelse tekst en voor de uitvoering van reprotechnisch werk.

Het typewerk werd, met nimmer aflatende toewijding, verzorgd door Agnes Verguld en Yvonne van den Kerkhof, waarvoor ik hen zeer erkentelijk ben.

Tenslotte wil ik graag vermelden dat dit proefschrift grotendeels tot stand is gekomen in tijd die eigenlijk toekwam aan Saapke en Annet.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1. INTRODUCTORY REMARKS

In this thesis we shall study the analysis of count data and dicho(poly)-tomous data classified by some fixed or random factors; the data may result from a sampling procedure or from a designed experiment. Some of the standard and well-known results on this subject will be extended, to cover more general situations.

The known results mainly concern the fixed factor case; results for random factor designs are quite sparse. It is this very important random factor case for which new results will be presented.

In this introduction, three important concepts from discrete data analysis will be discussed. Firstly, in section 1.1.1. some standard results on log-linear models for count data analysis will be reported. This class of models forms a subclass of the class of Generalized Linear Models (GLM's, see section 1.1.2.) for the analysis of data having a distribution which belongs to an exponential family of distributions of a specific type. A third rather new and promising concept is the concept of quasi-likelihood (see section 1.1.3.), which is closely related to GLM. Now, distributional assumptions are abandoned and the Iterative Weighted Least Squares (IWLS) algorithm for estimating the parameters of the GLM is borrowed from GLM theory to obtain maximum quasi-likelihood estimators of these model parameters.

Section 1.2. of this introduction deals with some practical examples from various fields for a further motivation of this study. It will be shown that the well-known techniques for count data analysis do not cover all such problems from practice. This fact is known more generally, and some quotations from the literature confirm our opinion that there is a need for a more general approach.

Section 1.3. summarizes some new results from the literature on this subject. Results beyond those that are already established are sparse; some of them will be used in this thesis to form a basis for building further work on.

Finally, section 1.4. reports new results which form the main part of this thesis. In a nutshell: problems on independent data from fixed factor classifications will be treated in cases of extreme variation observed in the data. Also, problems with dependent count data will be considered. It concerns those types of experimental designs where in the corresponding model random model components are to be introduced at several levels. Examples are the split-plot design and the random factor design. Some models with random components will be proposed and the analysis of data by these models will be treated.

1.1.1 Loglinear models

A class of models well-suited for the analysis of cross-classified count data is the class of loglinear models. On a general level, these models can be formulated as follows.

Let $\{ X_{i_1,i_2, \ldots, i_J} \}_{i_1=1}^{I_1} {}_{i_2=1}^{I_2} \ldots {}_{i_J=1}^{I_J}$ be a vector of independent random variables being classified by J nominal fixed factors with levels i_1, i_2, ..., i_J, where $i_j = 1, 2, \ldots, I_j$ for $j = 1, 2, \ldots, J$.
Furthermore, let

- $X_{i_1, i_2, \ldots, i_J}$ have a Poisson distribution with mean value $m_{i_1, i_2, \ldots, i_J}$;
- $\log m_{i_1, i_2, \ldots, i_J} = u + u_{1(i_1)} + u_{2(i_2)} + \ldots + u_{12(i_1 i_2)} + \ldots + u_{12..J\,(i_1 i_2 .. i_J)}$

which is a (saturated) linear model on log-scale.

For polytomous data having the multinomial distribution with parameters n and $p_{i_1,\, i_2, .., i_J}$ the model holds as well with $m_{i_1,\, i_2, .., i_J} = np_{i_1, i_2, .., i_J}$. Estimation and testing the model-parameters of the loglinear model and testing goodness-of-fit of reduced models are some aspects of statistical inference on count data with the loglinear model. The theory is consolidated in the books by Plackett (1974) and Everitt (1977), and somewhat more recently by Bishop c.s. (1975), Fienberg (1977) and Haberman (1974). Bishop c.s. (1975) in particular give an extensive treatment of the theory and practice, and bring together many results which could be found in the literature only at scattered places.

The estimation and testing of model parameters is carried out using the likelihood principle, and the distributional assumptions are Poisson and multinomial. Not many results are known about the consequences of violating these assumptions. The coherent log-likelihood ratio test statistic G^2 is usually preferred to Pearson's X^2 for testing goodness-of-fit and model reductions; however, Pearson's X^2 seems to be more robust against violation of the distributional assumptions, and has a better small sample behaviour. Both statistics will be shown to be of more general use, also for non-Poisson distributed data. Asymptotic distributional results will be derived under nonstandard conditions.

1.1.2. Generalized Linear Models

The class of loglinear models for independent Poisson data is a sub-class of the class of Generalized Linear Models (GLM's) for independent data. The basic distribution of the data then belongs to a more general exponential family of distributions with probability density function

$$f\,(x;\,\theta,\,\phi) = \exp\left[\alpha(\phi)\,\{\,x\theta - g(\theta) + h(x)\} + \beta(\phi,\,x)\right], \qquad \ldots\ (1)$$

where $E(X) = g'(\theta)$, $\alpha(\phi)\mathrm{var}(X) = g''(\theta)$ and ϕ is a nuissance parameter. The class of GLM's was introduced by Nelder and Wedderburn (1972). Three basic assumptions were made by these authors:

- Let X be a vector of independent random variables, each having distribution (1);
- Let $y = Z\beta$ be a linear model, where Z is a (design) matrix of fixed qualitative and quantitative covariates, and β is a parameter vector;
- Let ϕ be a function such that $\theta = \phi(y)$. This function ϕ is called a link function.

For the loglinear model (see 1.1.1.) and the Poisson distribution with mean m it is seen after doing some algebra that $\theta = \log m$ for this distribution, so that the loglinear model fits in the GLM framework with the link function being the identical function.
By the GLM theory, the classical linear modeling of normal response variables is extended to the linear modeling of a much wider class of response variables, having a distribution from the family (1). Examples are the Poisson distribution, and the gamma and binomial distribution.

Estimation and testing procedures are based on likelihood. Partial derivatives of the loglikelihood function are equated to zero and parameters are estimated by the Newton-Raphson algorithm modified by using the Fisher-score approximation, taking the expectation of the Hessian matrix. It was shown by Nelder and Wedderburn (1972) that this technique is equivalent to Iteratively Weighted Least Squares (IWLS) see also McCullagh and Nelder (1984). This latter procedure is implemented in the GLIM (Generalized Linear Interactive Modelling) computer program of Baker and Nelder (1978) for interactive data modeling. Later, Green (1984) showed IWLS to be of much wider use, also for problems where distributions are not of type (1).

With GLM, there is no need to transform non-normal, non-homoscedastic data to homoscedastic (and, preferably, normal) data, imposing the classical linear modeling framework in an unnatural way. It is widely known that doing so is wrong. Nevertheless, this transformation is widely applied.

Recent literature on GLM's is summarized by McCullagh and Nelder (1984) and a review is given by Pregibon (1984).

1.1.3. Quasi-likelihood

Essentially, to obtain linear model parameter estimators by IWLS, no distribution needs to be specified. It is sufficient to express the variance of the response variable X as a function of the mean E(X) in the form of the so-called variance function $\mathrm{var}(X) = c\ \phi(E(X))$, where ϕ is some known positive function and c is some unknown positive constant; in IWLS, $\mathrm{var}^{-1}(X)$ is used as a weight function. This is the idea behind the concept of quasi-likelihood, introduced by Wedderburn (1976); see also McCullagh (1983) and McCullagh and Nelder (1984).
Originally, the idea of Wedderburn seemed to be only to estimate model parameters; it was shown by McCullagh (1983) that quasi-likelihood also provides for testing procedures. A more extensive introduction to the method can be found in chapter 8 of this thesis. Now, we shall mention two important aspects of the use of quasi-likelihood.

- Sometimes (often?), no distributional assumptions can reasonably be made, but a variance function of type $\mathrm{var}(X) = c\ \phi(E(X))$ can be based on (sparse) experimental results. Quasi-likelihood may then do the estimation job.
- Distributional assumptions are made, boldly, but they lead to a complicated analysis. This is often the case in count data analysis when several random components are introduced in the model: likelihood procedures are very unattractive. If variance functions can be obtained, quasi-likelihood can lead to interesting results as described in chapter 8 of this thesis.

In spite of some fine techniques being available for discrete data analysis, there are problems that remain to be solved.
Some of these problems will be mentioned in section 1.2.

1.2. PROBLEMS THAT REMAIN AND THEIR ORIGIN

1.2.1. Problems that remain

The analysis of count data and polytomous data has proved to be of major theoretical interest, and the results obtained are of great practical value. However, for many practical problems the theory as it stands fails to give a proper solution, and there is a clear need for extensions. Two examples will illustrate this.

Example 1.
Univariate responses. Over(under)dispersion of independent data, e.g. count data showing "extra-Poisson variation", and dichotomous data showing "extra-binomial variation"; some generalization of the classical Anova theory is needed to accommodate an extra dispersion parameter. Another problem is the analysis of dependent count data from a random factor experiment, modeled in some way by several random components.

Example 2.
Multivariate responses. Cluster sampling and multiple categorical responses are examples of a dependent classification of data.

To illustrate example 1: consider the experimental design with two nested, random factors and some replicates per cell; see Scheffé (1959).
For normal responses X_{ijk}, the Anova model is $X_{ijk} = \mu + a_i + b_{j(i)} + e_{k(ij)}$ where a_i, $b_{j(i)}$ and $e_{k(ij)}$ are independent random model components. Certainly a model of this type is needed for count data analysis.

Example 2 may be illustrated by cluster sampling, where objects are classified not independently into one of K classes. Clearly, the multinomial model based on independent classification is of little use in this case: what we need is some alternative model, in which the dependence of the classification of objects is included.

1.2.2. Their origin

The problems that remain have their origin in various fields of human activity and the solutions to these problems have therefore a wide applicability.
Some of those fields of application are summarized under industrial manufacturing, marketing research and biomedical sciences.

Industrial manufacturing

Many industrial experiments have count response data or dichotomous (binary) response data instead of measured data. Count response data is often met as the number of (unwanted) particles (defects) on products. Dichotomous data arise as the number of good products out of a fixed number of n products included in the experiment.
We shall mention two disciplines of technology where count response data is encountered.

When soldering chip components on printed circuit boards, the number of soldering failures on the circuit board and the number of properly soldered chip components are measures of the quality of the soldering process. The question of the technologists to be answered is: what is the influence of the process factors in the experiment on the quality of the soldering process? A related question is: what is the influence of the process factors on the failure rate of the soldering process and what is the setting of these factors to give maximum process yield?

In the manufacture of Integrated Circuits (IC's) one of the major problems is the presence of small particles (dust) that can cause defects in IC-components. It is of vital importance to find settings for the process factors that will minimize the occurence of such defects. One question in this context is: what are the major sources of variation in the manufacturing process that influence the number of these particles?

Both the soldering problem and the IC manufacturing problem involve overdispersion and dependence of count data and there are no general statistical methods to analyse the data.

Marketing research

The number of units of a certain product, e.g. margarine, bought during a fixed time interval by an individual consumer can be modeled by the Poisson distribution. However, different consumers have different Poisson parameters and overdispersion of the number of units is observed.
Given the total number of units bought by an individual consumer, the result of classifying these units by branches can be modeled by the multinomial distribution. However, the probability vector of the multinomial is different for different consumers and an extension of the multinomial model is wanted.

Biomedical sciences

An example of an experiment with random nested factors and count responses is the following.
Suppose that some trees are sampled from a forest. Suppose that for each tree in the sample, branches are sampled from the branches of that tree. The number of insects is counted on each of the sampled branches. This experiment may serve to answer questions like: what is the influence of the variation of trees and branches on the number of insects? With count response data we would like to test and estimate variance components, and this problem cannot be solved by standard techniques.

An example from medicine relates to the frequent need for a dependent classification of results (cluster sampling).
Consider the following experiment: let each of J treatments be carried out K times on each of I patients and let the response variable be polytomous and ordered having L levels, such as health improvement at the levels no improvement, some improvement and substantial improvement. Usually, the I patients are considered as a random sample from a large population of patients and then the classification of treatment results is dependent. To answer questions like: is there any difference between the treatments, we need a model for dependently classified data.

Finally, if more motivation is needed, some literature will be refered to. Three quotations can speak for themselves.

... however, we have not considered the analysis of data using loglinear models in situations corresponding to nesting and random effects ANOVA models. Several research problems related to this topic require solution.

Bishop c.s. (1975), 371.

Problem 20. Give a general discussion for analysing interactions in log linear models for Poisson and binomial data in the presence:
(a) of overdispersion, (b) of underdispersion.

Cox (1984), 23.

... the important and difficult extension of the theory of generalized linear models to random effects models.

Pregibon (1984), 1592.

1.3. SOLUTIONS THAT HAVE BEEN GIVEN

Results available on the extensions of the standard theory of count data analysis can be classified into four categories:

1. Over(under)dispersed independent count data;
2. Over(under)dispersed independent dichotomous data;
3. Over(under)dispersed independent polytomous data;
4. Dependent count data, dependent dicho(poly)tomous data.

For each category, we shall summarize the results known from the literature.

1.3.1. Over(under)dispersed independent count data

The problem of overdispersed count data was studied by Paul and Plackett (1978), with respect to the behaviour of test statistics for testing the equality of Poisson parameters. A gamma compounded Poisson (or negative binomial) distribution is assumed and for this distribution the conclusion is: when using standard tests, the probability of rejecting the null hypothesis increases because of the increased variation in the data which is not accommodated in the standard Poisson model.

There is a paper by Breslow (1984) which treats the modeling of extra-Poisson variation in count data. An extra model parameter for overdispersion is estimated by two, somewhat heuristic, approaches: an approach assuming normality of the logarithm of large count data, and a quasi-likelihood approach for small counts. In both cases, the extra parameter and the linear model parameters are estimated iteratively.

A recent paper by Ross and Preece (1985) discusses the fitting of the negative binomial distribution to a set of data from a single population, estimating the model parameters by maximum likelihood. They give examples of and references to the application of this distribution in a biological context. Other applications in this context have been given by Manton c.s. (1981) and by Nedelman (1983).

Applications in the modeling of consumer purchasing behaviour are from Chatfield and Goodhardt (1970, 1973) and Dunn c.s. (1983).

Cox (1983) studies the efficiency of maximum likelihood estimation in the presence of modest amounts of overdispersion.

1.3.2. Over(under)dispersed independent dichotomous data

D.W. Finney observed overdispersion when fitting a linear model to 'the probit $\Phi^{-1}(\hat{p})$' of $\hat{p} = X/n$ where X is the number of successes out of n trials and Φ is the standard normal distribution function. He constructed the heuristic "heterogeneity correction" X^2/df for (co)variances of estimated model parameters, where X^2 is the Pearson goodness-of-fit statistic with df degrees of freedom.

Generalizations of the binomial distribution for overdispersion were studied by Skellam (1948), and later by Altham (1978) and Tarone (1979).

An application of such a generalization was described by Segreti and Munson (1981).

The problem of overdispersion was treated by Crowder (1978) in a more profound way. He assumes a beta-binomial distribution for the overdispersed dichotomous data. A likelihood procedure for the estimation and testing of model parameters is presented for a general (fixed factor) design matrix. The extra model parameter for overdispersion can be estimated if some replicates "per cell" are available.

A treatment by the quasi-likelihood method is presented by Williams (1982) who does not make any distributional assumptions. Only the variance function is needed; it is assumed to be of the "beta-binomial type" $\mathrm{var}(X) = \{1 + \phi(n-1)\}\ np\ (1-p)$. Estimation of the linear model parameters and the parameter ϕ for extra-binomial variation is carried out by IWLS and, respectively, by equating Pearson's X^2 for a full model to its expected value. The computer program GLIM can be used to do this, avoiding awkward calculations based on likelihood; see Crowder (1978). William's paper formed the basis for Breslow's treatment on extra-Poisson variation, see 1.3.1.

Also Brooks (1984) contributed to the likelihood approach of overdispersed dichotomous data. He proposes heuristic and approximate likelihood ratio tests based on the beta-binomial distribution that can be carried out by the standard program GLIM, contrary to Crowder's likelihood ratio procedure for which a special computer program is needed. Fitting linear models to correlated binary data by Gaussian estimation is discussed by Crowder (1985). Finally, Prentice (1986) has extended the beta-binomial distribution to allow for underdispersion and for dependence of the parameter ϕ on covariate measurements.

1.3.3. Over(under)dispersed independent polytomous data

Overdispersed polytomous data shows up at cluster sampling and survey sampling; it has received some attention in the literature.

Some early results are those of Mosimann (1962, 1963), Cohen (1976) and Altham (1976), mostly considering single populations. More general are those of Brier (1980), who derives asymptotic testing results for a vector $X = \{ X_i \}_{i=1}^{r}$ of numbers of classified objects, assumed to have the Dirichlet multinomial distribution. This distribution is obtained by giving the vector $(p_1, p_2, \ldots, p_r)$ of the multinomial $(n, p_1, p_2, .., p_r)$ distribution a Dirichlet distribution with parameter vector $(\beta_1, \beta_2, .., \beta_r)$. If classified data is available for N independent clusters it is shown that G^2 and X^2 statistics from standard loglinear model theory have a $C\ \chi^2$ type of limiting distribution for $N \rightarrow \infty$, C being some constant depending on n and the amount of heterogeneity between clusters.

Other results on cluster sampling that deserve mentioning are those of Plackett and Paul (1978) for testing symmetry in a squared contingency table, again under the assumption of a Dirichlet multinomial distribution, and those of Gleser and Moore (1983, 1985), who present some general results on the asymptotic distributions of goodness-of-fit tests under positive dependence of observations, showing that the asymptotic distributions of these test statistics are large-tailed as compared with the chi-squared distribution.
A survey of the use of chi-squared statistics is given by Fienberg (1979), who also mentions cluster sampling.

Problems from survey sampling are discussed, among others, by Bedrick (1983), Fellegi (1980) and Rao and Scott (1984). In this case, distributions cannot be established properly. Therefore, only assumptions of large sample normality are made by these authors to derive limiting distributions of test statistics.

1.3.4. Dependent count data, dependent dicho(poly)tomous data

For dependent count data, not many results can be found.
An extensive treatment of the negative multinomial distribution was given by Sibuya c.s. (1964); it is the multivariate analogue of the negative binomial distribution, see section 1.3.1. It can be of use for modeling dependent count data if the dependence structure is simple. The extension of Nelson (1985) allows for a more complicated dependence structure.

Forcina (1984) in an unpublished communication proposes a model for count data from a restricted version of the nested design with two random factors. McCullagh and Nelder (1984), p. 255, treat a simple nested structure.

On dependent dichotomous data, three papers appeared recently. In the context of questionnaires a paper by Anderson and Aitkin (1985) discusses models for dependent binary data with associated fixed covariates; the data is classified in groups by random nested factors. The EM-algorithm is used for maximum likelihood estimation of the model parameters.

A mixed model for categorial data is presented by Beitler and Landis (1985), which is similar to the corresponding two-way Anova model for quantitative data. Gilmour, Anderson and Rae (1985) deal with dependent dichotomous data formed by classifying samples from underlying normal distributions and it is their aim to estimate location and scale parameters of these underlying normal distributions.

These results on dependent count data analysis have only very recently become available. Possibly random factors have not always been recognized as such in count data problems and random factors were taken for fixed factors. On the other hand, the problem is quite hard to tackle.

In this thesis we shall present new results on this subject. A survey of these results can be found in section 1.4. of this introduction.

1.4. NEW RESULTS

In this section we shall summarize new results that were obtained recently, forming the basic contents of this thesis. We shall first mention briefly the types of experimental designs for which the new analysis methods for count response data have become available. These types of experimental designs are the following.

I. Fixed factor designs

1. Completely randomized fixed factor design with nominal factors, where count data show overdispersion (chapters 2, 3 and 4); in the model an additional random component is introduced for overdispersion.

2. Split-plot design (chapter 5) with fixed factors and with additional random components for whole plot error and for the interaction between whole plot error and sub-plot factors.

3. Paired comparisons design with two treatments and ordered response categories for treatment difference (chapter 6); additional random components are introduced for the interaction between blocks and treatments.

II. Random factor designs

4. Random factor design with crossed factors (chapter 7) and with nested factors (chapters 7 and 8) using a different approach in chapter 8; random components are introduced for main effects, for interactions and for error.

For each of these experimental designs the approach to data analysis will now briefly be discussed. An extensive discussion can be found in the relevant chapters of this thesis.

1.4.1. Fixed factor design

For the fixed factor design, there are new results for the following two situations:

- Count data restricted to a maximum of say n (dichotomous data, see chapter 2) in the case of overdispersion with respect to the binomial distribution; n does not depend on the levels of the design factors. Approximate χ^2-tests and F-tests are presented for testing linear models for the logit of the probability parameter π. This approximation holds for large n and for small overdispersion. For designed experiments a constant parameter n is often realizable and then the method is useful. It has some advantages over other relatively new methods published in the literature.

- Count data not restricted to a maximum (chapters 3 and 4), in the case of overdispersion with respect to the Poisson distribution. Approximate χ^2-tests and F-tests are presented where we assume homogeneity with respect to the extra parameter for overdispersion (compare this with the assumption of equal variances for the normal case). The results were obtained for large numbers of replicates and for large counts. In both cases two models will be discussed, which differ in the way extra-Poisson variation is introduced.

The case of large counts deserves some special attention, because the analysis is then particularly simple and elegant. The analysis is based on the Poisson deviance (for the model with a linear variance function) or the gamma deviance (for a quadratic variance function). As in the Anova theory, approximate F-tests are constructed to eliminate the extra parameter for overdispersion. For practical applications it is important to know that both deviances are available in a statistical package like GLIM.

1.4.2. Split-plot design

For the split-plot design a model is constructed (in chapter 5) allowing for dependent count data within the whole plot, and for the interaction between whole plot error and sub-plot factors. Again, approximate χ^2-tests and F-tests are used for the analysis of whole plot and sub-plot factorial effects; the approximate distribution holds for large counts and for a large number of replicates as in section 1.4.1.

1.4.3. Paired comparisons design

Let two treatments be applied pairwise, n times at each of N random blocks and let the observed differences of treatment responses be classified in K ordered response categories (see chapter 6).

Now the question is how to test for difference between treatments if an interaction between random blocks and treatments is present. For this situation no solution is given in the literature. We propose the use of a rank test of Wilcoxon as the instrument for testing the hypothesis of no treatment effect. In this case of ordered polytomous response data in random blocks where an interaction is present between treatments and blocks, the limiting distribution of e.g. Wilcoxon's signed rank test for symmetry is not the standard normal distribution.

Under a Dirichlet-multinomial model assumption the limiting distribution of this test of Wilcoxon is obtained in chapter 6.

1.4.3. Paired comparisons design

Let two treatments be applied pairwise, n times at each of N random blocks and let the observed differences of treatment responses be classified in K ordered response categories (see chapter 6).

Now the question is how to test for difference between treatments if an interaction between random blocks and treatments is present. For this situation no solution is given in the literature. We propose the use of a rank test of Wilcoxon as the instrument for testing the hypothesis of no treatment effect. In this case of ordered polytomous response data in random blocks where an interaction is present between treatments and blocks, the limiting distribution of e.g. Wilcoxon's signed rank test for symmetry is not the standard normal distribution.

Under a Dirichlet-multinomial model assumption the limiting distribution of this test of Wilcoxon is obtained in chapter 6.

1.4.4. Random factor design

The random factor design is the subject of chapters 7 and 8.

In chapter 7 we shall build a model by assuming lognormality for the random model components for main effects and interactions. The product of these components forms the random intensity of the Poisson process modeling the data generating process. For large counts we are now able to analyse the data for any factorial design with nested and crossed random factors.

Effectively, this analysis amounts to performing a standard Anova on the log-transform of the data.

Much less restrictive are the assumptions made in chapter 8. There we shall follow the quasi-likelihood approach to analyse the data from nested designs with random factors.

For this type of random factor design, a new approach to the analysis of data is proposed which is not even restricted to the analysis of count data. In the literature the quasi-likelihood method is only applied to independent data; we shall use it for dependent data.

The difference between the model building for nested designs in chapter 7 and in chapter 8 can be sketched as follows.

In chapter 8 we only assume that we know the variance function for each factor (and for error) in the nested design which expresses the variance of the random level of this factor as a function of its mean value given the level of the factor which is at one stage higher in the design.
For example, for two random factors the variance functions to be specified can be the following.

First,

$\mathrm{var}(M_i) = \sigma_1^2\mu$,

where M_i is the random level of the first factor;
μ is the general mean;
σ_1^2 is a positive parameter.

Secondly, given $M_i = m_i$,

$\mathrm{var}(M_{ij}) = \sigma_2^2 m_i$,

where M_{ij} is the random level of the second factor;
σ_2^2 is a positive parameter.

Thirdly, given $M_{ij} = m_{ij}$,

$\mathrm{var}(X_{ijk}) = \sigma_3^2 m_{ij}$,
where X_{ijk} is the k^{th} observation at the levels i and j of the two factors;
σ_3^2 is a positive parameter.

For large sample sizes, approximate χ^2-tests and F-tests will be obtained for testing the main effects. On a set of data, this method was compared with the method of chapter 7; no essential differences in the conclusions were observed.

Chapter 8 is concluded by suggesting some further research which has to be done. Specifically, it is the quasi-likelihood approach based only on very simple assumptions with respect to the mean and variance of the data that should be explored further to provide us with solutions to problems which have not yet been solved.

REFERENCES

Altham, P.M.E. (1976), Discrete variable analysis for individuals grouped into families, Biometrika 63, 263-269.

Altham, P.M.E. (1978), Two generalizations of the binomial distribution, Appl. Statist. 27, 162-167.

Anderson, D.A. and Aitkin, M. (1985), Variance component models with binary response: interviewer variability, J.R. Statist. Soc. B 47, 203-210.

Baker, R.J. and Nelder, J.A. (1978), The GLIM system, release 3, generalised linear interactive modelling manual, Oxford: NAG.

Bedrick, E.J. (1983), Adjusted chi-squared tests for cross-classified tables of survey data, Biometrika 70, 591-595.

Beitler, P.J. and Landis, J.R. (1985), A mixed-effects model for categorical data, Biometrics 41, 991-1000.

Bishop, Y.M.M.; Fienberg, S.E. and Holland, P.W. (1975).
Discrete multivariate analysis, theory and practice, The MIT-Press, Cambridge.

Breslow, N.E. (1984), Extra-Poisson variation in log-linear models, Appl. Statist. 33, 38-44.

Brier, S.S. (1980), Analysis of contingency tables under cluster sampling, Biometrika 67, 591-596.

Brooks, R.J. (1984), Approximate likelihood ratio tests in the analysis of beta-binomial data, Appl. Statist. 33, 285-289.

Chatfield, C. and Goodhardt, G.J. (1970), The beta-binomial model for consumer purchasing behaviour, Appl. Statist. 19, 240-250.

Chatfield, C. and Goodhardt, G.J. (1973), A consumer purchasing model with erlang inter-purchase times, JASA 68, 828-835.

Cohen, J.E. (1976), The distribution of the chi-squared statistic under clustered sampling from contingency tables, J.Am. Statist. Assoc. 71, 665-670.

Cox, D.R. (1983), Some remarks on overdispersion, Biometrika 70, 269-274.

Cox, D.R. (1984), Interaction, Int. Statist. Rev. 52, 1-32.

Crowder, M.J. (1978), Beta-binomial anova for proportions, Appl. Statist. 27, 34-37.

Crowder, M.J. (1985), Gaussian estimation for correlated binomial data, J.R. Statist. Soc. B 47, 229-237.

Dunn, R., Reader, S. and Wrigley, N. (1983), An investigation of the assumptions of the NBD model as applied to purchasing at individual stores, Appl. Statist. 32, 249-259.

Everitt, B.S. (1977), The analysis of contingency tables, Chapman and Hall, London.

Fellegi, I.P. (1980), Approximate tests of independence and goodness of fit based on stratified multistage samples, J. Am. Statist. Assoc. 75, 261-268.

Fienberg, S.E. (1977), The analysis of cross-classified categorical data, The MIT-Press, Cambridge.

Fienberg, S.E. (1979), The use of chi-squared statistics for categorical data problems, J.R. Statist. Soc. B 41, 54-64.

Forcina, A. (1984), Estimating variance components with discrete data, personal communication.

Gilmour, A.R., Anderson, R.D. and Rae, A.L. (1985), The analysis of binomial data by a generalized linear mixed model, Biometrika 72, 593-599.

Gleser, L.J. and Moore, D.S. (1983), The effect of dependence on chi-squared and empiric distribution tests of fit, Ann. Statist. 11, 1100-1108.

Gleser, L.J. and Moore, D.S. (1985), The effect of positive dependence on chi-squared tests for categorical data, J.R. Statist. Soc. B, 47, 459-465.

Green, P.J. (1984), Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives, J.R. Statist. Soc. B 46, 149-192.

Haberman, S.J. (1974), The analysis of frequency data, The University of Chicago Press, Chicago.

Manton, K.G., Woodbury, M.A. and Stallard, E. (1981), A various components approach to categorical data models with heterogeneous cell populations: analysis of spatial gradients in lung cancer mortality rates in North Carolina counties, Biometrics 37, 259-270.

McCullagh, P. (1983), Quasi-likelihood functions, Ann. Statist. 11, 59-67.

McCullagh, P. and Nelder, J.A. (1984), Generalized linear models, Chapman and Hall, London.

Mosimann, J.E. (1962), On the compound multinomial distribution, the multivariate beta-distribution, and correlations among proportions, Biometrika 49, 65-82.

Mosimann, J.E. (1963), On the compound negative multinomial distribution and correlations among inversely sampled pollen counts, Biometrika 50, 47-54.

Nedelman, J. (1983), A negative binomial model for sampling mosquitoes in a malaria survey, Biometrics 39, 1009-1020.

Nelder, J.A. and Wedderburn, R.W.M. (1972), Generalized linear models, J.R. Statist. Soc. A 135, 370-384.

Nelson, J.F. (1985), Multivariate gamma-Poisson models, JASA 80, 828-834.

Paul, S.R. and Plackett, R.L. (1978), Inference sensitivity for Poisson mixtures, Biometrika 65, 591-602.

Plackett, R.L. (1974), The analysis of categorical data, Griffin, London.

Plackett, R.L. and Paul, S.R. (1978), Dirichlet models for square contingency tables, Commun. Statist. - Theor. Meth., A7, 939-952.

Pregibon, D. (1984), Book Review of: P. McCullagh and J.A. Nelder, Generalized linear models, 1983; Ann. Statist. 12, 1589-1596.

Prentice, R.L. (1986), Binary regression using an entended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors, JASA 81, 321-327.

Rao, J.N.K. and Scott, A.J. (1984), On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data, Ann. Statist. 12, 46-60.

Ross, G.J.S. and Preece, D.A. (1985), The negative binomial distribution, The Statistician 34, 323-336.

Scheffé, H. (1959), The analysis of variance, John Wiley & Sons, New York.

Segreti, A.C. and Munson, A.E. (1981), Estimation of the median lethal dose when responses within a litter are correlated, Biometrics 37, 153-156.

Sibuya, M., Yoshimura, I. and Shimizu, R. (1964), Negative multinomial distribution, Ann. Inst. Statist. Math. 16, 409-426.

Skellam, J.G. (1948), A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials, J.R. Statist. Soc. B 10, 257-261.

Tarone, R.E. (1979), Testing the goodness of fit of the binomial distribution, Biometrika 66, 585-590.

Williams, D.A. (1982), Extra-binomial variation in logistic linear models, Appl. Statist. 31, 144-148.

Wedderburn, R.W.M. (1974), Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method, Biometrika 61, 439-447.

CHAPTER 2

ON THE ANALYSIS OF VARIANCE FOR BETA-BINOMIAL RESPONSES

Abstract

The beta-binomial distribution is reported in literature as a useful generalization of the binomial in case of heterogeneous binomial sampling. An extra model parameter is introduced to accommodate for extra-binomial variation. Some additions to results already available will be given by presenting approximate F-tests for factorial designs, where the response variable is of 0-1 type and sampling is heterogeneous binomial. These tests can be used when sample sizes are large and equal and some degrees of freedom are left from replicates or negligible interactions to estimate the extra model parameter.

(Published in Statistica Neerlandica 39 (1985), 27-34).

1. Introduction

The analysis of binomial and multinomial response data, classified by several fixed factors, is surveyed in the books of Bishop c.s. (1975) and Fienberg (1977). The log-linear model has proved to be a flexible tool for the analysis of this data.
However, sometimes it is observed that variation in 0-1 responses cannot be explained by the full log-linear model and binomial error. A distribution accommodating for this heterogeneous binomial sampling or extra-binomial variation is the beta-binomial distribution (BBD). Recently it was discussed by Paul and Plackett (1978), Brier (1980) and Crowder (1978) reverting to earlier results of Mosimann (1962).

Some contributions to the discussion will be made by presenting approximate F-tests for the analysis of beta-binomial responses or in general, Dirichlet multinomial responses. The approximate tests can be used for i) large and equal sample sizes, ii) small extra-binomial variation and iii) factorial designs, where replicates or negligible interactions deliver remainder degrees of freedom to estimate an extra parameter.

Further, these tests are based on standard statistics, known from loglinear model analysis, which makes them attractive in a computational way. The analysis of a 2^4 complete factorial design, with 0-1 classified responses is presented as an application at the end of the paper.

2. Approximate test statistics

2.1. Dirichlet multinomials

The Dirichlet compounded multinomial distribution (DMD, see Mosimann (1962) and Brier (1980)), and the bivariate special case, the beta-binomial distribution (BBD, see Crowder (1978)) will be basic to the models considered.

The DMD is obtained in the following way. Suppose that random vector $X = \{X_i\}_{i=1}^{I}$ has a multinomial (n, p) distribution, conditionally on probability vector p. If p is a random vector having a Dirichlet distribution with parameter vector $(\beta_1, \beta_2, \ldots, \beta_I)$, then the marginal distribution of X is DM.

Frequently, a more useful parameterization is by vector $\pi = (\pi_1, \ldots, \pi_I)$, where $\pi_i = \beta_i/(\beta_1 + \ldots + \beta_I)$ and sum $\beta = \beta_1 + \ldots + \beta_I$. Note that $E\ p_i = \pi_i$, and $\mathrm{var}\ p_i = \pi_i(1 - \pi_i)/(1 + \beta)$. If $\beta \to \infty$ and π is fixed, the multinomial (n, π) is obtained.

From Paul and Plackett (1978) we shall recall a limit property of the DMD. Later, this property will be needed to derive a limit theorem for test statistics.

