Limburgs Universitair Centrum
Faculteit Wetenschappen

Modellering van Niet-Normale Longitudinale Data in Continue Tijd, Gebaseerd op de Likelihoodfunctie

Likelihood Based Approaches to Modelling Non-Normal Series in Continuous Time

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen
aan het Limburgs Universitair Centrum te verdedigen door

PHILIPPE LAMBERT

Promotor:
Prof. dr. J.K. Lindsey

1995

= longitudinale data

Limburgs Universitair Centrum
Faculteit Wetenschappen

Modellering van Niet-Normale Longitudinale Data in Continue Tijd, Gebaseerd op de Likelihoodfunctie

Likelihood Based Approaches to Modelling Non-Normal Series in Continuous Time

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen
aan het Limburgs Universitair Centrum te verdedigen door

PHILIPPE LAMBERT

Promotor:
Prof. dr. J.K. Lindsey

1995

Contents

1 Introduction 7

1.1 Preamble and scope of this work 7

1.2 Rationale for using conditional methods 12

1.3 Characteristics of longitudinal data 15

1.4 Some key conditional models 16

1.4.1 Normal autoregression models 16

1.4.2 Non-normal autoregression models 20

1.4.3 The dynamic linear model and the Kalman filter 21

1.4.4 Dynamic generalized linear models 25

1.4.5 Other applications of the Kalman filter: models for count data 28

1.4.6 The multivariate dynamic generalized linear model 30

1.4.7 Exponential dispersion models 32

1.5 Random effects models 35

1.6 Marginal models 37

1.7 Further reading 40

1.7.1 Autoregression models 41

1.7.2 Dynamic models 43

1.7.3 Random effects models 47

1.7.4 Marginal models 48

1.7.5 Missing data 51

1.7.6 Other models 51

2 Positive longitudinal data modelling 53

2.1 Data set of interest 53

2.2 Generalized autoregression models 55

2.3 Model selection 60

2.4 Analysis of the triglyceride data set 60

3 Longitudinal count data in continuous time 67

3.1 Problem setting with two examples 67

3.2 Gamma-Poisson model 69

3.3 Analysis of data set 1 71

3.4 Robustification of the gamma-Poisson model 74
3.4.1 The Poisson distribution as a limiting case 74
3.4.2 Robustification 75
3.5 Modelling the growth curve 75
3.6 Analysis of data set 2 76

4 Other Applications of the GARM 79
4.1 Series of binary and multinomial data 79
4.1.1 Binomial series 81
4.1.2 Analysis of the fly data set 82
4.1.3 Multinomial series 85
4.1.4 Analysis of the pollen data set 86
4.2 The GARM and generalized linear models 87
4.3 An approximate predictive likelihood 88
4.3.1 Likelihood prediction envelopes 89
4.3.2 Application 92

5 Conclusion 95
5.1 Conditional models 95
5.2 Models in continuous time 96
5.3 Nonlinear regression 96
5.4 Non-normal models 97
5.5 Model selection 97
5.6 Further research 98

6 Summary 101

7 Samenvatting 107

8 Résumé 113

Acknowledgements

First, I would like to express my deep gratitude to Prof. J.K. Lindsey. Jim first helped me to put my ideas in order after a one year stay at the University of St-Andrews, where my numerous courses in statistics gave me a real start in the subject. At that time, I did my graduating project with him on dynamic generalized linear models, which turned out to be the starting point of my PhD thesis. After graduating in Mathematics, he proposed me to work as his assistant in the Department of Social Sciences, at the University of Liège. Although I might have been of great help to him in Liège, he strongly encouraged me to do a MSc in Biostatistics at the LUC, where I learned much of what I know in statistics. After this very demanding one year training, I started my PhD under his supervision. Again, during these last two years, I felt his complete trust, and his discrete but always careful guidance. Jim also introduced me to the world of consulting, which is another very important and fascinating aspect of our job. He taught me to listen, to ask the right question, and to realize the limits of any theoretical development. This deeply influenced all my future work, where I constantly wonder whether my contributions can have practical consequences in applied statistics. For all this and many other reasons, I am deeply indebted to him.

I would also like to thank the Diepenbeek biostatistics team, who offers Belgian students the possibility to get a high quality training in their home country.

More particularly, I would like to acknowledge the help of Prof. N. Veraverbeke for his kind and efficient mediation with the Diepenbeek academic authorities at the final stage of this work. My sincere thanks also go to him and Prof. G. Molenberghs for the Dutch translation of the summary, to Marianne Diez and Bruno Génicot for respectively the triglyceride and calf data sets of Chapters 2 and 3.

I am also very grateful to my fiancée, Evelyne, who suffered my moody humour during these two years. Finally, I would like to dedicate this work to my Grandmother, Léonie.

Philippe Lambert
Liège, October 1995.

1
Introduction

1.1 Preamble and scope of this work

Longitudinal studies consist of observing the evolution of one or several quantities through time on one or several units. Such settings often arise in practice in a wide range of fields. For example, in economy, one could observe the evolution with time of wage indices in different countries, and try to understand how other recorded variables such as the current leading politic party, the unemployment rate, the power of Labor Unions, etc., influence the response.

In animal ecology, the inventory of some animal population in different areas of a wood could be observed over several years, and explained in terms of dynamic elements such as the season, the number of predators, the severity of the preceding winter, in addition to more static explanatory variables such as the geography of the area (light or dark area, type of soil), the proximity of human paths, etc.

Human and veterinary medicine are other subjects where longitudinal data arise. One example, that will be treated in Ch. 2, studies the effect of different types of diets on several blood variables observed on two groups of four adult Beagle dogs. These variables include the level of α-amino-N (mg/l), urea (mg/l), glucose (mg/l), insulin (g/l), cholesterol (g/l), non esterified fatty acid (μEq/l) and triglycerides (g/l), observed at irregularly spaced time points on a six hour period after feeding. The latter response *profiles* — or plots of the triglyceride response against time — for the four types of diets considered, are given on Figures 2.1, 2.2, 2.3 and 2.4. One can see that the animals were observed at unequally spaced times and that the response is bounded below by zero. In addition to time and the type of diet administered, the sex of the animal is another variable that might help to explain the variability in the data. One further example in veterinary medicine is presented in Section 3.1, where the respiratory rate of calves is observed over time under different doses of a receptor blocker. The corresponding count profiles are plotted for six calves in Figures 3.1 and 3.2. The drug administration lasted for 30 minutes, explaining the drop in the response after that time. One question of interest was to develop an equation for the respiratory rate profile taking the drug dose into account.

Experimental biology is still another subject where longitudinal data are commonly gathered. In Ch. 3, we shall study the evolution of the sizes of *Paramecium aurelium* colonies on a twenty day period (see Fig. 3.3). One equation, common to the three series, modelling the number of individuals in each colony at any time, was desired. The stabilization of the colony size around day 10 may be one major concern when developing a suitable model for these data.

In this work, all the treated examples will be related to biological and biomedical problems. However this does not mean that the scope of the longitudinal models that we shall present in the coming chapters, is restricted to these sciences. The technical problems in longitudinal data modelling in other areas, such as economy, are identical. The choice of the above two particular subjects to illustrate our findings, was simply due to our close collaboration with researchers in these fields.

In order to define clearly the scope and the goal of this work, we now briefly review the main steps in the development of a longitudinal study:

(1) First a suitable experimental design is set up to answer the question at hand. This problem is extremely complex because any sensible estimation of things like the sample size required to detect any covariate effect, should be based on a model. Of course such a model might not be available before the experiment is performed, and, hence, rough approximations based on distributions like the normal are often used. Such guidelines can be found in the literature (Diggle *et al.*, 1994, Ch. 2). They should be used with caution. Note that we have not considered the problem in our framework. Likelihood based methods for deriving sample size formulas in a longitudinal data setting could mimic what Lindsey (1995b, pp. 88-91) suggested in an independence context. One has just to remember the decomposition of the likelihood as a product of conditionally independent random variables.

(2) When analysing longitudinal data, the first thing to do is generally to plot the data against time and the covariates. This is usually extremely instructive. It will give a first idea about the important explanatory variables and about the shape of the response profile. It might also put the focus on sometimes unsuspected results at the basis of a scientific discovery. Indeed even after a well-written protocol, where all the questions of interest (together with the statistical techniques required to answer these) have been described, surprises cannot be avoided, and the analyst should be ready to deal with such situations. Note that these discoveries might completely invalidate what has been carefully prescribed in the protocol. This subject will not be developed in this work. The reader is referred to Diggle *et al.* (1994, Ch. 3) and the references at the end of the corresponding chapter for a description of the possible ways of exploring longitudinal

data.

(3) The nature of the data should help to select a set of sensible candidate distributions for describing the stochastic mechanism underlying our statistical model. For example, in the case of the heart beat frequency data, one could choose the Poisson distribution or the negative binomial if overdispersion is likely to arise. If one wants to analyse triglyceride records, distributions, like the gamma, Weibull, or more generally the generalized gamma, all excluding negative values, are sensible choices. This set of distributions should be sufficiently large to be able to assess the sensitivity of the conclusions to the stochastic component. It could be specified in the protocol (Lindsey and Jones, 1995).

