Mixed models for repeated count data.
van Duijn, Maria

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1993

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Repeated count data can be found in various fields of scientific research. Such data contain the number of certain events in multiple observations (or measurement occasions) on a sample of subjects (or measurement units) and they may be represented by a two-way contingency table. An empirical data set concerning a study on spelling errors made by children is used as an example. In that study the number of errors for four different types of words was counted. Repeated measures analysis is aimed both at the estimation of the parameters belonging to the subjects (the measurement units, represented by the rows in the contingency table) and of those belonging to the experimental conditions (the measurement occasions, the columns of the contingency table), and at the estimation of possible interactions between subjects and experimental conditions.

If the data contain many small counts the usual methods such as analysis of variance, which assume a normal distribution of the data, are not appropriate. Therefore, three alternative models for the analysis of repeated counts are presented: the Gamma–Poisson, the Gamma–Gamma–Poisson, and the Dirichlet–Gamma–Poisson model. For each cell of the contingency table a Poisson distribution is assumed, conditional on an intensity parameter. This intensity parameter has a multiplicative structure: it is the product of a row parameter and a column parameter in the Gamma–Poisson model. In the Dirichlet–Gamma–Poisson and Gamma–Gamma–Poisson models the structure is expanded by introducing an interaction parameter that depends on both row and column.

In all three models it is assumed that the rows (subjects) are a random sample from some population. Therefore, the row parameters are considered to be random. For the marginal distribution of the cell elements the introduction of random row parameters results in a variance larger than the mean, whereas the Poisson distribution implies an equal mean and variance. This is the well-known overdispersion often observed in count data. The interaction parameters are also assumed to be random. The column parameters are supposed to be fixed. Thus a mixed model is obtained, with both fixed and random effects. It is possible to incorporate a within-subjects and between-subjects design into the models, and to test the column parameters (as well as structural parameters belonging to the experimental
design) by means of analysis of deviance (or, equivalently, by Likelihood Ratio tests).

Chapter One presents an introduction to the models for repeated measures considered here, and their relation with existing methods. It also describes examples of applications in various disciplines.

The Gamma–Poisson model is the simplest model because it contains no interaction between rows and columns. It is presented in Chapter Two. In this model a gamma distribution is assumed for the random row parameters. The gamma distribution is conjugate to the Poisson distribution, which leads to nice mathematical properties and straightforward estimation procedures.

The Gamma–Poisson model is extended to the Gamma–Gamma–Poisson model in Chapter Three, by means of independent and identically gamma distributed interaction parameters. For the row parameters a gamma distribution is assumed again. The estimation of the model parameters for the Gamma–Gamma–Poisson model becomes complicated due to this choice of the random parameter distributions, because the product of two gamma distributions is not conjugate to the Poisson distribution.

In Chapter Four a different way of defining interaction is introduced, resulting in the Dirichlet–Gamma–Poisson model. Besides the gamma distribution for the row parameters a Dirichlet distribution is assumed for the row vectors of interaction parameters. The Dirichlet distribution is conjugate to the multinomial distribution, which is the conditional distribution of the row vectors given their marginal row totals. Therefore, the property of conjugate prior distributions is retained. This renders the estimation of the model much less strenuous than that of the Gamma–Gamma–Poisson model.

The fixed parameters, i.e., the column parameters and the parameters of the distributions of the random row and interaction parameters, are estimated with Maximum Likelihood methods. The random parameters can be estimated by using their conditional distribution given the data, where the Maximum Likelihood estimators are substituted for the fixed parameters. This so-called posterior distribution depends on the data as well as on the estimated fixed parameters indicated above. The posterior mean and posterior mode are often called empirical Bayes estimators for the random parameters.

In Chapter Five Score test statistics are developed to test the null hypothesis of a Poisson model against a Gamma–Poisson model, and a Gamma distribution for column totals of the data. A normal and a chi-square statistic are proposed.