A limit property of Dirichlet multinomials

If vector $\{X_i\}_{i=1}^{I}$ has a DMD with parameters n, $\{\pi_i\}_{i=1}^{I}$ and β, where $\beta = n\gamma$ for some fixed γ and $n \to \infty$, then the asymptotic distribution of $\{X_i\}$ after standardisation is normal. Approximately, for large n,

$E\ X_i = n\pi_i$, $\mathrm{var}\ X_i = \Lambda n\pi_i(1 - \pi_i)$, $\mathrm{cov}\ (X_i, X_{i'}) = -\Lambda n\pi_i\pi_{i'}$

for $i \neq i'$, where $\Lambda = (1 + \gamma)/\gamma$.

2.2. The model

We shall now consider independent, I variate response vectors $\{X_{ijk}\}_{i=1}^{I}$, being classified by J levels of a fixed factor A, K replicates being available for each level j. It is assumed that $\{X_{ijk}\}_{i=1}^{I}$ has the DMD with parameters n, $\{\pi_{ij}\}_{i=1}^{I}$ and β_j, not depending on k. A further homogeneity assumption is $\beta_j = \beta$, not depending on j. Then the DMD is parametrized by n and $\{\beta\ \pi_{ij}\}_{i=1}^{I}$.

A motivation for the DM model may be found in the following considerations. From replicates of the experiment at level j of design-factor A, large variation in the data may show that the multinomial distribution with fixed probability vector p is not an acceptable distribution. Sometimes a decent technical explanation can be given for this phenomenon. Then the extension of the multinomial model can be useful, which is obtained by letting the probability vector p vary between replicates according to the Dirichlet distribution, accommodating for extra-multinomial variation. The result is the DM distribution.

Referring to the limit property of 2.1., we shall derive a limit theorem for test statistics under the following assumption.

Assumption

$n \to \infty$ and $\beta \to \infty$, where $\gamma = \beta/n$ is fixed.

If p_{ijk} is the class i multinomial probability, then the assumption implies

$$\operatorname{var} p_{ijk} = \frac{1}{1+\beta}\,\pi_{ij}(1-\pi_{ij}) = \frac{1}{1+\gamma n}\,\pi_{ij}(1-\pi_{ij}).$$

This also shows that large n implies small var p_{ijk}. Also note that

$$\operatorname{var} X_{ijk} = n\,\pi_{ij}(1-\pi_{ij})\{1 + \frac{1}{1+\gamma n}(n-1)\} \text{ so that}$$

$$\operatorname{var} (n^{-1/2}X_{ijk}) \to (1+\gamma^{-1})\,\pi_{ij}\,(1-\pi_{ij}) \text{ for } n \to \infty,$$

where the limit variance was $\pi_{ij}(1-\pi_{ij})$ for $\{X_{ijk}\}$ having the multinomial (n, $\{\pi_{ij}\}$) distribution.

Throughout the paper, we shall use the notation Λ for $1 + \gamma^{-1}$.

2.3. Test statistics

We shall derive approximate F-statistics for testing the main effect of factor A or, to be more general, main effects and interactions of M crossed factors $A_1, A_2, \ldots, A_M$. For this purpose, the log-linear model is fitted to the expected probability vector $\{\pi_{ij_1j_2\ldots j_M}\}_{i=1}^{I}$, where $j_m = 1,2,\ldots,J_m$; $m = 1,2,\ldots,M$, treating vector $\{X_{ij_1\ldots j_Mk}\}_{i=1}^{I}$ as a response vector by conditioning on proper marginal sums. For $I = 2$ this reduces to the logit model, see Fienberg (1977), p. 77.

To compose the F-statistics, recall from Bishop c.s. (1975), Pearson's X^2 and the multinomial log likelihood-ratio G^2:

$$X^2 = \sum_{ijk} (X_{ijk} - n\,\hat{\pi}_{ij})^2 / n\,\hat{\pi}_{ij}$$

$$G^2 = 2 \sum_{ijk} X_{ijk} \log (X_{ijk}/n\,\hat{\pi}_{ij}).$$

Here $\hat{\pi}_{ij}$ is the MLE of π_{ij} under the multinomial distribution and the hypothesized log-linear model for π_{ij}.

Under the assumption of 2.2 we shall prove that X^2 and G^2 are distributed as Λ times chi-square under the hypothesis, asymptotically for $n \to \infty$. Important results are summarized in the following proposition.

Define nested log-linear models by indices p, q .. and the ordering $p < p + 1 \leq q < q + 1$ in the following sense: if $i < j$, then model j is a reduction of model i by deletion of one or more model parameters.

Proposition

Under the assumption of 2.2., and if the tested hypotheses are true, the following statements hold

1. G^2_p and X^2_p have the same asymptotic distribution, where G^2_p and X^2_p correspond with a model p.

2. $G^2_p \overset{D}{\to} \Lambda\chi^2_{v_p}$ asymptotically, where $\overset{D}{\to}$ denotes convergence in distribution and $\chi^2_{v_p}$ is a chi-square random variable with v_p degrees of freedom for model p.

   Also

   $G^2_{p+1} - G^2_p \overset{D}{\to} \Lambda\chi^2_{v_{p+1} - v_p}$, asymptotically.

3. $G^2_{q+1} - G^2_q$ and $G^2_{p+1} - G^2_p$ are independent, asymptotically.

4. $(G^2_{q+1} - G^2_q)\ d_p/(G^2_{p+1} - G^2_p)\ d_q \overset{D}{\to} F_{d_q,d_p}$ asymptotically, where $d_p = v_{p+1} - v_p$ and $F_{d_q,\ d_p}$ is an F random variable with d_q and d_p degrees of freedom.

We shall give an outline of proof.

Proof 1. Recall Theorem 14.9-2 from Bishop c.s. (1975):

Let $\hat{p}$ be a vector of observed fractions having an I-dimensional multinomial distribution with parameters (n, π), while $\hat{\pi}$ is any estimate of π such, that $\hat{p}$ and $\hat{\pi}$ have a joint limiting normal distribution, i.e.

$$\sqrt{n}((\hat{p}, \hat{\pi}) - (\pi, \pi)) \overset{D}{\to} N(0,\Sigma) \qquad \ldots (1)$$

for some covariance matrix Σ. Then G^2 and X^2 have the same limiting distribution.

As (1) holds for the multinomial distribution and the log-linear model, it can be proved that

$$\sqrt{n}((\hat{p}, \hat{\pi}) - (\pi, \pi)) \overset{D}{\to} N(0, \Lambda\Sigma)$$

for the DM, and analogous results hold for the product of independent DMD's with equal scale factor Λ. So for each true log-linear model, G^2 and X^2 have the same asymptotic distribution.

2. As to the asymptotic distribution of X^2, we revert to corollary 14.9-3 from Bishop c.s. (1975). Following the proof of this corollary, it can be shown that

$X^2 \overset{D}{\rightarrow} \Lambda\chi^2_v$ where v is the number of degrees of freedom.

Then by 1., $G^2 \overset{D}{\rightarrow} \Lambda\chi^2_v$.

Analogously, it is proved that $G^2_{p+1} - G^2_p \overset{D}{\rightarrow} \Lambda\chi^2_{v_{p+1} - v_p}$ following the proof of theorem 14.9-8 from Bishop c.s.

3. For the multinomial distribution this result is stated by Haberman (1974), p. 117; for the DMD it holds as well.

4. This follows from results 2. and 3. with the Mann-Wald theorem, see Billingsley (1968), p. 31.

We shall pay some more attention to part 4. of the proposition, which enables us to construct approximate F tests for large n. The denominator of this F-test consists of a remainder G^2 statistic, corresponding with a full log-linear model, where all relevant effects are included. The numerator is a statistic of type $G^2_{q+1} - G^2_q$ for testing some hypothesized reduction of the log-linear model.

Then under the hypothesis, $F = \dfrac{G^2_{q+1} - G^2_q}{G^2} \quad \dfrac{v}{v_{q+1} - v_q}$ has an approximate F distribution, with $v_{q+1} - v_q$ and v degrees of freedom.

Comparing this result with Crowder (1978) and Brier (1980), we can draw some conclusions.

A possible draw back of the method is the required equality of sample sizes say n_s, which may be no problem in case of a designed experiment, but which can be a fatal requirement for sampled data.

When sample sizes n_s do not differ widely, in practice we are tempted to use the same procedure as for equal n_s. From results of Rao and Scott (1984) it can be shown that X^2 and G^2 for testing a hypothesized model reduction are approximately distributed as $\bar{\Lambda} * \chi^2$, where $\bar{\Lambda}$ is a weighted mean

$$\bar{\Lambda} = \sum_s w_s \Lambda_s = \sum_s w_s (1 + n_s/\beta) = 1 + \sum_s w_s n_s/\beta, \ \sum_s w_s = 1,$$

weights w_s depending on the hypothesized reduction. Using the testing procedure somewhat heuristically as if sample sizes n_s were equal comes to neglecting the dependence of weights on the hypothesis. Obviously, some further research on the accuracy of this approach may be useful.

Of special interest is the case $K = 1$, only one replicate being available per cell. In factorial experiments, high-order interactions can often be neglected on technical grounds, and a denominator G^2 is then obtained by fitting a log-linear model, from which these interactions are excluded. Also fractional replication and confounding can be treated according to these principles. In fact, Crowder and Brier do not give a solution for these cases.

We may conclude that the approximate analysis can be of use where nominal response data are obtained from factorial experiments, where equal sample sizes n_s can be realized without much difficulty, and where some interactions can be neglected to obtain remainder degrees of freedom for the F-tests.

3. Application

To investigate the dependence of the yield of a resistor manufacturing process on process factors, an experiment was carried out. Four factors were included in the experimental design, say factor A, B, C and D, each having two levels.

At each combination of factorial levels, 500 resistors were manufactured and classified as $i = 0$ (rejected) or $i = 1$ (accepted) according to some quality measure. The experimental design is then a complete 2^4-factorial with 0 - 1 response. Observations are presented in table 1.

Factor A		a_0		a_1	
C	B D	b_0	b_1	b_0	b_1
c_0	d_0	172	406	180	440
	d_1	438	441	363	461
c_1	d_0	196	418	190	450
	d_1	406	431	349	495

Table 1. Numbers of resistors being classified as $i = 1$ (accepted), out of 500 resistors.

We may start the data analysis by fitting the logit model, assuming binomial responses, omitting the interaction ABCD. The G^2-result is $G^2 = 8.49$, with 1 degree of freedom (df). As the interaction ABCD can be neglected on technical grounds, we are on the alert for extra-binomial variation. As it seems, the random part of the model is not of binomial type.

Introducing the BBD, a primary estimate $\hat{\Lambda} = 8.49$ is obtained for parameter Λ. Next, for testing three-factor interactions we use $\hat{\Lambda}$ to correct statistics of type $G^2_{q+1} - G^2_q$. We adhere to the testing procedure proposed by Brown (1976), who introduced the concepts of partial and marginal association. The association, corrected by $\hat{\Lambda}$, is presented in table 2.

Factorial effect	df	Partial association	Marginal association
ABC	1	0.76	0.69
ABD	1	4.12	3.38
ACD	1	1.53	0.76
BCD	1	1.35	0.55

Table 2. Partial and marginal association for three factors, corrected by $\hat{\Lambda}$.

Comparing these values with 1 df chi-square fractiles, only ABD tends to be significant at 5%. We shall incorporate the other three-factor interactions into the remainder G^2 to obtain a new and final estimate $\hat{\Lambda} = G^2/4 = 8.12$, based on 4 degrees of freedom.

As $\gamma = (\Lambda - 1)^{-1}$ can be estimated by $\hat{\gamma} = (\hat{\Lambda} - 1)^{-1} = 0.14$,

$\beta = n\,\gamma$ leads to the estimate $\hat{\beta} = 70$, which may give some justification to the use of a testing procedure for large β.

A final testing of main effects and two-factor interactions by means of approximate F-tests shows, that B and D main effects and BD interaction are significant. Approximate F-tests, obtained by deviding partial and marginal association $G^2_{q+1} - G^2_q$ by the remainder G^2, correcting for df, are given in table 3.

Factorial effect	df	Partial F	Marginal F
B	(1,4)	137.76	127.47 ($P < 0.005$)
D	(1,4)	79.77	69.50 ($P < 0.005$)
BD	(1,4)	12.26	11.30 ($P < 0.05$)

Table 3. Approximate F-tests with 1 and 4 degrees of freedom.

Under the logit model

$$\log \frac{\pi_{1jklm}}{1-\pi_{1jklm}} = u + u_{2(k)} + u_{4(m)} + u_{24(km)}$$

π-parameters can simply be written as $\pi = \pi_{1km}$.

Moment estimates $\hat{\pi}_{1km}$ for π_{1km} are presented in table 4.

D \ B	b_0	b_1
d_0	0.369	0.857
d_1	0.778	0.914

Table 4. Moment estimates $\hat{\pi}_{1km}$ for expected probabilities.

Since $\hat{\pi}_{1km} = \dfrac{X_{1+k+m}}{4n}$, X_{1jklm} being approximately normal for large n, with mean value n π_{1km} and variance $\Lambda n\ \pi_{1km}\ (1 - \pi_{1km})$, an approximate $(1 - \alpha)$ confidence interval

$$\pi_{1km}\ \varepsilon\ \hat{\pi}_{1km} \pm u_{1-\alpha/2} \surd \left\{\frac{\hat{\Lambda}\ \hat{\pi}_{1km}\ (1-\hat{\pi}_{1km})}{4n}\right\}$$

for π_{1km} may give some additional information on the accuracy of the estimate.

Acknowledgement

The author wishes to thank both referees for their valuable comments and suggestions to improve the manuscript.

References

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975),
Discrete Multivariate Analysis: Theory and Practice, Massachusetts Institute of Technology Press.

Brier, S.S. (1980),
Analysis of contingency tables under cluster sampling, Biometrika 67, 591 - 596.

Brown, M.B. (1976),
Screening effects in multidimensional contingency tables,
Applied Statistics 25, 37 - 46.

Billingsley, P. (1968),
Convergence of Probability Measures, John Wiley & Sons, New York.

Crowder, M.J. (1978),
Beta-binomial Anova for proportions, Appl. Statist. 27, 34 - 37.

Fienberg, S.E. (1977),
The analysis of cross-classified categorical data, The MIT-Press, Cambridge.

Haberman, S.J. (1974),
The Analysis of Frequency Data, The University of Chicago Press, Chicago.

Mosimann, J.E. (1962),
On the compound multinomial distribution, the multivariate beta-distribution, and correlations among proportions, Biometrika 49, 65 - 82.

Paul, S.R. and Plackett, R.L. (1978),
Inference sensitivity for Poisson mixtures, Biometrika 65, 591 - 602.

Rao, J.N.K. and Scott, A.J. (1984),
On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data, Ann. Statist. 12, 46 - 60.

CHAPTER 3

MODELS FOR RESPONSE DATA SHOWING EXTRA-POISSON VARIATION

Abstract

When count data show extra-Poisson variation, standard log-linear techniques to analyse the data may fail. In this paper a generalization of the log-linear modelling technique is proposed for the negative binomial model, as an extension of the Poisson model. An illustration is given by the analysis of a two-way classification of soldering failure data; extensions to more general classifications are possible.

(published in Statistica Neerlandica 38 (1984), 159-167).

1. Introduction

To analyse count response data, linear models are usually fitted to the logarithm of the vector of expected values. Well-established techniques of log-linear modelling are given by e.g. Fienberg (1977) and Bishop c.s. (1975). In this literature it is assumed that the response variable has a Poisson distribution.
From practice however, it was noted before that sometimes the Poisson model is not a suitable model, as the data may show too much variation. A survey of literature is presented by Paul & Plackett (1978) in which the phenomenon of what may be called extra-Poisson variation is discussed.

Recently a problem was met in consultation practice for which no solution was found in literature. Counts of soldering failures on print panels were classified by two factors, and it was asked to test their main effects and interaction, while a clear extra-Poisson variation showed from the data. A solution was found by introducing the negative binomial distribution generalizing the Poisson, accommodating in this way for extra-Poisson variation. Two negative binomial models were fitted to the soldering data, which is a relatively straightforward operation if the GLIM computer program is available as a tool for doing the computing work.

2. Models for extra-Poisson variation

In the following we shall concentrate on a two-way crossed classification of a vector of counts $\{X_{ij}\}$, $i=1,2,\ldots,I$, $j=1,2,\ldots,J$, where i and j indicate the levels of two fixed factors.
One set of standard assumptions for $\{X_{ij}\}$ is as follows:

(i) $\{X_{ij}\}$ has independent components X_{ij}
(ii) X_{ij} has a Poisson (m_{ij}) distribution
(iii) $\log m_{ij}$ is linear in some unknown parameters.

Sometimes, the random variable X_{ij} shows more variation than is explained by the Poisson distribution and the log-linear model containing all explanatory variables. To accommodate for extra-Poisson variation, assumption (ii) may be replaced by two new assumptions, (ii)' and (ii)":

(ii)' given $M_{ij} = m_{ij}$, X_{ij} has a Poisson (m_{ij}) distribution, where M_{ij} is a positive random variable
(ii)" $M_{ij} = G(\alpha_{ij}, \theta_{ij})$, where $G(\alpha_{ij}, \theta_{ij})$ are independent gamma random variables, with shape parameter α_{ij} and scale parameter θ_{ij}.

The new distribution of X_{ij} is the negative binomial with parameters $(\alpha_{ij}, p_{ij} = \theta_{ij}/(1+\theta_{ij}))$ and with probability function

$$P(X_{ij} = x) = \binom{x+\alpha_{ij}-1}{x} p_{ij}^{x} (1-p_{ij})^{\alpha_{ij}} , \quad x=0,1,2,\ldots$$

(see Johnson & Kotz (1969), p. 122).

The Poisson distribution can be seen as a special case of the negative binomial just presented: for $\alpha_{ij} \rightarrow \infty$, $\theta_{ij} \rightarrow 0$ and $\alpha_{ij}\theta_{ij} \rightarrow m_{ij}$ it converges to the Poisson (m_{ij}) distribution.
We shall specialize the general negative binomial model to the following special models:

Model I, with parameters (α_{ij}, θ).
Then X_{ij} is negative binomial $(\alpha_{ij}, p = \frac{\theta}{1+\theta})$ and

$$E\ X_{ij} = \alpha_{ij}\theta = m_{ij}$$
$$\mathrm{var}\ X_{ij} = \alpha_{ij}\theta(1+\theta) = m_{ij}(1+\theta)$$

substituting parameter $m_{ij} = E\ X_{ij}$ for $\alpha_{ij}\ \theta$.

Note that only shape parameter α_{ij} of the gamma distribution depends on factorial effects.

Model II, with parameters (α, θ_{ij}).

Then X_{ij} is negative binomial $(\alpha, p_{ij} = \dfrac{\theta_{ij}}{1+\theta_{ij}})$ and

$$E\ X_{ij} = \alpha\ \theta_{ij} = m_{ij}$$
$$\mathrm{var}\ X_{ij} = \alpha\ \theta_{ij}(1+\theta_{ij}) = m_{ij}(1+\alpha^{-1}m_{ij}).$$

Note that only scale parameter θ_{ij} of the gamma distribution depends on factorial effects.

In literature, the phenomenon of extra-Poisson variation and the relationship between var X_{ij} and E X_{ij} were studied before; a survey is given by Paul and Plackett (1978). In practice, a relationship of the type var $X_{ij} = c(E\ X_{ij})^b$, where $1 < b < 2$ seems to cover most cases of extra-Poisson variation. If the ratio var $X_{ij}/E\ X_{ij}$ is about constant, Model I will sufficiently explain the extra-Poisson variation. If the ratio increases with E X_{ij}, Model II may be useful, being somewhat extreme in the light of the results of Paul and Plackett.

It is advised to start the analysis by testing the hypothesis of Poisson distributions, fitting the full factorial model based on the Poisson distribution for X_{ij}. If the hypothesis is rejected, a choice between Model I and Model II can be based on a plot of $(X_{ijk} - \hat{m}_{ij})/\sqrt{\hat{m}_{ij}}$ against the estimates $\hat{m}_{ij} = \bar{X}_{ij+}$ of m_{ij} under the full Poisson model. If the variance of these residuals is more or less constant, Model I can be chosen; if it increases with $\hat{m}_{ij}$, Model II may be more suitable. Of course, it is not impossible that neither Model I nor Model II is a satisfactory model in case of extra-Poisson variation.

3. Model I

3.1. Discussion

We shall now study Model I in some detail, showing its relationship with the standard Poisson model as regards its asymptotic behaviour. Here and further we suppose that for each level i and j of the two factors, K independent replicates X_{ijk}, being distributed as X_{ij}, are available from a properly designed experiment.

For Model I, two types of asymptotic behaviour are relevant and lead to interesting results, namely:

- $m_{ij} \to \infty$, where θ and K are fixed
- $K \to \infty$, where θ and m_{ij} are fixed.

Firstly, we consider the asymptotic situation for $m_{ij} \to \infty$, where θ and K are fixed. Then by the Central Limit Theorem, the standardized X_{ij} has a normal limiting distribution. Hence, limit properties of X_{ij} are those of a Poisson (m_{ij}) random variable, when the variance of the latter is multiplied by a fixed constant $(1+\theta)$. The consequence is, that when the full parameter log-linear model is fitted, the standard Pearson X^2 and log-likelihood ratio G^2 statistics, defined as:

$$X^2 = \sum_{ijk} (X_{ijk} - \hat{m}_{ij})^2/\hat{m}_{ij} \text{ and}$$

$$G^2 = 2 \sum_{ijk} X_{ijk} \log(X_{ijk}/\hat{m}_{ij})$$

(see Bishop c.s. (1975), section 4.2),

will be distributed asymptotically as $(1+\theta)$ times chi-square, with $\nu = IJ(K-1)$ degrees of freedom. Analogous results hold for restricted models. To test for interaction and main-effects we may proceed as follows:

Let G^2 be the log-likelihood ratio statistic for the full log-linear model, while $G_2^2 - G_1^2$ is a difference of G^2 statistics, where G_2^2 corresponds with a hypothesized restriction on a model with log-likelihood ratio statistic G_1^2. Then G^2 and $G_2^2 - G_1^2$ are asymptotically independent and both $(1+\theta)\ \chi^2$ distributed under the hypothesis.

An approximate F-test is based on

$$F = \frac{G_2^2 - G_1^2}{G^2} \quad \frac{v}{v_2 - v_1},$$

having an asymptotic F-distribution with $v_2 - v_1$ and v degrees of freedom for $m_{ij} \to \infty$ under the hypothesis. Elimination of nuisance parameter $(1+\theta)$ is then on the lines of the analysis of variance.
The above statements can be proved following arguments from Engel (1983), where analogous results for Dirichlet multinomial responses are given.

Secondly, we consider the case $K \to \infty$, where θ and m_{ij} are fixed. A consistent estimate for $1 + \theta$ is the estimate $1 + \theta^* = X^2/v$, where X^2 corresponds with the full model, and $v = IJ(K-1)$. If preferred, θ can be estimated by $\hat{\theta}$, obtained by maximizing the log-likelihood

$$\sum_{ijk} \{\log \binom{X_{ijk} + \theta^{-1} m_{ij} - 1}{X_{ijk}} + X_{ijk} \log (\tfrac{\theta}{1+\theta}) + \theta^{-1} m_{ij} \log(\tfrac{1}{1+\theta})\}$$

as a function of θ, substituting the moment estimate $\hat{m}_{ij} = \bar{X}_{ij+}$ for m_{ij}.

Models are tested by statistics $G_2^2 - G_1^2$, again being distributed as $(1+\theta)\chi^2$. To see this, note that these statistics only depend on X_{ijk} via X_{ij+}, being distributed as a negative binomial $(\theta^{-1}Km_{ij}, p)$ random variable. Therefore $K \to \infty$ and $m_{ij} \to \infty$ asymptotics are equivalent, the latter being considered just before. To close this section, we propose to test model parameters by statistics $(G_2^2 - G_1^2)/(1+\tilde{\theta})$, for $K \to \infty$ asymptotically distributed as χ^2 with $v_2 - v_1$ degrees of freedom, where $\tilde{\theta}$ is a consistent estimate of θ, like θ^* or $\hat{\theta}$.

3.2. Application

By an industrial soldering-team an experiment was carried out to study the soldering quality of print panels. Two factors were varied in the experiment:

- factor L: soldering location, with levels L_1, L_2 and L_3
- factor M: soldering method, with levels M_1 and M_2.

For each combination of factorial levels, 5 panels were soldered, so 5 replicates per cell were obtained. The experiment was carried out by complete randomization over the 5 x 6 = 30 experimental units. Afterwards, the number of soldering failures was counted for each print panel. The results are displayed in table 1.

Soldering Method \ Soldering Location	L_1	L_2	L_3
M_1	10 15 11 11 16	2 11 4 14 25	13 10 29 16 25
M_2	22 12 5 8 11	12 35 11 12 17	12 27 25 15 15

Table 1; numbers of soldering failures on print panels.

Obvious from the data is a within-cell heterogeneity, suggesting extra-Poisson variation. We shall analyse the data by Model I.

3.3. The analysis

The full log-linear model

$\log m_{ij} = u + u_{1;i} + u_{2;j} + u_{12;ij}$

was fitted to the data, as well as reduced models, containing parameters for main effects only. The value of X^2 and G^2 test statistics for each (reduced) model is presented in table 2.

Factorial effect included in model	df ν	X^2	G^2
L * M	24	93.63	90.28
L , M	26	96.67	95.65
L	27	98.66	97.26
M	28	111.71	110.44
Ø	29	113.41	112.06

Table 2; X^2 and G^2 test statistics for fitted log-linear models.

Evidently, the full model L * M gives a bad fit, showing extra-Poisson variation in the data.

Firstly we shall consider results of approximate tests for large m_{ij}. We shall follow the first testing procedure from 3.1.

Results of testing model terms by approximate F-tests are given in the Anova table 3.

Source	degrees of freedom v	$G_2^2-G_1^2$	F
L * M interaction	2	5.37	0.71
L given M	2	14.79	1.96
M given L	1	1.61	0.42
Error	24	90.28	

Table 3; F-statistics for model terms.

As is seen from table 3, values of $G_2^2-G_1^2$ can be very misleading, because of a dominant error factor. From the correct approximate F-tests no effects are shown to be significant.

Secondly, results of approximate tests for large K are obtained by test statistics $(G_2^2-G_1^2)/(1+\tilde{\theta})$, with the estimate $(1+\theta^*) = 3.90$ or $(1+\hat{\theta}) = 3.02$. Here also no significant effects show up.

4. Model II

4.1. Discussion

Under Model II, X_{ij} has a negative binomial $(\alpha,\ p_{ij} = \frac{m_{ij}}{\alpha+m_{ij}})$ distribution, with $E\ X_{ij} = m_{ij}$ and $\text{var}\ X_{ij} = m_{ij}(1 + \alpha^{-1}m_{ij})$.

For Model II we shall consider $K \to \infty$ asymptotics, where α and m_{ij} are fixed; because of a non-additivity in parameter m_{ij}, the case $m_{ij} \to \infty$ does not lead to equivalent results, which is contrary to Model I.

If α were known, the family of negative binomial distributions is parameterized by p_{ij} only:

$$P(X_{ij} = x) = \exp \{\alpha(x/\alpha \log p_{ij} + \log (1-p_{ij})) + \log \binom{x+\alpha-1}{x}\}.$$

This is an exponential family and therefore of the type being studied by Nelder and Wedderburn (1972). In their paper they show how to fit a (generalized) linear model for $\log m_{ij}$ by an iterative procedure, and how to test factorial effects by an analysis of deviance, based on likelihood ratio tests. We shall follow their approach and first pay some attention to the likelihood function.

The log-likelihood function has the form:

$$L_{m}(\alpha) = \sum_{ijk} \{X_{ijk} \log \left(\frac{m_{ij}}{\alpha+m_{ij}} \right) + \alpha \log \left(\frac{\alpha}{\alpha+m_{ij}} \right) + \log \binom{X_{ijk}+\alpha-1}{X_{ijk}}\}$$

where index m indicates a specific linear model for $\log m_{ij}$.

For the saturated model the log-likelihood is

$$L_{s}(\alpha) = \sum_{ijk} \{X_{ijk} \log \left(\frac{X_{ijk}}{\alpha+X_{ijk}} \right) + \alpha \log \left(\frac{\alpha}{\alpha+X_{ijk}} \right) + \log \binom{X_{ijk}+\alpha-1}{X_{ijk}}\}.$$

The deviance, defined by Nelder and Wedderburn as $d_{m}(\alpha) = 2(L_{s}(\alpha) - L_{m}(\alpha))$ results in

$$d_{m}(\alpha) = 2 \sum_{ijk} \{X_{ijk} \log \left(\frac{X_{ijk}}{\alpha+X_{ijk}} \cdot \frac{\alpha+m_{ij}}{m_{ij}} \right) + \alpha \log \left(\frac{\alpha+m_{ij}}{\alpha+X_{ijk}} \right)\}.$$

We shall estimate α by maximum likelihood, maximizing $L_{m}(\alpha)$ as a function of α, when we have substituted $\hat{m}_{ij} = \bar{X}_{ij+}$ for m_{ij}, the m.l.e. under the full log-linear model. If needed, the moment estimate may serve as a starting value for α. We shall discuss it briefly.

Following Johnson and Kotz (1969), p. 134, for factorial levels i and j we obtain the "truncated" moment estimate

$$\alpha^*_{ij} = \frac{\bar{X}^2_{ij+}}{S^2_{ij} - \bar{X}_{ij+}} \qquad \text{if } S^2_{ij} > \bar{X}_{ij+}$$

$$\alpha^*_{ij} = \infty \qquad \text{if } S^2_{ij} \leq \bar{X}_{ij+}$$

where $\bar{X}_{ij+}$ and S^2_{ij} are sample mean and variance from K replicates per cell. Note that $\alpha = \infty$ (with finite m_{ij}) corresponds to the Poisson case. The approximate variance of α^*_{ij} being $\frac{2\alpha(\alpha+1)}{K\, p^2_{ij}}$, it is suggested to use weights

$$w_{ij} = p^2_{ij} = \left(\frac{m_{ij}}{\alpha+m_{ij}}\right)^2 \text{ to obtain the estimate } \alpha^* = \frac{\sum_i \sum_j w^*_{ij}\, \alpha^*_{ij}}{\sum_i \sum_j w^*_{ij}}$$

where w^*_{ij} is the estimate $\left(\frac{\bar{X}_{ij+}}{\alpha^*_{ij} + \bar{X}_{ij+}}\right)^2$ of w_{ij}; in the sum, terms for which w^*_{ij} equals zero are omitted.

By GLIM, $d_m(\alpha)$ is minimized, and therefore L_m is maximized under some restrictions on the full linear model for log m_{ij}. If a linear model m_2 is a hypothesized restriction of a model m_1, the corresponding set of model parameters is tested by the difference $d_{m_2}(\alpha) - d_{m_1}(\alpha)$ which is, for known α and $K \to \infty$ asymptotically chi-square if the hypothesis is true. If $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are the m.l.e.'s of α under model m_1 and model m_2 respectively,

$d_{m_2}(\hat{\alpha}_2) - d_{m_1}(\hat{\alpha}_1)$ has an asymptotic chi-square distribution under the hypothesis. Finally, this is true for $d_{m_2}(\hat{\alpha}) - d_{m_1}(\hat{\alpha})$, where $\hat{\alpha}$ is the m.l.e. under the full model, as it has the same asymptotic properties as $\hat{\alpha}_1$ and $\hat{\alpha}_2$, if the hypothesized restrictions on the full model are true. *It is this estimate of* α *we shall use in the testing procedure of section* 4.2.

4.2. The analysis

We shall reanalyse by Model II the data from table 1, using the computer program GLIM from Baker and Nelder (1978). Within GLIM, our model is defined by specifying four model properties:

1. the link function, which is the natural logarithm
2. its derivative
3. the variance function var $X_{ij} = m_{ij}(1+\alpha^{-1}m_{ij})$
4. the deviance $d_m(\alpha)$.

In a first step, the "moment" estimate $\alpha^* = 5.14$ was obtained for α. It was used as a starting value to estimate α by maximum likelihood, resulting into a value $\hat{\alpha} = 7.32$.

The second step in the testing procedure is the formation of an analysis of deviance table 4.

Model	Degrees of freedom	Deviance
L * M	24	31.43
L + M	26	33.28
L	27	33.82
M	28	38.10
Ø	29	38.63

Table 4; Analysis of deviance table for the crossed-classification of soldering failures table 1.

Log-likelihood ratio test statistics for factorial effects are obtained by subtraction of deviances. No significant effects show up, which confirms our earlier conclusions.

5. Discussion

Two negative binomial models for the analysis of the structure of counts showing extra-Poisson variation have been discussed. An application was given on a two dimensional classification of counts, with some replicates per cell.

A generalization of most of the results exists for i) more dimensional classifications and for ii) unequal numbers of replicates per cell. Note that for Model I when assuming large m_{ij}, only one replicate per cell is needed to estimate θ if some more-factor interactions are negligible.

Finally, it was observed by Paul & Plackett (1978) that the negative binomial $(\alpha, p=\frac{\theta}{1+\theta})$ distribution is the limiting distribution of the beta-binomial (n,α,β) where $\beta = n\theta^{-1}$ and $n \to \infty$. The beta-binomial distribution applies when the response variable is the number of successes out of n Bernoulli trials, the probability of success not being constant between replicates of the experiment (see Crowder (1978)). Then the negative binomial distribution seems a good approximation of the beta-binomial for a large number of trials when the response is small. The log-linear analysis can be performed by the methods presented in this paper.

Acknowledgement

The author wishes to thank ir. P. Banens for his help with the use of the GLIM program, and prof. dr. P. van der Laan and the referees for some valuable suggestions.

References

Baker, R.J. and Nelder, J.A. (1978),
The GLIM system. Release 3. Generalized linear interactive modelling manual, Oxford: NAG.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975),
Discrete Multivariate Analysis, Theory and Practice,
The MIT-Press, Cambridge.

Crowder, M.J. (1975),
Beta-binomial Anova for proportions,
Appl. Statist. 27, 34 - 37.

Engel, J. (1983),
On the analysis of variance for beta-binomial responses,
CQM-note.

Fienberg, S.E. (1977),
The analysis of cross-classified categorical data,
The MIT-Press, Cambridge.

Johnson, N.L. and Kotz, S. (1969),
Discrete distributions,
Houghton Mifflin Company, Boston.

Nelder, J.A. and Wedderburn, R.W.M. (1972),
Generalized linear models,
J.R. Statist. Soc. A 135, 370 - 384.

Paul, S.R. and Plackett, R.L. (1978),
Inference sensitivity for Poisson mixtures,
Biometrika 65, 591 - 602.