(4) Using the step 2 hints, express some location parameter as a function of the explanatory variables. This can be done in the same way as with models assuming independence. If, for example, one chooses a member of the exponential family, then one could just view the canonical parameter as a linear combination of the regressors (if this is judged sensible). Of course non-linear models should be used if the main data features are better described in such a way. An underlying biological theory should be at the basis of the statistical model each time this is found possible. Non-linear models are then more likely to appear.

One could further model some scale or shape parameter, allowing, for example, different types of distributions in different strata of the data.

(5) Inferences can then be made. We have restricted our approach to conditional models for which likelihood methods are readily usable (see Equation (1.1)). The reason for that choice will be explained further in Section 1.4. However, we shall also present the fundamental ideas of the marginal model school (see Section 1.6) where approximate methods based on asymptotic arguments of the resulting parameter estimators are often used.

(6) Now that the model has been built, we should enquire about its adequacy. Residuals constitute an important tool for assessing the quality of a fit. Deviance residuals are easy to compute with likelihood methods. Influential observations as well as model lack of fit can be pointed out. We shall not discuss this aspect of data modelling in the present work, although its importance cannot be ignored in the building of any statistical description. However note that the arguments used to develop the tools (Davison and Gigli, 1989; Davison and Tsai, 1992) commonly used in procedures assuming data independence, are also applicable in the context of our research where the likelihood is factored as the product of conditionally independent terms.

Our work will focus on points 3, 4 and 5. We shall restrict our attention

to non-normal data.

Our first contribution (see Ch. 2) is devoted to the development of models — the *generalized autoregressive models* or GARM — for irregularly sampled profiles of non-negative data under different distributional assumptions (Lambert, 1996c). The starting point of this paper is the set of dog triglyceride profiles observed under four different diets, that were provided by the veterinary team of the University of Liège Nutrition Department. The main challenges with that data set were:

- to release the too restrictive normality hypothesis in another way than just taking the log of the response to be able to use standard statistical packages. A large family of non-negative random variable distributions, the generalized gamma family, including the exponential, gamma, Weibull and log-normal as well-known members, will be used as the stochastic component of the model.
- to cope with the irregular sampling procedure where, as very often in veterinary studies, the animal is more frequently observed at the start of the study. Note also, that because of the possibly noncomplying behaviour of animals, some measurements often cannot be performed at the planned time (simply e.g. because the animal was moving, or some other nonexperimented animal required urgent care). Therefore the statistician has to deal with incomplete series of data, where, it is hoped, the missing data mechanism is not informative. More adequate methods (see Section 1.7.5) have to be used if that condition is not met. We strongly believe that a detailed discussion with the experimenter should bring the necessary light on the problem.
- to compute estimates of the parameters involved in the stochastic and systematic parts of the model. For that use, given that all our methods were likelihood based, we found the procedure Proc OPTMUM in GAUSS (which is now available on UNIX platforms) together with the programming facilities of that software incomparably efficient [1].

Note that we invariably used the AIC (Akaike, 1973) to compare the performances of possibly non-nested models. That method is most adapted in a model selection context (Lindsey, 1994).

Our second concern was to develop methods for analysing discrete longitudinal data (see Ch. 3). We propose a model for series of overdispersed counts measured at unequally-spaced time points (Lambert, 1996a). This model is an extension of the discrete time Harvey and Fernandes (1989) (see also Ord *et al.*, 1993) gamma-Poisson model in an empirical Bayesian setting.

[1] We could not imagine the state of our work at this time if we only had access to a FORTRAN compiler. We strongly encourage people unsatisfied with their everyday statistical package, or researchers wanting to try their own undocumented statistical discovery, to use this efficient tool.

It has been further refined in Lambert (1996b), yielding a more robust model. The data set of interest, describing the growth of three colonies of *Paramecium aurelium* in a nutritive medium (Gause, 1934; Diggle, 1990), clearly shows a non-linear pattern: the colony size seems to stabilize after about ten days, thereby indicating the need for a non-linear systematic part for the model. A generalized form of the logistic growth curve (Nelder, 1961 and 1962) further developed by Heitjan (1991a and b) and including the Mitscherlich, Gompertz, logistic and exponential forms as well-known members, was used to allow for that asymptotic behaviour.

Using the same kind of empirical Bayesian method to model binary or binomial data turned out to be technically more difficult. An approximation to the posterior density in the beta-binomial model is required to keep working with beta distributions. We did this by using a beta approximation with the same mode and Fisher information at that point as the exact posterior. But the complexity and the quasi-likelihood aspect of the procedure convinced us not to proceed further. That might be the subject of further research.

This deceiving conclusion led us to further develop the generalized autoregression model to deal with binary, binomial, multinomial (see Section 4.1) and count data series observed in continuous time. All these models have been applied to data sets from the literature and seem to give encouraging results (Lambert, 1995d). Note that, when the distributions are restricted to the exponential family and the regression parameters appear in a linear way in the description of the location parameter of these GARMs, the IWLS algorithm, available in most statistical packages, can be used to make inferences about most of the parameters, except for the shape, scale and two autoregression parameters for which a grid of values has to be considered (see Section 4.2). In our work, we just used a non-linear optimizer on that small set of unknowns, with, at each step, an estimation of the linear parameters using the IWLS algorithm. We shall show that a transformation of the classical design matrix is all that is required, together with the computation of an offset. That means that the interested reader just could write a GLIM or S-PLUS macro for modelling series of non-normal longitudinal data. Of course, other methods have to be used outside the exponential family, or when non-linear systematic parts are desired in the model. Notice that no major assumption has been made about the stochastic element of the model, thereby showing the wide applicability of the GARM.

Finally, we concluded our research by looking at methods suitable for making predictions using the GARM (see Section 4.3). The usual likelihood based prediction method, the predictive likelihood by Fisher (1959, pp. 128–133), appears not to be satisfactory because the profile predictive likelihood (where the observation to come plays the role of the parameter of interest) just assumes that the 'nuisance' parameters are fixed at their MLEs. Therefore we decided to enquire about methods allowing for the

randomness of the conditioning elements. Intuitively, these should yield wider (predictive) likelihood intervals. The only computable method outside the exponential family that we could find, is described in Butler (1986, Rejoinder). We have considered it to predict the behaviour of an artificially truncated series of the *Paramecium aurelium* data set described above (see Section 3.1). Note that we have used a polynomial model instead of the generalized logistic growth curve that has been advocated in Lambert (1996b) to be able to use the GLM rewriting of the GARM in the particular case of a negative binomial distribution with fixed overdispersion parameter. This highly simplified the computational implications of the procedure, which, in non-linear settings, requires a lot of patience, while still being usable.

1.2 Rationale for using conditional methods

As already mentioned in the last section, conditioning will be our tool of predilection when developing models for longitudinal data. Hence that might be the source of a biased exposure which should largely be balanced by the abundant literature on marginal models. However we shall try to convince the reader that our choice is a sensible one and not only the consequence of some dogmatism.

Statistics, whatever the refinement of the methods involved, is usually just an elaborate way of comparing means or proportions in different strata of a population. When, say, a clinical trial is set up to assess the efficacy of a given treatment in raising the white blood cell count (or CD4 level) of AIDS patients, we expect to improve the health state of each individual on the active treatment arm. Therefore building a statistical model that puts the subject at the centre of the study instead of some sample averaged quantity that does not take the patient specificity (defined by as much personal information as possible on each studied unit such as its individual history of the illness) into account, seems to be the most important goal in any study of the data structure. We do not think that people will care about an average positive treatment effect on the proportion of people with an improved health state, if, individually, this is not followed by the same kind of prospect for each treated patient. That apparently aberrant conclusion can arise if a suitable conditioning is not done on the important patient specificities.

But before giving an example of such a paradox, let us point out the specificity of longitudinal studies as far as randomization is concerned. If the data considered are representative of some larger population, which is generally ensured by random sampling, then the marginal average obtained from a model should tell us something about that population as a group. However, randomization only ensures that variable values at the moment of sample selection are 'representative', or, in other words, that both the observed and unobserved variables which have an influence on the response,

are reasonably balanced among the different treatment branches, say. Unfortunately, as soon as repeated measurements are collected, any time-varying covariates of individual subjects are no longer randomized, even if both observed and unobserved influential factors were suitably spread among treatments. This unbalance can become more and more marked as time goes by. Because of this, successive responses become conditional on the previous patient history and not just on the variable values at randomization.

A marginal model which ignores such time-varying covariates can produce conclusions which, although correct for a population average, might not be valid at the patient level, thereby giving rise to the announced paradox. As a simple fictitious example, consider a trial to compare placebo to treatment with a binary outcome. The favourable marginal results of the repeated measures might be summarized as

	Recover			Recovery
Treatment	No	Yes	Total	rate
Placebo	200	200	400	50%
Active	160	240	400	60%

We conclude that, in the population as a whole, the active treatment should be adopted.