The main difference in structure. In the Gamma–Gamma–Poisson model the interaction parameters are assumed to be normally distributed. The Gamma–Gamma–Poisson model in smaller covariance changes in variance parameters. The interrelation of the interaction parameters is measured by the Gamma–Gamma–Poisson model.

The usefulness of the study, presented in this chapter, is based on the estimation of the parameters of the Gamma–Gamma–Poisson model. The study designs and three replications in each of the three models. For each study, the three models were used for the estimation of the parameters.
Mixed models for repeated count data

Summary

ance (or, equivalently, by Likelihood Ratio
tion to the models for repeated measures
with existing methods. It also describes
lines.

The simplest model because it contains no
it is presented in Chapter Two. In this model
the random row parameters. The gamma
poisson distribution, which leads to nice
ard estimation procedures.

I to the Gamma-Gamma-Poisson model in Chapter
identically gamma-distributed interaction
gamma distribution is assumed again. The
the Gamma-Gamma-Poisson model becomes
random parameter distributions, because the
't conjugate to the Poisson distribution.

y of defining interaction is introduced,
model. Besides the gamma distribution for
tion is assumed for the row vectors of
tribution is conjugate to the multinomial
tribution of the row vectors given their
etry of conjugate prior distributions is
model much less strenuous than that of

changes because of a distinction between the in-

parameters and the parameters of the inter-
action parameters, are estimated with
parameters can be estimated by using their
ere the Maximum Likelihood estimators are
's so-called posterior distribution depends
fixed parameters indicated above. The
en called empirical Bayes estimators for
are developed to test the null hypothesis
of a Poisson model against a Gamma-Poisson model, and to test the null hypothesis of
a Gamma-Poisson model against the alternative hypotheses of a Dirichlet-Gamma-
Poisson, and a Gamma-Gamma-Poisson model, respectively. For the latter test
statistics the conditional mean and variance are derived, given the marginal row and
column totals of the contingency table. With the conditional mean and variance, both
a normal and a chi-square approximation of the conditional distribution of the test
statistics are proposed.

The main differences between the three models concern the variance-covariance
structure. In the Gamma-Gamma-Poisson model the covariances between the cell elements
are the same as in the Gamma-Poisson model because of the independence of the
interaction parameters whereas the variances are increased. The assumption of a
Dirichlet distribution for row vectors in the Dirichlet-Gamma-Poisson model results
in smaller covariances and larger variances than in the Gamma-Poisson model. The
changes in variances and covariances depend upon the values of the various
parameters. The intraclass correlation reflects these differences and can be used as
a measure of interaction. Therefore, the intraclass correlation is an important
characteristic of the models.

The usefulness of the intraclass correlation is established in a simulation
study, presented in Chapter Six, where a comparison between the various models is
made based on the variability of the data (as measured by the variances of the cell
elements). The study is used to examine the behavior of estimators and tests for the
three models. For a simulation design containing 72 cells, cross-classified with
respect to the three models, two sample sizes (50 and 500), two within-subjects
designs, and three different expected values of the marginal row totals, 1000
replications in each cell were obtained. The parameter values were selected with the
aim of producing data with mostly small counts.

Data sets generated according to the three models were estimated using the
Gamma-Poisson and Dirichlet-Gamma-Poisson models. The Gamma-Gamma-Poisson model
was not used for the analysis because of the large computational costs involved. The
estimation of the parameters belonging to the distribution of the random parameters,
and of the column parameters, as well as the behavior of the Score test and
Likelihood Ratio test statistics were investigated. The results for the Gamma-
Poisson and Dirichlet-Gamma-Poisson models are favorable, concerning the power of
the tests, and the estimation of column parameters and of interaction parameters. It was found that the estimation of the parameters belonging to the gamma distribution of the row parameters is best performed on a log-scale, due to asymmetry in the distribution of their estimators. Because the Gamma–Gamma–Poisson model was not used to analyze the data, the results concerning that model are incomplete.

Although the scope of the simulation study is limited, and the generalizability of its results therefore restricted it may be concluded that repeated count data containing many small counts can very well be analyzed using the estimators and tests developed for the models presented in this thesis.