CHAPTER 4

A LIMITING PROPERTY OF MODELS FOR OVERDISPERSED COUNT RESPONSE DATA

1. INTRODUCTION

In chapter 3 we presented models for extra-Poisson variation in independent count data which was accounted for by introducing the gamma distribution in the parameter of the Poisson distribution as an extra component of variation. Two models were thus obtained in which either the shape parameter of the gamma distribution depends on the levels of the design factors (this is Model I) or the scale parameter of the gamma distribution depends on the levels of these factors (this is Model II). In chapter 3, Model I and Model II were studied for a large number of replicates and Model I was studied for a large value of the shape parameter, which implies large mean values of the count data. In this chapter we shall study Model II for a large value of the scale parameter, which also implies large mean values of the data.

A limit theorem will be presented in section 2 which is a generalization of a theorem of Pessin (1961). By this limit theorem a simplification of the gamma-Poisson model for large values of the scale parameter of the gamma distribution is obtained.
More concretely, the limit theorem has the following implication.

Let M_i, $i = 1,2, \ldots, I$, be a set of I random variables. Given $M_i = m_i$ let independent random variables X_i have Poisson (m_i) distributions. Suppose that M_i can be written as $M_i = \theta H_i$, $i=1,2, \ldots, I$, where θ is a positive parameter and $\{H_i\}_{i=1}^{I}$ is a vector of jointly distributed, positive and non-degenerate random variables. Then for large θ the limit theorem implies that the vector $\{X_i\}_{i=1}^{I}$ is approximately distributed as the vector $\{M_i\}_{i=1}^{I}$.

The consequence of the limit theorem for the gamma-Poisson model (Model II) for independent random variables X_i is that the approximate distribution of X_i for large values of the scale parameter of the gamma distribution is the gamma distribution itself. We shall discuss the analysis of count data by this approximate gamma model in section 3.

For a single classification of count data having the gamma-Poisson model, simulation results are obtained for the true significance level of six test statistics for testing the hypothesis of no main effect of the classifying factor, where these test statistics are based on the approximate gamma distribution.

An important conclusion from these simulation results is that an approximate F-test which is based on the relatively simple gamma deviance and which eliminates the shape parameter of the gamma distribution, is quite reasonable for testing the hypothesis of no main effect of the factor if some replicates are available per cell and if the scale parameter of the gamma distribution is not too small. For the nominal level of significance of 5%, the true level of significance is no more than 8%.

Finally, an application is given in section 4, revisiting the data of the soldering experiment from chapter 3. These data will be reanalysed by the use of test statistics based on the approximate gamma distribution.

2. A LIMIT THEOREM

Let random variables X_i and M_i, i = 1,2, ..., I, be given. Given $M_i = m_i$, let X_i have a Poisson (m_i) distribution for i = 1,2, ..., I. If $M_i = \theta H_i$, then X_i is approximately distributed as M_i for large θ. This result reduces much of the complexity of e.g. the gamma-Poisson model. In a multivariate version it is formally stated in the following limit theorem.

Theorem 1

Let $X = \{X_i\}_{i=1}^{I}$ and $M = \{M_i\}_{i=1}^{I}$ be I variate random vectors, where random vector X given M = m, $m = \{m_i\}_{i=1}^{I}$, is distributed as a vector of I independent Poisson (m_i) random variables.

Let $M = \theta H$, where $H = \{H_i\}_{i=1}^{I}$ is a vector of jointly distributed, positive and non-degenerate random variables, having finite second moments.

Then $X^* \rightarrow M^*$ in distribution for $\theta \rightarrow \infty$ where X^* and M^* are vectors of standardized components X_i^* and M_i^*, respectively, where
$X_i^* = (X_i - E(X_i))/\sqrt{var(X_i)}$ and $M_i^* = (M_i - E(M_i))/\sqrt{var(M_i)}$.

For an outline of proof, see appendix.

Corollary

Let H be a vector of independent gamma (α, Ψ_i) distributed random variables, i=1,2,...,I, so that X is a vector of independent random variables having negative binomial $(\alpha, p_i = \theta\ \Psi_i/(1 + \theta\ \Psi_i))$ distributions. Then for large θ the components of vector X have, approximately, independent gamma $(\alpha, \theta\Psi_i)$ distributions for i = 1,2,...,I.

This is the practical interpretation of Theorem 1: for a large value of the general level θ, vector X is approximately distributed as vector M.
Note that $E(X) = \theta\ E(H)$, so the condition in Theorem 1 implies large expectations.

It is important to realize that the components of vector H may be dependent but should all be non-degenerate. If all components are degenerate, the components of vector X are independent, having the normal limiting distribution, as is known from standard results.

We shall use Theorem 1 for deriving simplified models for fixed (in section 3) and random (in chapter 7) factor designs with count response data.

3. ANALYSIS OF DATA BY MODEL II

3.1. The Model

By the corollary of Theorem 1, we are able to present the analysis of count data by Model II (see Engel (1984)) for large θ_i. This Model II was defined as follows.

Let M_i, $i = 1,2, \dots, I$, be a set of random variables.
Given $M_i = m_i$, $i = 1,2,\dots,I$, a random vector X has I independent components X_i having Poisson (m_i) distributions. Further $M_i = G(\alpha, \theta_i)$, where $G(\alpha, \theta_i)$ are independent gamma random variables for $i = 1,2,\dots,I$, with shape parameter α and scale parameter θ_i.

Under Model II, the distribution of X_i is gamma-Poisson; the variable M_i can be written in the form $M_i = \theta H_i$, where $H_i = G_i(\alpha, \Psi_i)$ is a gamma (α, Ψ_i) random variable, $\Psi_i = \theta_i/\theta$ and $\theta = \sqrt[I]{\prod_i \theta_i}$.
For large θ the variable X_i has, approximately, the same distribution as $M_i = G_i(\alpha, \theta_i) = m_i\, G_i(\alpha, \alpha^{-1})$, where $m_i := E(X_i) = \alpha\theta_i$; further $\mathrm{var}(X_i) = \alpha^{-1} m_i^2$ and the coefficient of variation of X_i is $cv(\alpha) = \alpha^{-1/2}$.

In section 3.2 we shall let $i = 1,2,\dots,n$, where n is the number of independent counts X_i.
The analysis of data by Model II for large θ_i and so for large m_i, $i=1,2,\dots,n$, will be as follows.
Introducing $G_i(\alpha) := G_i(\alpha, \alpha^{-1})$, the variable X_i is distributed, approximately, as

$X_i \sim m_i\, G_i(\alpha)$ for $i=1,2,\dots,n$,

and the analysis of responses X_i fits into the framework of Nelder and Wedderburn (1972) for the analysis of generalized linear models, the gamma distribution being optional in the GLIM-system (Baker & Nelder (1978)). As α is an unknown form parameter to be estimated, the analysis falls apart into two stages:

- estimate α, e.g. by fitting a 'full' model;
- fit reduced models, and choose the best model.

Both stages of the analysis will be discussed in section 3.2.

3.2. Estimation and testing

3.2.1. Estimation of α by (modified) maximum likelihood

Three methods for the estimation of parameter α will be presented:

1) maximum likelihood (ml);
2) modified ml;
3) the 'naive' method from GLIM.

Firstly, parameter α can be estimated by fitting a full linear model to m_i, with log link function and gamma error. Such a full linear model, which contains all parameters of relevance to the problem, is fitted to log m_i, thus taking the logarithm as a link function between m_i and the linear model. For more details here and further, see Nelder and Wedderburn (1972). If m_i is estimated by the maximum likelihood estimator (mle) $\hat{m}_i$, the mle for α is obtained as the solution $\hat{\alpha}$ of the equation (see also Dunn c.s. (1983) for the one-sample case)

$$n\,(\log \alpha - \Psi(\alpha)) = \sum_i \log \hat{m}_i - \sum_i \log X_i + \sum_i (X_i/\hat{m}_i - 1) \qquad \ldots (1)$$

using tables of $\log \alpha - \Psi(\alpha)$ from Chapman (1956).
Here, $\Psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$, the digamma function, whereas $\sum_i (X_i/\hat{m}_i - 1)$ vanishes for the log link function, see Nelder and Wedderburn (1972).
Then (1) results in

$$n(\log \alpha - \Psi(\alpha)) = \sum_i \log \hat{m}_i - \sum_i \log X_i. \qquad \ldots (2)$$

The approximate variance of $\hat{\alpha}$ for large n equals $\{n(\Psi'(\alpha) - 1/\alpha)\}^{-1}$, the number of estimated parameters m_i being small.

Secondly, there are some reasons to consider a modified mle for α.
The modified mle α^* is obtained by equating the deviance D to its expectation E(D) under a full model, which is
$E(D) = 2n\alpha\,\{\log \alpha - \Psi(\alpha)\} - k + O(n^{-1})$; see Cordeiro (1983).
Here $D = -2\alpha \sum_i \log X_i/\hat{m}_i$ is the deviance for the gamma distribution and k is the number of estimated model parameters. The deviance D is defined as the log likelihood ratio statistic for testing goodness-of-fit of a linear model for log m_i against the saturated model having a parameter m_i for each observation X_i.

Setting D equal to E(D) we obtain

$$n(\log \alpha - \Psi(\alpha)) = \sum_i \log \hat{m}_i - \sum_i \log X_i + k/2\alpha + O((\alpha n)^{-1}). \qquad \ldots (3)$$

The solution α^* of (3) may also be considered as a kind of a moment estimator, setting deviance D equal to its first moment E(D). The estimator α^* has the same asymptotic ($n \to \infty$) properties as $\hat{\alpha}$, e.g. consistency.

The difference between the equations (2) and (3) is the correcting term $k/2\alpha$ in (3), which is important if it does not hold that $2\alpha >> k$. As $\log \alpha - \Psi(\alpha)$ is a decreasing function of α, see Abramowitz and Stegun (1965), p. 259, 6.3.21, one always has $\alpha^* < \hat{\alpha}$ (with probability one), hence $cv(\hat{\alpha}) < cv(\alpha^*)$. The term $k/2\alpha$ can be seen as a correction for a loss of k degrees of freedom (df) by estimating k model parameters. This procedure is familiar from the estimation of normal variances. To estimate σ^2 for normal responses, the mle $\hat{\sigma}^2$ equals SSE/n, where SSE is the sum of squares for error. The mle $\hat{\sigma}^2$ is biased, whereas $\sigma^{2}* = SSE/(n-k)$ is a is a modified version of $\hat{\sigma}^2$ to obtain the unbiased estimator σ^2* with the property $\sigma^2* > \hat{\sigma}^2$, compensating for the loss of k df for estimating model parameters.

As a <u>third</u> method, the GLIM method for estimating α should be mentioned, which is related to the modified ml method. It is based on naively equating the deviance D to its remainder degrees of freedom n-k, which is a correct procedure for the normal error case. However, note that E(D) = n-k only approximately for large α. A <u>third</u> α-estimator α^{**} is then obtained within GLIM, simply by solving D = n-k for α, so that $\alpha^{**} = (n-k)/S$, defining $S := (D/\alpha)$, and no iteration is needed. However, it can be shown that $2\alpha(\log \alpha - \Psi(\alpha)) > 1$, the inequality being substantial for small α; then always $\alpha^{**} < \alpha^*$, so that α^{**} inflates the estimate of $cv(\alpha)$.

To conclude, we summarize that $\alpha^{**} < \alpha^* < \hat{\alpha}$ with probability one, preferring the modified mle α^* as an estimator for α, having the same approximate variance for large n as the mle $\hat{\alpha}$. The estimator α^{**} is second best for α not too small, having the advantage of a simple computation.

The mle $\hat{\alpha}$ seems not to be too attractive, unless n is large: for $n \to \infty$, α being fixed, the methods one and two are equivalent. For $\alpha \to \infty$, n being fixed, the methods two and three are equivalent, as $2\alpha\ (\log \alpha - \Psi(\alpha)) = 1 + O(\alpha^{-1})$ for $\alpha \to \infty$; see Abramowitz and Stegun (1965), p. 259, 6.3.18.

3.2.2. Testing by (modified) deviances

The testing of reduced models can be based on the deviance D for the gamma distribution.

It is Cordeiro's advice to use modified deviances $D^* = (n-k)D/E(D)$ instead of deviances D for model testing: for α known they are distributed as χ^2_{n-k} reasonably well, even for small n. Differences $D^*_2 - D^*_1$ for hierarchic loglinear models are then approximately $\chi^2_{k_1-k_2}$. For large n, the approximate $\chi^2_{k_1-k_2}$ distribution of $D_2 - D_1$ is supported by well-known large sample results on log-likelihood ratio tests. For small n, e.g. singly replicated factorial designs, the testing procedure with $D_2 - D_1$ should be considered in a more informal way. The modification $D^*_2 - D^*_1$ should behave more like a $\chi^2_{k_1-k_2}$ statistic than $D_2 - D_1$ does.

In section 3.3 some relevant test statistics based on D and D* are defined and studied by simulation.

3.2.3. Estimation and testing by generalized Pearson's X^2

A statistic which is not based on the likelihood ratio is the generalized Pearson's $X^2 = \sum_i (X_i - \hat{m}_i)^2/(\alpha^{-1}\ \hat{m}_i^2)$, where $\hat{m}_i$ is the mle of m_i and $\mathrm{var}(X_i) = \alpha^{-1} m_i^2$; see also McCullagh and Nelder (1984), p. 26.

For large n, the distribution of X^2 can be approximated by the χ^2-distribution with n-k degrees of freedom, where k is the number of estimated model parameters. Also, X^2 can be used (in GLIM) to estimate the parameter α. An α^{**}-type estimator $\alpha^{**}_{X^2}$ is defined as $\alpha^{**}_{X^2} = (n-k)/[\sum_i (X_i - \hat{m}_i)^2/\hat{m}_i^2]$.

Generalized Pearson's X^2 corresponding to the NBD is given by $X^2 = \sum_i (X_i - \hat{m}_i)^2/(\hat{m}_i(1 + \alpha^{-1}\ \hat{m}_i))$ which is, for large $\hat{m}_i$, near to the X^2

for the gamma distribution. For normally distributed random variables X_i, Pearson's X^2 and the deviance (loglikelihood ratio statistic) are equivalent and are both equal to $X^2 = \sum_i (X_i - \hat{m}_i)^2/\sigma^2$, where $\sigma^2 = \text{var}(X_i)$.

This result will be used in section 3.3.

3.3. Simulation results for test statistics

3.3.1. Analysis of NB data by gamma distribution

To investigate the quality of the approximate analysis discussed in section 3.2, approximating the negative binomial or gamma-Poisson distribution by the gamma distribution, we carried out a simulation study. We used the NAG library to obtain negative binomial data, simulated by the Monte Carlo method. In the simulation study we consider the problem of testing the hypothesis of zero main effect in a one-way classification of data, with a total number of $n = 12$ observations (counts), classified by one factor having

- $I = 2$ levels, leaving $J = 6$ replicates per cell;
- $I = 6$ levels, leaving $J = 2$ replicates per cell.

Further, the parameters α and θ (note that under the hypothesis, θ does not depend on the index i, $i=1,2,\ldots,I$) were each set at three levels, that is

- $\alpha = 2, 5, 10$;
- $\theta = 5, 10, 15$,

so that simulation results were obtained for 9 combinations of parameter values. Results concerning the estimated tail probability of the χ^2_{I-1}-fractile at a nominal significance level of 5% were obtained under the hypothesis of no main effect, for the following test statistics.

T_1 : $D_2 - D_1$, where D_1 is the residual deviance for the gamma distribution:

$D_1 = -2\,\alpha \sum_{ij} \log X_{ij}/\bar{X}_{i+}$ and D_2 is the deviance under the hypothesis:

$D_2 = -2\,\alpha \sum_{ij} \log X_{ij}/\bar{X}_{++}$, so that $D_2 - D_1 = -2\alpha \sum_{ij} \log \bar{X}_{i+}/\bar{X}_{++}$.

T_2 : $D_2^* - D_1^*$, where D_1^* and D_2^* are Cordeiro modifications of D_1 and D_2, respectively, see section 3.2.2.

T_3 : $X_2^2 - X_1^2$, where $X_1^2 = \sum_{ij} (X_{ij} - \bar{X}_{i+})^2/(\alpha^{-1}\ \bar{X}_{i+}^2)$ and

$X_2^2 = \sum_{ij} (X_{ij} - \bar{X}_{++})^2/(\alpha^{-1}\ \bar{X}_{++}^2)$ are generalized Pearson statistics.

Parameter α was estimated by α^{**} when testing by $D_2 - D_1$ and $D_2^* - D_1^*$, and by $\alpha_{X^2}^{**}$ in the case of $X_2^2 - X_1^2$. Remember that $\alpha^{**} = I(J-1)/S_1$, where $S_1 = D_1/\alpha$ and $\alpha_{X^2}^{**} = I(J-1)/(X_1^2/\alpha)$.

Results concerning the estimated tail probability of the $F_{I-1,I(J-1)}$-fractile, at a nominal level of significance of 5%, were obtained for the test statistics

T_4 : $F_D = (D_2 - D_1)/D_1 * I(J-1)/(I-1)$;

T_5 : $F_{D*} = (D_2^* - D_1^*)/D_1^* * I(J-1)/(I-1)$;

T_6 : $F_{X^2} = (X_2^2 - X_1^2)/X_1^2 * I(J-1)/(I-1)$,

where it is seen that T_4, T_5 and T_6 are the "F-test modifications" of the test statistics T_1, T_2 and T_3, respectively.

For the approximate F-distribution of F_D there is some support from results of Jørgensen (1983). It is suggested that a reasonable approximate distribution of F_D for large J is the F-distribution with I-1 and I(J-1) degrees of freedom. A limit theorem (see Jørgensen (1984)) establishes the asymptotic F-distribution of F_D for $\alpha \to \infty$, J being fixed. So the results concerning F_D are expected to improve for increasing J and α.

The simulation results are based on 1000 MC-trials; they are summarized in tables $1^{a,b}$ for I = 2, and in tables $2^{a,b}$ for I = 6 factorial levels.

α θ	2		5		10	
5	17	12	17	13	19	14
	16	11	17	13	19	14
	16	11	10	7	10	7
10	10	7	11	8	11	7
	9	7	10	7	10	7
	15	10	11	8	10	7
15	10	6	9	7	10	7
	9	6	9	6	9	7
	15	10	11	7	10	7

Table 1^a: Simulation results (1000 trials; independence between cells) of the analysis with the gamma distribution of negative binomial data for I = 2. In cells: estimated tail probabilities (%) for T_1, T_2 and T_3 (column 1), T_4, T_5 and T_6 (column 2) at the 5% level.

α θ	2		5		10	
5	27	14	24	11	24	11
	26	14	23	11	23	11
	32	12	23	9	24	9
10	20	9	19	8	19	8
	18	8	19	8	19	8
	28	8	21	6	20	6
15	18	7	18	7	19	8
	17	7	18	7	19	8
	26	7	21	6	20	6

Table 2^a: Simulation results (1000 trials; independence between cells) of the analysis with the gamma distribution of negative binomial data for I = 6. In cells: estimated tail probabilities (%) for T_1, T_2 and T_3 (column 1), T_4, T_5 and T_6 (column 2) at the 5% level.

θ \ α	2		5		10	
5	17	12	16	10	17	12
	15	11	15	10	17	12
	16	10	11	7	10	6
10	11	6	11	7	11	7
	9	6	11	7	11	7
	14	10	11	7	10	5
15	8	5	9	6	9	6
	7	4	9	6	9	6
	15	10	11	7	10	6

Table 1^b: Simulation results (1000 trials; dependence between cells) of the analysis with the gamma distribution of negative binomial data for I = 2. In cells: estimated tail probabilities (%) for T_1, T_2 and T_3 (column 1), T_4, T_5 and T_6 (column 2) at the 5% level.

θ \ α	2		5		10	
5	27	14	24	11	25	11
	25	14	24	10	24	11
	30	10	23	8	22	7
10	20	8	22	8	20	7
	18	8	21	8	20	7
	28	8	22	6	20	6
15	18	6	19	6	20	6
	17	6	19	6	20	6
	27	7	21	5	19	5

Table 2^b: Simulation results (1000 trials; dependence between cells) of the analysis with the gamma distribution of negative binomial data for I = 6. In cells: estimated tail probabilities (%) for T_1, T_2 and T_3 (column 1), T_4, T_5 and T_6 (column 2) at the 5% level.

3.3.2. Analysis of NB data by NB distribution

For reasons of comparison, simulation results were also obtained, for some values of I, α and θ, for the analysis of NB data by the negative binomial distribution (NBD). Results were obtained concerning the estimated tail probability of the χ^2_{I-1}-fractile at a nominal significance level of 5%, under the hypothesis, for the following two test statistics:

- $D_2 - D_1$, where D is the deviance of the NBD,

  $$D = 2 \sum_{ij} [X_{ij} \log[(X_{ij}/(\alpha + X_{ij}))((\alpha + \hat{m}_{ij})/\hat{m}_{ij})] + \alpha \log[(\alpha + \hat{m}_{ij})/(\alpha + X_{ij})]];$$

  see also Engel (1984). Here $\hat{m}_{ij} = \bar{X}_{i+}$ for D_1, and $\hat{m}_{ij} = \bar{X}_{++}$ for D_2.

- $X_2^2 - X_1^2$, where X^2 is the generalized Pearson X^2

  $$X^2 = \sum_{ij} (X_{ij} - \hat{m}_{ij})^2/(\hat{m}_{ij}(1 + \alpha^{-1} \hat{m}_{ij}))$$

  for the NBD, and $\hat{m}_{ij} = \bar{X}_{i+}$ for X_1^2, $\hat{m}_{ij} = \bar{X}_{++}$ for X_2^2; see also section 3.2.3.

Parameter α was estimated by solving α from the equality

$$D_1 = 2 \sum_{ij} [X_{ij} \log[(X_{ij}/(\alpha + X_{ij}))((\alpha + \bar{X}_{i+})/\bar{X}_{i+})] + \alpha \log[(\alpha + \bar{X}_{i+})/(\alpha + X_{ij})]] = I(J-1)$$

obtained by equating the deviance D_1 to its degrees of freedom, iteratively at each MC-trial in the case of the test $D_2 - D_1$, taking $\alpha_0 = 5$ as a starting value; this is an α^{**}-type of estimator.
In the case of the $X_2^2 - X_1^2$ -test, α was estimated by solving α from

$$X_1^2 = \sum_{ij} (X_{ij} - \bar{X}_{i+})^2/[\bar{X}_{i+}(1 + \alpha^{-1} \bar{X}_{i+})] = I(J-1)$$

obtained by equating Pearson's X_1^2 to its degrees of freedom, iteratively at each MC-trial starting with $\alpha_0 = 5$, and an $\alpha^{**}_{X^2}$- type of estimator is obtained (see also Breslow (1984)).
The simulation results are summarized in tables $3^{a,b}$ for I = 2 and in tables $4^{a,b}$ for I = 6.

α θ	2	10
5	7 16	7 9
15	6 14	7 9

Table 3^a: Simulation results (1000 trials; independence between cells) for I = 2, NB-analysis of NB-data. In cells: estimated tail probabilities (%) for $D_2 - D_1$ and for $X_2^2 - X_1^2$ (α is estimated by α^{**} and $\alpha^{**}_{X^2}$, respectively) at the 5% level.

α θ	2	10
5	17 32	17 19
15	16 29	17 19

Table 4^a: Simulation results (1000 trials; independence between cells) for I = 6, NB-analysis of NB-data. In cells: estimated tail probabilities (%) for $D_2 - D_1$ and for $X_2^2 - X_1^2$ (α is estimated by α^{**} and $\alpha^{**}_{X^2}$, respectively) at the 5% level.

θ \ α	2	10
5	7 17	8 10
15	6 15	8 10

Table 3^b: Simulation results (1000 trials; dependence between cells) for $I = 2$, NB-analysis of NB-data. In cells: estimated tail probabilities (%) for $D_2 - D_1$ and for $X_2^2 - X_1^2$ (α estimated by α^{**} and $\alpha^{**}_{X^2}$, respectively) at the 5% level.

θ \ α	2	10
5	17 31	17 20
15	16 29	18 20

Table 4^b: Simulation results (1000 trials; dependence between cells) for $I = 6$, NB-analysis of NB-data. In cells: estimated tail probabilities (%) for $D_2 - D_1$ and for $X_2^2 - X_1^2$ (α estimated by α^{**} and $\alpha^{**}_{X^2}$, respectively) at the 5% level.

3.3.3. Analysis of lognormal-Poisson data by lognormal distribution

Anticipating later needs (see chapter 7), some simulation results were also obtained for the analysis of data having the lognormal-Poisson distribution (this distribution is very similar to the gamma-Poisson distribution), using the lognormal approximation. The lognormal-Poisson distribution is obtained as follows.

Let X_{ij}, i=1,2,...,I, j=1,2,...,J, be independent random variables. Given $M_{ij} = m_{ij}$, the random variables X_{ij} have Poisson (m_{ij}) distributions. Further, random variables M_{ij}, i=1,2,...,I, j=1,2,...,J, are independent, having the lognormal distribution with mean $\alpha\ \theta_i$ and variance $\alpha\ \theta_i^2$; remember that these are the moments of the gamma distribution with parameters (α, θ_i); see sections 3.2. and 3.3.1. For large θ_i, i = 1,2, ..., I, the independent X_{ij}, i=1,2,...,I, j=1,2,...,J, have, approximately, lognormal distributions because of Theorem 1, so that $Y_{ij} = \log X_{ij}$ has the normal distribution with

mean $\mu_i = \log[\alpha\ \theta_i/(1 + \alpha^{-1})^{1/2}]$ and variance $\sigma^2 = \log(1 + \alpha^{-1})$.

On simulated data, the hypothesis $H_0 : \theta_1 = \theta_2 = \ldots = \theta_I$ of no main effect in a one-way classification was tested by the well-known F-test

$$F_D = (D_2 - D_1)/D_1 * I(J - 1)/(I - 1),$$

where D_1 is the deviance $D_1 = \sum_{ij} (Y_{ij} - \bar{Y}_{i+})^2/\sigma^2$ for the normal distribution and $D_2 = \sum_{ij} (Y_{ij} - \bar{Y}_{++})^2/\sigma^2$. The hypothesis H_0 was also tested by the statistic $D_2 - D_1$ having a chi-squared distribution, estimating σ^2 by $\sigma^{2**} = \sum_{ij} (Y_{ij} - \bar{Y}_{i+})^2/I(J-1)$. The simulation results are summarized in tables $5^{a,b}$ and tables $6^{a,b}$.

α θ	2		10	
5	9	6	8	4
15	7	4	6	4

Table 5^a: Simulation results (1000 trials; independence between cells) for $I = 2$, lognormal analysis of lognormal-Poisson data. In cells: estimated tail probabilities (%) for $D_2 - D_1$ (σ^2 estimated by σ^{2**}) and F_D, respectively, at the 5% level.

α θ	2		10	
5	19	6	22	6
15	19	7	19	6

Table 6^a: Simulation results (1000 trials; independence between cells) for $I = 6$, lognormal analysis of lognormal-Poisson data. In cells: estimated tail probabilities (%) for $D_2 - D_1$ (σ^2 estimated by σ^{2**}) and F_D, respectively, at the 5% level.

α θ	2		10	
5	8	3	7	4
15	9	4	9	4

Table 5^b: Simulation results (1000 trials; dependence between cells) for $I = 2$, lognormal analysis of lognormal-Poisson data. In cells: estimated tail probabilities (%) for $D_2 - D_1$ (σ^2 estimated by σ^{2**}) and F_D, respectively, at the 5% level.

α θ	2		10	
5	12	3	16	3
15	16	3	16	2

Table 6^b: Simulation results (1000 trials; dependence between cells) for $I = 6$, lognormal analysis of lognormal-Poisson data. In cells: estimated tail probabilities (%) for $D_2 - D_1$ (σ^2 estimated by σ^{2**}) and F_D, respectively, at the 5% level.

3.3.4. Conclusions drawn from simulation results

From the simulation results of sections 3.3.1, 3.3.2 and 3.3.3, the following conclusions can be drawn.

Ad 3.3.1. Analysis of NB data by gamma distribution

From tables $1^{a,b}$ and tables $2^{a,b}$ we conclude:

A. χ^2-approximation of test statistics T_1, T_2 and T_3 where α is estimated by the estimator α^{**} for T_1 and T_2 and by the estimator $\alpha^{**}_{\chi^2}$ for test T_3.

- Only for I=2 are the α-estimators α^{**} and $\alpha^{**}_{\chi^2}$ sufficiently accurate so that, when they are substituted for α in test statistics T_1, T_2 and T_3, these statistics have tail probabilities good enough for further study (no results for I = 6 are useful).

- In the case of I=2, tests T_1 and T_2 behave not too badly for large θ (which is supported by Theorem 1). It is known that likelihood ratio tests like T_1 are sensitive to wrong distributional assumptions which are made when assuming a gamma distribution for the data for small θ.

- In the case of I=2, test T_3 is not too bad for all values of θ (the quality of this test does not depend too much on distributional assumptions) if at least α is not too small: increasing α seems to improve the performance of this test.

B. F-approximation of test statistics T_4, T_5 and T_6.

- The results are now acceptable for I=2 and for I=6. The variability of the estimator α^{**} (or $\alpha^{**}_{\chi^2}$) of α, being large for I=6, is accommodated by the F-test as in fact α is eliminated. The approximate F-distribution of the test statistic T_4 for large J and α is supported by Jørgensen (1983, 1984).

- The tests T_4 and T_5 are reasonable for large θ.

- The test T_6 appears to be reasonable for all θ, parameter α not being too small.

Ad 3.3.2. Analysis of NB data by NB distribution

From tables $3^{a,b}$ and tables $4^{a,b}$ we come to the following conclusions:

A. χ^2-approximation of tests $D_2 - D_1$ and $X_2^2 - X_1^2$ (α is estimated by α^{**}-type estimator).

- Only for I=2 is parameter α estimated sufficiently accurate, so that these tests deserve further consideration (for I = 6 these tests are of no use).
- For the case I = 2, the test based on D is reasonable for all values of α and θ; the test based on X^2 is reasonable for all θ and for large values of α.

B. No F-approximation of F-modifications of the above tests was studied. The reason is that, contrary to the gamma distribution, no F-test can be constructed to eliminate the parameter α in order to avoid the estimation of parameter α.

Ad 3.3.3. Analysis of lognormal-Poisson data by lognormal distribution

From tables $5^{a,b}$ and tables $6^{a,b}$ we conclude:

A. χ^2-approximation based on the statistic $D_2 - D_1$.

- Only for the case I=2 do the estimated tail probabilities have reasonable values. However, in the Anova this χ^2-approximation is never used.

B. F-approximation of the test statistic F_D.

- For both cases I=2 and I=6 the test F_D has good tail probabilities for all values of θ and α.

From the (of course, restricted!) simulation results it is tempting to conclude that not many replicates per cell and no large values of θ and α are needed for the F_D-test of section 3.3.3 to have the approximate F-distribution.

Summarizing two important conclusions from the simulation results

1. For the case of two or so replicates per cell the analysis of NB data by the gamma distribution becomes attractive. An approximate F-test is available, having reasonable properties when θ is not too small. By constructing this F-test the parameter α is eliminated so that there is no need to compute an estimate for this parameter α. For the "exact" NB analysis there is no such F-test.

2. Assuming the lognormal-Poisson distribution (instead of the gamma-Poisson distribution) for the data, the analysis based on the approximate lognormal distribution is advised. The corresponding F-test has good properties for all values of α and θ in the study, even when only a few replicates are available per cell.

4. APPLICATION

As an application, we first calculate estimates of α for the soldering failure problem from Engel (1984), using the gamma approximation for the gamma-Poisson model, based on Theorem 1. The results are $\alpha^{**} = 3.34$ (GLIM estimate), $\alpha^* = 3.53$ (estimate based on Cordeiro's results), $\hat{\alpha} = 4.34$ (mle).

All α-estimates are obtained from $D/\alpha = 7.19$ for the full model L * M fitted to the data. Note the difference between this estimate $\hat{\alpha}$ and the mle $\hat{\alpha} = 7.32$ based on the gamma-Poisson model. This difference is explained by the approximation used where the Poisson-variation is ignored. Then the gamma variance $\alpha^{-1}m_i^2$, $i = 1,2, \ldots, 30$, explains all the variation in the data, so that the estimate $\hat{\alpha}$ is smaller for the gamma model than for the gamma-Poisson model.

Secondly, we shall consider the analysis of deviance of soldering failures by the gamma distribution, modifying deviances as proposed by Cordeiro (1983). Deviances D are presented in table 7, as well as expected deviances E(D), and modified deviances D*.

Model	df	Deviance D	Expected deviance E(D)	Modified deviance D*
L*M	24	25.38	25.35	24.03
L,M	26	26.75	27.35	25.43
L	27	27.14	28.35	25.85
M	28	30.21	29.35	28.82
Ø	29	30.59	30.35	29.23

Table 7. Deviances, expected deviances and modified deviances for soldering failure data, where α is estimated by α^*.

Model component	df	$D_2 - D_1$	$D_2^* - D_1^*$	F_D	F_{D*}
LM-interaction	2	1.37	1.40	0.65	0.70
L given M	2	3.46	3.39	1.64	1.69
M given L	1	0.39	0.42	0.37	0.42

Table 8. Tests $D_2 - D_1$ $(D_2^* - D_1^*)$ and F_D (F_{D*}) for testing model components.

In table 8, the results of the tests

$$F_D = (D_2 - D_1)/D_{L*M} * 24/df$$

and

$$F_{D*} = (D_2^* - D_1^*)/D_{L*M}^* * 24/df$$

are displayed, where D_{L*M} is the deviance for the model L*M. Assuming approximate F-distributions for F_D and F_{D*} with degrees of freedom (df, 24) it is seen that no interaction and no main effects are significant at the 5% level. At the values at hand of, roughly, $\alpha = 4$ and $\theta = 5$ the F-tests are liberal, see the simulation results of section 3.3.

5. DISCUSSION

In this chapter a limit theorem was presented to simplify compounded Poisson models in the case of large scale parameters. This theorem was applied to simplify Model II of chapter 3, a gamma-Poisson model, to obtain the gamma distribution as an approximate distribution for count data, if the scale parameter is large. To some extent, test statistics based on this gamma distribution were studied.

Applications of the method are found in industrial practice (see section 4) and also in consumer purchasing behaviour (see Chatfield and Goodhardt (1970, 1973), Dunn c.s. (1983)) and in medical statistics (see Manton c.s. (1981)). Another remark is that the results of section 3 are of direct use for responses having gamma distributions. Applications can be found in the field of reliability and survival analysis, where the gamma distribution is used for modeling lifetime data.