Now let us condition on one of the time-varying covariates which the marginal model ignores: the number of infections over the period of collecting the repeated measures. The summary table for these same data is now

	One infection				Several infections			
	Recover			Recovery	Recover			Recovery
Treatment	No	Yes	Total	rate	No	Yes	Total	rate
Placebo	20	80	100	80%	180	120	300	40%
Active	90	210	300	70%	70	30	100	30%

Within each of these two groups, active treatment is worse. Note how realistic this example is: patients under placebo have more infections and these with more infections have a lower recovery rate.

This is just a simple example of Simpson's (1951) paradox which illustrates one of the inherent dangers of any marginal modelling. But the conclusion obtained using the marginal table, is not as wrong as it seems to be, as the following argument tends to show. Assuming that the patients were randomized to the two treatment branches, we can see that the active treatment is estimated to prevent the occurrence of more than one infection in 75% of the randomized units, against 25% for the placebo group. Because the probability to recover is largely improved when the patient has not had more than one infection, whatever the treatment branch considered, the active treatment is marginally more efficient than the the placebo, although

this last treatment is conditionally preferable. What the conditional analysis brings in addition to the marginal approach, is a deeper understanding of the biological mechanism which generates the data. In the present case, it suggests that predicting the number of infections that a patient is likely to have further in time, would be interesting information in order to decide to assign or not a patient to the active treatment. Indeed, the population under study seems to be made up of three types of individuals:

(1) frail patients, who, even under the active treatment, will be subject to several infections. They represent 25% of the population under study. For these people, the placebo would be preferable. Note that speaking of 'population' in a clinical trial, is rather dangerous, because the individuals involved in such trials, are not a random sample from the larger population that we actually aim to study. Only the treatment assignment is random.

(2) robust patients, who, even on the placebo branch, will not suffer from more than one infection. They represent 25% of the sample under study.

(3) the other 50% of the sample, who should be treated on the active branch, to prevent them from suffering from several infections, thereby improving their chance of recovery.

The third group is the only one where the treatment is of interest. Some kind of surrogate marker would then be useful to determine if a patient belongs to this last category (which is half way between the two extreme ones). Note that, even if such an indicator is not available, a possible policy would be to give the treatment to all the patients, and, as information on their number of infections is gathered, decide or not to continue the treatment. The probability to recover for treated patients observed to have several infections during the course of the treatment (and hence switching to the placebo branch), would then probably be between 30 and 40%. Consequently, the marginal recovery rate resulting from this policy would lie between 60 and 62.5%, which is a bit better than a blind application of the active treatment, as recommended by the marginal analysis. Of course, this kind of hypothesis would require testing on new data. Another possible (but more hazardous) way to improve that rate, would be to stop giving the active treatment to patients with only one infection, when their state is judged (by some criteria to be defined) to be stable from an infection point of view. The recovery rate for these patients would then lie between 70 and 80%, if the end of the treatment does not lead to a second infection. Note that this elaborate way to treat patients only concerns the patients in the two extreme categories, namely frail and robust individuals.

From this rather subtle discussion, we hope that the reader is more aware of the possible unfortunate consequences of a blind use of marginal models to analyse longitudinal data. Conditional models, by trying to understand

the data generating mechanism, can suggest other ways of interpreting the data, which, in situations similar to above, have practical consequences on the treatment policy.

A technical argument for using conditional inference is the 'easy' computation of the likelihood function once the dependence on previous patient history has been suitably specified. Indeed, if we denote by $\boldsymbol{y}_i$ the n_i observations $\{y_{i1}, \ldots, y_{in_i}\}$ on unit i $(i = 1, \ldots, I)$ made at times $\{t_{i1}, \ldots, t_{in_i}\}$, then one can factor the likelihood function as

$$L(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_I) = \prod_i f(y_{i1}|\boldsymbol{\beta}_i, \upsilon) \prod_{j=2}^{n_i} f(y_{ij}|\boldsymbol{\beta}_i, \upsilon; \mathcal{F}_{ij-}) \qquad (1.1)$$

where

- $f(.)$ denotes the chosen density (or probability function in the discrete case) for the response.
- $\boldsymbol{\beta}_i$ stands for a vector of regression parameters. There is, of course, associated with each unit and at any time point, a vector of explanatory variables, say $\boldsymbol{x}_{ij}$, that could indicate the treatment branch to which a patient was affected, his age, his blood pressure, etc., and anything that can be thought to have an incidence on the response variable under study.
- υ stands for nuisance parameters such as scale, shape, or autoregression parameters. It could also include a parameter indexing several possible choices among the elements of a family of stochastic components (Lambert, 1996c).
- $\mathcal{F}_{ij-}$ denotes the history of the unit up to time t_{ij}, except its response at that time point. Hence it can include past responses, past and present regressor values, etc., and anything in the process history that can be thought of interest to model the response.

Note that one usually makes the assumption that the dependence on the past history of the subject can be limited to a small number of past time points (Markov assumption).

1.3 Characteristics of longitudinal data

Two types of dependence appear in longitudinal data. Firstly, data observed on the same unit tend to show more consistency than data across units. For example, when recording the heart beat frequency profile of patients in a clinical trial (say), the records of a young and active patient will naturally be less variable than a mixture of these records with those of their seventy year old neighbours. The homogeneity among data from the same unit is most often easily detectable with biological data when the response is plotted against time. This phenomenon — also called *heterogeneity* or *frailty* in a survival context — is often responsible for most of the observed variability

in longitudinal data. It can be accounted for using *random effects* models, or even more elaborate, *random coefficients* models (see Section 1.5).

Secondly, observations made close in time on the same unit will tend to be more closely related than those made at more distant time points. This *serial association* is particularly clear in the above example where heart beat frequencies tend to evolve in a smooth way with time. This is mainly due to the inertia of biological variables. Note that some variables, such as the level of luteinizing hormone in the blood, exhibit occasional large increases in value. However the behaviour of such pulsatile quantities can be modelled stochastically (Diggle and Zeger, 1989). This is rather important because, once that the pulses have been modelled, the notion of inertia is still sensible, Serial association can be modelled by working on either the systematic part, or directly one some higher order component of the model structure such as the covariance matrix in the normal multivariate distribution. The two approaches are equivalent in this latter setting. Sections 1.4 and 1.6 will expose the main techniques used to this end in the literature.

Modelling heterogeneity and serial association are usually the extra effort required with respect to the well-known analysis of independent data. In the next section, we review some key conditional models from the literature, which have been used to account for one or both of these two features.

1.4 Some key conditional models

1.4.1 NORMAL AUTOREGRESSION MODELS

We shall first start with the normal distribution, which, in longitudinal data, as in many other statistical areas, has been the subject of extreme attention. The main reasons for this are its remarkably simple properties and the easy implementation of its estimates using least squares arguments. For example, the marginal and conditional distributions of the elements of a multivariate normal (MVN) random vector are both normal. In addition to this, the specification of the mean vector and of the covariance matrix completely defines the distribution of a MVN vector. This is of course an important element when the modelling of the dependence between successive observations in a longitudinal study is of interest.

Normal autoregression models are probably the most well-known conditional models. As a first step in the building of these models, suppose that the mean response μ_{ij} is expressed as a linear combination of explanatory variables and past responses, thereby modelling the serial association mentioned in Section 1.3. If, for example, a first order Markov process is assumed, we have the *autoregressive process* of order one

$$\mu_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i + \rho y_{i,j-1} \tag{1.2}$$

Extension to higher order models is easy. The response is then assumed to be influenced not only by the last state of the process, but also by events further in the past. That helps to assess if first order dependence is a reasonable hypothesis. Note that the contribution of the first observations on which we condition is fundamental to start the recursive implementation of the conditional means. They are often given a stationary distribution, because, with these observations, there is no way to condition on the past. When we have an autoregressive process of order two, we 'lose', for each series, the first two observations. Hence a high order dependence quickly becomes unrealistic with short series of data.

Using Equation (1.1), and under the hypothesis that the model is an acceptable approximation to reality, we see that the data can just be treated as independent normal elements with a mean defined conditionally on past events. That makes the use of traditional regression software a convenient choice for estimating the model parameters. Note that the influence of the explanatory variable can be introduced in a more complex way, using for example non-linear systematic parts motivated by theoretical arguments.

Another refinement of the model, yielding the well-known *autoregression model* of order one, or AR(1), is to express the mean response as its regression part plus a correction proportional to the last raw residual:

$$\mu_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i + \rho(y_{i,j-1} - \boldsymbol{x}_{i,j-1}^T \boldsymbol{\beta}_i) \tag{1.3}$$

The autoregression element now corrects for the sequences of residuals with identical signs resulting from a 'naive' independence model. Indeed a typical consequence of serial association is the inertia of the unmodelled part of the data. In other words, an observation that tends to be above (below) the curve defined by $\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i$ at one time, tends to be followed by one with the same characteristic. But that refinement has its cost. Indeed, we do not have a linear model anymore because the regression and autoregression parameters now appear multiplicatively. Therefore iterative methods have to be used with, at each step, the regression parameters estimated at the last updated value of ρ.