Finally some special attention will be given to the simulation results for count data having the lognormal-Poisson distribution. The analysis of this data is based on the approximate lognormal distribution and is performed by carrying out a standard Anova on the log-transform of the data. The simulation results show that the F-test from the Anova for a one-way classification behaves well enough in this unorthodox situation.
We shall return to this subject in chapter 7. There we shall study lognormal-Poisson distributions having a more complex structure than the one considered here; the random mean of the Poisson distribution is then a product of independent lognormal components.

REFERENCES

Abramowitz, M. and Stegun, I.A. (1965), Handbook of mathematical functions, Dover Publications, New York.

Baker, R.J. and Nelder, J.A. (1978), The GLIM system. Release 3. Generalized linear interactive modelling manual, Oxford: NAG.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975), Discrete multivariate analysis, Theory and practice, The MIT-Press, Cambridge.

Breslow, N.E. (1984), Extra-Poisson variation in log-linear models, Appl. Statist. 33, 38 - 44.

Chapman, D.G. (1956), Estimating the parameters of a truncated gamma distribution, Ann. Math. Statist. 27, 498 - 506.

Chatfield, C. and Goodhardt, G.J. (1970), The beta-binomial model for consumer purchasing behaviour, Appl. Statist. 19, 240 - 250.

Chatfield, C. and Goodhardt, G.J. (1973), A consumer purchasing model with Erlang inter-purchase times, J. Am. Statist. Ass. 68, 828 - 835.

Cordeiro, G.M. (1983), Improved likelihood ratio statistics for generalized linear models, J.R. Statist. Soc. B 45, 404 - 413.

Dunn, R., Reader, S. and Wrigley, N. (1983), An investigation of the assumptions of the NBD model as applied to purchasing at individual stores, Appl. Statist. 32, 249 - 259.

Engel, J. (1984), Models for response data showing extra-Poisson variation, Statistica Neerlandica 38, 159 - 167.

Feller, W. (1971), An introduction to probability theory and its applications, volume II, John Wiley & Sons, New York.

Jørgensen, B. (1983), Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models, Biometrika 70, 19 - 28.

Jørgensen, B. (1984), Exponential dispersion models, Preprint 9, Odense University, Department of Mathematics.

McCullagh, P. and Nelder, J.A. (1984), Generalized linear models, Chapman and Hall, London.

Manton, K.G., Woodbury, M.A. and Stallard, E. (1981), A variance components approach to categorical data models with heterogeneous cell populations: analysis of spatial gradients in long cancer mortality rates in North Carolina counties, Biometrics 37, 259 - 270.

Nelder, J.A. and Wedderburn, R.W.M. (1972), Generalized linear models, J.R. Statist. Soc. A 135, 370 - 384.

Pessin, V. (1961), Some asymptotic properties of the negative binomial distribution, Ann. Math. Statist. 32, 922 - 923.

Outline of proof of Theorem 1

Denote by $\phi_X(t) = \sum_x \exp(it.x)P(X = x)$ the characteristic function (cf) of random vector X (see Feller (1971), chapter 15), where t.x denotes the inner product of vectors x and t, each having J components.
Given vector M = m, vector X has the Poisson distribution and

$$\phi_{X|M=m}(t) = \prod_{j=1}^{J} e^{-m_j(1 - e^{it_j})}$$

, t_j being the j^{th} component of vector t,

because $P(X = x|M = m) = \prod_j e^{-m_j} \frac{m_j^{x_j}}{x_j!}$ which represents the Poisson distribution of the independent components of vector X given vector M = m.

Since

$P(X = x) = \int P(X = x|M = m)\, dP(M \leq m)$, the cf of X results in

$$\phi_X(t) = \int \prod_{j=1}^{J} e^{-m_j(1 - e^{it_j})}\, dP(M \leq m),$$

where we applied Fubini's theorem when interchanging sum and integral.

For j = 1,2, ..., J, let $a_j := E(H_j)$, $b_j := var(H_j)$; further,
$P(M \leq m) = P(\theta H \leq m) = G(\varphi)$, where $G(h) := P(H \leq h)$ and $\varphi := m/\theta$.
Then $E(X_j) = a_j\theta$, $var(X_j) = a_j\theta + b_j\theta^2$ so that the standardized of component X_j is $X^*_j = (X_j - a_j\theta)/\surd\{\theta(a_j + b_j\theta)\}$, j = 1,2, ..., J.

The cf of vector $X^* = \{X^*_j\}_j$ is

$$\phi_{X*}(t)=\exp(-i \sum_j t_j a_j \theta/\surd\{\theta(a_j+b_j\theta)\})\int_0^\infty \prod_j \exp(-\theta\varphi(1-\exp(it_j/\surd\{\theta(a_j+b_j\theta)\})))dG(\varphi)$$

Letting $\theta \to \infty$ and applying Lebesque's theorem,

$$\lim_{\theta\to\infty} \phi_{X*}(t) = \exp\left(-i \sum_j t_j a_j/\sqrt{b_j}\right) \int_0^\infty \prod_j \exp(it_j\phi_j/\sqrt{b_j})dG(\phi) =$$

$$= \int_0^\infty \prod_j \{\exp(i\ t_j(\phi_j - a_j)/\sqrt{b_j})\}dG(\phi). \qquad \ldots (1)$$

Further, the cf of M is

$$\phi_M(t) = \int_0^\infty \exp(i\sum_j t_j m_j)\ dP(M \leqslant m) = \int_0^\infty \exp(i\sum_j t_j \theta\phi_j)\ dG(\phi).$$

As $M_j{}^* = (\theta\sqrt{b_j})^{-1} M_j - a_j/\sqrt{b_j}$, $j = 1,2, \ldots, J$, the cf of vector

$M^* = \{M_j{}^*\}_j$ is

$$\phi_{M*}(t) = \exp(-i \sum_j t_j a_j/\sqrt{b_j}) \int_0^\infty \exp(i \sum_j t_j\ \theta\ \phi_j/(\theta\sqrt{b_j}))dG(\phi) =$$

$$= \int_0^\infty \prod_j \{\exp(i\ t_j(\phi_j - a_j)/\sqrt{b_j})\}dG(\phi). \qquad \ldots (2)$$

This cf is a continuous function of t at t=0 so that, see Feller (1971), chapter xv.3, theorem 2, (1) and (2) being identical cf's, the random vector X* tends in distribution to the random vector M* for $\theta \to \infty$, which was to be proved.

CHAPTER 5

SPLIT-PLOT DESIGN: MODEL AND ANALYSIS FOR COUNT DATA

Abstract

The analysis of count response data from designed experiments is well-known for independent response variables having the Poisson distribution. For experimental designs where responses are dependent, no general results seem to be available. An example of this type of design is the split-plot design, where sub-plot responses are essentially dependent within whole plots.

In this paper, a model will be proposed for split-plot count data and a separate analysis for whole plot and sub-plot data will be presented. It is interesting to note, that the same model is used in the quite different context of consumer buying behaviour. It was derived by Goodhardt, Ehrenberg and Chatfield and it was called the 'Dirichlet model', as the Dirichlet multinomial distribution, together with the negative binomial distribution, build up the model.

(Published in Statistica Neerlandica 40 (1986), 21-33).

1. INTRODUCTION

An extension of the Poisson model for classified, count response data showing more variation than could be explained by the Poisson model was given by Engel (1984). In this paper, the gamma distribution was introduced into the Poisson model to accommodate for extra-Poisson variation, and the analysis of variance was carried out by means of the resulting negative binomial model.

Closely related are problems of the following type. Suppose that experimental units (e.u.), which are classified by fixed and crossed factors, are split up into smaller units by other factors; also, suppose that count data is available for the smaller experimental units. A typical result of this experimental design is that counts are dependent by their nature, so that any of the models for independent counts is not suitable, see e.g. De Roos and Schaafsma (1981) for an example.

As another example, consider large print panels for electrical components, which are split up into smaller sub-panels. On each sub-panel, a certain type of copper pattern was mounted. Types of copper pattern have to be compared with respect to their quality, by means of observed numbers of soldering failures, resulting from a soldering experiment. The large print panel can be regarded as a whole plot, and the sub-panels as sub-plots. The experimental design considered above is usually called a split-plot design. The effect of sub-plot factors, which is here the factor 'type of copper-pattern', is analysed apart from whole plot factors. It will be the subject of this paper to analyse count data from a split-plot design.

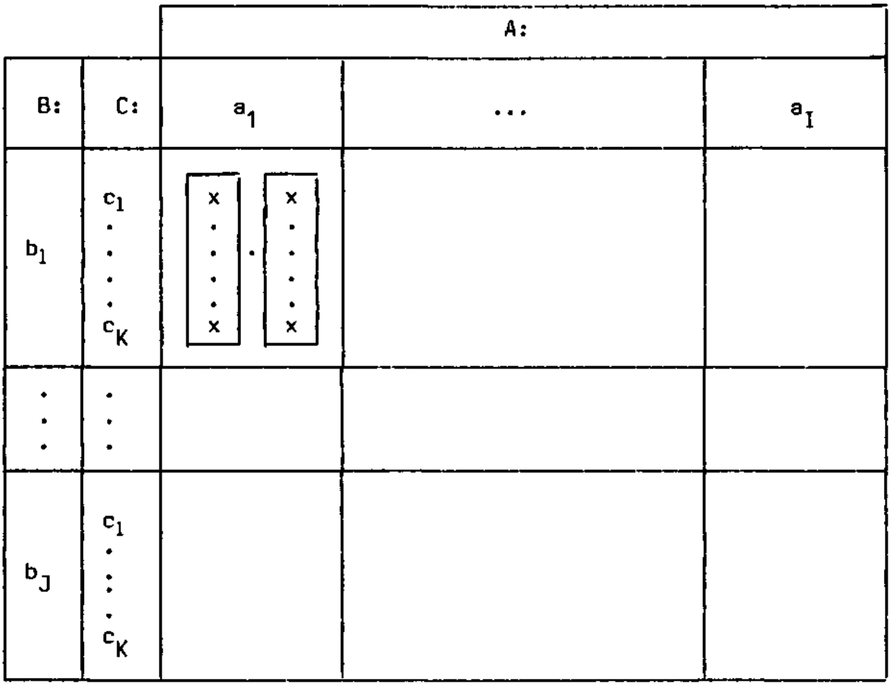
A model based on the loglinear model is proposed for count data from the split-plot design (see section 2). The analysis of data (in section 3) is rather straightforward for a simplified version of the model, which is often adequate in practice. Also for the more comprehensive model, some approximations are almost inevitable to meet with requirements of manageability.

Finally, an application of the split-plot design will be given.
The analysis for whole plots was worked out by Engel (1984). What is left is the analysis of sub-plot factorial effects, which is carried out in section 4.

2. THE MODEL

To fix our minds, we shall build up a model for the split-plot design in case of two whole plot factors, say factor A (index i) and factor B (index

j) and one sub-plot factor C (index k). Of course, any number of fixed and crossed factors can be treated in a similar way. We shall consider equal numbers of replicates per cell, which will be indicated by index ℓ. Table 1 shows the experimental design.



		A:		
B:	C:	a_1	...	a_I
b_1	c_1 . . . . c_K	x . . . . x · x x		
. . .	. . .			
b_J	c_1 . : . c_K			

Table 1. A split-plot experimental design with whole plots and sub plots.

The count response variable for replicate ℓ of sub-plot k of whole plot (i,j) will be denoted by $X_{ijk\ell}$, where i=1,2,...,I, j=1,2,...,J, k=1,2,...,K and ℓ=1,2,...,L.

Some components, which should be included in the model for $X_{ijk\ell}$ are the following; see Montgomery (1976), p.292:

- whole plot error; then $X_{ijk_1\ell}$ and $X_{ijk_2\ell}$ are dependent random variables, for any k_1 and k_2
- interaction between sub-plot factor C and whole plot error
- sub-plot error.

The Poisson distribution will be the basis for this model, and model components of random type are introduced into the Poisson parameter.
Taking into consideration the above requirements, the following model for $X_{ijk\ell}$ is proposed:

(i) Let $M_{ijk\ell}$ be a positive random variable, then given $M_{ijk\ell} = m_{ijk\ell}$, $X_{ijk\ell} \sim \text{Poisson}\ (m_{ijk\ell})$

(ii) $M_{ijk\ell} = G_{ij\ell}(\alpha_{ij}, \theta) \cdot H_{ijk\ell}(\beta_{ijk})$

where

$$H_{ijk\ell}(\beta_{ijk}) = G_{ijk\ell}(\beta_{ijk}, \eta)/\sum_k G_{ijk\ell}(\beta_{ijk}, \eta).$$

All G(a,b)'s are independent random variables having the gamma distribution with form parameter a and scale parameter b.
It is assumed that β_{ijk} can be written as $\beta_{ijk} = \beta\pi_{ijk}$, where $\sum_k \pi_{ijk} = 1$, β not depending on i and j. (1)

In model (1), the effect of whole plot and sub-plot factors is modelled separately. Therefore, the random Poisson mean M is the product of two factors $G_{ij\ell}$ and $H_{ijk\ell}$. Because of the normation $\sum_k H_{ijk\ell} \equiv 1$, the whole plot level $M_{ij+\ell}$ equals $G_{ij\ell}$. This factor describes the effect of whole plot factors A and B (via α_{ij}), and the whole plot error is included (random $G_{ij\ell}$). The second factor describes the effect of sub-plot factor C (via β_{ijk}) and the random interaction between factor C and whole plot error (random $H_{ijk\ell}$).

A further motivation for taking the product of $G_{ij\ell}$ and $H_{ijk\ell}$ is that these factors can be considered as whole plot and sub-plot random intensities of a Poisson failure process. If the whole plot intensity $G_{ij\ell}$ tends to zero, then so does $M_{ijk\ell}$, as it should be.

Omitting indices i, j and ℓ just for a moment, the resulting distribution of vector $(X_1, X_2, \ldots, X_K)'$, denoted by $\{X_k\}_{k=1}^{K}$, given $H_k = h_k$, $k=1,2,\ldots,K$, is the negative multinomial distribution (see Sibuya c.s. (1964)), with probability distribution function:

$$P(\{X_k\}_{k=1}^{K} = \{x_k\}_{k=1}^{K}) = \binom{x_+ + \alpha-1}{x_1, x_2, \ldots, x_K, \alpha-1} p_0^{\alpha} \prod_k p_k^{x_k}.$$

Here $p_k = \theta h_k/(1+\theta \sum_k h_k) = \theta h_k/(1+\theta)$, $p_0 = 1/(1+\theta \sum_k h_k) =$
$= 1/(1+\theta)$, as $\sum_k h_k \equiv 1$.

Unconditionally, $P_k = (\theta/(1+\theta))\, H_k$, where random vector $\{H_k\}_{k=1}^{K}$ has the Dirichlet distribution, $p_0 = 1/(1+\theta)$ being fixed.

It can be shown that the distribution of $\{X_k\}_{k=1}^{K}$ is given by

$$P(\{X_k\}_{k=1}^{K} = \{x_k\}_{k=1}^{K}) =$$

$$= \binom{x_+ + \alpha-1}{x_1, x_2, \ldots, x_K, \alpha-1} \left(\frac{1}{1+\theta}\right)^{\alpha} \left(\frac{\theta}{1+\theta}\right)^{x_+} \frac{\Gamma(\Sigma\, \beta_k)}{\Gamma(x_+ + \Sigma\, \beta_k)} \; \frac{\prod_k \Gamma(x_k+\beta_k)}{\prod_k \Gamma(\beta_k)} =$$

$$= \binom{x_+ + \alpha-1}{x_+} \left(\frac{1}{1+\theta}\right)^{\alpha} \left(\frac{\theta}{1+\theta}\right)^{x_+} * \binom{x_+}{x_1, x_2, \ldots, x_K} \frac{\Gamma(\Sigma\, \beta_k)}{\Gamma(x_+ + \Sigma\, \beta_k)} \; \frac{\prod_k \Gamma(x_k+\beta_k)}{\prod_k \Gamma(\beta_k)} =$$

$$= P(X_+ = x_+) * P(\{X_k\}_{k=1}^{K} = \{x_k\}_{k=1}^{K} \mid X_+ = x_+)$$

all sums Σ being over $k=1,2,\ldots,K$, where

- the distribution of the marginal sum X_+ is the negative binomial distribution (NBD) with parameters α and $p = \theta/(1+\theta)$, see Engel (1984);
- the distribution of the vector $\{X_k\}_{k=1}^{K}$ given X_+ is the Dirichlet multinomial distribution (DMD) with parameter vector $\{\beta_k\}_{k=1}^{K}$, see Mosimann (1962), Brier (1980), Engel (1985). Note that the DMD has dependent components.

This model for vector $\{X_k\}_{k=1}^{K}$ was called the 'Dirichlet' by Goodhardt c.s. (1984). The authors apply the model to the field of consumer purchase behaviour, and it was derived in a different way. Interesting is also a characterization, which was given by the authors for the gamma distribution (for random variable G) and the Dirichlet distribution (for random vector $\{H_k\}_{k=1}^{K}$) and which is also relevant in our case. This characterization may give a justification for the model used.

A simplification of model (1) is obtained by letting $\beta_{ijk} \to \infty$, $\eta \to 0$ in $G_{ijk\ell}(\beta_{ijk},\eta)$, while $\eta\,\beta_{ijk} = \theta_{ijk}$ is fixed. Then $G_{ijk\ell} \to \theta_{ijk}$ in probability, hence $H_{ijk\ell}(\beta_{ijk}) \to \theta_{ijk}/\sum_k \theta_{ijk} = \pi_{ijk}$ in probability.

The simplified model is:

(i) Let $M_{ijk\ell}$ be a positive random variable, then given
$M_{ijk\ell} = m_{ijk\ell}$, $X_{ijk\ell} \sim$ Poisson $(m_{ijk\ell})$

(ii) $M_{ijk\ell} = G_{ij\ell}(\alpha_{ij}, \theta) \cdot \pi_{ijk}$, where $\sum_k \pi_{ijk} = 1$. (2)

All G(a,b)'s are independent random variables having the gamma distribution with form parameter a and scale parameter b.

Note that model (2) differs from model (1) only by the absence of the interaction between factor C and whole plot error. It may be interesting to compare model (1) with the classical Anova split-plot model for a normal response variable $Y_{ijk\ell m}$, see Montgomery (1976), p. 292.

The usual model for Y is, in a shorthand notation,

$$Y_{ijk\ell m} = \mu + (\alpha\beta)_{ij} + e_{\ell(ij)} + (\alpha\beta\gamma)_{ijk} + e_{k\ell(ij)} + e_{m(ijk\ell)}.$$

For a comparison of both models, see table 2. It is seen that there is a one to one correspondence, of model parameters and error components, for the two models.

Sometimes, $e_{k\ell(ij)}$ is set equal to zero beforehand, which may be justified on 'technical grounds', or by the result of a testing procedure. The corresponding effect on the models for discrete X is that model (2) is reduced to model (1).

Model component for X	Model component for Y	Description
θ	μ	general level
α_{ij}	$(\alpha\beta)_{ij}$	whole plot parameter
$G_{ij\ell}$	$e_{\ell(ij)}$	whole plot error
β_{ijk}	$(\alpha\beta\gamma)_{ijk}$	sub-plot parameter
$H_{ijk\ell}$	$e_{k\ell(ij)}$	random interaction between factor C and whole plot error
$X_{ijk\ell}$ given $M_{ijk\ell}$	$e_{m(ijk\ell)}$	sub-plot error

Table 2. Comparison of model components of discrete X and continuous Y split-plot model.

3. ANALYSIS

3.1. Some introductory remarks

Intuitively, it is not unreasonable that the analysis of the split-plot data is performed as a separate analysis on whole plot data, via plot total $X_{ij+\ell}$, and on sub-plot data, via $X_{ijk\ell}$ given $X_{ij+\ell}$.

A separate analysis is further motivated by noting that the conditional distribution of $\{X_{ijk\ell}\}_{k=1}^{K} | X_{ij+\ell} = x_{ij+\ell}$, which is the Dirichlet multinomial, does not depend on (α_{ij},θ). Then $X_{ij+\ell}$ is a sufficient statistic for (α_{ij},θ) and an ancillary statistic for β_{ijk} (see Cox & Hinkley (1974), p.31). Inference on α_{ij} will be based on $X_{ij+\ell}$, inference on β_{ijk} will be based on $\{X_{ijk\ell}\}_{k=1}^{K} | X_{ij+\ell}$.

The analysis of model (2) and model (1) will be discussed in the following sections. We shall start with the simpler model (2) in section 3.2.

3.2. Analysis of model (2)

The analysis of model (2) needs a short discussion only.

To summarize model concepts,

(i) $X_{ij+\ell} \sim \mathrm{NBD}(\alpha_{ij},\theta)$

(ii) $\{X_{ijk\ell}\}_{k=1}^{K} \mid X_{ij+\ell} = x_{ij+\ell} \sim$ multinomial $(x_{ij+\ell}, \{\pi_{ijk}\}_{k=1}^{K})$.

A whole plot analysis on $X_{ij+\ell}$ was discussed by Engel (1984). For the subplot analysis, the literature on the analysis of multinomials is extensive, see Bishop c.s. (1975) and Fienberg (1979). For $L > 1$, we may simply add over replicates to obtain $\{X_{ijk+}\}_{k=1}^{K} | X_{ij++} = x_{ij++}$, which is a sufficient statistic for $\{\pi_{ijk}\}_{k=1}^{K}$, having the multinomial $(x_{ij++}, \{\pi_{ijk}\}_{k=1}^{K})$ distribution. Asymptotic results for test statistics hold for x_{ij++} tending to infinity.

For $L > 1$, the hypothesis H: model (2) is a suitable model, can be tested against the wide alternative hypothesis A: model (2) is not a suitable model, by testing the hypothesis of the equality of the L multinomial probability vectors within each cell (i,j) of the design. Pearson's X^2_{ij}'s, having (K-1)(L-1) degrees of freedom (df) may be added to an overall $X^2 = \sum_{ij} X^2_{ij}$ with IJ(K-1)(L-1) df, having the approximate χ^2-distribution for large marginal sums. If H is rejected, the consequence should be that model (2) is rejected and then model (1) is a possible alternative.
The analysis of count data by this model is more complicated; it will be discussed in the following section.

3.3. Analysis of model (1)

Summarizing the concepts of model (1),

(i) $X_{ij+\ell} \sim \text{NBD}\ (\alpha_{ij},\theta)$

(ii) $\{X_{ijk\ell}\}_{k=1}^{K} \mid X_{ij+\ell} = x_{ij+\ell} \sim \text{DM}\ (x_{ij+\ell},\ \{\beta_{ijk}\}_{k=1}^{K})$.

Then π_{ijk} is the parameter for main effect and interactions of factor C, an β parameterizes the random interaction between factor C and whole plot error. The whole plot analysis on $X_{ij+\ell}$ is just as for model (2), so we may concentrate on the sub-plot analysis.

Two cases will be distinguised, namely,
case 1: $L = 1$ (one replicate)
case 2: $L \to \infty$ (many replicates, practically spoken).

Again, the sub-plot analysis is carried out on
$\{X_{ijk+}\}_{k=1}^{K} \mid X_{ij++} = x_{ij++}$, obtained by addition over replicates.

For case 1 as well as case 2 it will be proved that the statistics $X^2 = \Sigma(X-m)^2/m$ and $G^2 = 2\ \Sigma\ X \log X/m$ have an approximate $C * \chi^2$ distribution under certain conditions, where constant C is to be specified later. The results, which are based on those of Brier (1980), Rao and Scott (1984) and Fellegi (1980) are stated in theorem 1, but first two lemma's will be presented, which are needed to prove the theorem.
We shall introduce the vectorial notation $X_* = \{X_k\}_{k=1}^{K}$.

Lemma 1 (case 1: $L = 1$)

Let $x_{ij+1} \to \infty$, $\beta \to \infty$, where $\gamma_{ij} = \beta/x_{ij+1}$ is fixed.
Then, conditionally on $X_{ij+1} = x_{ij+1}$, the distribution of

$\surd x_{ij+1}(X_{ij*1} / x_{ij+1} - \pi_{ij*})$ tends to the

K-variate normal distribution with mean 0 and covariance matrix $V_{ij} = \Gamma_{ij}\ \Pi_{ij}$; $\Gamma_{ij} = 1+\gamma^{-1}_{ij}$; $\Pi_{ij} = D_{\pi_{ij}} - \pi_{ij*}\ \pi'_{ij*}$; $D_{\pi_{ij}}$ is a diagonal matrix with entries $\pi_{ij1}, \ldots, \pi_{ijK}$.
For a proof, see Paul and Plackett (1978).

Lemma 2 (case 2: L is large)

Let $1 \leqslant n_{ij} \leqslant x_{ij+\ell} \leqslant N_{ij}$ for certain numbers n_{ij} and N_{ij}, $\ell=1,2,\ldots,L$ and let
limits $x_{ij} = \lim\limits_{L\to\infty} (\sum\limits_{\ell} x_{ij+\ell}/L)$ and $y_{ij} = \lim\limits_{L\to\infty} (\sum\limits_{\ell} x^2_{ij+\ell}/\sum\limits_{\ell} x_{ij+\ell})$ exist.

For $L \to \infty$, conditionally on $X_{ij+\ell} = x_{ij+\ell}$, the distribution of

$\surd\{x_{ij}\ L\}(X_{ij*+} / x_{ij++} - \pi_{ij*})$ tends to

the K-variate normal distribution with mean 0 and covariance matrix $V_{ij} = \beta_{ij}\ \Pi_{ij}$; $\beta_{ij} = (y_{ij} + \beta) / (1 + \beta)$ and Π_{ij} is as in lemma 1.

For a proof, see Appendix 1.

The results of lemma 1 and lemma 2 are used to prove the following Theorem 1.

Theorem 1 (case 1 and case 2)

The approximate distribution of the statistic X^2 as well as G^2 is $\Gamma^W * \chi^2$ under the conditions of lemma 1 and $\beta^W * \chi^2$ under the conditions of lemma 2, where $\Gamma^W = \Sigma_{ij} w_{ij} \Gamma_{ij}$ and $\beta^W = \Sigma_{ij} w_{ij} \beta_{ij}$, $\Sigma_{ij} w_{ij} = 1$, are weighted averages of $\Gamma_{ij} = 1 + \gamma_{ij}^{-1}$ and $\beta_{ij} = (y_{ij}+\beta)/(1+\beta)$, respectively, weights depending on the hypothesis and on parameters π_{ijk}.

Proof

The Theorem can be proved by a straightforward extension of results of Rao & Scott (1984) on a single distribution of classified numbers, to a set of independent distributions, all having the multivariate normal limiting distribution. The asymptotic covariance matrix has a block structure where blocks are $\Gamma_{ij}\, \Pi_{ij}$ or $\beta_{ij}\, \Pi_{ij}$, respectively, i=1,2,...,I; j=1,2,...,J. Using Theorem 1 from Rao & Scott on the set of independent distributions the result is obtained that X^2 and G^2 are $\Gamma^W * \chi^2$ or $\beta^W * \chi^2$, approximately. Weights depend on the hypothesized model reduction and on the true value of parameter vectors π_{ij*}.

For practical purposes, we have no better solution than to replace β^W by the unweighted mean $\beta = \Sigma_{ij} \beta_{ij} / IJ$ following suggestions of Fellegi (1980), treating Γ^W in a similar way. As it seems, a conservative test is then obtained (Rao & Scott (1984)), but the advantage for practice is evident.

A final problem is the estimation of the extra parameter Γ and β. We suggest an estimation procedure in the following.

For case 1 ($L=1$), Γ can be estimated by fitting a full log-linear model, including all relevant effects of split-plot factor C. If $\upsilon>0$ df are left, Γ is estimated by $\hat{\Gamma} = G^2/\upsilon$, where G^2 is for the full model. Finally, $\hat{\Gamma}$ is used to correct G^2-statistics for testing model reductions, which are obtained in the usual way.

For case 2 (L is large) we can use replicates to estimate B. Following Brier (1980), X^2_{ij} is calculated as in section (3.2). If, for certain n_{ij}, N_{ij}: $1 < n_{ij} < x_{ij+\ell} < N_{ij}$, $\ell=1,2,\dots, L$ and if $x_{ij} = \lim_{L\to\infty} (\sum_{\ell} x_{ij+\ell}/L)$ exists, then $X^2_{ij}/(K-1)(L-1)$ is a consistent estimator of $(x_{ij} + \beta)/(1 + \beta)$, which can be proved by a simple extension of Brier's results.

Combining IJ estimators, $\bar{X}^2 = \sum_{ij} X^2_{ij} \,/\, IJ(K-1)(L-1)$ is an estimator of $(\bar{x}_{++} + \beta) \,/\, (1 + \beta)$, then $\hat{\beta} = (\bar{X}^2 - \bar{x}_{++})/\,(1 - \bar{X}^2)$ is a consistent estimator of β.
Substituting $\hat{\beta}$ for β in B to obtain $\hat{B}$, corrected G^2-statistics have the form $G^2/\hat{B}$.

3.4. Some discussion

In the previous sections we have seen that sub-plot factorial effects can be tested by statistics $G^2/\hat{\Gamma}$ and $G^2/\hat{B}$ for $L=1$ and for large L, respectively.
Here, $\hat{\Gamma}$ is an estimate of $\Gamma = \sum_{ij} \Gamma_{ij} \,/\, IJ$ and $\hat{B} = \sum_{ij} \hat{B}_{ij} \,/\, IJ$, where $\hat{B}_{ij} = (y_{ij} + \hat{\beta}) \,/\, (1 + \hat{\beta})$ so that $\hat{B} = (\bar{y}_{++} + \hat{\beta}) \,/\, (1 + \hat{\beta})$.
For $\bar{X}^2 < 1$, it seems reasonable to set $\hat{B} = 1$. If whole plot totals $X_{ij+\ell}$, for $\ell=1,2,\dots,L$ do not differ widely, then $x_{ij} \approx y_{ij}$ and $\bar{X}^2$ can be used directly as an estimator for B. This may decrease conservativity of tests, as always $\bar{x} < \bar{y}$, hence, in distribution for $L \to \infty$ $\bar{X}^2 \to (\bar{x}_{++} + \beta) \,/\, (1 + \beta) < (\bar{y}_{++} + \beta) \,/\, (1 + \beta) = B$.

4. APPLICATION

As an application, we shall consider the split-plot design for a soldering experiment of print panels, see Engel (1984).
Whole plot factors are: factor A, soldering location and factor B, soldering method. Sub-plot factor is: factor C, type of copper pattern, with levels c_1 and c_2.

Five replicates per cell are available and all factors may have some influence on the response data, which is the number of soldering failures; see table 3 for the data.

		A: soldering location		
B: soldering method	C:copper pattern	a_1	a_2	a_3
b_1	c_1	3 7 3 6 7	1 8 3 6 7	8 7 12 9 14
	c_2	7 8 8 5 9	1 3 1 8 18	5 3 17 7 11
	print total	10 15 11 11 16	2 11 4 14 25	13 10 29 16 25
b_2	c_1	12 9 3 4 7	3 13 5 5 6	4 15 16 11 8
	c_2	10 3 2 4 4	9 22 6 7 11	8 12 9 4 7
	print total	22 12 5 8 11	12 35 11 12 17	12 27 25 15 15

Table 3. Numbers of soldering failures, counted on print panels.

The analysis of whole plot effects has already been discussed and we can restrict ourselves to sub-plot effects.

Following the procedure from section 3, we shall first test for interaction between factor C and whole plot error.
A Pearson X^2_{ij} statistic for homogeneity was calculated for the 2x5 table at each cell (i,j) of the design. The results are shown in table 4.

Level i of factor A	1		2		3	
Level j of factor B	1	2	1	2	1	2
Pearson X^2_{ij} =	2.434	1.794	7.846	1.213	3.256	4.985

Table 4. X^2-statistics for interaction between factor C and whole plot error.

Using the χ^2-approximation with df = 4, no X^2_{ij}-statistic is significant at the 5% level. The sum value of X^2_{ij}'s, devided by total df = 24 equals $\bar{X}^2 = 0.897$, which is even less than 1.

The conclusion is that the interaction between factor C and whole plot error may be ignored, so that the simpler model (2) is suitable for our purposes.

The addition of data over replicates leads to the result of Table 5.

		A: soldering location		
B: soldering method	C:copper pattern	a_1	a_2	a_3
b_1	c_1	26	25	50
	c_2	37	31	43
b_2	c_1	35 (61)	32 (57)	54 (104)
	c_2	23 (60)	55 (86)	40 (83)

Table 5. Data added over replicates (between brackets: data added over levels of factor B).

On this data, a conditional analysis was carried out by the standard loglinear model, including the interaction AB in each model that was fitted to the data. In table 6, loglikelihood ratio G^2-statistics and their corresponding df's are presented.

model term	df	G^2	P
ABC	0	0	1.000
AB, AC, BC	2	4.71	0.095
AB, AC	3	5.56	0.135
AB, BC	4	13.34	0.010
AB, C	5	13.74	0.017
AB	6	13.85	0.031

Table 6. Loglikelihood ratio G^2-statistics, df and tail probability P.

From the methods available for model searching, we tried backward elimination, leading to the model AB, AC where $P = 0.135$, which does not give an excellent fit to the data, but which is the best we have. The interaction between factor A, soldering location and factor C, type of copper pattern seems to be important, which is somewhat surprising from a technical point of view. Addition of the data over levels of factor B was carried out, see table 5. Neither type c_1 or c_2 of copper pattern seems to be uniformly best over locations.

Acknowledgement

The author wishes to thank prof. dr. P. van der Laan, from the Agricultural University Wageningen, and the two referees for critically reading the manuscript and for their valuable suggestions.

REFERENCES

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975)
Discrete multivariate analysis, Theory and practice,
The MIT-Press, Cambridge.

Brier, S.S. (1980),
Analysis of contingency tables under cluster sampling,
Biometrika 67, 591 - 596.

Cox, D.R. and Hinkley, D.V. (1974),
Theoretical Statistics,
Chapman and Hall, London.

Engel, J. (1984),
Models for response data showing extra-Poisson variation,
Statistica Neerlandica 38, 159 - 167.

Engel, J. (1985),
On the analysis of variance for beta binomial responses,
Statistica Neerlandica (to appear).

Fellegi, I.P. (1980),
Approximate tests of independence and goodness of fit based on stratified multistage samples,
JASA 75, 261 - 268.

Fienberg, S.E. (1979),
The analysis of cross-classified categorical data,
The MIT-Press, Cambridge.

Goodhardt, G.J., Ehrenberg, A.S.C. and Chatfield, C. (1984),
The Dirichlet: A comprehensive model of buying behaviour,
Internal Report.

Hájek, J. and Sidák, Z. (1967),
The theory of rank tests,
Academic Press, New York.

Montgomery, D.C. (1976),
Design and analysis of experiments,
Wiley, New York.

Mosimann, J.E. (1962),
On the compound multinomial distribution, the multivariate beta distribution, and correlations among proportions,
Biometrika 49, 65 - 82.