Finally note that, because the correction implied by the autoregression term with respect to the systematic part in Equation (1.3) is a quantity taking values in a neighbourhood of zero, omitting it will be acceptable with the first observations for which 'no past' is available. Hence the marginal contribution of these first observations can be included in the likelihood, thereby allowing the computation of the likelihood on the whole data set. That means that the deviances[2] of models of different orders are now comparable. Hence a model selection using for example the AIC is now possible.

[2]There exist different definitions for the deviance in the literature. Here, the deviance will stand for -2 log Likelihood.

But note that assuming the same regression parameters to describe the distribution of the first data as that of the other ones, imposes stationarity.

We cannot close this section without speaking about inter-unit heterogeneity, which is often an important source of variability in the data. Two units with apparently exactly the same characteristics can respond at completely different levels. However, the response profiles for such subjects may often be assumed to be parallel, which seems to be a reasonable hypothesis in most practical situations, as can be seen from a plot of the response against time. One explanation for that apparently incoherent behaviour is that a large set of variables, either unknown to the experimenter, or too large to be recorded, has a combined influence on the response. This is particularly true with biological and economical processes where an experiment is only partially under control. Patients involved in a clinical trial do not constitute a batch of homozygotic rats! Hence the complex mixture of uncontrolled influences could drown the treatment effect if the inter-unit variability is not taken into account. In contrast, laboratory experiments in physics are usually under full control, which means that any departure from the theoretical model is often just a measurement error. In these very particular situations random effects are unnecessary.

Because the uncontrolled variability between units usually results from the additive influence of multiple quantities on the response, it is often modelled as a $N(0, \sigma_U^2)$ normal random variable U_i. Different combinations between random effects and serial association terms are then possible to construct a model:

(1) the naive 'independence' model ignoring inter-unit variability and serial association. Note that, even with longitudinal data, this model might turn out to be sensible (cfr. laboratory experiments in physics, cross-over trials with a long washout period or with few observations per individual, etc.).

(2) the pure random effects model, for unit i,

$$\mu_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i + u_i$$

for which

$$\boldsymbol{\Sigma}_i = \mathrm{var}(\boldsymbol{Y}_i) = \begin{pmatrix} \sigma^2 + \sigma_U^2 & \sigma_U^2 & \dots & \sigma_U^2 \\ \sigma_U^2 & \sigma^2 + \sigma_U^2 & \dots & \sigma_U^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_U^2 & \sigma_U^2 & \dots & \sigma^2 + \sigma_U^2 \end{pmatrix} \qquad (1.4)$$

where σ^2 is the intra-unit variance. Therefore we can see that observations on the same unit are correlated with pairwise correlation $\sigma_U^2/(\sigma^2 + \sigma_U^2)$ independently of the time separating the data.

(3) the pure 'serial association' or autoregression models described in

Equations (1.2) and (1.3) with possible extensions to a higher order dependence. The correlation matrix for the vector of observations made on unit i is, for an AR(1),

$$\operatorname{corr}(\boldsymbol{Y}_i) = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n_i} \\ \rho & 1 & \rho & \dots & \rho^{n_i-1} \\ \rho^2 & \rho & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{n_i} & \rho^{n_i-1} & \rho^{n_i-2} & \dots & 1 \end{pmatrix} \tag{1.5}$$

Because $0 < \rho < 1$, we conclude that the correlation between any two observations made on the same unit only depends on the time lag between them and is decreasing in an exponential way with it.

(4) a mixing of models 1 and 2, taking both heterogeneity between units and serial association within units into account:

$$\mu_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i + \rho y_{i,j-1} + u_i$$

or

$$\mu_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i + \rho(y_{i,j-1} - \boldsymbol{x}_{i,j-1}^T \boldsymbol{\beta}_i) + u_i$$

with possible extensions to a higher order dependence.

Of course, we make the assumption that observations from different units are independent, i.e.

$$\operatorname{corr}(\boldsymbol{Y}_i, \boldsymbol{Y}_k) = 0$$

for different values of i and k. The pure measurement error model could also be 'mixed' with the above models, if the measurement process is suspected to add a component of variation to the data (Diggle *et al.*, 1994, Ch. 5), as when simultaneous samples are taken on the same unit.

We shall show in Section 1.6 how to derive analogous models using marginal arguments.

Note that, until now, we have assumed that we had no missing data. Things can become very complicated if 'holes' occur in the data set. Indeed, with autocorrelation models, we cannot condition on something that we have not been able to observe. One way out of this problem is to build a model in continuous time instead of the regular time spacing that has been assumed. However, if the reason for one datum to be missing is related to its unknown value, then the last procedure is not satisfactory anymore. Special techniques have to be used that explicitly model the missing data mechanism. The literature on the subject is now rapidly growing. We refer the reader to these papers for more information (see Section 1.7.5).

Throughout this work, we shall assume that missing data are *missing at random*, i.e. the probability for these data to be missing is not related to their unobserved values.

In the next section, we shall see how to extend autoregression models to deal with non-normal data.

1.4.2 NON-NORMAL AUTOREGRESSION MODELS

As in the normal case, the idea underlying autoregression models is to express some location parameter μ_{ij} as a function of covariates and past responses. Of course the chosen function must fulfil some conditions to avoid, for example, having a negative value modelled for the mean of a counting process. One way to solve this problem, is to work with a function of μ_{ij} taking values on the real line. For example, if the probability of success is the parameter of interest in successive Bernouilli trials, then one could choose to model the logit of that probability as a linear function of the covariates and of past responses. More generally, using the generalized linear model vocabulary, if $g(.)$ denotes the link function, then a non-normal (linear) autoregression model of order P would be of the form

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i + \sum_{p=1}^{P} \rho_p y_{i,j-p} \tag{1.6}$$

Here the value of the current location parameter is expressed as a linear combination of the P last responses. The effect of these past responses on the state is measured by the parameters ρ_p. Most of the time, the further the past response, the lesser influence it has on the current state of the process. Therefore we expect the parameter estimates $\widehat{\rho}_p$ to be a decreasing function of p.

Of course conditioning on, say, two past responses, is impossible at the first two observation times. Hence their contributions cannot be included in the likelihood function. Remember that this last function can easily be computed using a conditioning argument similar to the one used to construct Equation (1.1).

In practice, we restrict conditioning to only a few past observations precisely to reduce the incidence of the last problem. Of course, it is particularly acute with experiments recording a lot of short series of data in order to control the inter-unit variability. Assessing the order of an autoregression model is not an easy task because deviances computed on different order autoregression models will not be based on the same number of observations.

Note that it is not clear if conditioning should be done on the past responses or on some transformation of them. For example, one might feel more comfortable with the expression

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_i + \sum_{p=1}^{P} \rho_p g(y_{i,j-p}) \tag{1.7}$$

although the last quantity in that equation is not always defined. For example, the log of a zero count will often arise in a Poisson model with the log link.

An alternative approach, giving a more satisfactory answer to the problem of the autoregression order, expresses the transformed location parameter as a function of the covariates and past residuals:

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_i + \sum_{p=1}^{P} \rho_p \left[g(y_{i,j-p}) - \boldsymbol{x}_{i,j-p}^T\boldsymbol{\beta}_i\right] \tag{1.8}$$

The definition of the residual is obviously arbitrary. Here, as in the last model, some quantities might be undefined for particular values of the response. Consider now an autoregression model of order two. For the first observation, no past residuals can be computed. Because residuals are quantities that vary around zero when the systematic part of the model is properly defined, dropping the residual correction, or in other words, setting it equal to zero, will just express the fact that the only available information on the process before the first observation, is the value of the covariates. For the second observation, only one residual can be computed, the one related to the lag two observation before being set to zero. From the third observation time on, the process is completely defined and can correct for the model insufficiencies. In summary, a 'classical' type of model is assumed when no past history is available, and corrections are added to that quantity when the model is observed to under or over-estimate the past data. The important advance in this setting, compared to the usual autoregression models, is the possibility to compare models of different autoregression orders, because the same number of observations enters the likelihood. Of course the order of the autoregression will be limited by the size of the series. But this is not a technical limitation anymore.

Non-normal autoregression models will be extended to deal will continuous time longitudinal data in Chapter 2.

There exist alternative methods to model series of normal data in continuous time. One of these, based on *state-space* models and the *Kalman filter*, allows to derive the likelihood in both a flexible and numerically convenient manner. This is the subject of the next section.

1.4.3 THE DYNAMIC LINEAR MODEL AND THE KALMAN FILTER

What follows summarizes the main ideas of the paper by Jones and Boadi-Boateng (1991) who use the Kalman filter to model unequally-spaced (normal) longitudinal data. Related references include Jones and Ackerson

(1990) and Jones (1993). The starting point is the Laird and Ware (1982) marginal model for the vector of observations made through time on unit i:

$$\mathrm{E}(\boldsymbol{Y}_i|\boldsymbol{\gamma}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i \tag{1.9}$$

with

$$\begin{aligned}\mathrm{var}(\boldsymbol{Y}_i|\boldsymbol{\gamma}_i) &= \sigma^2\boldsymbol{\Sigma}_i \\ \boldsymbol{\gamma}_i &\sim N_{r_i}(\mathbf{0}, \sigma^2\boldsymbol{\Gamma})\end{aligned}$$

As usual, the vector $\boldsymbol{\beta}$ stands for the regression parameters related to the covariates in the design matrix $\boldsymbol{X}_i$. The parameters in the $r_i \times 1$ vector $\boldsymbol{\gamma}_i$ are assumed to be random. We have already justified the need for a random intercept to model inter-unit variability. We just do as if our patients were drawn at random from a larger population. Similar arguments have led to *random coefficient models* (see Section 1.5) involving other random parameters such as the slope of a line in a linear model. For example, in an experiment studying the evolution of rat weights under two different treatments, one might feel the necessity to assume that the growth rate of rats with time is varying randomly across the population. This is precisely the type of thing that the above equation is modelling. It can be easily shown that the corresponding deviance is

$$\sum_i \left[n_i \log(2\pi\sigma^2) + \log\left|\boldsymbol{Z}_i\boldsymbol{\Gamma}\boldsymbol{Z}_i^T + \boldsymbol{\Sigma}_i\right| + \frac{1}{\sigma^2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T(\boldsymbol{Z}_i\boldsymbol{\Gamma}\boldsymbol{Z}_i^T + \boldsymbol{\Sigma}_i)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right] \tag{1.10}$$

The classical methods of estimation would directly derive the score vector and set it equal to zero to obtain equations to find the MLEs. This is not without problems when the number n_i of data recorded on patient i is large. Indeed the above procedure requires the computation of matrix inverses of size equal to n_i.

Note that we have not yet discussed the problem of serial association modelling. Basically, in a marginal context, this can be done by giving some structure to the variance $\sigma^2\boldsymbol{\Sigma}_i$ of the observation random vector. For example, if an autoregression of order one is desired, then one can set $\boldsymbol{\Sigma}_i$ equal to the matrix in Equation (1.5). That would add one more parameter to the deviance function. An algorithm to perform non-linear optimization would then be required with, at each step of the process, the other linear parameters estimated using non-iterative methods. Concretely, the value of $\boldsymbol{\beta}$ and σ^2 maximizing Equation (1.10) respectively are

$$\widehat{\boldsymbol{\beta}} = \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{tot,i}^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{tot,i}^{-1} \boldsymbol{X}_i \right) \tag{1.11}$$

and

$$\widehat{\sigma}^2 = \frac{1}{\sum_i n_i} \sum_i \left[(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}) \boldsymbol{\Sigma}_{tot,i}^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}) \right]$$

where

$$\boldsymbol{\Sigma}_{tot,i} = \boldsymbol{Z}_i \boldsymbol{\Gamma} \boldsymbol{Z}_i^T + \boldsymbol{\Sigma}_i$$

Replacing the above estimated parameters by their values in the deviance function, we finally get the profile deviance

$$\sum_i n_i \left[1 + \log(2\pi\widehat{\sigma}^2) \right] + \sum_i \log |\boldsymbol{\Sigma}_{tot,i}| \tag{1.12}$$

which is a sole function of ρ if Equation (1.5), defining an AR(1) structure, used to model serial association. That last parameter is precisely the one for which a non-linear optimizer is required.

With long time series, the above procedure is numerically very demanding. Therefore we shall present an alternative method based on the Kalman filter that computes the parameter estimates and the likelihood function recursively. We shall use the convenient notation ϵ_{ij} to describe the residual part (*innovation* in econometrics) of the model. According to our above assumption, it is normally distributed with mean zero and variance σ^2.

Autocorrelation in a discrete time setting can be defined, similarly to Equation (1.3), by relating ϵ_{ij} to its value at the previous observation time:

$$\mathrm{E}(\epsilon_{ij} | \mathcal{F}_{i,j-1}) = \rho \epsilon_{i,j-1} \tag{1.13}$$

where again $\mathcal{F}_{i,j-1}$ denotes the history of subject i up to the last observation time. More generally, when observations are sampled at irregularly spaced time points, or when missing (completely) at random data arise in a discrete time setting, one can define a continuous AR(1) which assumes that two observations made on the same subject $\Delta t_{ij} (= t_{ij} - t_{i,j-1})$ units of time apart are correlated by a quantity of $\zeta(\Delta t_{ij}) = \mathrm{e}^{-\rho \Delta t_{ij}}$. Equation (1.13) then becomes

$$\mathrm{E}(\epsilon_{ij} | \mathcal{F}_{ij-}) = \zeta(\Delta t_{ij}) \epsilon_{i,j-1} \tag{1.14}$$

The expression

$$\epsilon_{ij} - \mathrm{E}(\epsilon_{ij} | \mathcal{F}_{ij-}) \tag{1.15}$$

is the difference between the residual observed at time t_{ij} and its 'expected' (or modal) value. The 'imprecision' of that prediction is estimated by the conditional variance of the last expression:

$$\mathrm{var}(\epsilon_{ij}|\mathcal{F}_{ij-}) = \sigma^2 \left[1 - \zeta(\Delta t_{ij})^2\right]$$

Note that it grows away from zero as the time lag between the two observations increases.

Consider now the *state vector*

$$\boldsymbol{\theta}_{ij} = \begin{bmatrix} \epsilon_{ij} \\ \boldsymbol{\gamma}_i \end{bmatrix}$$

where, as explained above, ϵ_{ij} models the residual part of the response that is not explained by either the systematic part or the random effects. Given the AR(1) structure of the latter random variable (and Equation (1.14) in particular), one can easily show that the prediction for the state vector at time t_{ij} given its past history is

$$\boldsymbol{\theta}_{i,j|j-1} = \mathrm{E}(\boldsymbol{\theta}_{ij}|\mathcal{F}_{ij-}) = \begin{bmatrix} \zeta(\Delta t_{ij}) & 0 \\ 0 & \boldsymbol{I}_{r_i} \end{bmatrix} \boldsymbol{\theta}_{i,j-1|j-1} \tag{1.16}$$

where $\boldsymbol{I}_{r_i}$ denotes the $r_i \times r_i$ identity matrix. Given that the random effects are not evolving with time, but just vary from one individual to the other, we do not predict any systematic change to them. Of course, our knowledge about them will evolve, becoming more and more precise as we accumulate observations on each unit.

The conditional covariance matrix of the state vector is

$$\Theta_{i,j|j-1} = \begin{bmatrix} \Theta^{(1,1)}_{i,j-1|j-1}\zeta(\Delta t_{ij}) - \zeta(\Delta t_{ij})^2 + 1 & \Theta^{(1,2)}_{i,j-1|j-1}\zeta(\Delta t_{ij}) \\ \Theta^{(2,1)}_{i,j-1|j-1}\zeta(\Delta t_{ij}) & \Theta^{(2,2)}_{i,j-1|j-1} \end{bmatrix} \tag{1.17}$$

Note that the state vector is initialized to zero, whereas the initial conditional variance is chosen to be

$$\Theta_{i,1|0} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \Gamma \end{bmatrix}$$

Consider now the *innovation vector*

$$\boldsymbol{e}_{ij}^T = \left[\boldsymbol{x}_{ij}^T \; y_{ij}\right] - \left[1 \; \boldsymbol{z}_{ij}^T\right] \boldsymbol{\theta}_{i,j|j-1}$$

The conditional covariance of that row vector is given by

$$v_{i,j|j-1} = \Theta^{(1,1)}_{i,j|j-1} + 2\Theta^{(1,2)}_{i,j|j-1}\boldsymbol{z}_{ij} + \boldsymbol{z}_{ij}^T\Theta^{(2,2)}_{i,j|j-1}\boldsymbol{z}_{ij}$$

Once the observation at time t_{ij} is available, the state vector and its variance matrix can be updated using, for example, Bayes theorem, yielding

$$\boldsymbol{\theta}_{i,j|j} = \boldsymbol{\theta}_{i,j|j-1} + \boldsymbol{\Psi}_{i,j|j-1}\boldsymbol{e}_{ij}^T$$
$$\Theta_{i,j|j} = \begin{bmatrix} \Theta^{(1,1)}_{i,j|j-1} & \Theta^{(1,2)}_{i,j|j-1} \\ \Theta^{(2,1)}_{i,j|j-1} & \Theta^{(2,2)}_{i,j|j-1} \end{bmatrix} - v_{i,j|j-1}\boldsymbol{\Psi}_{ij}\boldsymbol{\Psi}_{ij}^T$$

where

$$\boldsymbol{\Psi}_{ij} = \frac{1}{v_{i,j|j-1}} \begin{bmatrix} \Theta^{(1,1)}_{i,j|j-1} + \Theta^{(1,2)}_{i,j|j-1}\boldsymbol{z}_{ij} \\ \Theta^{(2,1)}_{i,j|j-1} + \Theta^{(2,2)}_{i,j|j-1}\boldsymbol{z}_{ij} \end{bmatrix}$$

The quantities $(1/v_{i,j|j-1})\boldsymbol{e}_{ij}\boldsymbol{e}_{ij}^T$ and $\log(v_{i,j|j-1})$ are cumulated at each step of the recursion, yielding respectively a matrix and a scalar. The matrix contains $\sum_i \boldsymbol{X}_i^T\boldsymbol{\Sigma}_{tot,i}^{-1}\boldsymbol{X}_i$ completed by the column vector $\sum_i \boldsymbol{X}_i^T\boldsymbol{\Sigma}_{tot,i}^{-1}\boldsymbol{y}_i$, required to compute $\widehat{\boldsymbol{\beta}}$ from Equation (1.11). The scalar can be used to compute the deviance, as justified by a factorization of the likelihood similar to Equation (1.1),

The Kalman filter technique is easy to compute for a linear normal model. It is typically dynamic, in the sense that, as soon as one new observation is available, the new MLEs for the parameters of interest can simply be computed (in one iteration) through an update of the old parameter estimate. A short review of the literature on related subjects is proposed in Section 1.7.2.