Paul, S.R. and Plackett, R.L. (1978),
Inference sensitivity for Poisson mixtures,
Biometrika 65, 591 - 602.

Rao, J.N.K. and Scott, A.J. (1984),
On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data,
Ann. Statist 12, 46 - 60.

Roos, G. Th. de and Schaafsma, W. (1981),
Is recreation affecting the number of bird's nests? A case study,
Statistica Neerlandica 35, 69 - 90.

Scheffé, H. (1959),
The analysis of variance,
Wiley, New York.

Sibuya, M., Yoshimura, I. and Shimizu, R. (1964),
Negative multinomial distribution,
Ann. Inst. Statist. Math. 16, 409 - 426.

APPENDIX 1

Proof of Lemma 2

The expression

$$\sum_{\ell} (X_{ij*\ell} - x_{ij+\ell}\pi_{ij*}) \Big/ \Big\{\sum_{\ell} \Big[\frac{x_{ij+\ell} + \beta}{1+\beta}\Big] x_{ij+\ell}\Big\}^{1/2} \qquad \ldots (3)$$

can be written as

$$\sum_{\ell} \frac{\Big\{\Big[\frac{x_{ij+\ell} + \beta}{1+\beta}\Big] x_{ij+\ell}\Big\}^{1/2}}{\Big\{\sum_{\ell} \Big[\frac{x_{ij+\ell} + \beta}{1+\beta}\Big] x_{ij+\ell}\Big\}^{1/2}} \; \frac{(X_{ij*\ell} - x_{ij+\ell}\pi_{ij*})}{\Big\{\Big[\frac{x_{ij+\ell} + \beta}{1+\beta}\Big] x_{ij+\ell}\Big\}^{1/2}} \, . \qquad \ldots (4)$$

For L → ∞, this result tends to the K-variate normal distribution with mean 0 and covariance matrix Π_{ij} if $\frac{\Sigma\, a_{\ell}^2}{\max\, a_{\ell}^2} \to \infty$, where a_{ℓ} is the former factor in the expression (4). This follows from a multivariate version of a Theorem of Hájek and Sidák (1967), p. 153. As $\Sigma a_{\ell}^2 = 1$, the condition is equivalent with

$$\sum_{\ell} (x_{ij+\ell} + \beta)\, x_{ij+\ell} \Big/ \max_{\ell}\, (x_{ij+\ell} + \beta)\, x_{ij+\ell} \to \infty$$

which is satisfied under the condition of boundedness for $x_{ij+\ell}$.

Finally, (3) is equal to

$$\sqrt{L} \sum_{\ell} \Big[\frac{X_{ij*\ell} - x_{ij+\ell}\pi_{ij*}}{\sum_{\ell} x_{ij+\ell}}\Big] \Big/ \Big\{\sum_{\ell} \Big[\frac{\Big(\frac{x_{ij+\ell} + \beta}{1+\beta}\Big) x_{ij+\ell}}{\sum_{\ell} x_{ij+\ell}}\Big]\Big[\frac{L}{\sum_{\ell} x_{ij+\ell}}\Big]\Big\}^{1/2} \, .$$

As the denominator has the limit $\sqrt{[\beta_{ij}\ x_{ij}^{-1}]}$ for $L \to \infty$, it follows that

$$\sqrt{\{L\ x_{ij}\}}(X_{ij*+} / x_{ij++} - \pi_{ij*}) \to N(0, \beta_{ij}\ \Pi_{ij})$$

which is the required result.

CHAPTER 6

RANK TESTS AND RANDOM BLOCKING OF CLASSIFIED DATA

1. INTRODUCTION

Random blocking of classified, categorical data may have some consequences for the distribution of rank tests in testing for symmetry (Wilcoxon) and for treatment effect (Friedman). Essentially, the presence of a random interaction between blocks and treatments increases the variance of the asymptotic distribution of these rank tests. This influence of a random interaction on testing main-effects of fixed factors is familiar from the Anova mixed model, e.g. with a fixed factor A and a random factor B. By an approximate F-test, the A main effect is tested 'against' AB interaction, which has the status of a model term for error.

There are some relationships with Brier (1980), who studies the classification of objects by nominal categorical variables under cluster sampling, obtaining asymptotical results for the distribution of X^2-tests for loglinear models. In our case, one of the classifying variables is an ordered response variable, and rank tests instead of X^2-tests are preferred for testing effects. It will be shown that, under the Dirichlet multinomial model used by Brier, similar results are obtained: the asymptotic distribution of the square of Wilcoxon's rank test is of type $\beta * \chi^2$, where β is a constant (see section 2 for the details).
The examples (see sections 3 and 5) consider experiments in which the quality of two manufacturing processes for a certain equipment is compared by the judgement of critical judges. The block factor 'judges' may have an interpretation as a fixed factor; however, the interpretation as a random factor seems to be much more to the point. Then, the levels of the factor 'judges' are considered as a random sample from some large population of judges, e.g. all potential buyers of the equipment. This has some consequences for modeling and hypotheses testing and for the interpretation of the test results. This paper presents some results on this subject.

2. TESTING FOR SYMMETRY

2.1. Model assumptions

We shall consider problems of the following type. Suppose that for each of N levels of a factor "blocks", n objects are classified into 2J+1 classes, which are ordered by numbers -J, ..., 0, 1, ..., J. The hypothesis is that the classification of the objects into the 2J+1 classes is symmetrical with respect to class 0. As to the type of block factor, two cases can be distinguished.

Case 1

The block factor is considered as a factor with N fixed levels.
Our interest then lies in these N levels only, and each block level forms a population by itself. The vector of numbers $X_i = \{X_{ij}\}_{j=-J}^{J}$, i=1,2,..., N, of classified objects for block level i has a multinomial $(n, \{\pi_{ij}\}_{j=-J}^{J})$ distribution, where $\Sigma_{j=-J}^{J} \pi_{ij} = 1$ for each i.

Case 2

Now we are interested in one large population of block levels, and the N block levels represent a random sample from this population. The factor blocks is a random factor (non-specific factor, see Cox (1984)), and it has a different interpretation. As was noticed by Brier (1980), there is a dependence in the classification of objects, and we assume that a reasonnable model for vector $X_i = \{X_{ij}\}_{j=-J}^{J}$ is the Dirichlet multinomial distribution (DMD), see Mosimann (1962), Brier (1980), and Engel (1985), with parameters $(n, \{\beta\ \pi_j\}_{j=-J}^{J})$, where $\Sigma_{j=-J}^{J} \pi_j = 1$. Parameters β and $\pi = \{\pi_j\}_{j=-J}^{J}$ do not depend on the index i of blocks, as they are parameters of the entire population.

The DMD is generated by giving the probability vector $P = \{P_j\}_{j=-J}^{J}$ of the multinomial distribution a Dirichlet distribution, with parameter vector $\{\beta\pi_j\}_{j=-J}^{J}$. In the case of random blocks, for each block sampled from the population of blocks a random P-vector is sampled from this Dirichlet distribution, representing a random probability vector for the classification of the n objects.

The vector π has the interpretation of an average probability vector over the population of vectors, in the sense that $E(P) = \pi$.

Furthermore, parameter β measures the variability of the random vector P. For $\beta \to \infty$, this variability reduces to zero, the DMD reduces to the multinomial distribution of case 1 with probability vector π.

Hypotheses for symmetry can be formulated and tested for both cases.

<u>Case 1</u> H_1: $\pi_{ij} = \pi_{i\ -j}$, $j=1,2, ..,J$, $i=1,2, ..,N$;

<u>Case 2</u> H_2: $\pi_j = \pi_{-j}$ for $j=1,2,...,J$.

The random probability vector P of the DMD will deviate with probability one from its mean value π, for every finite value of the parameter β. This deviation of P from π can be interpreted qualitatively as "random interaction" between random blocks and the factor "difference between treatments" or treatment effect. The overall treatment effect is expressed by the probability vector π (for which the hypothesis H_2 is formulated), but the treatment effect may show local variations from block to block, which is expressed by the random probability vector P with mean π and parameter β for dispersion, which is interpreted as a parameter for interaction. The distribution of e.g. Wilcoxon's rank test for symmetry will be shown to depend on this interaction parameter β.

The presence of this random interaction does not make testing for treatment effect useless, as the treatment effect is defined as an average effect, represented by vector π, over the entire population of blocks, of which only a few are sampled in the experiment.

2.2. Test Statistic

For this type of problem, a typical test statistic for hypothesis H_1 of case 1 is a rank test, e.g. Wilcoxon's test for symmetry with correction for ties (see Lehmann (1975), p. 123; Conover (1971), p. 206).
It is applied straight to the sum vector of data $X_+ = \{X_{+,j}\}_{j=-J}^{J}$,
where $X_{+,j} = \sum_i X_{ij}$ for all j.
Let D_j, j=1,2,...,J, be defined as $D_j = X_{+,-j} + X_{+,j}$ and $D_0 = X_{+,0}$.
Given $D_j = d_j$ for all j, midranks $r_j = d_0 + d_1 + .. + d_{j-1} + (d_j + 1)/2$ are introduced, because ties occur by the nature of the problem.
Then Wilcoxon's test statistic is defined as

$$W = \sum_{j=1}^{J} X_{+,j}\, r_j \qquad \ldots (1)$$

with conditional moments, under the hypothesis H_1,

$$E_M(W) = (Nn\,(Nn+1) - d_0(d_0+1))/4;$$

$$\begin{aligned} \mathrm{var}_M(W) &= \left[Nn\,(Nn+1)(2Nn+1) - d_0(d_0+1)(2d_0+1)\right]/24 \\ &\quad - \left[\sum_{j=1}^{J} d_j(d_j-1)(d_j+1)\right]/48; \qquad \ldots (2) \end{aligned}$$

see Lehmann (1975), p. 130, where the index M denotes that these moments are computed under the multinomial distribution of case 1.

Under the hypothesis H_1, the distribution of the standardized statistic $W^* = (W - E_M(W))/\sqrt{\mathrm{var}_M(W)}$ tends to the standard normal distribution for $Nn - d_0 \to \infty$, which has the following practical interpretation: for large N or large n, d_0 not being dominant, the distribution of W can be approximated by the normal distribution.

We shall now turn to the hypothesis H_2 of case 2.
We propose W as a test statistic for testing the hypothesis H_2. Under the basic DMD and the hypothesis H_2, $E_{DMD}(W) = E_M(W)$ and $var_{DMD}(W) = \beta * var_M(W)$, where $\beta = (n+\beta)/(1+\beta)$. Under the hypothesis H_2, the limiting distribution for $N \to \infty$ of $W^* = (W-E_M(W))/\sqrt{var_M(W)}$ is the $N(0,\sqrt{\beta})$ distribution.
This result is formulated more precisely in the following theorem 1.

Theorem 1

Let $X = \{X_j\}_{j=-J}^{J}$ be a random vector having the 2J+1 variate DMD with parameters n and $\{\beta\pi_j\}_{j=-J}^{J}$, $\sum_{j=-J}^{J} \pi_j=1$. Let $X_i = \{X_{ij}\}_{j=-J}^{J}$, $i=1,2,\dots,N$, be a set of N independent random vectors, having the distribution of X.
Let W be Wilcoxon's test for symmetry (1) with moments (2).
Then under hypothesis H_2 and given $D_j = d_j$ for $j=1,2,\dots,J$, the distribution of $W^* = (W-E_M(W))/\sqrt{var_M(W)}$ tends to the $N(0,\sqrt{\beta})$ distribution for $N \to \infty$, where $\beta = (n+\beta)/(1+\beta)$.

Proof For multinomial random vectors Y_i, parameterized by n and $\{\pi_j\}_{j=-J}^{J}$, the result is true with $\beta = 1$, see Lehmann (1975). Like sum vector Y_+, sum vector X_+ has a multivariate normal limiting distribution for $N \to \infty$ when standardized by

$$E(X_{+,j}) = N\,n\,\pi_j = E(Y_{+,j});\ var(X_{+,j}) = N\,n\,\pi_j(1-\pi_j) * \beta = var(Y_{+,j})*\beta;$$

$$cov(X_{+,j},X_{+,k}) = -\,N\,n\,\pi_j\,\pi_k * \beta = cov(Y_{+,j},Y_{+,k}) * \beta.$$

As W is a linear function of the components of vector X_+, the limiting distribution of W* based on X_+ for $N \to \infty$ is that of W* based on Y_+ for $N \to \infty$, both given $D_j = d_j$ for $j=1, 2, \dots, J$, and under H_2, except for the factor β in the covariance matrix of X_+.
So the distribution of W* tends to the $N(0,\sqrt{\beta})$ distribution for $N \to \infty$, given $D_j = d_j$, $j=1, 2, \dots, J$, and under the hypothesis H_2.

Under the DMD the W-test is in fact a parametric test, as its distribution depends on the unknown parameter β (only) even under H_2.

Following Brier (1980), the constant β is estimated consistently by $\hat{\beta} = X^2/2J(N-1)$, where $X^2 = \sum_{ij} (X_{ij} - X_{+,j}/N)^2/(X_{+,j}/N)$ is Pearson's statistic. If $\hat{\beta}$ turns out to be less than one, setting $\beta = 1$ seems to be a reasonable truncation. For large N the distribution of the corrected W^*, which is $W^*/\sqrt{\hat{\beta}}$, is approximately the standard normal distribution.

2.3. Earlier results for J=1

A more direct relationship with earlier results of Brier exists for objects classified into three classes, hence J=1. Then Wilcoxon's W is equivalent to the sign test (Lehmann (1975), p. 120), with test statistic $S = X_{+,1}$. Under multinomial sampling, given $D_1 = d_1$, moments are $E_M(S) = d_1\pi_1/(\pi_1 + \pi_{-1})$ and $\text{var}_M(S) = d_1\pi_1\pi_{-1}/(\pi_1 + \pi_{-1})^2$.
Further, it holds that the distribution of $S^* = (S-E_M(S))/\sqrt{\text{var}_M(S)}$ tends to the $N(0,\sqrt{\beta})$ distribution for $N \to \infty$ under DM sampling, which is essentially implied by Theorem 1.

Then, under the hypothesis H_2: $\pi_1 = \pi_{-1}$,

$$(S^*)^2 = (X_{+,1}-1/2\ d_1)^2/(1/4\ d_1) \to \beta * \chi_1^2$$

in distribution for $N \to \infty$. On the other hand, $(S^*)^2$ is equivalent to Bowker's X_B^2 test for symmetry (Bishop c.s. (1975), p. 283)
$X_B^2 = (X_{+,1} - X_{+,-1})^2/D_1$, given $D_1 = d_1$.

We can formulate the problem for J=1 in terms of the loglinear model

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

for a 2x2 table, where $u_{1(i)} = u_{2(i)}$, $u_{12(ij)} = u_{12(ji)}$ under the hypothesis of symmetry. It is also implied by Brier's results that $X_B^2 \to \beta * \chi_1^2$ in distribution for $N \to \infty$, and we have a link with these earlier results. Contrary to the sign test, Wilcoxon's test for symmetry cannot be formulated in terms of the standard loglinear model.

3. EXAMPLE 1

In the development of Video Long Play (VLP) discs, the quality of two processes P_1 and P_2 for manufacturing VLP discs is to be compared by visual means. Each of 10 critical judges is asked to compare the quality of 32 types of images recorded on VLP discs from process P_1, with the quality of the same 32 types of images on VLP discs from process P_2. We shall consider these 32 types of images not as a sample but as a fixed and established population of representative images, so that no dependence is introduced between judges of the classification results. In the experiment carried out, the images are presented pairwise on two identical monitors and judges are asked to classify each of the 32 observed differences of image quality into one of the following classes:

number	-3	-2	-1	0	1	2	3
class	P_1 worse than P_2	P_1 less than P_2	P_1 slightly less than P_2	no difference	P_2 slightly less than P_1	P_2 less than P_1	P_2 worse than P_1

For each judge, the result of his classification of images is a vector of numbers $X = \{X_j\}_{j=-3}^{3}$, where $\sum_j X_j = 32$ is fixed by design. Results for 10 judges are presented in table 1.

Judge	P_1 worse than P_2	P_1 less than P_2	P_1 sl. less than P_2	no diff.	P_2 sl. less than P_1	P_2 less than P_1	P_2 worse than P_1	Tot.
1	1	3	6	16	3	3	0	32
2	0	1	4	19	4	4	0	32
3	0	1	6	19	1	4	1	32
4	1	4	6	8	9	4	0	32
5	0	0	6	22	4	0	0	32
6	0	3	8	17	2	2	0	32
7	0	3	9	18	2	0	0	32
8	0	3	5	19	5	0	0	32
9	1	4	6	15	1	5	0	32
10	0	0	9	18	5	0	0	32

Table 1. Results from a VLP comparison experiment with 10 judges.

From the data, some heterogeneity between judges is observed, as well as some slight asymmetry to the left. Formal testing by Pearson's X^2 of the hypothesis of homogeneity, resulting in the value $X^2 = 68.54$ with $2J(N-1) = 54$ degrees of freedom (df), does lead to a rejection of this hypothesis at the 10% level, so that there is an indication for a random interaction between judges and treatments.

If the factor judges is interpreted as a fixed factor, Wilcoxon's statistic can be used for testing hypothesis H_1 of symmetry. The value is W = 14789.50 where $E_M(W) = 18327.00$, $\sqrt{var_M(W)} = 1516.54$, hence $W^* = -2.33$. By the standard normal approximation of the distribution of W* the conclusion is that W is significant at the 5% level, so that H_1 is rejected. Some preference seems to exist among the 10 judges for process P_2.

Interpreting judges as a random block factor and allowing for random interaction, a correction $\sqrt{\beta}$ is needed for W*, where β is estimated by $\hat{\beta} = X^2/54 = 1.27$. Then $W^*/\sqrt{\hat{\beta}} = -2.07$, which is still significant at 5%. Averaged over the large number of judges in the population, i.e. all potential buyers of VLP equipment, an asymmetry between the manufacturing processes P_1 and P_2 appears to exist. However, note that many 'no difference' classifications were given by judges, so that formal testing results should be interpreted with some care.

The results of table 1 can be condensed to those of table 2, which could have been obtained if only three categories had been available for a classification.

Judge	P_1 less than P_2	no difference	P_2 less than P_1
1	10	16	6
2	5	19	8
3	7	19	6
4	11	8	13
5	6	22	4
6	11	17	4
7	12	18	2
8	8	19	5
9	11	15	6
10	9	18	5

Table 2. Condensed results from the VLP comparison experiment with 10 judges.

The sign test for the condensed data results in S = 59, where $E_M(S) = 74.50$, $\sqrt{var_M(S)} = 6.10$ and $S^* = -2.54$, which is significant at the 5% level. From $X^2 = 26.46$ with df = 18, we obtain $\hat{\beta} = 26.46/18 = 1.47$ as an estimate of β. The corrected S* is $S^*/\sqrt{\hat{\beta}}$, which results in $S^*/\sqrt{\hat{\beta}} = -2.10$, and this leads to a rejection of the hypothesis of symmetry, confirming our earlier results.

4. TESTING FOR TREATMENT EFFECT

4.1. Model assumptions

Closely related to the comparison problem for two treatments by testing for symmetry is the following problem. Suppose that a factor blocks has N levels. At each level, let J treatments be applied to n x J objects by complete randomisation, n objects being available for each treatment j. The result of the application of the j^{th} treatment on an object is classified in one of K ordered classes. The question is, how can we test the hypothesis of no treatment effect? Friedman's test seems to be a good candidate. Again we shall make a distinction between fixed and random blocks.

Case 1

If blocks are considered as fixed, a standard model for the data is the product multinomial distribution for each of the N independent matrices $X_i = \{X_{ijk}\}_{j,k}$ of observations, i=1,2,...,N. Then vectors $\{X_{ijk}\}_{k=1}^{K}$, j=1,2,...,J, have independent multinomial $(n,\{\pi_{ijk}\}_{k=1}^{K})$ distributions, where $\Sigma_{k=1}^{K} \pi_{ijk} = 1$.

The hypothesis of no treatment effect is formulated as

$H_1: \pi_{ij_1k} = \pi_{ij_2k}$ for all i, j_1, j_2 and k.

Case 2

For random blocks the random component for block effect should be included in the model. Also a random interaction between blocks and treatments can be present, which may have its influence on the distribution of Friedman's test for treatment effect.

In general, for each block the J probability vectors of the multinomial distributions of case 1 will be dependent random vectors in case 2, with a vector of mean values $\{\pi_{jk}\}_k$ for probability vector j, j=1,2,...,J.

The hypothesis of no treatment effect is now formulated as

$H_2: \pi_{j_1k} = \pi_{j_2k}$ for all j_1, j_2 and k.

We shall first consider Friedman's test for case 1, and then we shall see how to use it in an example with random blocks.

4.2. Test statistic

A rank test for testing hypothesis H_1 is obviously Friedman's test (Lehmann (1975), p. 262, Conover (1971), p. 273) for n objects per treatment and with correction for ties. Given

$D_{ik} = d_{ik}$, where $D_{ik} = \sum_j X_{ijk}$, the numbers $r_{ik} = d_{i1} + d_{i2} + \dots +$

$+ d_{ik-1} + (d_{ik}+1)/2$ are midranks for i=1, 2, ..., N, k=1,2,.., K.

Asymptotically for $N \to \infty$ the Friedman statistic

$$Q = \frac{\frac{12}{N n^2 J(n J+1)} \sum_{j=1}^{J} \left(\sum_{i=1}^{N} \sum_{k=1}^{K} X_{ijk} r_{ik} \right)^2 - 3N(n J+1)}{1 - \sum_i \sum_k (d_{ik}^3 - d_{ik})/\{N n J\{(nJ)^2 - 1\}\}} \qquad \dots (3)$$

has the χ^2_ν distribution with ν=J-1 degrees of freedom under the hypothesis H_1, given $D_{ik} = d_{ik}$ for i=1, 2, ..., N, k=1, 2, ...,K.

For n = 1 the proof of this limiting result can be found e.g. in Lehmann (1975).

We shall see by an example how to use Q for fixed and for random blocks.

5. EXAMPLE 2

As a variant of example 1 we shall consider the following experiment with judges, which is different from the former one. In this new experiment each of 10 judges is asked to give his judgement on the process quality of each manufacturing process P_1 and P_2 separately.

The quality of each process is judged on 32 types of images recorded on VLP discs from the process and the judges are asked to classify each of their

32 judgements in one of four ordered classes (very good, good, not good, less good). As in example 1, the 32 types of images are considered to be the entire population of images.
The experiment is carried out as a randomized blocks experiment and the results are found in table 3.

Block differences appear from the data, but an interaction between blocks and treatments seems not to be present. Applying Friedman's Q (see section 4.2) to the data for testing the effect of the factor process, correcting for ties, gives us the result Q = 2.64.

*In the case of fixed blocks (case 1), the distribution of Q is approxi-*mately χ^2 with df = 1 under hypothesis H_1 and for large N, so that H_1 is not rejected at the 5% level.

In the case of random blocks (case 2) we proceed as follows.
The hypothesis to be tested is

$$H_2: \pi_{1k} = \pi_{2k} \text{ for all } k;$$

see section 4.1. Remember that we have to do with dependent random probability vectors, say vectors

$$P_{ij} = \{P_{ijk}\}_k, \ j = 1,2, \text{ for each block } i \text{ with}$$

$E(P_{ij}) = \pi_j$, where $\pi_j := \{\pi_{jk}\}_k$, j = 1,2, are vectors of mean probabilities.

If we condition on the levels of the random (block) factor judges we are back in the situation which was called case 1 in section 4.1. In fact we condition on the random probability vectors P_{i1} and P_{i2} for all i so that we obtain fixed probability vectors p_{i1} and p_{i2}.

As in the case 1 we can test the hypothesis

$$\tilde{H}_2: p_{i1k} = p_{i2k} \text{ for all } i \text{ and } k$$

by applying Friedman's test Q to this conditioned problem. This hypothesis $\tilde{H}_2$ is of course different from the hypothesis

$H_2: \pi_{1k} = \pi_{2k}$ for all k

of no overall process effect.
Generally speaking, there can be

1. differences between the blocks as regards to their response levels (under $\tilde{H}_2$ as well as under H_2);

2. an influence of the factor manufacturing process on these differences (only under H_2).

We tested hypothesis $\tilde{H}_2$ before in the case 1, and it was not rejected. We shall not formally test the hypothesis H_2. However, as the former hypothesis was not rejected, there does not seem to be any reason to reject the latter, much less restrictive, hypothesis.

Of course, some more formal procedure for testing the hypothesis H_2 is needed; it is an interesting subject for further research.

Judge	Man. process	Response class			
		1 very good	2 good	3 less good	4 not good
1	P_1	6	10	8	8
	P_2	8	9	7	8
2	P_1	0	22	10	0
	P_2	0	24	8	0
3	P_1	0	25	5	2
	P_2	0	22	7	3
4	P_1	2	19	11	0
	P_2	3	21	8	0
5	P_1	13	19	0	0
	P_2	16	16	0	0
6	P_1	5	18	9	0
	P_2	7	16	9	0
7	P_1	0	17	15	0
	P_2	0	19	12	1
8	P_1	21	11	0	0
	P_2	24	6	2	0
9	P_1	0	24	8	0
	P_2	0	28	4	0
10	P_1	2	22	8	0
	P_2	3	24	5	0

Table 3. Results from a VLP comparison experiment with 10 judges; all row totals are equal to 32.

6. SOME DISCUSSION

For ordered categorical data in random blocks where a random interaction between blocks and treatments may be present a simple correction to the limiting distribution of the sign test and the rank test of Wilcoxon was given in the previous sections. The Friedman test too was applied to a problem with random blocks.

The asymptotic approximation to the distribution of these rank tests is valid for a large number N of blocks. For small N, exact permutation tests are popular in nonparametrics, and large computers are helpful for doing calculative work. Unfortunately, under the DM-model the null distribution of e.g. the sign test depends on the unknown interaction parameter β. Another handicap is that the class of DMD's is not closed under addition: if vectors X_i, $i = 1,2, \ldots, N$, are as in Theorem 1, the sum vector X_+ has no DMD.

Estimating β by $\hat{\beta}$ is a way out, doing calculations on independent DMD's where $\hat{\beta}$ is substituted for β and conditioning on relevant marginal sums. However, much of the elegance of nonparametric permutation tests is lost.

REFERENCES

Bishop, Y.M.M.; Fienberg, S.E.; Holland, P.W. (1975),
Discrete multivariate analysis, Theory and practice,
The MIT-Press, Cambridge.

Brier, S.S. (1980), Analysis of contingency tables under cluster sampling,
Biometrika 67, 591 - 596.

Conover, W.J. (1971), Practical nonparametric statistics,
John Wiley, New York.

Cox, D.R. (1984), Interaction, Intern. Statist. Review 52, 1 - 32.

Engel, J. (1985),
On the analysis of variance for beta-binomial responses,
Statistica Neerlandica 39, 27-34.

Lehmann, E.L. (1975),
Nonparametrics, Statistical methods based on ranks,
Holden-day, San Francisco.

Mosimann, J.E. (1962),
On the compound multinomial distribution, the multivariate beta-distribution, and correlations among proportions,
Biometrika 49, 65 - 82.

CHAPTER 7

RANDOM MODELS FOR COUNT RESPONSE DATA

1. INTRODUCTION

In chapters 2 to 6 we considered the analysis of count data for experimental designs where all primary factors were fixed, and where random factors, modeled by random components, were only of secondary importance. However, in many situations the primary factors are random factors and we need models for random factor designs.

In this chapter and in chapter 8, models for count response data from random factor designs will be proposed. Models for random factor designs with continuous (normal) response data are very well known and the analysis of such data (the analysis of variance) can be found in standard books like that by Scheffé (1959), chapter 7. No results have been published, however, on such models for count data; for some remarks in this direction see Cox (1984), 21.

Fixed factors and random factors have different interpretations. The levels of a fixed factor by themselves are considered as a population and the levels of a random vector are considered as a random sample from a population. In both cases we are interested in the population that should be characterized and this characterization is the basis for statistical inference. It is typical for random factors that it is still relevant to test for main effects in the case of random interaction, which is not true in the fixed factor case. For the random factor design the estimated values of the variance components enable us to assign the total variation observed in the data to the various sources of variation.

The models proposed in this chapter for the random factor design are of a multiplicative type. By conditioning on the levels of the random factors we obtain fixed factors and the loglinear model, extended to allow for overdispersion as in the chapters 3 and 4. In fact, Model II from chapter 3 is reobtained if overdispersion is modeled by the gamma distribution.

The limit theorem stated in chapter 4 is used here too for model simplification. The following result is obtained: in the case of large counts and with a lognormal assumption for the random model components the analysis of the data can be carried out by performing a standard Anova on the log-transform of the data. This establishes one of the heuristic practical approaches to count data analysis.

An application of the theory is given by the analysis of data from a man-machine experiment in two random blocks where the response data concerns the number of defect products manufactured by each man on each machine. We assumed the lognormal-Poisson model and the application is worked out in section 3. Some discussion in section 4 concludes this chapter.

2. MODELS FOR RANDOM FACTOR DESIGNS

We shall propose a model for the analysis of count data from random factor designs, and we shall use the result of Theorem 1 of chapter 4 for model simplification. Actually, no results seem to have been published on the subject of random factor designs for count data. It seems unlikely however, that the problem has not been met before in a practical situation.
In agreement with the usual models for fixed factor designs, a multiplicative model will be proposed for the random Poisson mean M_{ij} for cell (i,j) in, say, a two-way classification of crossed and random factors, with levels $i=1,2,\ldots,I$, $j=1,2.,\ldots,J$, and where $k=1,2,\ldots,K$, replicates per cell are available. The model will have the following structure.

(i) Given $M_{ijk} = m_{ijk}$, $i=1,2,\ldots,I$, $j=1,2,\ldots,J$, $k=1,2,\ldots,K$, random variables X_{ijk} are independent, having the Poisson (m_{ijk}) distribution;

(ii) $M_{ijk} = M_{ij}\, F_{k(ij)}$, where M_{ij} is the positive random mean for cell (i,j), $i=1,2,\dots,I$, $j=1,2,\dots,J$, and $F_{k(i,j)}$ is the positive error factor for the k^{th} replicate, $k=1,2,\dots,K$, at this cell. It is assumed that the random variables $F_{k(ij)}$, $i = 1,2, \dots, I$, $j = 1,2, \dots, J$, $k = 1,2, \dots, K$, are independent and independent of the random variables M_{ij}.

Applying the results of Scheffé (1959), section 7.4, where the decomposition of classified continuous response variables into model components is considered onto the variables $\log M_{ij}$, $i=1,2,\dots,I$, $j=1,2,\dots,J$, as if these variables were response variables, the result is that these random variables are decomposed into a sum of model components as follows:

$$\log M_{ij} = \mu + A_i + B_j + (AB)_{ij},$$

where $E(A_i) = 0$, $\operatorname{var}(A_i) = \log(1+\alpha_1^{-1})$;

$$E(B_j) = 0,\ \operatorname{var}(B_j) = \log(1+\alpha_2^{-1});$$

$$E((AB)_{ij}) = 0,\ \operatorname{var}((AB)_{ij}) = \log(1+\alpha_{12}^{-1}).$$

Here A_i and B_j are the model components for main effects and $(AB)_{ij}$ is the model component for interaction; the variances are expressed as they are for reasons that will become clear later. It was shown by Scheffé that the covariance matrix of the vector $\{A^T, B^T, (AB)^T\}$, where e.g. $A = \{A_i\}_{i=1}^{I}$ and A^T is the transposed of A, is diagonal. Further, M_{ij}, $i=1,2,\dots,I$, $j=1,2,\dots,J$, can be written as

$$M_{ij} = \exp(\mu + A_i + B_j + (AB)_{ij}) = \Psi F_i(\alpha_1)\, F_j(\alpha_2)\, F_{ij}(\alpha_{12}),$$

where $\Psi := e^{\mu}$;

$$F_i(\alpha_1) := e^{A_i} ;$$

$$F_j(\alpha_2) := e^{B_j} ;$$

$$F_{ij}(\alpha_{12}) := e^{(AB)_{ij}} .$$

Note that these product terms are not necessarily uncorrelated.

Incidentally, by conditioning on the levels of both random factors we obtain

(ii)' $M_{ijk} = \Psi\ f_i\ f_j\ f_{ij}\ F_{k(ij)}.$

Combining (i) and (ii)' we obtain a multiplicative model for the fixed factor design; see also chapters 3 and 4 where this model was discussed for $F_{k(ij)}$ having the gamma distribution.

The model (i), (ii) is specified by assigning a distribution to the random variables F_i, F_j, F_{ij} and $F_{k(ij)}$, $i=1,2,\ldots,I$, $j=1,2,\ldots,J$, $k=1,2,\ldots,K$. Possible choices of a distribution are the following:

(iii)' the gamma distribution;

(iii)'' the lognormal distribution.

It would be consistent to study model (i), (ii), (iii)' as a natural extension of previous results: model (i), (ii)', (iii)' was studied in chapter 4. Unfortunately, even in the two-factor case the analysis is complicated, as products of gamma distributed random variables no longer have a gamma distribution.

If the random variables F have lognormal distributions instead of gamma distributions, it will be shown that the analysis is greatly simplified.
Some support for the similarity of the two types of distributions for a large shape parameter is given by Johnson & Kotz (1970), p. 196.
We shall make the assumption of joint normality for the uncorrelated components of vector $\{A^T, B^T, (AB)^T\}$, which implies that these components are independent random variables having normal distributions.
The consequence is that the components of vector $\{F^T(\alpha_1), F^T(\alpha_2), F^T(\alpha_{12})\}$ are independent random variables, having lognormal distributions with mean values $\surd(1 + \alpha^{-1})$ and variances $(\alpha^{-1} + \alpha^{-2})$, where $\alpha = \alpha_1, \alpha_2, \alpha_{12}$, respectively; see e.g. Aitchison and Brown (1957) for some properties of the lognormal distribution. Assuming the error factor $F_{k(ij)}$ to have a lognormal distribution with parameter α_3, so that

$E(F_{k(ij)}) = \surd(1 + \alpha_3^{-1})$ and $\mathrm{var}(F_{k(ij)}) = \alpha_3^{-1} + \alpha_3^{-2}$,

the result is that the variables

$$M_{ijk} = \Psi\, F_i(\alpha_1)\, F_j(\alpha_2)\, F_{ij}(\alpha_{12})\, F_{k(ij)}(\alpha_3) \quad \text{for } i=1,2,\ldots,I,\ j=1,2,\ldots,J,\ k=1,2,\ldots,K$$

have lognormal distributions in the following sense: the vector log M where $M = \{M_{ijk}\}_{ijk}$ has a multivariate normal distribution. Further, $E(X_{ijk}) = E(M_{ijk})$ is a constant.
Also, vector M has the form $M = \Psi\, H$, where H is a vector of jointly distributed positive and non-degenerate random components, so that the assumptions of Theorem 1 of chapter 4 are satisfied. Then for large Ψ, vector X is distributed as vector M, approximately, and we shall write

$$X_{ijk} = \Psi\, F_i(\alpha_1)\, F_j(\alpha_2)\, F_{ij}(\alpha_{12})\, F_{k(ij)}(\alpha_3).$$

The equivalent form

$$\log X_{ijk} = \mu + A_i + B_j + (AB)_{ij} + E_{k(ij)}, \qquad \ldots (1)$$

where $E_{k(ij)} = \log F_{k(ij)}$, is the Anova model for the crossed design with two random factors.