The dynamic linear model (DLM) makes the assumption that the data are normally distributed. In the same manner as in an independence setting, where the generalized linear model transposes the ideas underlying the normal linear regression to a non-normal context, it would be interesting to build a model generalizing the DLM. *Dynamic generalized linear models* are a possible answer to this problem.

1.4.4 DYNAMIC GENERALIZED LINEAR MODELS

Dynamic generalized linear models (DGLM) extend the normal dynamic linear model, presented in section 1.4.3, to deal with distributions in the exponential family. One main contribution to the subject is due to West, Harrison and Migon (1985) in a forecasting context. We refer to Section 1.7.2 for further references.

Unfortunately, the original paper assumed that only one time series was available. Therefore we shall drop the index i in the following lines. Generalizations of this approach to deal with multiple series will be the subject

of the coming chapters. But note that the West *et al.* (1985) approach can be applied when several units are available if no parameter is common to any two times series.

Denote by

$$\Pr(y_j|\eta_j,\phi) = \exp\{\phi\,[y_j\eta_j - a(\eta_j)]\}\,b(y_j,\phi) \tag{1.18}$$

an exponential dispersion family distribution for the response at time t_j. Consider the linear model

$$\eta_j = g(\mu_j) = \boldsymbol{x}_j^T\boldsymbol{\beta}_j \tag{1.19}$$

for the location parameter μ_j. Note that ϕ is just a scale parameter such as the variance in a normal setting. The basic idea is to assume that the regression parameters are allowed to vary with time in a dynamic way. That kind of assumption is fundamental in economic contexts where a static model is neither convenient, nor realistic. Indeed, when studying economic indices, quick changes of behaviour are likely to arise at any time, making the current model obsolete to the current trend in the data.

One way to get these desirable properties is to give to all the regression parameters a distribution that is updated with time as new observations become available. This is precisely a task for which the Bayesian alternative (if you are frequentist or Fisherian) is more than tempting. To enable closed forms for both the marginal distribution of the response and for the posterior of the parameters, a conjugate prior will be preferred to any other arbitrary choice. Instead of directly working with a multivariate prior for $\boldsymbol{\beta}_j$, a univariate prior for η_j will be chosen. Denote this last prior by

$$\begin{aligned} p(\eta_j|\mathcal{F}_{j-1}) &= c(\alpha_j,\xi_j)\exp\left[\alpha_{j|j-1}\eta_j - \xi_{j|j-1}a(\eta_j)\right] \\ &\sim \mathrm{CP}(\alpha_{j|j-1},\xi_{j|j-1}) \end{aligned} \tag{1.20}$$

Suppose now that the first two moments of the regression parameter evolve according to the relations

$$\mathrm{E}(\boldsymbol{\beta}_j|\mathcal{F}_{j-1}) = \boldsymbol{\beta}_{j|j-1} = \boldsymbol{G}_j\boldsymbol{\beta}_{j-1|j-1} \tag{1.21}$$

$$\mathrm{var}(\boldsymbol{\beta}_j|\mathcal{F}_{j-1}) = \Theta_{j|j-1} = \boldsymbol{G}_j\Theta_{j-1|j-1}\boldsymbol{G}_j^T + \boldsymbol{B}_j \tag{1.22}$$

These expressions just result from a stochastic relation predicting the future regression parameters as a linear combination of their past values plus some random noise to indicate our ignorance of what will happen to them in the future. For example, if one wants to build a linear expression in time, then Equation (1.21) could be chosen to be

$$\mathrm{E}\left[\begin{pmatrix}\beta_{0j}\\ \beta_{1j}\end{pmatrix}\middle|\,\mathcal{F}_{j-1}\right] = \begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}\beta_{0,j-1|j-1}\\ \beta_{1,j-1|j-1}\end{pmatrix} \tag{1.23}$$

with β_{0j} and β_{1j} respectively denoting the intercept and the slope of the regression at time t_j. The linear regression in time is then obtained by setting

$$\boldsymbol{x}_j^T = (1\ 0)$$

in Equation (1.19). Similar state-space equations can be built to define seasonal components for example. The choice of the matrix $\boldsymbol{B}_j$ in Equation (1.22) is of course totally arbitrary, as are the initial moments of the regression parameters. But the influence of a vague prior distribution, as the one resulting from a large $\boldsymbol{B}_1$ matrix, will quickly become negligible as observations become available.

Of course, defining relations between subsequent values of the first two moments of the regression parameter will not remain without influence on the distribution of η_j, particularly on the coefficients $\alpha_{j|j-1}$ and $\xi_{j|j-1}$ in Equation (1.20). Indeed, from Equations (1.19) and (1.20), we conclude that

$$\begin{aligned}
\mathrm{E}\left[\eta_j|\mathcal{F}_{j-1}\right] &= \boldsymbol{x}_j^T \boldsymbol{G}_j \boldsymbol{\beta}_{j-1|j-1} \\
\mathrm{var}\left[\eta_j|\mathcal{F}_{j-1}\right] &= \boldsymbol{x}_j^T (\boldsymbol{G}_j \Theta_{j-1|j-1} \boldsymbol{G}_j^T + \boldsymbol{B}_j)\boldsymbol{x}_j
\end{aligned}$$

which clearly will put constraints on $\alpha_{j|j-1}$ and $\xi_{j|j-1}$.

The response conditional distribution in Equation (1.18) can be rewritten as $\Pr(y_j|\eta_j, \phi, \mathcal{F}_{j-1})$ to stress the dependence of the regression parameters on past responses. As soon as the observation at time t_j becomes available, the posterior distribution of η_j can be computed using Bayes theorem

$$\begin{aligned}
p(\eta_j|\mathcal{F}_j) &\propto p(\eta_j|\phi, \mathcal{F}_{j-1}) \Pr(y_j|\eta_j, \phi, \mathcal{F}_{j-1}) \\
&\sim \mathrm{CP}(\alpha_{j|j-1} + \phi y_j, \xi_{j|j-1} + \phi)
\end{aligned} \tag{1.24}$$

As predicted, the posterior distribution for η_j has a closed analytic form and is still in the conjugate family. The main problem will be to compute the posterior distribution of $\boldsymbol{\beta}_j$ while taking Equation (1.24) into account. Note that Bayes theorem cannot simply be used, because we only have made assumptions on the first two moments of $\boldsymbol{\beta}_j$, leaving its distribution undefined. West *et al.* (1985) overcome this problem by using the quadratic risk function

$$\mathrm{trace}\mathrm{E}_{\eta_j|\mathcal{F}_{j-1}}\left[(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j|j})(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j|j})^T \middle| \eta_j, \mathcal{F}_{j-1}\right] \tag{1.25}$$

Their first assumption is that the posterior mean $\boldsymbol{\beta}_{j|j-1}$ of $\boldsymbol{\beta}_j$ is a linear function of η_j, as suggested by Equation (1.19). Injecting this linear expression into the risk function, and determining the corresponding linear

coefficients by minimizing that quantity, we finally get the posterior moments

$$\boldsymbol{\beta}_{j|j} = \boldsymbol{\beta}_{j|j-1} + \Theta_{j|j-1}\boldsymbol{x}_j \left\{ \mathrm{E}\left[g(\eta_j)|\mathcal{F}_j\right] - \mathrm{E}\left[g(\eta_j)|\mathcal{F}_{j-1}\right] \right\} / \mathrm{var}\left[g(\eta_j)|\mathcal{F}_{j-1}\right]$$
$$\mathrm{var}(\boldsymbol{\beta}_j|\mathcal{F}_j) = \Theta_{j|j-1} - \Theta_{j|j-1}\boldsymbol{x}_j\boldsymbol{x}_j^T\Theta_{j|j-1} \left\{ 1 - \mathrm{var}\left[g(\eta_j)|\mathcal{F}_j\right] / \mathrm{var}\left[g(\eta_j)|\mathcal{F}_{j-1}\right] \right\} / \mathrm{var}\left[g(\eta_j)|\mathcal{F}_{j-1}\right]$$

by using the relation

$$\mathrm{var}\left[\eta_j, \boldsymbol{\beta}_j|\mathcal{F}_{j-1}\right] = \boldsymbol{x}_j^T\Theta_{j|j-1}$$

Note that Equation (1.25) really provides the posterior moments, because it can easily be shown that $\boldsymbol{\beta}_j$ is conditionally independent of y_j given $(\eta_j|\mathcal{F}_j)$.