The analysis of variance on the data by this model is presented by e.g. Scheffé (1959), chapter 7. To simplify further calculations as regards to the estimation of $\mathrm{var}(X_{ijk})$ and its components, we shall make the following transformation:

$$L_i(\alpha_1) \;:= F_i\, \surd(\alpha_1(\alpha_1 + 1)^{-1});$$

$$L_j(\alpha_2) \;:= F_j\, \surd(\alpha_2(\alpha_2 + 1)^{-1});$$

$$L_{ij}(\alpha_{12}) \;:= F_{ij}\, \surd(\alpha_{12}(\alpha_{12} + 1)^{-1});$$

$$L_{k(ij)}(\alpha_3) \;:= F_{k(ij)}\, \surd(\alpha_3(\alpha_3 + 1)^{-1});$$

$$\theta \;:= \Psi\surd[(\alpha_1+1)(\alpha_2+1)(\alpha_{12}+1)(\alpha_3+1)\, \alpha_1^{-1}\, \alpha_2^{-1}\, \alpha_{12}^{-1}\, \alpha_3^{-1}].$$

Then we obtain

$$X_{ijk} = \theta\, L_i(\alpha_1)\, L_j(\alpha_2)\, L_{ij}(\alpha_{12})\, L_{k(ij)}(\alpha_3), \qquad \ldots (2)$$

where

$$E(L_i(\alpha_1)) = 1,\ \mathrm{var}(L_i(\alpha_1)) = \alpha_1^{-1};$$

$$E(L_j(\alpha_2)) = 1,\ \mathrm{var}(L_j(\alpha_2)) = \alpha_2^{-1};$$

$$E(L_{ij}(\alpha_{12})) = 1,\ \mathrm{var}(L_{ij}(\alpha_{12})) = \alpha_{12}^{-1};$$

$$E(L_{k(ij)}(\alpha_3)) = 1,\ \mathrm{var}(L_{k(ij)}(\alpha_3)) = \alpha_3^{-1}.$$

Between the model components of the models (1) and (2), relationships exist of type $\mathrm{var}(A_i) = \log(1 + \alpha_1^{-1}) = \alpha_1^{-1} + O(\alpha_1^{-2}) = \mathrm{var}(L_i(\alpha_1)) + O(\alpha_1^{-2})$ for large α_1. Furthermore, as

$$E(L_i(\alpha_1)) = E(L_j(\alpha_2)) = E(L_{ij}(\alpha_{12})) = E(L_{k(ij)}(\alpha_3)) = 1\ ,$$

it holds that

$$\mathrm{var}(X_{ijk}) = \theta^2\{\alpha_1^{-1} + \alpha_2^{-1} + \alpha_{12}^{-1} + \alpha_3^{-1} + O(\alpha^{-2})\}$$

for large α_1, α_2, α_{12} and α_3, where $O(\alpha^{-2})$ stands for $O(\alpha_1^{-1}\ \alpha_2^{-1})$ + + $O(\alpha_1^{-1}\ \alpha_2^{-1})$ + ... + $O(\alpha_1^{-1}\ \alpha_2^{-1}\ \alpha_{12}^{-1}\ \alpha_3^{-1})$. We shall use the approximation

$$\mathrm{var}(X_{ijk}) = \theta^2\{\alpha_1^{-1} + \alpha_2^{-1} + \alpha_{12}^{-1} + \alpha_3^{-1}\},$$

so that estimators of the variance components $\theta^2\ \alpha^{-1}$, for $\alpha = \alpha_1, \alpha_2, \alpha_{12}, \alpha_3$, can be added to obtain an estimator of $\mathrm{var}(X_{ijk})$. Estimators of $\mathrm{var}(A_i)$, $\mathrm{var}(B_j)$, $\mathrm{var}((AB)_{ij})$ and $\mathrm{var}(E_{k(ij)})$ follow directly from the results of Anova applied to $\log X_{ijk}$. Ignoring terms

$O(\alpha^{-2})$, for $\alpha = \alpha_1, \alpha_2, \alpha_{12}, \alpha_3$, these are estimators of α_1^{-1}, α_2^{-1}, α_{12}^{-1} and and α_3^{-1}, respectively. To estimate the variance components of type $\theta^2\alpha^{-1}$, it remains to estimate the parameter θ for the general level.

This parameter θ will be estimated by the geometric mean $\hat{\theta}$ of, say, the $N := IJK$ observations; in a short notation

$$\hat{\theta} = \sqrt[N]{\prod_{n=1}^{N} X_n}.$$

This choice is motivated by the fact that a reasonable estimator for $\mu = \log \theta + O(\alpha^{-1})$, where $O(\alpha^{-1})$ stands for $O(\alpha_1^{-1}) + O(\alpha_2^{-1}) + O(\alpha_{12}^{-1}) + O(\alpha_3^{-1})$ for $\alpha_1, \alpha_2, \alpha_{12}$ and α_3 being large, is the arithmetic mean $\hat{\mu} = \overline{\log X} = \log \sqrt[N]{\prod_n X_n}$. As $\hat{\theta} = e^{\hat{\mu}}$ it holds that $E(\hat{\theta}) = \theta(1+O(\alpha^{-1}))$, so that for large α's, the estimator $\hat{\theta}$ is approximately unbiased.

This final result follows from the Taylor series approximation

$$\hat{\theta} = e^{\hat{\mu}} = e^{\mu} + (\hat{\mu}-\mu)\, e^{\mu} + \frac{1}{2}(\hat{\mu}-\mu)^2\, e^{\mu}$$

from which we obtain, using that $E(\hat{\mu}) = \mu$,

$$E(\hat{\theta}) = e^{\mu} + \frac{1}{2} e^{\mu} \operatorname{var}(\hat{\mu}) \text{ with } \operatorname{var}(\hat{\mu}) = O(\alpha^{-1}).$$

Then

$$E(\hat{\theta}) = e^{\mu}(1 + O(\alpha^{-1})) = \theta\, e^{O(\alpha^{-1})}(1 + O(\alpha^{-1})) = \theta(1 + O(\alpha^{-1}))^2 =$$
$$= \theta(1 + O(\alpha^{-1})),$$

and this is the result that was to be obtained.

3. APPLICATION

The classical man-machine experiment was carried out in a factory hall. Let three machines be randomly sampled from a large population of machines, i.e. all machines in a factory hall. Four men (say: workers) are randomly sampled from a population of workers, i.e. all potential operators of the machines. All workers are supposed to manufacture large and equal numbers of products on each machine, and the number of defects is counted. See table 1 for the data from this experiment in two randomized blocks.

block	1				2			
worker / machine	1	2	3	4	1	2	3	4
1	13	17	11	8	15	19	9	10
2	16	18	19	12	20	17	18	15
3	5	6	8	8	7	9	10	7

Table 1. Numbers of defects from the worker-machine experiment in two blocks.

All three factors in the design are typically random factors, and we shall analyse the data by the methods presented before. Labeling the factor machines with M, the factor workers with W and blocks with B, the random and linear model for log $X_{ijk\ell}$ is

$$\log X_{ijk\ell} = \mu + M_i(\alpha_1) + W_j(\alpha_2) + B_k(\alpha_3) + MW_{ij}(\alpha_{12}) + \dots$$

$$\dots + MWB_{ijk}(\alpha_{123}) + E_{\ell(ijk)}(\alpha_4),$$

where $E(M_i) = 0$, $\mathrm{var}(M_i) = \log(1 + \alpha_1^{-1})$, etc., see section 2.

Some important aspects of the analysis of this data by the random model are

- testing hypotheses of type H: $\alpha^{-1} = 0$, $\alpha = \alpha_1, \alpha_2, \dots, \alpha_{23}$, for main effects and interactions, identifying important sources of variation.
- estimating variance components of type $\theta^2 \alpha^{-1}$, assigning the variation observed in the data to important sources of variation identified before.

The Anova results for the data of table 1 are as follows, see table 2.

Model term	SS	df	MS	$F(\nu_1,\nu_2)$	P
M	2.7484	2	1.3742	15.07 (2,4)	< 0.05
W	0.3057	3	0.1019	1.07 (3,5)	> 0.10
B	0.0868	1	0.0868	4.00*)(1,6)	> 0.05
MW	0.6091	6	0.1015	4.69 (6,6)	< 0.05
MB	0.0232	2	0.0116	0.54 (2,6)	> 0.10
WB	0.0472	3	0.0157	0.73 (3,6)	> 0.10
MWB+E	0.0216	6	0.0216		

Table 2. Anova results for the worker-machine experiment. *) No reasonable denominator degrees of freedom are obtained for the F-test for the B-effect. As the interactions MB and WB are not significant, the B-effect is tested against MWB+E.

A synthesis of variances is needed to test main effects via approximate F-tests; see Scheffé (1959) and Cox (1984). Remember that the approximation we made amounts to ignoring the Poisson part (i) of the model of section 2. The within-cell variation in the data is then fully explained by the model component for error $E_{\ell(ijk)}$ with parameter α_4.

From the Anova table 2 it is seen that the main effect of the factor machines is significant. Also the interaction between machines and workers is important. There seem to be differences between machines in the factory hall, and the variance component $\theta^2\alpha_1^{-1}$ is non-vanishing. Differences between machines depend on workers and certain machines seem to be favourite only for some workers, as is seen from the data.

Estimates of the variance components that contribute to the variance in the data are shown in table 3.

Model term	Estimate of α^{-1}	Estimate of $\theta^2\alpha^{-1}$
M	0.160	20.95
MW	0.040	5.24
MWB+E	0.022	2.88

Table 3. Estimates of variance components.

An estimate of $\mathrm{var}(X_{ijk\ell}) = \theta^2(\alpha_1^{-1} + \alpha_{12}^{-1} + \alpha_{123}^{-1} + \alpha_4^{-1})$ is $\widehat{\mathrm{var}}(X_{ijk\ell}) = (11.444)^2 * 0.222 = 29.07$, estimating θ by the geometric mean $\hat{\theta}$. The major part of the variation in the data should be assigned to machines and, to a much smaller extent, to the interaction between machines and workers.

4. DISCUSSION

A class of lognormal-Poisson models was proposed for modeling count data from random factor designs. For large counts the analysis of the data is carried out simply by performing standard Anova on the log-transform of the data.

It is recalled that some simulation results were obtained in chapter 4 for the experimental design with one fixed factor and the lognormal-Poisson model for the data. From these results we know that the approximate F-statistic for testing the main effect of the single factor behaves quite reasonably under the null hypothesis. For the random factors design, the analysis method is not essentially different from the method used in the fixed factor case. In both cases the lognormal-Poisson model is approximated by the lognormal model, and ratios of sums of squares form F-tests for testing the hypotheses. Thus the quality of the F-tests used in this chapter is not expected to be less than that of the F-test used in the fixed factor case.

Generalizations of the results of section 2 to the case of three and more crossed and random factors are straightforward. Also nested designs with random factors can be treated by this method. An example will be presented in chapter 8, where the results of the analysis of data by this method are compared with those obtained by a quasi-likelihood approach. Important too are extensions to the "mixed model" case where fixed and random crossed factors are present in the design. These may be performed without involving too many difficulties; it is a potential subject for further research.

REFERENCES

Aitchison, J. and Brown, J.A.C. (1957), The lognormal distribution, University Press, Cambridge.

Cox, D.R. (1984), Interaction, Intern. Statist. Review 52, 1 - 32.

Johnson, N.L. and Kotz, S.K. (1970), Distributions in statistics, Continuous univariate distributions - 1, Houghton Mifflin Company, Boston.

Scheffé, H. (1959), The analysis of variance, John Wiley & Sons, New York.

CHAPTER 8

RANDOM MODELS FOR COUNT RESPONSE DATA, A QUASI-LIKELIHOOD APPROACH FOR NESTED DESIGNS

1. INTRODUCTION

Breslow (1984) and Engel (1984) presented some methods for the analysis of independent count data from fixed-factor designs, showing non-Poisson distributional behaviour (extra-Poisson variation). A class of models for count data from random-factor designs was presented in chapter 7. The analysis of this data appeared to be rather straightforward for i) an approximate version of the model for large expections and ii) a lognormal assumption made for model components. The result is that the random variable X of counts has a lognormal distribution, approximately, so that an analysis of variance can be carried out on log X, having a constant variance.

We shall now try to tackle one of the problems from the analysis of count data for random factor designs by the quasi-likelihood approach, which was formally introduced by Wedderburn (1974), although some aspects of it had been used before. The essential part of this method is that assumptions are made for the random variable X only regarding its mean and variance; that is, a known mean-variance relationship is assumed, expressing the variance of X as a known function of the mean. This mean-variance relationship is used in the estimation procedure by Iteratively Weighted Least Squares for the parameters of a linear model assumed for some function of the mean value of X.

From the mean-variance relationship a quasi-likelihood function too can often be made explicit. It plays the role of the likelihood function and it is a basis for deriving test statistics for hypotheses on the parameters of the linear model. Note that no distributional assumptions have to be made. This can be seen as an advantage of quasi-likelihood: the formulation of a complicated model, leading to a complicated analysis, is now avoided.

In the following sections, the use of quasi-likelihood will be proposed for the analysis of count data from nested designs with several random factors. Variance components will be estimated and hypotheses will be tested by statistics based on quasi-likelihood functions, having asymptotic distributions (for many replicates) of chi-squared type.

To justify the use of the quasi-likelihood method, the assumed mean-variance relationships should be verified by the data, if possible. On the other hand, it is shown that one set of quasi-likelihood assumptions (out of two that will be made) is approximately satisfied for the class of models proposed in chapter 7 for the case of a nested design.

In section 4 the results are compared of the analyses of one set of data under the two sets of quasi-likelihood assumptions made in section 2, and also under the approximate lognormal model from chapter 7 for the data. Not much difference is observed between these three methods with respect to the estimates of variance components, and the testing procedures lead to the same conclusions. Of course, not everything is said from only one set of data. It was, however, also mentioned by McCullagh and Nelder (1984), p. 132, that in general the results do not heavily depend on the specific quasi-likelihood assumptions made.
Some discussion and suggestions for further research make up the concluding section 5 of this chapter.

2. THE QUASI-LIKELIHOOD APPROACH FOR NESTED DESIGNS

2.1. Some introductory remarks

The quasi-likelihood method was formally defined by Wedderburn (1974), and a more theoretical foundation was laid by McCullagh (1983).
With the usually sophisticated models for count data there is certainly a problem in analysing data by a formal likelihood approach. It was the idea of Wedderburn not to make any distributional assumptions, but only to make

assumptions of finite expectation and variance, and of a so-called mean-variance relationship, expressing the variance as a known function (times an unknown scale parameter) of the mean value. This function is called the variance function; it is denoted by $V(\mu)$, where $\mu = E(X)$.
Parameters of a linear model assumed for some (link) function of the mean value can then be estimated by Iteratively Weighted Least Squares; see Nelder and Wedderburn (1972) and McCullagh (1984).
Some aspects of the quasi-likelihood method have been known for some time. It was Finney who used the method in an informal way in probit analysis. On the other hand, no results appeared on the asymptotic distribution and optimality of estimators for the linear model parameters until 1983; see McCullagh (1983).

An example of a mean-variance relationship for random variable X is simply $var(X) = \mu$, which is true for e.g. the Poisson random variable. Slightly more general is $var(X) = \sigma^2\mu$, where $\sigma^2>0$ is an unknown parameter. In the case of 'overdispersion', $\sigma^2>1$, and the analysis of overdispersed count data by quasi-likelihood may be compared with the analysis of this data by one of the models for overdispersed count data (extra-Poisson variation) from Engel (1984). Note that 'underdispersion', where $\sigma^2<1$, can be studied as well. Another example of a mean variance relationship is $var(X) = \sigma^2\mu^2$, corresponding to a constant coefficient of variation of X.

Often it is possible to define a log quasi-likelihood function as a substitute for a log likelihood function. For univariate data this function $\ell(\mu,x)$ is defined by $\dfrac{\partial\ell(\mu,x)}{\partial\mu} = (x-\mu)/V(\mu)$, where $V(\mu)$ is the variance function. For certain $V(\mu)$, explicit solutions of $\ell(\mu,x)$ can be found from this differential equation. Some examples can be found in McCullagh (1983). From the log quasi-likelihood function $\ell(\mu,x)$ obtained explicitly, test statistics can be derived for testing hypotheses concerning generalized linear models for the mean value μ. Then often an estimator is needed for the dispersion parameter σ^2, which can be obtained e.g. from replicated data.

We shall follow the quasi-likelihood approach to analyse count data from nested designs with random factors. It gives us the opportunity to analyse this data without making any distributional assumptions. In the following section, mean-variance relationships will be assumed which are reasonable for count data from well-designed experiments. This data will be analysed by quasi-likelihood and variance components will be estimated and tested. For the nested design, two types of mean-variance relationships will be studied, namely $\mathrm{var}(X) = \sigma^2\mu$ and $\mathrm{var}(X) = \sigma^2\mu^2$. Two log quasi-likelihood functions can be derived corresponding to these two relationships, having the form of the Poisson log likelihood $\ell(\mu,x) = x \log \mu - \mu$ for the first relationship and the gamma log likelihood $\ell(\mu,x) = -x/\mu - \log \mu$ for the second; see Wedderburn (1974) and McCullagh (1983) for further details. From these log quasi-likelihood functions deviances are obtained in the usual way, following Nelder and Wedderburn (1972), as $D = 2 \Sigma x \log (x/\mu)$ and $D = -2 \Sigma \log (x/\mu)$, respectively. By taking differences of these deviances test statistics are obtained for testing variance components. These test statistics will be called log quasi-likelihood ratio test statistics. Finally, a generalized Pearson's X^2 useful for estimating variance components is defined as $X^2 = \Sigma(X - \mu)^2/V(\mu)$, where $V(\mu)$ is the variance function (see McCullagh and Nelder (1984)).

2.2. Quasi-likelihood for nested designs with random factors

As an example we shall study a nested design with two random factors A and B, with levels $i=1,2,\dots,I$ and $j=1,2,\dots,J$, respectively, where factor B is nested within factor A and where K replicates are available 'per cell', with levels $k=1,2,\dots,K$. As an orientation, consider the design from table 2, section 4.

Let X_{ijk} be the k^{th} replicate belonging to cell (i,j); let $E(X_{ijk})$ equal μ, the overall mean value. We shall make the following set of quasi-likelihood assumptions considering variance functions and independence of random variables. These latter assumptions are directly related to the properties of the nested experimental design; see Scheffé (1959), chapter 7.

Quasi-likelihood assumptions

1. Let M_i be independent random variables for i=1,2,...,I, where

   $$E(M_i) = \mu;$$

   $$\mathrm{var}(M_i) = \sigma_1^2\,\mu^r,\ r > 0 \text{ is some constant.}$$

2. Given $M_i = m_i$, i=1,2,...,I, let M_{ij} be independent random variables for j=1,2,...,J, where

   $$E(M_{ij}|M_i=m_i)= m_i;$$

   $$\mathrm{var}(M_{ij}|M_i=m_i)= \sigma_2^2\,m_i^r,\ r > 0 \text{ is some constant.}$$

3. Given $M_{ij}=m_{ij}$, i=1,2,...,I, j=1,2,...,J, let X_{ijk} be independent (response) random variables for k=1,2,...,K, where

   $$E(X_{ijk}|M_{ij}=m_{ij}) = m_{ij};$$

   $$\mathrm{var}(X_{ijk}|M_{ij}=m_{ij}) = \sigma_3^2\,m_{ij}^r,\ r > 0.$$

Note that these quasi-likelihood assumptions are not necessarily restricted to the case of count data X_{ijk}; they do also make sense for other types of data, like continuous data.

It is seen from the assumptions that m_{ij} is the mean value of X_{ijk}, given $M_{ij} = m_{ij}$, that is given level j and given level i of the random factors B and A, respectively. Also, m_i is the mean value of X_{ijk}, given level i of factor A.

It can be proved that, given $M_i = m_i$, i=1,2,...,I, the sample means $\bar{X}_{ij+}$ are independent random variables, for j=1,2,...,J. It can also be proved that sample means $\bar{X}_{i++}$ are independent random variables for i=1,2,...,I.

Three parameters, σ_1^2, σ_2^2 and σ_3^2 were introduced to describe the variation in the data at three levels:

- replicates level (σ_3^2), with index k;
- factor B level (σ_2^2), with index j;
- factor A level (σ_1^2), with index i.

At each level, the variation in the data is described by means of a variance function type of relationship as described in section 2.1.

Note that for r=1 the assumptions are those of Poisson overdispersion ($\sigma_1^2>1$) or underdispersion ($\sigma_1^2<1$). For r=2, the parameters σ_1^2, σ_2^2 and σ_3^2 represent constant squared coefficients of variation. We shall estimate these parameters from the data for certain values of r, and test hypotheses of type $H_2 : \sigma_2^2 = 0$ (no effect of random factor B) and $H_1 : \sigma_1^2 = 0$ (no effect of random factor A) in section 2.4.
To obtain the estimators of σ_1^2, σ_2^2 and σ_3^3 and test statistics of the hypotheses H_1 and H_2 with their asymptotic distributions, firstly some implications of the quasi-likelihood assumptions will be derived in section 2.3. These implications give expressions for:

I1. Mean value and variance function of $\bar{X}_{ij+}$, given $M_i = m_i$;
I2. Mean value and variance function of $\bar{X}_{i++}$;
I3. The variance of X_{ijk}.

The implications I1 and I2 are useful for the estimation and testing of the parameters σ_1^2 and σ_2^2, respectively. The implication I3 is needed for estimating the variance of X_{ijk}.

2.3. Implications of the quasi-likelihood assumptions

2.3.1. Two Lemmas

Firstly, we shall mention a general and familiar lemma for calculating variances from conditional variances and expectations in the form of lemma 1.

Lemma 1. Let X and Y be random variables having a joint distribution. Let f(x,y) be a real-valued function of $(x,y) \in R^2$. Then

$$\operatorname{var}[f(X,Y)] = E_Y \operatorname{var}_X[f(X,Y)|Y] + \operatorname{var}_Y E_X[f(X,Y)|Y].$$

A slightly more general version of lemma 1 is lemma 2.

Lemma 2. Let X, Y and Z be random variables having a joint distribution. Let f(x,y) be a real-valued function of $(x,y) \in R^2$. Then

$$\mathrm{var}[f(X,Y)|Z=z] = E_{Y|z}\,\mathrm{var}_{X|z}[f(X,Y)|Y] + \mathrm{var}_{Y|z}E_{X|z}[f(X,Y)|Y].$$

Lemmas 1 and 2 will be used once or more in sections 2.3.2, 2.3.3 and 2.3.4 when deriving some implications of the quasi-likelihood assumptions.

2.3.2. Restricting values of parameter r

From $\mathrm{var}(\bar{X}_{ij+}|M_{ij}=m_{ij}) = K^{-1}\,\sigma_3^2\,m_{ij}^r$ and lemma 2, with $Z = M_i$, $Y = M_{ij}$, $f(X,Y) = \bar{X}_{ij+}$, it follows that

$$\mathrm{var}(\bar{X}_{ij+}|M_i=m_i) = E_{M_{ij}|m_i}\,\mathrm{var}(\bar{X}_{ij+}|M_{ij}) + \mathrm{var}_{M_{ij}|m_i}E(\bar{X}_{ij+}|M_{ij}) =$$

$$= K^{-1}\,\sigma_3^2\,E(M_{ij}^r|M_i=m_i) + \sigma_2^2\,m_i^r.$$

The statistical inference with respect to parameter σ_2^2 will be performed on the statistic $\bar{X}_{ij+}$ given $M_i=m_i$; then the variance function of $\bar{X}_{ij+}$ given $M_i=m_i$ has to be a known function, so that $E(M_{ij}^r|M_i=m_i)$ has to be a known expectation.
Because of the limited amount of information we have concerning the moments of M_{ij} given $M_i=m_i$, the expectation is known for only three values of r:

$r=0$, $E(M_{ij}^0|M_i=m_i) = 1$;

$r=1$, $E(M_{ij}|M_i=m_i) = m_i$;

$r=2$, $E(M_{ij}^2|M_i=m_i) = \sigma_2^2\,m_i^2 + m_i^2$.

The case r=0 corresponds to the Anova-like situation of constant variances:

$$\mathrm{var}(M_i) = \sigma_1^2;$$

Given $M_i = m_i$, $\mathrm{var}(M_{ij}) = \sigma_2^2$;

Given $M_{ij} = m_{ij}$, $\mathrm{var}(X_{ijk}) = \sigma_3^2$.

In this case the quasi-likelihood analysis is based on the usual sums of squares known from Anova, not making the assumption of normality of distributions so common in Anova. It is possible to derive some interesting approximate results on χ^2-tests and F-tests under these non-normal conditions, shedding some new light on the Anova tests. However, we shall not explore the r=0 case any further because, for count data, our primary interest lies in the cases r=1 and r=2. For the case r=1 we denote the quasi-likelihood assumptions by Assumptions I, and for the case r=2 by Assumptions II.

2.3.3. Assumptions I, three implications

I1. From 2.3.2. (for r=1) we obtain the result

$$\mathrm{var}(\bar{X}_{ij+}|M_i=m_i) = K^{-1}\sigma_3^2\, E(M_{ij}|M_i=m_i) + \sigma_2^2 m_i = K^{-1}(\sigma_3^2 + K\sigma_2^2)m_i.$$

If we define $\sigma_{23}^2 := \sigma_3^2 + K\sigma_2^2$, then σ_2^2 can be expressed as

$$\sigma_2^2 = (\sigma_{23}^2 - \sigma_3^2)/K.$$

It follows that $\mathrm{var}(\bar{X}_{ij+}|M_i=m_i) = K^{-1}\sigma_{23}^2\, m_i$, and because $E(\bar{X}_{ij+}|M_{ij}=m_{ij}) = m_{ij}$, we obtain $E(\bar{X}_{ij+}|M_i=m_i) = m_i$.
Now we have expressed the variance of $\bar{X}_{ij+}$, given $M_i=m_i$, as a linear variance function of the conditional mean value m_i.

I2. Firstly, $\mathrm{var}(\bar{X}_{i++}|M_i=m_i) = (JK)^{-1}(\sigma_3^2 + K\sigma_2^2)\, m_i = (JK)^{-1}\sigma_{23}^2\, m_i$, which is the variance of the average of $\bar{X}_{ij+}$ for j=1,2,...,J, and given $M_i=m_i$ these random variables are independent. Then it follows by lemma 1 that

$$\mathrm{var}(\bar{X}_{i++}) = E_{M_i}\mathrm{var}(\bar{X}_{i++}|M_i) + \mathrm{var}_{M_i}E(\bar{X}_{i++}|M_i) = (JK)^{-1}\sigma_{23}^2\, \mu + \sigma_1^2\, \mu =$$

$$= (JK)^{-1}(\sigma_{23}^2 + JK\,\sigma_1^2)\, \mu.$$

Defining $\sigma^2_{123}: = \sigma^2_{23} + JK\ \sigma^2_1$ $(= \sigma^2_3 + K\sigma^2_2 + JK\ \sigma^2_1)$, we can express σ^2_1 as

$\sigma^2_1 = (\sigma^2_{123} - \sigma^2_{23})/JK$.

Further, $\mathrm{var}(\bar{X}_{i++}) = (JK)^{-1}\ \sigma^2_{123}\ \mu$, and from $E(\bar{X}_{i++}|M_i=m_i) = m_i$ we obtain the result $E(\bar{X}_{i++}) = \mu$.

We have now obtained the variance function of $\bar{X}_{i++}$ as a linear function of the mean value μ.

I3. Finally, from $\mathrm{var}(X_{ijk}|M_{ij}=m_{ij}) = \sigma^2_3\ m_{ij}$ it follows by lemma 2 that $\mathrm{var}(X_{ijk}|M_i=m_i) = (\sigma^2_3 + \sigma^2_2)m_i$ and by lemma 1 we obtain

$\mathrm{var}(X_{ijk}) = (\sigma^2_3 + \sigma^2_2 + \sigma^2_1)\ \mu$.

It is seen that the variance of X_{ijk} is split up into three variance components $\sigma^2_1\mu$, $\sigma^2_2\mu$ and $\sigma^2_3\mu$ for the three levels of random variation.

The expressions for σ^2_1 and σ^2_2 derived above will be needed in section 2.4 where estimators will be presented for these parameters. These estimators are obtained from estimators of σ^2_3, σ^2_{23} and σ^2_{123}.

Three similar implications for Assumptions II (r=2) will be derived in section 2.3.4.

2.3.4. Assumptions II, three implications

I1. From $\mathrm{var}(\bar{X}_{ij+}|M_{ij}=m_{ij}) = K^{-1}\sigma^2_3\ m^2_{ij}$ and lemma 2, it follows that

$$\mathrm{var}(\bar{X}_{ij+}|M_i=m_i) = E_{M_{ij}|m_i}\mathrm{var}(\bar{X}_{ij+}|M_{ij}) + \mathrm{var}_{M_{ij}|m_i}E(\bar{X}_{ij+}|M_{ij}) =$$

$$= K^{-1}\sigma^2_3\ (\sigma^2_2 + 1)\ m^2_i + \sigma^2_2\ m^2_i = K^{-1}\ (\sigma^2_3\ \sigma^2_2 + \sigma^2_3 + K\ \sigma^2_2)\ m^2_i.$$

Defining $\sigma^2_{23}: = (\sigma^2_3\ \sigma^2_2 + \sigma^2_3 + K\ \sigma^2_2)$, we can express σ^2_2 as

$\sigma^2_2 = (\sigma^2_{23} - \sigma^2_3)/(K + \sigma^2_3)$.

Further, $\mathrm{var}(\bar{X}_{ij+}|M_i=m_i) = K^{-1}\ \sigma^2_{23}\ m^2_i$, and also $E(\bar{X}_{ij+}|M_i=m_i) = m_i$.

Now we have obtained a quadratic variance function of $\bar{X}_{ij+}$, given $M_i=m_i$.

I2. From $\mathrm{var}(\bar{X}_{i++}|M_i=m_i) = (JK)^{-1}\ \sigma^2_{23}\ m_i^2$ and lemma 1, it follows that

$$\mathrm{var}(\bar{X}_{i++}) = E_{M_i}\mathrm{var}(\bar{X}_{i++}|M_i) + \mathrm{var}_{M_i}E(\bar{X}_{i++}|M_i) =$$

$$= (JK)^{-1}\ \sigma^2_{23}\ (\sigma^2_1 + 1)\ \mu^2 + \sigma^2_1\ \mu^2 = (JK)^{-1}\ (\sigma^2_{23}\ \sigma^2_1 + \sigma^2_{23} + JK\ \sigma^2_1)\ \mu^2.$$

Defining $\sigma^2_{123} := \sigma^2_1\ \sigma^2_{23} + \sigma^2_{23} + JK\ \sigma^2_1$, we can express σ^2_1 as

$$\sigma^2_1 = (\sigma^2_{123} - \sigma^2_{23})/(JK + \sigma^2_{23}).$$

Further, $\mathrm{var}(\bar{X}_{i++}) = (JK)^{-1}\ \sigma^2_{123}\ \mu^2$, and also $E(\bar{X}_{i++}) = \mu$, so that a quadratic variance function is obtained for $\bar{X}_{i++}$.

I3. Finally, from $\mathrm{var}(X_{ijk}|M_{ij}=m_{ij}) = \sigma^2_3\ m^2_{ij}$ and lemma 2 it follows that $\mathrm{var}(X_{ijk}|M_i=m_i) = (\sigma^2_3\ \sigma^2_2 + \sigma^2_3 + \sigma^2_2)\ m^2_i$ and by applying lemma 1 and ordering terms,

$$\mathrm{var}(X_{ijk}) = (\sigma^2_1 + \sigma^2_2 + \sigma^2_3 + \sigma^2_1\ \sigma^2_2 + \sigma^2_1\ \sigma^2_3 + \sigma^2_2\ \sigma^2_3 + \sigma^2_1\ \sigma^2_2\ \sigma^2_3)\ \mu^2 =$$

$$= (\sigma^2_1 + \sigma^2_2 + \sigma^2_3)\ \mu^2$$

where the final equality holds approximately for small σ^2_i, i=1,2,3. We now have again expressed the variance of X_{ijk} as the sum of three variance components which are now $\sigma^2_1\ \mu^2$, $\sigma^2_2\ \mu^2$ and $\sigma^2_3\ \mu^2$.

Having derived (conditional) variance functions and mean values for the sample means $\bar{X}_{ij+}$ and $\bar{X}_{i++}$ and an expression for $\mathrm{var}(X_{ijk})$, we are now able to derive estimation and testing procedures for the parameters σ^2_1, σ^2_2 and σ^2_3 in these variance functions. In section 2.4.1. estimators will be presented, and the results for the tests of the hypotheses H_1 and H_2 will be discussed in section 2.4.3.. Firstly, some more notation will be introduced.

2.4. Estimation and testing procedures

Before discussing estimation and testing procedures, some more notation will be introduced.

Some notation for Assumptions I

Let D be the Poisson deviance, i.e.

$D = 2 \sum_{ijk} X_{ijk} \log (X_{ijk}/\hat{E}(X_{ijk}))$, where $\hat{E}(X_{ijk})$ is an estimator for $E(X_{ijk})$.

More concretely, $\hat{E}(X_{ijk})$ is supposed to be the maximum quasi-likelihood estimator for $E(X_{ijk})$ and indices will be used for the deviance to express the conditions under which this maximum quasi-likelihood estimator is computed. Then,

- for deviance D_{AB}, $\hat{E}(X_{ijk})$ is computed given $M_{ij} = m_{ij}$, so that $\hat{E}(X_{ijk}) = \hat{m}_{ij} = \bar{X}_{ij+}$;

- for deviance D_A, $\hat{E}(X_{ijk})$ is computed given $M_i = m_i$, assuming $M_{ij} \equiv M_i$, so that $\hat{E}(X_{ijk}) = \hat{m}_i = \bar{X}_{i++}$;

- for deviance D_0, $\hat{E}(X_{ijk})$ is computed as the grand mean $\hat{\mu} = \bar{X}_{+++}$, assuming that $M_{ij} \equiv M_i \equiv \mu$.

The first result is obtained as follows. Given $M_{ij}=m_{ij}$, the variance function of X_{ijk} is $\text{var}(X_{ijk}|M_{ij}=m_{ij}) = \sigma_3^2 m_{ij}$; see section 2.2. The log quasi-likelihood function associated with this variance function is $\sum_k (X_{ijk} \log m_{ij} - m_{ij})$; see McCullagh (1983). Maximizing this function with respect to m_{ij} gives the required result. In a similar way, the other results can be obtained.

It follows that

$$D_A - D_{AB} = 2 \sum_{ijk} X_{ijk} \log (X_{ijk}/\bar{X}_{i++}) - 2 \sum_{ijk} X_{ijk} \log (X_{ijk}/\bar{X}_{ij+}) =$$

$$= 2 \sum_{ijk} X_{ijk} \log (\bar{X}_{ij+}/\bar{X}_{i++}) = 2K \sum_{ij} \bar{X}_{ij+} \log (\bar{X}_{ij+}/\bar{X}_{i++})$$

and

$$D_0 - D_A = 2JK \sum_i \bar{X}_{i++} \log (\bar{X}_{i++}/\bar{X}_{+++}).$$

These statistics will be used as log quasi-likelihood ratio test statistics (see section 2.4.3) under Assumptions I; they form the basic material for testing hypotheses H_2: $\sigma_2^2 = 0$ and H_1: $\sigma_1^2 = 0$, respectively.

Further, for estimating σ_1^2, σ_2^2 and σ_3^2, estimators will be defined in section 2.4.1. which are based on generalized Pearson statistics, defined as follows (see also section 2.1):

$$X^2_{AB} = \sum_{ijk} (X_{ijk} - \bar{X}_{ij+})^2/\bar{X}_{ij+};$$

$$X^2_{A;AB} = K \sum_{ij} (\bar{X}_{ij+} - \bar{X}_{i++})^2/\bar{X}_{i++};$$

$$X^2_{0;A} = JK \sum_i (\bar{X}_{i++} - \bar{X}_{+++})^2/\bar{X}_{+++}.$$

Some notation for Assumptions II

Let D be the gamma deviance, i.e.

$$D = -2 \sum_{ijk} \log (X_{ijk}/\hat{E}(X_{ijk})),$$ where $\hat{E}(X_{ijk})$ is an estimator of $E(X_{ijk})$.

Again, $\hat{E}(X_{ijk})$ is the maximum quasi-likelihood estimator for $E(X_{ijk})$ and indices A, AB and 0 are used for the deviance D as in Assumptions I. The log quasi-likelihood ratio test statistics for testing H_2 and H_1 are as follows:

$$D_A - D_{AB} = -2 \sum_{ijk} \log (X_{ijk}/\bar{X}_{i++}) + 2 \sum_{ijk} \log (X_{ijk}/\bar{X}_{ij+}) = -2 \sum_{ijk} \log (\bar{X}_{ij+}/\bar{X}_{i++}) =$$

$$= -2K \sum_{ij} \log (\bar{X}_{ij+}/\bar{X}_{i++});$$

$$D_0 - D_A = -2JK \sum_i \log (\bar{X}_{i++}/\bar{X}_{+++}).$$

For estimating σ_1^2, σ_2^2 and σ_3^2 let generalized Pearson statistics be defined as

$$X^2_{AB} = \sum_{ijk} (X_{ijk} - \bar{X}_{ij+})^2/\bar{X}^2_{ij+};$$

$$X^2_{A;AB} = K \sum_{ij} (\bar{X}_{ij+} - \bar{X}_{i++})^2/\bar{X}^2_{i++};$$

$$X^2_{0;A} = JK \sum_{i} (\bar{X}_{i++} - \bar{X}_{+++})^2/\bar{X}^2_{+++}.$$

2.4.1. Estimators

We shall present consistent estimators for the parameters σ^2_1, σ^2_2 and σ^2_3 in this section. For this purpose we need the generalized Pearson statistics X^2_{AB}, $X^2_{A;AB}$ and $X^2_{0;A}$ which were introduced at the beginning of this section. Also, additional assumptions will be made that certain moments of orders 3 and 4 are finite and these assumptions are assumed to hold wherever needed throughout this section.

Estimator of σ^2_3

As an estimator of σ^2_3, we propose

$\hat{\sigma}^2_3 = X^2_{AB}/\nu_3$, where $\nu_3 = IJ(K-1)$;

see also McCullagh (1983), p. 63.

Under the additional assumption that the distribution of X_{ijk} given $M_{ij}=m_{ij}$, for $i=1,2,\dots,I$, $j=1,2,\dots,J$, has finite moments of orders 3 and 4, the estimator $\hat{\sigma}^2_3$ is a consistent estimator of σ^2_3 for IJK tending to infinity. This will be proved in section 2.4.2.

Estimator of σ^2_2

Firstly, we propose the estimator

$\hat{\sigma}^2_{23} = X^2_{A;AB}/\nu_2$, where $\nu_2 = I(J-1)$

as an estimator of σ^2_{23}.

Under the additional assumption that the distribution of $\bar{X}_{ij+}$ given $M_i = m_i$ for $i=1,2,\dots,I$, has finite moments of orders 3 and 4, the estimator $\hat{\sigma}^2_{23}$ is a consistent estimator of σ^2_{23} for IJ tending to infinity; see section 2.4.2.

Secondly, the estimator of σ^2_2 is proposed as follows.

Under Assumptions I we derived in section 2.3.3 that $\sigma^2_2 = (\sigma^2_{23} - \sigma^2_3)/K$.

The estimator

$\hat{\sigma}_2^2 = (\hat{\sigma}_{23}^2 - \hat{\sigma}_3^2)/K$

is proposed as an estimator of σ_2^2 under Assumptions I. It is a consistent estimator of σ_2^2 for IJ tending to infinity.

Under Assumptions II we derived in section 2.3.4. that
$\sigma_2^2 = (\sigma_{23}^2 - \sigma_3^2)/(K + \sigma_3^2)$.
The estimator

$\hat{\sigma}_2^2 = (\hat{\sigma}_{23}^2 - \hat{\sigma}_3^2)/(K + \hat{\sigma}_3^2)$

is then proposed as an estimator of σ_2^2 under Assumptions II. It is a consistent estimator of σ_2^2 for IJ tending to infinity.

Estimator of σ_1^2

Firstly, we propose

$\hat{\sigma}_{123}^2 = X_{0;A}^2/\upsilon_1$, where $\upsilon_1 = I - 1$

as an estimator of σ_{123}^2. Under the additional assumption that the distribution of $\bar{X}_{i++}$ has finite moments of orders 3 and 4, the estimator $\hat{\sigma}_{123}^2$ is a consistent estimator of σ_{123}^2 for I tending to infinity (see section 2.4.2.).

Secondly, the estimator of σ_1^2 is proposed as follows.
Under Assumptions I we derived in section 2.3.3. that $\sigma_1^2 = (\sigma_{123}^2 - \sigma_{23}^2)/JK$.
The estimator

$\hat{\sigma}_1^2 = (\hat{\sigma}_{123}^2 - \hat{\sigma}_{23}^2)/JK$

is then proposed as an estimator of σ_1^2 under Assumptions I. It is a consistent estimator of σ_1^2 for I tending to infinity.

Under Assumptions II we derived in section 2.3.4. that
$\sigma_1^2 = (\sigma_{123}^2 - \sigma_{23}^2)/(JK + \sigma_{23}^2)$.

The estimator

$$\hat{\sigma}_1^2 = (\hat{\sigma}_{123}^2 - \hat{\sigma}_{23}^2)/(JK + \hat{\sigma}_{23}^2)$$

is proposed as an estimator of σ_1^2 under Assumptions II. It is a consistent estimator of σ_1^2 for I tending to infinity.

Unfortunately, the variances of the estimators $\hat{\sigma}_3^2$, $\hat{\sigma}_{23}^2$ and $\hat{\sigma}_{123}^2$ and thus the variances of the estimators $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ depend on fourth moments of the data (see also McCullagh and Nelder (1983), p. 173), which are assumed to be finite but which are unknown in general. Therefore, there seems to be no way to construct confidence intervals for the parameters σ_1^2, σ_2^2 and σ_3^2 so that only point estimators are available.

In section 2.4.2. a proof will be given for the results of section 2.4.1. as regards the consistency of the estimators $\hat{\sigma}_3^2$, $\hat{\sigma}_{23}^2$ and $\hat{\sigma}_{123}^2$.

2.4.2. Proof of the consistency of the estimators $\hat{\sigma}_3^2$, $\hat{\sigma}_{23}^2$ and $\hat{\sigma}_{123}^2$.

We shall prove that $\hat{\sigma}_3^2$ is a consistent estimator of σ_3^2 for IJK → ∞ under Assumptions I. The consistency under Assumptions II and the consistency of the estimators $\hat{\sigma}_{23}^2$ and $\hat{\sigma}_{123}^2$ under Assumptions I and II can be proved in a similar way.

Under Assumptions I, the estimator $\hat{\sigma}_3^2$ equals

$$\hat{\sigma}_3^2 = X_{AB}^2/\upsilon_3 = \{\sum_{ijk} (X_{ijk} - \bar{X}_{ij+})^2/\bar{X}_{ij+}\}/\upsilon_3, \text{ where } \upsilon_3 = IJ(K-1).$$

Firstly, we shall consider the distribution of X_{ijk} given $M_{ij}=m_{ij}$, for $i=1,2,\ldots,I$, $j=1,2,\ldots,J$.

The assumptions I imply that $E(X_{ijk}|M_{ij}=m_{ij}) = m_{ij}$ and $\text{var}(X_{ijk}|M_{ij}=m_{ij}) = \sigma_3^2 m_{ij}$; given $M_{ij}=m_{ij}$, the random variables X_{ijk}, $k=1,2,\ldots,K$ are independent (see section 2.2).

We introduce $Q_{AB}(m_{ij}) := \sum_k (X_{ijk} - m_{ij})^2/m_{ij}$ as a useful quadratic form; X_{AB}^2 is obtained from $Q_{AB}(m_{ij})$ as $X_{AB}^2 = \sum_{ij} Q_{AB}(\bar{X}_{ij+})$.

It can be proved that the expected value of $Q_{AB}(\bar{X}_{ij+}) = \sum_k (X_{ijk}-\bar{X}_{ij+})^2/\bar{X}_{ij+}$ equals $(K-1)\sigma_3^2$ so that

$$E(\sum_{ij} Q_{AB}(\bar{X}_{ij+})) = IJ\ (K-1)\sigma_3^2. \qquad \ldots (1)$$

From standard results on variances of quadratic forms (see e.g. Seber (1977), theorem 1.8) it follows that the variance of $Q_{AB}(m_{ij})$ equals

$$\mathrm{var}(Q_{AB}(m_{ij})) = (\mu_{4;ij} - 3\mu_{2;ij}^2)m_{ij}^{-2}K + 2\mu_{2;ij}^2 m_{ij}^{-2}K = (\mu_{4;ij}-\mu_{2;ij}^2)m_{ij}^{-2}K$$

where $\mu_{4;ij}$ and $\mu_{2;ij}$ are finite fourth and second moments of X_{ijk} about its mean m_{ij} given $M_{ij} = m_{ij}$.
Further,

$$\mathrm{var}(\sum_{ij} Q_{AB}(m_{ij})/IJ\ (K-1)) = O((IJK)^{-1}) \text{ for } IJK \to \infty$$

so this variance tends to zero for $IJK \to \infty$.

Also,

$$\mathrm{var}(\sum_{ij} Q_{AB}(\bar{X}_{ij+})/IJ\ (K-1)) \text{ tends to zero for } IJK \to \infty. \qquad \ldots (2)$$

Combining the results (1) and (2) it follows that

$$\hat{\sigma}_3^2 = X_{AB}^2/IJ\ (K-1) = \sum_{ij} Q_{AB}(\bar{X}_{ij+})/IJ\ (K-1)$$

is a consistent estimator of σ_3^2 for $IJK \to \infty$.

Secondly, the estimator $\hat{\sigma}_3^2$ is a consistent estimator of σ_3^2 also unconditionally (not given $M_{ij} = m_{ij}$ for $i=1,2,\ldots, I$ and $j=1,2,\ldots, J$), as the (degenerate) limiting distribution of $\hat{\sigma}_3^2$ for $IJK \to \infty$ given $M_{ij} = m_{ij}$ does not depend on m_{ij}.
The proof has now been completed.

2.4.3. Test statistics

We shall present test statistics for testing the hypotheses H_1: $\sigma_1^2 = 0$ and H_2: $\sigma_2^2 = 0$. To obtain the asymptotic distribution of these test statistics, the following three properties of means of counts X_{ijk} are very useful.

Property 1

- Given $M_i=m_i$, $i=1,2,\dots,I$, and under H_2, the distribution of $\bar{X}_{ij+}$, $j=1,2,\dots,J$, tends to the normal distribution for K tending to infinity. This result follows from the Central Limit Theorem applied to the mean $\bar{X}_{ij+}$ of (under H_2) independent X_{ijk}, $k=1,2,\dots,K$.

  We recall that, given $M_i=m_i$, $\bar{X}_{ij+}$, $j=1,2,\dots,J$, are independent random variables; see section 2.2.

Property 2

- Under both H_1 and H_2, the distribution of $\bar{X}_{ij+}$, for $i=1,2,\dots,I$, $j=1,2,\dots,J$, tends to the normal distribution for K tending to infinity. This follows, again, from the CLT applied to the mean $\bar{X}_{ij+}$ of (under H_1 and H_2) independent X_{ijk}, $k=1,2,\dots,K$.

  Note that under H_1 the random variables X_{ijk}, $k=1,2,\dots,K$, are dependent if H_2 is not true; thus property 2 is not true under H_1 only.

Property 3

- Under H_1, the distribution of $\bar{X}_{i++}$, $i=1,2,\dots,I$, tends to the normal distribution for J tending to infinity.
  This follows from the CLT applied to the mean $\bar{X}_{i++}$ of (under H_1) independent $\bar{X}_{ij+}$, $j=1,2,\dots,J$.

We have derived three basic properties, and now we shall present statistics for testing H_1 and H_2 and investigate their approximate distributions. The results will be stated without proof in this section. They will be proved in section 2.4.4.

Testing the hypothesis H_2: $\sigma_2^2 = 0$

Hypotheses equivalent to H_2 are

H_2': $\sigma_{23}^2 = \sigma_3^2$;

H_2'': $M_{ij} \equiv M_i$, $i=1,2,\dots,I$, $j=1,2,\dots,J$.

Using the result that $\mathrm{var}(\bar{X}_{ij+} | M_i = m_i) = K^{-1}\ \sigma_{23}^2\ m_i^r$, where $r=1$ under Assumptions I and $r=2$ under Assumptions II, we shall prove in section 2.4.4. that under H_2

$$D_A - D_{AB} \to \sigma_3^2\ \chi^2_{\upsilon_2} \text{ in distribution for } K \to \infty, \qquad \dots (3)$$

where $\upsilon_2 = I\ (J-1)$.

From section 2.4.1. we use the estimator $\hat{\sigma}_3^2 = X^2_{AB}/\upsilon_3$ with $\upsilon_3 = IJ\ (K-1)$, which is a consistent estimator of σ_3^2 for $K \to \infty$, to obtain the following test statistic T_2 for H_2:

$$T_2 = (D_A - D_{AB})/\hat{\sigma}_3^2.$$

Under H_2, the asymptotic distribution for $K \to \infty$ of T_2 is the distribution of the $\chi^2_{\upsilon_2}$- statistic, with $\upsilon_2 = I\ (J-1)$.

Testing the overall hypothesis H_1 and H_2: $\sigma_1^2 = \sigma_2^2 = 0$

A hypothesis which is equivalent to H_1 and H_2 is

$(H_1 \text{ and } H_2)'$: $\sigma_3^2 = \sigma_{23}^2 = \sigma_{123}^2$.

We shall test H_1 and H_2 against the alternative K_1: $\sigma_1^2 \neq 0$. An interpretation is: testing hypothesis H_1: $\sigma_1^2=0$ if H_2: $\sigma_2^2 = 0$ was tested and was not rejected.

By the result $\mathrm{var}(\bar{X}_{i++}) = (JK)^{-1}\, \sigma^2_{123}\mu^r$, where r=1 under Assumptions I and r = 2 under Assumptions II, we shall prove in section 2.4.4. that under H_1 and H_2

$$D_0 - D_A \to \sigma_3^2\, \chi^2_{\upsilon 1} \text{ in distribution for } K \to \infty, \qquad \dots (4)$$

where $\upsilon_1 = I-1$.

By combining the results (3) and (4) and by using the asymptotic independence for $K \to \infty$ of $D_0 - D_A$ and $D_A - D_{AB}$ (see section 2.4.4.) we shall prove that under H_1 and H_2 the asymptotic distribution for $K \to \infty$ of the test statistic T_{12} defined as

$$T_{12} = (D_0 - D_A)/(D_A - D_{AB}) * \upsilon_2/\upsilon_1$$

is the distribution of the $F_{\upsilon_1,\upsilon_2}$-statistic, $\upsilon_1 = I-1$, $\upsilon_2 = I(J-1)$. ... (5)

We shall make some remarks on this testing procedure.

Under H_1 and H_2, the random variables X_{ijk}, $i=1,2,\dots,I$, $j=1,2,\dots,J$, $k=1,2,\dots,K$, are independent, where $E(X_{ijk}) = \mu$ and $\mathrm{var}(X_{ijk}) = \sigma_3^2\, \mu^r$, where r=1 and r=2 under Assumptions I and II, respectively. An alternative test statistic for H_1 and H_2 against the alternative hypothesis $K_1 : \sigma_1^2 \neq 0$ can be proposed as

$$\tilde{T}_{12} = (D_0 - D_A)/\tilde{\sigma}_3^2$$

where $\tilde{\sigma}_3^2$ is some consistent estimator of σ_3^2 for $K \to \infty$, such as $\hat{\sigma}_3^2$. However, $\tilde{T}_{12}$ is quite sensitive to the alternative hypothesis $K_2 : \sigma_2^2 \neq 0$, and we may expect T_{12} to be much less sensitive to this alternative hypothesis so that T_{12} is preferred to $\tilde{T}_{12}$. There is a similarity with the analysis of variance for nested designs and normal data, where the test statistic corresponding to T_{12} is only sensitive to the alternative $\sigma_1^2 \neq 0$. Note that the distribution of T_{12} under the hypothesis $H_1 : \sigma_1^2 = 0$ alone is not necessarily the $F_{\upsilon_1,\upsilon_2}$-distribution, as the distribution of $\bar{X}_{ij+}$, for $i=1,2,\dots,I$, $j=1,2,\dots,J$ is not necessarily normal for large K.

Testing the hypothesis $H_1 : \sigma_1^2 = 0$

Hypotheses which are equivalent to H_1 are

$H_1' : \sigma_{23}^2 = \sigma_{123}^2$;
$H_1'' : M_i \equiv \mu$, $i=1,2,\dots,I$.

From the result $\mathrm{var}(\bar{X}_{i++}) = (JK)^{-1} \sigma_{123}^2 \mu^r$, where r=1 and r=2 under Assumptions I and II, respectively, it will be proved in section 2.4.4 that under hypothesis H_1

$$D_0 - D_A \to \sigma_{23}^2 \chi_{\upsilon_1}^2 \text{ in distribution for } J \to \infty, \qquad \dots (6)$$

where $\upsilon_1 = I-1$.

Introducing the estimator $\hat{\sigma}_{23}^2 = X_{A;AB}^2/\upsilon_2$ from 2.4.1 where $\upsilon_2 = I(J-1)$, which is a consistent estimator of σ_{23}^2 for $J \to \infty$, the following test statistic T_1 for H_1 is obtained:

$$T_1 = (D_0 - D_A)/\hat{\sigma}_{23}^2.$$

Under H_1, the asymptotic distribution for $J \to \infty$ of T_1 is the distribution of the $\chi_{\upsilon_1}^2$-statistic, with $\upsilon_1 = I-1$.

2.4.4. Proof of the results (3), (4), (5) and (6) of section 2.4.3

In this section we shall prove:

I. the result (3) concerning the limiting distribution under H_2 of the statistic $D_A - D_{AB}$ for $K \to \infty$;

II. the result (5) concerning the limiting distribution under H_1 and H_2 of the test statistic
$T_{12} = (D_0 - D_A)/(D_A - D_{AB}) * \upsilon_2/\upsilon_1$ for $K \to \infty$.

The proof of the results (4) and (6) is completely similar to the proof of the result (3) and we shall not give it here.
A reference is made to the proof sketched by McCullagh (1983), p. 62 concerning similar asymptotic results on log quasi-likelihood ratio test statistics.

I. Proof of the result (3)

In some steps we shall prove the result (3) under Assumptions I. The proof of the result (3) under Assumptions II is similar to the proof of this result under Assumptions I.

1. Firstly we shall condition on $M_i = m_i$; later this will be relaxed.

   Given $M_i=m_i$, $i=1,2,\dots,I$, the random variables $\bar{X}_{ij+}$, $j=1,2,\dots,J$, are independent, with $E(\bar{X}_{ij+}|M_i=m_i) = m_i$ and conditional variance function $\text{var}(X_{ij+}|M_i=m_i) = K^{-1}\,\sigma_3^2\, m_i$ as $\sigma_{23}^2 = \sigma_3^2$ under H_2; see section 2.3.3.
   From McCullagh (1983), p. 66 we find that the log quasi-likelihood function is $\bar{X}_{ij+}\log m_i - m_i$ for one sample mean $\bar{X}_{ij+}$ and

   $$\sum_{ij}[\bar{X}_{ij+}\log m_i - m_i]$$

   for the whole set of sample means. If we maximize this function with respect to m_i, the result is

   $$\sum_{ij}[\bar{X}_{ij+}\log \bar{X}_{i++} - \bar{X}_{i++}].$$

2. More generally, the log quasi-likelihood function for one sample mean $\bar{X}_{ij+}$ is $\bar{X}_{ij+}\log m_{ij} - m_{ij}$, and for the whole set it is

   $$\sum_{ij}[\bar{X}_{ij+}\log m_{ij} - m_{ij}].$$

   Maximizing this function with respect to m_{ij}, we obtain

   $$\sum_{ij}[\bar{X}_{ij+}\log \bar{X}_{ij+} - \bar{X}_{ij+}].$$

3. We take the difference of the two maximized log quasi-likelihoods of 2. and 1. and we multiply it by 2K, so that we obtain

$$2K \{\sum_{ij} [\bar{X}_{ij+} \log \bar{X}_{ij+} - \bar{X}_{ij+}] - \sum_{ij} [\bar{X}_{ij+} \log \bar{X}_{i++} - \bar{X}_{i++}]\} =$$

$$= -2K \{\sum_{ij} [\bar{X}_{ij+} \log \bar{X}_{i++} - \bar{X}_{i++}] - \sum_{ij} [\bar{X}_{ij+} \log \bar{X}_{ij+} - \bar{X}_{ij+}]\}.$$

In steps 4. and 5. it will be proved that under H_2 and for $K \to \infty$ the distribution of the difference just obtained tends to the distribution of the $\sigma_3^2 \chi^2_{\upsilon_2}$ statistic, where $\upsilon_2 = I\ (J-1)$.

4. Again, consider the log quasi-likelihood function

$$l(m) = \sum_{ij} [\bar{X}_{ij+} \log m_{ij} - m_{ij}].$$

This function is a function of the vector $m := \{m_{ij}\}_{ij}$.
By $l(\hat{m})$ we shall denote this log quasi-likelihood function with $\hat{m}$ substituted for m, vector $\hat{m}$ being defined by
$\hat{m} = \{\hat{m}_{ij}\}_{ij}$ and $\hat{m}_{ij} = \bar{X}_{ij+}$, $i=1,2,\ldots,I$, $j=1,2,\ldots,J$.

We also need the Fisher information matrix. From $l(m)$ we obtain the observed information matrix $I_m := -\delta^2 l/\delta m^2$ as follows.
The vector of first derivatives of $l(m)$ is the vector $\delta l/\delta m$ with IJ components $\bar{X}_{ij+}/m_{ij} - 1$.
The matrix of second derivatives is then a diagonal matrix with elements $-\bar{X}_{ij+}/m^2_{ij}$ on the diagonal, so that the observed information matrix I_m is diagonal with elements $\bar{X}_{ij+}/m^2_{ij}$. The expected information matrix $E(I_m)$ is then a diagonal matrix with elements $1/m_{ij}$; we shall denote this matrix by i_m.

Next, we expand $l(m)$ in a Taylor series about $\hat{m}$ up to terms of second order; in vector notation, the vector of first derivatives being zero:

$$l(m) - l(\hat{m}) = -1/2\ (\hat{m}-m)^T\ I_{m*}(\hat{m}-m).$$

Here I_{m*} is the observed diagonal information matrix with diagonal elements $\bar{X}_{ij+}/m^{*2}_{ij}$, where m^{*}_{ij} is on a line segment joining m_{ij} and $\hat{m}_{ij} = \bar{X}_{ij+}$.

Note that I_{m*} is a consistent estimator of i_m for $K \to \infty$, because $\bar{X}_{ij+}$ is a consistent estimator of m_{ij}. Incidentally, in sum notation, the above equation is written as

$$\sum_{ij}\left[\bar{X}_{ij+}\log m_{ij}-m_{ij}\right] - \sum_{ij}\left[\bar{X}_{ij+}\log \bar{X}_{ij+}- \bar{X}_{ij+}\right] =$$

$$= -1/2 \sum_{ij} (\bar{X}_{ij+}- m_{ij})^2\ \bar{X}_{ij+}/m^{*}_{ij}{}^{2}.$$

To get further, we multiply both sides of the equation in vector notation by -2K:

$$-2K\ (l(m) - l(\hat{m})) = \surd K\ (\hat{m}-m)^T\ I_{m*}\ \surd K\ (\hat{m}-m).$$

Under H_2 and for $K \to \infty$ the distribution of vector $\surd K\ (\hat{m}-m)$ tends to the IJ-variate normal distribution with mean vector 0 and diagonal covariance matrix $\sigma_3^2\ i_m^{-1}$ with diagonal elements $\sigma_3^2\ m_{ij}$, $i=1,2,\ldots,I$, $j=1,2,\ldots,J$ (here property 1 is relevant). Then the distribution of the statistic

$$-2K\ \left[l(m) - l(\hat{m})\right] = \surd K\ (\hat{m}-m)^T\ I_{m*}\ \surd K\ (\hat{m}-m)$$

or, in sum notation, the statistic

$$-2K\ \{\sum_{ij}\left[\bar{X}_{ij+}\log m_{ij}- m_{ij}\right] - \sum_{ij}\left[\bar{X}_{ij+}\log \bar{X}_{ij+}- \bar{X}_{ij+}\right]\}$$

tends to the distribution of the $\sigma_3^2\ \chi_\upsilon^2$ statistic with $\upsilon=IJ$ for $K \to \infty$, because I_{m*} is a consistent estimator of i_m for $K \to \infty$.

By arguments similar to those of McCullagh the result can be proved that under H_2 and for $K \to \infty$ the distribution of the statistic (see 3.)

$$-2K\ \{\sum_{ij}[\bar{X}_{ij+}\log\bar{X}_{i++}-\bar{X}_{i++}]-\sum_{ij}[\bar{X}_{ij+}\log\bar{X}_{ij+}-\bar{X}_{ij+}]\}$$

tends to the distribution of the $\sigma_3^2\ \chi^2_{\upsilon_2}$ statistic with $\upsilon_2=I(J-1)$. Here $\bar{X}_{i++}$ is substituted for m_{ij} because $m_{ij}=m_i$, being the conditional expectation of $\bar{X}_{ij+}$ given $M_i=m_i$, estimating m_i by $\bar{X}_{i++}$ for $i=1,2,\ldots,I$.

Further, this result is also true unconditionally (so not given $M_i=m_i$ for $i=1,2,\ldots,I$) as this limiting distribution does not depend on m_i.

5. Finally, the statistic D_A-D_{AB} which was introduced for testing the hypothesis H_2 in section 2.4.3. was defined before as

$$D_A-D_{AB}=2K\sum_{ij}\bar{X}_{ij+}\log(\bar{X}_{ij+}/\bar{X}_{i++})$$

and this statistic can also be written as

$$D_A-D_{AB}=-2K\ \{\sum_{ij}[\bar{X}_{ij+}\log\bar{X}_{i++}-\bar{X}_{i++}]-\sum_{ij}[\bar{X}_{ij+}\log\bar{X}_{ij+}-\bar{X}_{ij+}]\}.$$

Here we recognize the above difference of log quasi-likelihoods of steps 3 and 4 for which we proved that under H_2 and for $K\to\infty$ the distribution tends to the distribution of the $\sigma_3^2\ \chi^2_{\upsilon 2}$ statistic with $\upsilon_2=I\ (J-1)$. Now we have proved the result (3).

II. Proof of the result (5)

To prove the result (5) it is sufficient to prove the asymptotic independence under H_1 and H_2 for $K\to\infty$ of the statistics D_0-D_A and D_A-D_{AB}. Under Assumptions I (the proof is similar under Assumptions II) following a similar reasoning as in the proof of the result (3), a Taylor series expansion is obtained of D_0-D_A and of D_A-D_{AB}. So

$$D_0-D_A=2JK\sum_i\bar{X}_{i++}\log(\bar{X}_{i++}/\bar{X}_{+++})=JK\sum_i(\bar{X}_{i++}-\bar{X}_{+++})^2\ \bar{X}_{i++}/\mu_i^{*2},\qquad\ldots(7)$$

where μ_i^* is on a line segment joining $\bar{X}_{i++}$ and $\bar{X}_{+++}$. Note that $\bar{X}_{i++}/\mu_i^{*2}$ is a consistent estimator of μ^{-1} for $K \to \infty$. Further

$$D_A - D_{AB} = 2K \sum_{ij} \bar{X}_{ij+} \log (\bar{X}_{ij+}/\bar{X}_{i++}) = K \sum_{ij} (\bar{X}_{ij+} - \bar{X}_{i++})^2 \, \bar{X}_{ij+}/\mu_{ij}^{*2}, \quad \dots (8)$$

where μ_{ij}^* is on a line segment joining $\bar{X}_{ij+}$ and $\bar{X}_{i++}$. Here $\bar{X}_{ij+}/\mu_{ij}^{*2}$ is a consistent estimator of μ^{-1} for $K \to \infty$.

Under H_1 and H_2 the random variables X_{ijk}, $i=1,2,\dots,I$, $j=1,2,\dots,J$, $k=1,2,\dots,K$, are independent, with $E(X_{ijk}) = \mu$, $\mathrm{var}(X_{ijk}) = \sigma_3^2\mu$. The vector with elements $\sqrt{K}\,(\bar{X}_{i++} - \bar{X}_{+++})$ and the vector with elements $\sqrt{K}\,(\bar{X}_{ij+} - \bar{X}_{i++})$ are uncorrelated vectors in orthogonal subspaces of the IJK-dimensional Euclidean space. Asymptotically for $K \to \infty$ these vectors have a joint multivariate normal distribution so that, asymptotically for $K \to \infty$, these vectors are independent random vectors. Then the right-hand sides of the equalities (7) and (8) are independent, asymptotically for $K \to \infty$ (see below) and so are the left-hand sides, which was to be proved.
To see the stated asymptotic independence of the right-hand sides of (7) and (8), write

$$D_0 - D_A = 2JK \sum_i \bar{X}_{i++} \log (\bar{X}_{i++}/\bar{X}_{+++}) = JK \sum_i (\bar{X}_{i++} - \bar{X}_{+++})^2 \, \bar{X}_{i++}/\mu_i^{*2} =$$

$$= JK \sum_i (\bar{X}_{i++} - \bar{X}_{+++})^2 \, (\bar{X}_{i++}/\mu_i^{*2} - \mu^{-1}) + JK \sum_i (\bar{X}_{i++} - \bar{X}_{+++})^2 \, \mu^{-1} \quad \dots (9)$$

and

$$D_A - D_{AB} = 2K \sum_{ij} \bar{X}_{ij+} \log (\bar{X}_{ij+}/\bar{X}_{i++}) = K \sum_{ij} (\bar{X}_{ij+} - \bar{X}_{i++})^2 \, \bar{X}_{ij+}/\mu_{ij}^{*2} =$$

$$= K \sum_{ij} (\bar{X}_{ij+} - \bar{X}_{i++})^2 \, (\bar{X}_{ij+}/\mu_{ij}^{*2} - \mu^{-1}) + K \sum_{ij} (\bar{X}_{ij+} - \bar{X}_{i++})^2 \, \mu^{-1}. \quad \dots (10)$$

The result to be proved follows from the asymptotic independence for $K \to \infty$ of the latter terms on the right-hand sides of (9) and (10) and the convergence to zero in probability for $K \to \infty$ of the former terms.

3. THE QUASI-LIKELIHOOD METHOD FOR NESTED DESIGNS UNDER THE ASSUMPTION OF A PROBABILITY MODEL

3.1. Introduction

At the moment, two probability models for nested designs with random factors for count data are known. The problem of data analysis by these models is only partly solved. In this section we shall consider the use of a quasi-likelihood approach when these models can be regarded as reasonable for the data. For both models we shall try to verify the quasi-likelihood assumptions.

3.2. Two models for the data

3.2.1. Model 1

The first model, for two random nested factors and a Poisson distribution for errors, was communicated by Forcina (1984). As in section 2.2, the model is defined in three stages, as follows, using the notation from 2.2.

i) Let M_{ij} be positive random variables for $i=1,2,\dots,I$, $j=1,2,\dots,J$.
Given $M_{ij}=m_{ij}$, $i=1,2,\dots,I$, $j=1,2,\dots,J$, the random variable X_{ij} has the Poisson (m_{ij}) distribution.

ii) Let M_i be positive random variables for $i=1,2,\dots,I$.
Given $M_i=m_i$, $i=1,2,\dots,I$, let $M_{ij}= G_{ij}$, $j=1,2,\dots,J$, where G_{ij} are independent random variables having a gamma distribution with
$E(G_{ij}) = m_i$, $\operatorname{var}(G_{ij}) = \alpha^{-1} m_i^2$, $\alpha > 0$.

Then given $M_i=m_i$ the random variables X_{ij}, $j=1,2,\dots,J$, are independent, having the negative binomial distribution with parameters α and $p_i = m_i/(\alpha + m_i)$.

iii) Random variables $P_i = M_i/(\alpha + M_i)$, $i=1,2,\dots,I$, are i.i.d. random variables having the beta (β_1, β_2) distribution.