The recursion is now completely defined. The only thing still required is the likelihood. The total likelihood can be decomposed as the product of the conditional (or predictive) distributions

$$\Pr(y_j|\mathcal{F}_{j-1}) = \frac{c(\alpha_{j|j-1}, \xi_{j|j-1})}{c(\alpha_{j|j-1} + \phi y_j, \xi_{j|j-1} + \phi)} b(y_j, \phi)$$

A grid of values can be considered to determine the value of ϕ maximizing the likelihood. More systematically, a gamma prior distribution can be given to ϕ. We refer the interested reader to the original paper for more information.

Other authors have tried to generalize the dynamic linear model to deal with discrete data. The next section is devoted to one of these alternatives in discrete time. It will further be generalized, in Ch. 3, to model count data in continuous time.

1.4.5 OTHER APPLICATIONS OF THE KALMAN FILTER: MODELS FOR COUNT DATA

Harvey and Fernandes (1989), and later Ord *et al.* (1993), have proposed alternative methods for modelling series of count, binary and multinomial data. Basically, parameters, such as the mean or the proportion of successes in Bernouilli trials, are given a conjugate distribution that is updated using Bayes theorem as observations become available. Explanatory variables can be introduced, but the corresponding regression coefficients are assumed to be fixed. Thus this method is essentially comparable with the West *et al.* (1985) DGLM, although, here, no quasi-likelihood argument is required. This is at the cost of non-stochastic slope or seasonal components. However that last restriction should not be considered as a handicap because very often, data do not contain enough information to support non-fixed effects for parameters other than the intercept. As in the last section, they assume

that only one series is observed and that time is discrete, extensions to multiple units and continuous time being the subject of the coming chapters (see also Lindsey, 1993, pp. 58–60, 206–209 and Lindsey and Lambert, 1995).

Suppose that Y_j is Poisson distributed with mean μ_j and that the mean is given a conjugate gamma distribution:

$$\Pr(y_j|\mu_{j-1}, \mathcal{F}_{j-1}) = \mathrm{e}^{-\mu_j}\frac{\mu_{j-1}^{y_j}}{y_j!}$$

$$p(\mu_{j-1}|\mathcal{F}_{j-1}) = \frac{\upsilon_{j-1|j-1}^{\kappa_{j-1|j-1}}}{\Gamma(\kappa_{j-1|j-1})}\mathrm{e}^{-\upsilon_{j-1|j-1}\mu_{j-1}}\mu_{j-1}^{\kappa_{j-1|j-1}-1}$$

From the prior distribution of μ_{j-1}, we should build a predictive distribution for time t_j that expresses the loss of information between the times t_{j-1} and t_j. One way to ensure this is to make the predictive mean $\kappa_{j|j-1}/\upsilon_{j|j-1}$ equal to its previous (filtered) value and the predictive variance $\kappa_{j|j-1}/\upsilon_{j|j-1}^2$ larger than its last filtered estimation. Setting

$$\kappa_{j|j-1} = \omega\kappa_{j-1|j-1}$$
$$\upsilon_{j|j-1} = \omega\upsilon_{j-1|j-1}$$

with $0 < \omega < 1$ gives these properties. Note that the matrix $\boldsymbol{B}_j$ in Equation (1.22) in the West *et al.* (1985) approach plays a role similar to ω. Here we do not have a matrix, because, as mentioned at the start of this section, the regression parameters are assumed to be non-stochastic.

The prior distribution can be updated as soon as a new observation is available. The corresponding posterior is still a gamma with parameters

$$\begin{aligned}\kappa_{j|j} &= \kappa_{j|j-1} + y_j \\ \upsilon_{j|j} &= \upsilon_{j|j-1} + 1\end{aligned} \tag{1.26}$$

As usual, the influence of the first prior distribution on the final results can be made negligible by taking $\kappa_{0|0}$ and $\upsilon_{0|0}$ close to zero, yielding a flat prior (because of its large variance).

The distribution of Y_j conditional on its past history can be obtained by integrating out the mean parameter

$$\begin{aligned}\Pr(y_j|\mathcal{F}_{j-1}) &= \int_0^\infty \Pr(y_j|\mu_j, \mathcal{F}_{j-1})p(\mu_j|\mathcal{F}_{j-1})\, d\mu_j \\ &= \frac{\Gamma(\kappa_{j|j-1}+y_j)}{\Gamma(y_j+1)\Gamma(\kappa_{j|j-1})}\upsilon_{j|j-1}^{\kappa_{j|j-1}}(1+\upsilon_{j|j-1})^{-(\kappa_{j|j-1}+y_j)}\end{aligned}$$

giving a negative binomial distribution.

The likelihood can then be computed by multiplying (over j) these conditional distributions altogether. The unknown parameter in this last quantity is ω which can be estimated by minimizing the corresponding deviance (using a non-linear optimizer or by considering a grid of values for ω).

One way to introduce explanatory variables into the model is to work with a mean μ'_j, instead of μ_j, with distribution

$$p(\mu'_j|\mathcal{F}_j) \propto \mathrm{e}^{-v_{j|j}\mu'_j \exp(-\boldsymbol{x}_j^T\boldsymbol{\beta})}\mu_j'^{\kappa_{j|j}-1}$$

Equations (1.26) then become

$$\begin{aligned}\kappa'_{j|j} &= \kappa'_{j|j-1} + y_j \\ v'_{j|j} &= v'_{j|j-1} + \exp(\boldsymbol{x}_j^T\boldsymbol{\beta})\end{aligned}$$

Harvey and Fernandes (1989) also show how to deal with binomial and multinomial distributions. In the first case, they use a conjugate beta prior for the proportion of successes, yielding a beta-binomial compound distribution. Again, choosing a conjugate prior ensures that analytic solutions for posteriors and marginals result from the iterative process. In the second situation, a Dirichlet prior, generalizing the beta distribution to deal with multiple proportions, is chosen. The resulting compound distribution is the Dirichlet multinomial.

The DGLM can further be generalized to deal with a vector of observations. This is the subject of the next section.

1.4.6 THE MULTIVARIATE DYNAMIC GENERALIZED LINEAR MODEL

Generalizing the West *et al.* (1985) and the Harvey and Fernandes (1989) models to deal with multivariate data is rather complex, and traditional methods require the repeated computation of multidimensional integrals. An interesting answer to that problem was proposed by Fahrmeir (1992) who introduces a family of multivariate dynamic generalized linear models to analyse series of observations in the exponential family.

Using the same notation as in the previous sections, consider the conditional exponential density of the multivariate response $\boldsymbol{Y}_j$ at the j^{th} observation time

$$\Pr(\boldsymbol{y}_j|\boldsymbol{\beta}_j, \mathcal{F}_{j-1}) = \exp\left[\boldsymbol{\eta}_j^T\boldsymbol{y}_j - a(\boldsymbol{\eta}_j) - c(\boldsymbol{y}_j)\right]$$

together with the multivariate generalized linear model

$$\boldsymbol{\mu}_j = \boldsymbol{h}(\boldsymbol{X}_j^T\boldsymbol{\beta}_j)$$

where $\boldsymbol{\eta}_j$ is the natural parameter expressed as a function $\boldsymbol{\mu}_j$. The regressors in $\boldsymbol{X}_j$ can include any type of explanatory variable, including past

responses. The (normally distributed) regression parameters $\boldsymbol{\beta}_j$ are assumed to evolve according to the transition equations

$$\begin{aligned} \mathrm{E}(\boldsymbol{\beta}_j|\mathcal{F}_{j-1}) &= \boldsymbol{G}_j\boldsymbol{\beta}_{j-1|j-1} \\ \mathrm{var}(\boldsymbol{\beta}_j|\mathcal{F}_{j-1}) &= \boldsymbol{\Theta}_{j|j-1} = \boldsymbol{G}_j\boldsymbol{\Theta}_{j-1|j-1}\boldsymbol{G}_j^T + \boldsymbol{B}_j \end{aligned} \tag{1.27}$$

The posterior density of $\boldsymbol{\beta}_j$ can be computed using Bayes theorem. Unfortunately compounding of an exponential family distribution with a normal multivariate density usually does not give a closed analytic form for the posterior, unless the data are normally distributed. Therefore heavy numerical integration procedures are required. As an alternative, Fahrmeir (1992) proposes to estimate $\beta_{j|j}$ by its posterior mode. Choosing to work with posterior modes instead of posterior means avoids the repeated computation of multidimensional numerical integrals. Hence the posterior log-likelihood

$$\sum_{k=1}^{j}\left[\log \Pr(\boldsymbol{y}_k|\boldsymbol{\beta}_k, \mathcal{F}_{k-1}) - \frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\beta}_{k|k-1})^T \mathrm{var}(\boldsymbol{\beta}_k|\mathcal{F}_{k-1})^{-1}\right.$$
$$\left.(\boldsymbol{\beta}_k - \boldsymbol{\beta}_{k|k-1})\right] - \frac{1}{2}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_{0|0})^T \mathrm{var}(\boldsymbol{\beta}_0|\mathcal{F}_0)^{-1}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_{0|0})$$

has to be maximized. An approximate iterative solution to the above maximization problem can be derived using a simplified version of the Fisher scoring algorithm. The stepwise procedure can be summarized in three steps:

(1) The *prediction* step given by Equation (1.27) with arbitrary initial values for $\boldsymbol{\beta}_{0|0}$ and $\boldsymbol{\Theta}_{0|0}$.