The (marginal) distribution of X_{ij}, $i=1,2,\dots,I$, $j=1,2,\dots,J$, is called the generalized hypergeometric distribution with parameters α, β_1 and β_2;

see Sibuya c.s.(1964). Extensions to designs with more than two factors do not seem to be available. Note that this model is defined for only one replicate per cell.

3.2.2. Model 2

The second model is defined as follows:

i) Let M_{ijk} be positive random variables for $i=1,2,\ldots,I$, $j=1,2,\ldots,J$, $k=1,2,\ldots,K$.
Given $M_{ijk} = m_{ijk}$, random variables X_{ijk} are independent, having the Poisson (m_{ijk}) distribution.

ii) $M_{ijk} = \theta\, F_i(\alpha_1)\, F_{j(i)}(\alpha_2)\, F_{k(ij)}(\alpha_3)$, where the random variables $F_i(\alpha_1)$, $F_{j(i)}(\alpha_2)$ and $F_{k(ij)}(\alpha_3)$ are positive random variables with parameters α_1, α_2 and α_3, representing factorial effects of the two factors and error. This model type was proposed in chapter 7.
Extensions are possible for any number of factors. The lognormal random variable was proposed as a choice for F; the variables $F_i(\alpha_1)$, $F_{j(i)}(\alpha_2)$ and $F_{k(ij)}(\alpha_3)$ are then independent and Model 2 with this lognormal assumption will be considered in this section. For large θ,

$$X_{ijk} = \theta\, F_i(\alpha_1)\, F_{j(i)}(\alpha_2)\, F_{k(ij)}(\alpha_3) \qquad \ldots (11)$$

approximately.

3.3. An attempt to verify the quasi-likelihood assumptions

We shall try to verify the quasi-likelihood assumptions for Model 1 and Model 2. Firstly, some helpful results will be obtained for Model 1.

Model 1

1. From 3.2.1, iii) it is seen that random variables M_i, $i=1,2,\ldots,I$, are independent random variables, where

$E(M_i) = \mu$, say;

$$\mathrm{var}(M_i) = \frac{1}{1 + \beta_1 + \beta_2} (\mu^3/\alpha + 2\mu^2 + \alpha\mu), \text{ roughly, see later.}$$

2. From 3.2.1., ii) it is seen that,
   given $M_i = m_i$, $i=1,2,\dots,I$, random variables M_{ij}, $j=1,2,\dots,J$, are independent random variables, where

   $$E(M_{ij} | M_i = m_i) = m_i;$$
   $$\mathrm{var}(M_{ij} | M_i = m_i) = \alpha^{-1} m_i^2.$$

3. Finally, from 3.2.1., i),
   given $M_{ij} = m_{ij}$, $i=1,2,\dots,I$, $j=1,2,\dots,J$, the random variables X_{ijk}, $k=1,2,\dots,K$ are independent random variables having the Poisson (m_{ij}) distribution. Here we have extended Model 1 to allow for K replicates per cell. Further,

   $$E(X_{ijk} | M_{ij} = m_{ij}) = m_{ij};$$
   $$\mathrm{var}(X_{ijk} | M_{ij} = m_{ij}) = m_{ij}.$$

The expression for $\mathrm{var}(M_i)$ is obtained to get an impression of this variance function. Define $Q_i = 1-P_i$, and use the following crude linear approximation

$$Q_i = \frac{\alpha}{\alpha + M_i} = \frac{\alpha}{\alpha + \mu} + (M_i - \mu)\left(\frac{-\alpha}{(\alpha + \mu)^2}\right), \text{ where } \mu = E(M_i),$$

so that

$$E(Q_i) = \frac{\alpha}{\alpha + \mu}\ ; \ \mathrm{var}(Q_i) = \frac{\alpha^2}{(\alpha + \mu)^4} \mathrm{var}(M_i), \text{ approximately.}$$

Together with

$\mathrm{var}(Q_i) = \dfrac{1}{1+\beta_1+\beta_2} E(Q_i)(1 - E(Q_i))$, where we recall the assumption that P_i has the beta (β_1,β_2) distribution, the result is obtained by equating the two expressions for $\mathrm{var}(Q_i)$, resulting in

$$\frac{\alpha^2}{(\alpha + \mu)^4} \quad \mathrm{var}(M_i) = \frac{1}{1+\beta_1+\beta_2} \quad (\frac{\alpha}{\alpha + \mu}) \; (\frac{\mu}{\alpha + \mu})$$

and by performing some simple algebra.

Next, we shall obtain similar results for Model 2.

<u>Model 2</u>

1. Random variables $M_i := \theta\ F_i$, $i=1,2,\ldots,I$, are independent random variables, where

   $E(M_i) = \theta$; $\mathrm{var}(M_i) = \alpha_1^{-1}\ \theta^2$

   because

   $E(F_i) = 1$; $\mathrm{var}(F_i) = \alpha_1^{-1}$; see chapter 7.

2. Given $F_i = f_i$, $i=1,2,\ldots,I$, or, equivalently, given $M_i=m_i$, $i=1,2,\ldots,I$, where $m_i = \theta f_i$, the variables $M_{ij} := \theta\ F_i\ F_{j(i)}$, $j=1,2,\ldots,J$ are independent random variables, where

   $E(M_{ij}|F_i=f_i) = \theta\ f_i=m_i$, as $E(F_{j(i)}) = 1$;
   $\mathrm{var}(M_{ij}|F_i=f_i) = \mathrm{var}(\theta\ F_i\ F_{j(i)}|F_i=f_i) = \alpha_2^{-1}\ (\theta\ f_i)^2 = \alpha_2^{-1}\ m_i^2$
   because $\mathrm{var}(F_{j(i)}) = \alpha_2^{-1}$.

3. Given $M_{ij}=m_{ij}$, $i=1,2,\ldots,I$, $j=1,2,\ldots,J$, the random variables X_{ijk}, $k=1,2,\ldots,K$, are independent random variables, where

   $E(X_{ijk}|M_{ij}=m_{ij}) = E(M_{ijk}|M_{ij}= m_{ij}) = m_{ij}$, as $E(F_{k(ij)}) = 1$.

Further, by lemma 2, section 2.3.1,

$$\mathrm{var}(X_{ijk}|M_{ij}=m_{ij}) = E_{M_{ijk}|m_{ij}} \mathrm{var}_{X_{ijk}|m_{ij}} (X_{ijk}|M_{ijk}) +$$

$$+ \mathrm{var}_{M_{ijk}|m_{ij}} E_{X_{ijk}|m_{ij}} (X_{ijk}|M_{ijk}) = m_{ij} + \alpha_3^{-1}\, m_{ij}^2 = \alpha_3^{-1}\, m_{ij}^2 ,$$

where the final equality holds, approximately, for large m_{ij}.

3.4. Verification results

For Model 1: It is seen that the relationship for $\mathrm{var}(M_i)$ is not of the required type in general, unless e.g. $\mu = \alpha$ approximately, so that $\mathrm{var}(M_i) \approx \frac{1}{1+\beta_1+\beta_2} (\mu^2 + 2\,\mu^2 + \mu^2) = C \cdot \mu^2$, assuming $\beta_1 + \beta_2$ not to depend on μ. However, mean-variance relationships are of both linear and quadratic type so that no quasi-likelihood approach can be used here.

For Model 2: Given $M_{ij} = m_{ij}$ the mean-variance relationship for X_{ijk} is quadratic, approximately, for large m_{ij}. The other mean-variance relationships are quadratic as well. Under the condition of large m_{ij}, the quasi-likelihood approach based on Assumptions II can be used for this model.

For uniform notation, we shall substitute σ_i^2 for α_i^{-1}, $i=1,2,3$. It was shown in chapter 7 that for large θ, random variable X_{ijk} is approximately distributed as $\theta\, F_i\, F_{j(i)}\, F_{k(ij)}$. In this case all mean-variance relationships are of quadratic type, and Assumptions II are satisfied. On the other hand, in section 3.5 it will be shown that Assumptions II are equivalent to a generalization of the above approximate model.

3.5. Assumptions II: an equivalent model

The equivalence of the quasi-likelihood Assumptions II to a model for X_{ijk} which is more general than the model (11), is stated in the following theorem.

Theorem 1

The following two statements are equivalent:

1. Random variables X_{ijk}, $i=1,2,\dots,I$, $j=1,2,\dots,J$, $k=1,2,\dots,K$, satisfy the quasi-likelihood Assumptions II.

2. Random variables X_{ijk}, $i=1,2,\dots,I$, $j=1,2,\dots,J$, $k=1,2,\dots,K$, can be expressed as follows:

$$X_{ijk} = \mu \, F_i \, F_{j(i)} \, F_{k(ij)}, \qquad \dots (12)$$

where

$E(F_i) = E(F_{j(i)}) = E(F_{k(ij)}) = 1$;
$\mathrm{var}(F_i) = \sigma_1^2$, $\mathrm{var}(F_{j(i)}) = \sigma_2^2$, $\mathrm{var}(F_{k(ij)}) = \sigma_3^2$;
variables F_i, $i=1,2,\dots,I$, are independent random variables;
given F_i, the random variables $F_{j(i)}$, $j=1,2,\dots,J$, are independent random variables;
given $F_i F_{j(i)}$, the random variables $F_{k(ij)}$, $k=1,2,\dots,K$, are independent random variables.

Note that it is allowed that F_i, $F_{j(i)}$ and $F_{k(ij)}$ are dependent random variables.
Also note that the approximate version (11) for large θ of Model 2 (see section 3.2.2) is a special case of (12). Essentially, there are three differences between the models (11) and (12):

1) The model (11) has independent components, whereas the components of model (12) can be dependent.
2) The components of model (11) have lognormal distributions, whereas no distribution is assumed for the components of model (12).
3) The model (11) is for count data, the model (12) is for all data.

Proof of Theorem 1

I. The ← part of the equivalence is proved straightforward when defining

$M_{ij} := \mu\, F_i\, F_{j(i)}$ and $M_i := \mu\, F_i$.

Then the proof is as follows.

1) The random variables $M_i = \mu F_i$ are independent for $i=1,2,\ldots,I$, where

$E(M_i) = \mu E(F_i) = \mu;$

$\mathrm{var}(M_i) = \mu^2\ \mathrm{var}(F_i) = \mu^2\sigma_1^2.$

2) Given $M_i=m_i$, $i=1,2,\ldots,I$, or, equivalently, given $F_i=f_i$, $i=1,2,\ldots,I$, where $f_i=m_i/\mu$, the random variables $F_{j(i)}$ are independent random variables for $j=1,2,\ldots,J$, or, equivalently, the random variables M_{ij}, $j=1,2,\ldots,J$, are independent random variables. Also,

$E(M_{ij}|M_i=m_i) = m_i\ E(F_{j(i)}) = m_i;$

$\mathrm{var}(M_{ij}|M_i=m_i) = m_i^2\ \mathrm{var}(F_{j(i)}) = m_i^2\sigma_2^2.$

3) Given $M_{ij}=m_{ij}$ or, equivalently, given $F_i F_{j(i)} = m_{ij}/\mu$ for $i=1,2,\ldots,I$, $j=1,2,\ldots,J$, the random variables $F_{k(ij)}$, $k=1,2,\ldots,K$, are independent random variables, so that the random variables X_{ijk}, $k=1,2,\ldots,K$, are independent as well. Also,

$E(X_{ijk}|M_{ij}=m_{ij}) = m_{ij};$

$\mathrm{var}(X_{ijk}|M_{ij}=m_{ij}) = m_{ij}^2\sigma_3^2.$

II. The → part of the equivalence will be proved in five steps.

1) From the quasi-likelihood Assumptions II remember that, given $M_{ij} = m_{ij}$,

$E(X_{ijk}|M_{ij}= m_{ij}) = m_{ij}$ and $\text{var}(X_{ijk}|M_{ij}= m_{ij}) = \sigma_3^2\ m_{ij}^2$.

Then $E(X_{ijk}/M_{ij}|M_{ij}=m_{ij}) = E(X_{ijk}/m_{ij}|M_{ij}=m_{ij}) = 1$.
So given $M_{ij} = m_{ij}$, the expected value of X_{ijk}/M_{ij} does not depend on m_{ij}, so that $E(X_{ijk}/M_{ij}) = 1$, unconditionally.

In a similar way, $\text{var}(X_{ijk}/M_{ij}|M_{ij}=m_{ij})=\text{var}(X_{ijk}/m_{ij}|M_{ij}=m_{ij}) = \sigma_3^2$.
It appears that, given $M_{ij} = m_{ij}$, $\text{var}(X_{ijk}/M_{ij})$ does not depend on m_{ij}, so that $\text{var}(X_{ijk}/M_{ij}) = \sigma_3^2$, unconditionally.

Defining $F_{ijk} := X_{ijk}/M_{ij}$, we found that $E(F_{ijk})= 1$; $\text{var}(F_{ijk})= \sigma_3^2$.

Then

$X_{ijk} = M_{ij}\ F_{ijk}$ for all i,j and k,

where in general M_{ij} and F_{ijk} are dependent random variables.
However, given $M_{ij} = m_{ij}$, the random variables X_{ijk}, $k=1,2,\ldots,K$, are independent random variables, which follows from the quasi-likelihood assumptions.

2) By quasi-likelihood Assumptions II, in a similar way M_{ij} can be written as

$M_{ij} = M_i\ F_{ij}$ for all i and j,

with $E(F_{ij}) = 1$, $\text{var}(F_{ij}) = \sigma_2^2$, where in general M_i and F_{ij} are dependent random variables. However, given $M_i = m_i$, the random variables M_{ij}, $j=1,2,\ldots,J$, are independent.

3) Finally,

$M_i = \mu F_i$ for all i,

where F_i are independent random variables for all i, $E(F_i) = 1$, $var(F_i) = \sigma_1^2$.

4) Combining the results of 1), 2) and 3) it is seen that X_{ijk} can be expressed as

$X_{ijk} = \mu F_i F_{ij} F_{ijk}$ for all i, j and k

or, in a usual notation for nested designs,

$X_{ijk} = \mu F_i F_{j(i)} F_{k(ij)}$.

Finally, it needs to be proved that F_i, $F_{j(i)}$ and $F_{k(ij)}$ satisfy the statements in Theorem 1.

5) First, from 1), 2) and 3): $E(F_i) = E(F_{j(i)}) = E(F_{k(ij)}) = 1$ and $var(F_i) = \sigma_1^2$, $var(F_{j(i)}) = \sigma_2^2$, $var(F_{k(ij)}) = \sigma_3^2$.

Secondly, the variables F_i, i=1,2,...,I, are independent random variables (see 3));
given $F_i = f_i$, or given $M_i = \mu f_i$, i=1,2,...,I, the random variables $F_{j(i)}$, j=1,2,...,J, are independent (see 2) and 3));
given $F_i F_{j(i)}$, or given M_{ij}, i=1,2,...,I, j=1,2,...,J, the random variables X_{ijk}, k=1,2,...,K, are independent (see 1), 2) and 3)), so that the random variables $F_{k(ij)}$, k=1,2,...,K, are independent as well (see 1)).

Now the proof of Theorem 1 is completed.

4. APPLICATION

For a set of count data (see table 2) classified by two random nested factors A (with 2 levels) and B (with 3 levels) and with 5 replicates per cell, parameters σ_i^2, i=1,2,3 will be estimated, and hypotheses of type H_i: $\sigma_i^2 = 0$, i=1,2, will be tested.

A	1			2		
B	1	2	3	1	2	3
	8	12	9	17	23	11
	8	15	10	13	25	16
	4	9	11	15	14	10
	9	13	9	11	20	10
	5	8	14	18	18	13

Table 2. Count data results from random nested design with factors A and B.

4.1. Estimates

Estimates of parameters σ_i^2 were calculated for the set of data from table 2 under the quasi-likelihood Assumptions I and II (see section 2.4.1), and under Model 2 of section 3.2 with lognormal components, assuming θ to be large. The mean-squares from table 4 were used as input to calculate the estimates under Model 2. See table 3 for the results.

Source of variation	A	B	Error	Total
Assumptions I	$\hat\sigma_1^2 = 1.158$	$\hat\sigma_2^2 = 0.725$	$\hat\sigma_3^2 = 0.640$	
Variance components	$\hat\mu\;\hat\sigma_1^2 = 14.59$	$\hat\mu\;\hat\sigma_2^2 = 9.14$	$\hat\mu\;\hat\sigma_3^2 = 8.06$	$\widehat{\mathrm{Var}}(X) = 31.79$
Assumptions II	$\hat\sigma_1^2 = 0.093$	$\hat\sigma_2^2 = 0.058$	$\hat\sigma_3^2 = 0.055$	
Variance components	$\hat\mu^2\hat\sigma_1^2 = 14.76$	$\hat\mu^2\hat\sigma_2^2 = 9.20$	$\hat\mu^2\hat\sigma_3^2 = 8.73$	$\widehat{\mathrm{Var}}(X) = 32.69$
Model 2	$\hat\sigma_1^2 = 0.100$	$\hat\sigma_2^2 = 0.064$	$\hat\sigma_3^2 = 0.060$	
Variance components	$\hat\theta^2\hat\sigma_1^2 = 13.64$	$\hat\theta^2\hat\sigma_2^2 = 8.73$	$\hat\theta^2\hat\sigma_3^2 = 8.18$	$\widehat{\mathrm{Var}}(X) = 30.55$

Table 3. Estimates of parameters σ_i^2, i=1,2,3, under Assumptions I and II and under Model 2 (with lognormal components) for the data from table 2; $\hat\mu = 12.60$ (arithmetic mean); $\hat\theta = 11.68$ (geometric mean).

The conclusion is that factor B and the error component contribute about equally to the variance of X, and that factor A is slightly dominant over factor B. Note that the estimation results for variance components are quite similar for all three approaches, although Assumptions I and II are quite different with respect to the variance functions. Also note that $\hat{\sigma}_3^2$ is less than 1 under Assumptions I, although no underdispersion appears from the test results of section 4.2.

4.2. Test results

Under Assumptions I and II approximate χ^2-tests and F-tests (see section 2.4.3.) were computed, and approximate F-tests were obtained under Model 2. The results are presented in table 4.

Assumptions I

Deviance	Deviance result	df	χ^2-test H_2	df	F-test H_1	df
D_{AB}	15.58	24	26.62	4	5.07	(1,4)
D_A	32.64	28				
D_0	54.28	29				

Assumptions II

Deviance	Deviance result	df	χ^2-test H_2	df	F-test H_1	df
D_{AB}	1.397	24	25.34	4	4.99	(1,4)
D_A	2.801	28				
D_0	4.552	29				

Model 2 (with lognormal assumption)

Factors	Mean-square	df	F-test H_2	df	F-test H_1	df
error	0.060	24	6.34	(4,24)	4.94	(1,4)
B	0.379	4				
A	1.872	1				

Table 4. Test results for the data from table 2.

The uniform conclusion from the test results of table 4 is that the effect of factor B is significant at the 5% level and that the effect of factor A is significant at the 10% level.

4.3. Some verification of assumptions

From the K replicates available per cell, we can verify part 3 of Assumptions I:

- given $M_{ij}=m_{ij}$, i=1,2,...,I, j=1,2,...,J, the random variables X_{ijk}, k=1,2,...,K, are independent, with $E(X_{ijk}) = m_{ij}$ and $\mathrm{var}(X_{ijk}) = \sigma_3^2 m_{ij}$

to some extent by plotting standardized residuals $(X_{ijk} - \bar{X}_{ij+})/\sqrt{\bar{X}_{ij+}}$ versus $\bar{X}_{ij+}$; see figure 1. An increasing range (see table 5) of these residuals with increasing $\bar{X}_{ij+}$ may indicate that Assumptions I are violated. From figure 1 and table 5 no tendency for increasing ranges appears for the data from table 2, so that Assumptions I seem not to be violated.

Factor A	1			2		
Factor B	1	2	3	1	2	3
	0.46 0.46 -1.07 0.84 -0.69	0.18 1.07 -0.71 0.47 -1.01	-0.49 -0.18 0.12 -0.49 1.04	0.68 -0.55 0.06 -1.17 0.98	0.67 1.12 -1.34 0 -0.45	-0.29 1.15 -0.58 -0.58 0.29
Range	1.91	2.08	1.53	2.15	2.46	1.73
$\bar{X}_{ij+}$	6.80	11.40	10.60	14.80	20.00	12.00

Table 5. Standardized residuals $(X_{ijk} - \bar{X}_{ij+})/\sqrt{\bar{X}_{ij+}}$ and their ranges.

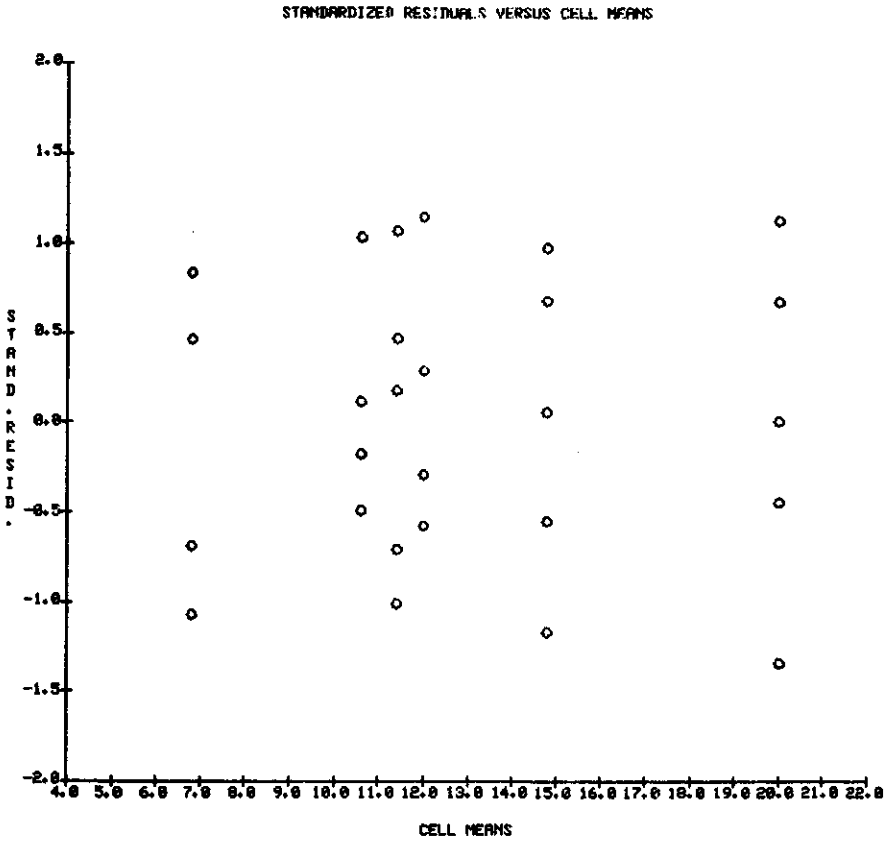


Figure 1. Standardized residuals $(X_{ijk} - \bar{X}_{ij+})/\sqrt{\bar{X}_{ij+}}$ plotted versus cell means $\bar{X}_{ij+}$.

5. DISCUSSION

A proposal was made for the analysis of count data from nested designs with random factors by the quasi-likelihood method. The quasi-likelihood method has an advantage over methods based on likelihood, which is that no distributional assumptions have to be made for the data to base the analysis on a likelihood function, as often such assumptions cannot be justified. We have derived asymptotic results for estimators and test statistics for large numbers of replicates.

From the application it appears that there are no large differences between estimated values of variance components calculated under Assumptions I and II and under Model 2 with the additional assumption of lognormality; the same conclusions were also drawn from test results of hypotheses. Some robustness seems to be present against improper choices of assumptions (see also McCullagh and Nelder (1984), p. 132).

Several extensions of the quasi-likelihood method just presented are still needed for practice. Really straightforward is the extension of the method of section 2 to random designs with more than two nested factors by extending the quasi-likelihood assumptions to more than three levels of variation. Less evident may be a treatment of unequal numbers of factorial levels and replicates, and a treatment of covariates 'explaining' part of the variation in the data. Also, a further comparison of Assumptions I and II by theory and practice is of interest. Certain optimality results could possibly be derived for the estimators of variance components under the quasi-likelihood assumptions, possibly within some restricted class of estimators. Finally, the quasi-likelihood approach for random designs with crossed factors needs some research.

REFERENCES

Breslow, N.E. (1984), Extra-Poisson variation in log-linear models, Appl. Statist. 33, 38 - 44.

Cox, D.R. (1984), Interaction, Intern. Statist. Review 52, 1 - 32.

Engel, J. (1984), Models for reponse data showing extra-Poisson variation, Statistica Neerlandica 38, 159 - 167.

Forcina, A. (1984), Estimating variance components with discrete data, personal communication.

McCullagh, P. (1983), Quasi-likelihood functions, Ann. Statist. 11, 59 - 67.

McCullagh, P. and Nelder, J.A. (1984), Generalized linear models, Chapman and Hall, London.

Nelder, J.A. and Wedderburn, R.W.M. (1972), Generalized linear models, J.R. Statist. Soc. A., 135, 370 - 384.

Scheffé, H. (1959), The analysis of variance, Wiley, New York.

Seber, G.A.F. (1977), Linear regression analysis, John Wiley, New York.

Sibuya, M., Yoshimura, I. and Shimizu, R. (1964), Negative multinomial distribution, Ann. Inst. Statist. Math. 16, 409 - 426.

Wedderburn, R.W.M. (1974), Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method, Biometrika 61, 439 - 447.

SUMMARY

In the literature, methods have been presented for the analysis of count data classified by fixed and crossed factors under the assumptions that this data can be modeled by independent binomial or Poisson distributions. In general, the mean value of these distributions depends on the levels of the classifying factors and a linear model is proposed for the logit transform or the log transform of these mean values.
In practice many situations occur which are different, such as:

- The counts are independent, but the observed variation in the data is more than can be explained by e.g. the Poisson distribution;

- The counts are dependent: the factors are not fixed but they are random.

For these situations no general analysis methods are available, and there is a strong need for extensions of the theory. In this thesis extensions of the theory will be presented to allow for the modeling of this count data.

In chapters 2, 3 and 4 of this thesis the situation is considered of overdispersion with respect to the binomial distribution and the Poisson distribution. In the case of overdispersion we may observe from the data that $\mathrm{var}(X) = \sigma^2 E(X)$ with $\sigma^2 > 1$, instead of $\mathrm{var}(X) = E(X)$ for the Poisson distribution. In chapter 2 we propose the beta-binomial distribution for modeling the overdispersed data, and limiting results for test statistics will be obtained for a large number of trials at each cell in the design.
A gamma-Poisson or negative binomial model is proposed for modeling overdispersed count data in the 3th and 4th chapter of this thesis. Here we obtain approximate distributions of test statistics for a large number of replicates and for large counts as well. In chapters 2, 3 and 4 the limiting results are obtained for standard test statistics known from the theory of loglinear and logitlinear models, like Pearson's X^2 statistic.

Chapter 5 deals with dependent count data in a split-plot situation. Here a model is proposed to allow for this dependence of the data from the split-plot experiment. Two separate analyses will be performed, namely for the whole plot and for the sub-plot factors, imitating the general Anova approach. The basic models are the gamma-Poisson model and the Dirichlet-multinomial model.

Data obtained by a dependent classification of objects in two or more ordered classes, testing hypotheses concerning the probabilities corresponding to these classes is a problem met e.g. in the context of questionnaires. In chapter 6 we study the signed rank test of Wilcoxon in the situation of such a dependent classification. It appears that the limiting distribution of this test statistic, under a Dirichlet-multinomial model assumption for the data is the normal distribution; there is an extra parameter for the dependence of classification.

The two final chapters 7 and 8 of this thesis deal with random factor problems for crossed and for nested designs (chapter 7) and for nested designs using a different method (chapter 8).

The approach in chapter 7 is as follows. Basically, we assume that the process which generates the counts can be modeled by the Poisson process. The intensity of this Poisson process is a random variable instead of a fixed parameter, and the random components for main effects and interactions of the factors are represented by this random intensity. We assume lognormality for the distributions of these random model components and we shall derive a limit theorem to simplify this complicated model. The result is a simple model for situations with large counts.

The quasi-likelihood approach for nested designs with random factors is the subject of chapter 8. The quasi-likelihood approach was proposed by Wedderburn in 1976 for the analysis of independent data, to be used if distributional assumptions are hard to make. It is an attractive method to use for the analysis of dependent count data as well, as the exact distribution of this data is rather intractable.

We shall use the quasi-likelihood approach to derive estimators and test statistics for the variance components in the case of a nested design with random factors, starting with a few very simple assumptions with respect to mean and variance of the data.
Interesting is, that the data which can be analysed is not restricted to count data. At the end of chapter 8 some topics for further research will be mentioned, advocating a further study of quasi-likelihood for the analysis of dependent (count) data for crossed designs with random factors.

SAMENVATTING

In de literatuur zijn methoden voorgesteld voor de analyse van tellingen, geklassificeerd door vaste en gekruiste factoren. Hierbij wordt de veronderstelling gemaakt dat de data gemodelleerd kunnen worden door onafhankelijke binomiale of Poisson verdelingen. In het algemeen hangt de gemiddelde waarde van deze verdelingen af van de niveaus van de klassificerende factoren; een lineair model wordt voorgesteld voor de logit transformatie of de log transformatie van deze gemiddelde waarden.

In de praktijk doen zich veel situaties voor die afwijkend zijn, zoals:

- De tellingen zijn onafhankelijk, maar de waargenomen variatie is groter dan door de Poisson verdeling verklaard wordt;

- De tellingen zijn afhankelijk: de factoren zijn niet vast, maar stochastisch.

Voor deze situaties zijn geen algemene analysemethoden beschikbaar en er is een sterke behoefte aan uitbreidingen van de theorie. In dit proefschrift worden uitbreidingen van de theorie gegeven die het modelleren en analyseren van dit soort gegevens mogelijk maken.

In de hoofdstukken 2, 3 en 4 wordt de situatie beschouwd van overdispersie met betrekking tot de binomiale verdeling en de Poisson verdeling. In het geval van overdispersie kan uit de waarnemingen blijken dat $\mathrm{var}(X) = \sigma^2 E(X)$ met $\sigma^2 > 1$, in plaats van $\mathrm{var}(X) = E(X)$ voor de Poisson verdeling.
In hoofdstuk 2 wordt de beta-binomiale verdeling gebruikt om overdispersie te modelleren en limietresultaten voor toetsingsgrootheden worden verkregen voor een groot aantal trials in iedere cel van het proefschema.
Een gamma-Poisson of negatief binomiaal model wordt voorgesteld voor het modelleren van overdispersie van tellingen in de hoofdstukken 3 en 4 van dit proefschrift. Benaderende kansverdelingen van toetsingsgrootheden worden verkregen voor een groot aantal herhalingen en voor grote tellingsuitkomsten. In de hoofdstukken 2, 3 en 4 worden de limietresultaten verkregen voor standaard toetsingsgrootheden, bekend uit de theorie van logitlineaire en loglineaire modellen, zoals de X^2-toets van Pearson.

In hoofdstuk 5 komen afhankelijke tellingen aan de orde in een split-plot situatie. Hier wordt een model voorgesteld dat rekening houdt met deze afhankelijkheid die inherent is aan het split-plot experiment. Twee separate analyses worden uitgevoerd, namelijk voor de whole plot en voor de subplot factoren, zoals in de variantieanalyse. De basismodellen zijn het gamma-Poisson model en het Dirichlet-multinomiale model.

Data die verkregen worden door op afhankelijke wijze objecten te klassificeren in twee of meer geordende klassen, daarbij hypothesen toetsend betreffende de kansen corresponderend met deze klassen is een probleem wat men b.v. in de context van enquetes ontmoet. In hoofdstuk 6 bestuderen we de symmetrietoets van Wilcoxon in de situatie van zo'n afhankelijke klassificatie. Het blijkt dat de limietverdeling van de corresponderende toetsingsgrootheid, onder de veronderstelling van een Dirichlet-multinomiaal model voor de data nog steeds de normale verdeling is, maar er is een extra parameter die de afhankelijkheid van klassificatie beschrijft.

De twee afsluitende hoofdstukken 7 en 8 van dit proefschrift behandelen problemen met stochastische factoren in gekruiste en in hiërarchische klassificaties (hoofdstuk 7) en in hiërarchische klassificaties, hierbij gebruik makend van een andere methode (hoofdstuk 8).

De aanpak in hoofdstuk 7 is de volgende. We veronderstellen dat het proces dat de tellingen genereert door het Poisson proces kan worden beschreven. De intensiteit van dit Poisson proces is een stochastische variabele in plaats van een vaste parameter en de stochastische componenten voor het hoofdeffect en de interactie van de factoren worden door deze stochastische intensiteit voorgesteld. We veronderstellen lognormaliteit voor de kansverdeling van deze stochastische modelcomponenten en we zullen een limietstelling afleiden om dit gecompliceerde model te vereenvoudigen. Het resultaat is een model dat eenvoudig hanteerbaar is en bij benadering geldig in het geval van grote tellingsuitkomsten.

De quasi-likelihood benadering voor hiërarchische klassificaties met stochastische factoren is het onderwerp van hoofdstuk 8. De quasi-likelihood aanpak werd voorgesteld door Wedderburn in 1976 voor de analyse van onafhankelijke data en is handig als het lastig is redelijk hanteerbare modellen te vormen. Aangezien dit inderdaad geldt voor het modelleren van afhankelijke tellingen, is het aantrekkelijk deze methode te gebruiken. We zullen schatters en toetsingsgrootheden onderzoeken voor de variantiecomponenten van de hiërarchische klassificatie met stochastische factoren; als uitgangspunt nemen we enkele eenvoudige veronderstellingen ten aanzien van verwachting en variantie van de data. Interessant hierbij is dat de data die geanalyseerd kunnen worden niet beperkt zijn tot tellingen. Aan het eind van hoofdstuk 8 worden enkele onderwerpen genoemd die voor verder onderzoek in aanmerking komen. Hierbij wordt ondermeer voorgesteld de quasi-likelihood aanpak te bestuderen voor de analyse van afhankelijke tellingen bij gekruiste klassificaties met stochastische factoren.

CURRICULUM VITAE

Jan Engel werd geboren op 4 januari 1952 te Amsterdam. Het h.b.s.-b-diploma behaalde hij in 1969 aan de Scholengemeenschap Noord te Amsterdam. Aansluitend studeerde hij aan de Universiteit van Amsterdam wis-, natuur- en scheikunde en deed hij in 1976 cum laude het doctoraalexamen mathematische fysica met bijvakken natuurkunde en mechanica bij prof. dr. E.M. de Jager. In 1977 trad hij als statisticus in dienst van de N.V. Philips Gloeilampenfabrieken bij de groep Statistics van de afdeling ISA-Research. Sedert 1979 werkt hij als consulterend statisticus bij de afdeling Centre for Quantitative Methods die onder leiding staat van Dr. J.P.M. de Kroon.