(2) The *correction* step

$$\boldsymbol{\beta}_{j|j} = \boldsymbol{\beta}_{j|j-1} + \boldsymbol{K}_j(\boldsymbol{y}_j - \boldsymbol{\mu}_j)$$
$$\boldsymbol{\Theta}_{j|j} = \left[\boldsymbol{I} - \boldsymbol{K}_j\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}_j^T}\boldsymbol{X}_j^T\right]\boldsymbol{\Theta}_{j|j-1}$$

with

$$\boldsymbol{K}_j = \boldsymbol{\Theta}_{j|j-1}\boldsymbol{X}_j\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}_j^T}$$
$$\left[\left(\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}_j^T}\right)^T \boldsymbol{X}_j^T\boldsymbol{\Theta}_{j-1|j-1}\boldsymbol{X}_j\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}_j^T} + \mathrm{var}(\boldsymbol{Y}_j|\boldsymbol{\beta}_j, \mathcal{F}_{j-1})\right]^{-1}$$

(3) The *smoother* step

$$\boldsymbol{\beta}_{k-1|t} = \boldsymbol{\beta}_{k-1|k-1} + \boldsymbol{\Theta}_{k-1|k-1}\boldsymbol{G}_k^T\boldsymbol{\Theta}_{k|k-1}^{-1}\left(\boldsymbol{\beta}_{k|t} - \boldsymbol{\beta}_{k|k-1}\right)$$

$$\Theta_{k-1|t} = \Theta_{k-1|k-1} + \Theta_{k-1|k-1} G_k^T \Theta_{k|k-1}^{-1} \left(\Theta_{k|t} - \Theta_{k|k-1}\right) \Theta_{k|k-1}^{-1} G_k \Theta_{k-1|k-1}$$

for values of k in $\{t, \ldots, 1\}$.

The smoother step cannot be used if the above model is aimed to compute predictions. Indeed parameters at a given time are corrected using (future) information that is not yet known at that time. Actually, everything works as if we had a 'multiple pass' Kalman filter. This is not unacceptable because maximum likelihood estimates are computed in order to fit the whole data set in the most acceptable way. The shape of the fitted curve at the beginning of the process depends on the values observed further ahead in time to avoid too important a lack of fit at future time points.

The choice of the initial conditions can be optimized using the EM algorithm (Anderson and Hinde, 1988).

Distributions outside the exponential family can be used to model non-normal longitudinal data. Flexible alternatives in continuous time are considered in Ch. 2 and in the next section.

1.4.7 EXPONENTIAL DISPERSION MODELS

Jørgensen (1987) introduces *exponential dispersion models* which are a multivariate generalization of generalized linear models. One basic characteristic property of these models is their interpretation in terms of stochastic processes with stationary and independent increments (Jørgensen, 1992), as will be shown below.

Consider the density

$$f(\boldsymbol{y}|\lambda, \boldsymbol{\theta}) = a(\lambda, \boldsymbol{y}) \mathrm{e}^{\lambda[\boldsymbol{y}^T \boldsymbol{\theta} - \kappa(\boldsymbol{\theta})]} \tag{1.28}$$

for the vector $\boldsymbol{Y}$ of responses where $a(\lambda, \boldsymbol{y})$ is a normalizing constant and $\kappa(\boldsymbol{\theta})$ is the *cumulant function*. The parameter $\sigma^2 = 1/\lambda$ is called the *dispersion parameter*. The moment generating function is

$$M(\boldsymbol{s}|\lambda, \boldsymbol{\theta}) = \mathrm{e}^{\lambda[\kappa(\boldsymbol{\theta} + \boldsymbol{s}/\lambda) - \kappa(\boldsymbol{\theta})]}$$

The mean response is thus $\boldsymbol{\mu} = \boldsymbol{\tau}(\boldsymbol{\theta})$, where $\boldsymbol{\tau}(\boldsymbol{\theta}) = \partial\kappa/\partial\boldsymbol{\theta}$, and the *variance function*

$$V(\boldsymbol{\mu}) = \left.\frac{\partial^2 \kappa}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \boldsymbol{\tau}^{-1}(\boldsymbol{\mu})}$$

The distribution defined by Equation (1.28) will be denoted by $ED(\boldsymbol{\mu}, \sigma^2)$.

An important property of exponential dispersion models is their closed form under convolution of densities with identical means. Consider the random variables $\boldsymbol{Y}_i \sim ED(\boldsymbol{\mu}, \sigma^2/w_i)$ $(i = 1 \ldots)$. Then we have

$$\frac{1}{w_.}\sum_i w_i \boldsymbol{y}_i \sim ED(\boldsymbol{\mu}, \sigma^2/w_.)$$

Hence the property also holds for the well-known normal, gamma and inverse Gaussian distributions which are examples of exponential dispersion distributions.

Similar things can be done with discrete random variables. Consider the family of *discrete exponential dispersion* models

$$f(\boldsymbol{y}|\lambda, \boldsymbol{\theta}) = a(\lambda, \boldsymbol{y})\mathrm{e}^{\boldsymbol{y}^T\boldsymbol{\theta} - \lambda\kappa(\boldsymbol{\theta})}$$

denoted by $\boldsymbol{Y} \sim ED^*(\lambda, \boldsymbol{\theta})$, where the components of the response can only take values in a discrete set. It can easily be shown that $\boldsymbol{Y}/\lambda \sim ED(\boldsymbol{\mu}, \sigma^2)$, with the same notation as above. Hence the mean and the variance functions related to $\boldsymbol{Y}$ are respectively given by $m = \lambda\boldsymbol{\mu}$ and $\lambda V(\boldsymbol{\mu})$ and we also have the convolution property

$$ED^*(\lambda_1, \boldsymbol{\theta}) * ED^*(\lambda_1, \boldsymbol{\theta}) = ED^*(\lambda_1 + \lambda_2, \boldsymbol{\theta})$$

Therefore, any stochastic process $\{Y(t) : \rho t \in \Lambda\}$ where

$$\begin{aligned} Y(t) &\sim ED^*(\lambda = \rho t, \theta) \\ Y(0) &= 0 \end{aligned}$$

with Λ denoting the set of possible values for λ, can be seen as the sum of independent increments

$$\Delta Z(t_i) = Z(t_i) - Z(t_{i-1}) \sim ED^*(\rho(t_i - t_{i-1}), \theta)$$

When $\lambda = \mathbb{R}^+$, the process is infinitely divisible, and can be considered in continuous time.

Univariate (discrete) exponential dispersion models are characterized by their variance function $V(\mu)$. These include the well-known families $ED^{(p)}(\mu, \sigma^2)$ of distributions with a power variance function $V(\mu) = \mu^p$, including the Poisson ($p = 1$), the gamma ($p = 2$), the normal, the inverse Gaussian and some stable distributions. An interesting property of these distributions is that they are closed under scale transformations:

$$cY \sim ED^{(p)}(c\mu, c^{2-p}\sigma^2)$$

for any positive constant c. Note also the form of their cumulant generating function

$$\kappa(\theta) = \begin{cases} \mathrm{e}^{\theta} & (p = 1) \\ -\log(-\theta) & (p = 2) \\ -\frac{1}{p-2} & (p \neq 1, 2) \end{cases} \tag{1.29}$$

Asymptotic properties of the MLEs of $\boldsymbol{\beta}$ (where $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$) and of the deviance

$$D(\boldsymbol{y};\boldsymbol{\mu}) = -2\left\{\left[\boldsymbol{y}^T\boldsymbol{\tau}^{-1}(\boldsymbol{\mu}) - \kappa\left(\boldsymbol{\tau}^{-1}(\boldsymbol{\mu})\right)\right] - \sup_{\boldsymbol{\theta}}\left[\boldsymbol{y}^T\boldsymbol{\theta} - \kappa(\boldsymbol{\theta})\right]\right\}$$

can be derived when either the sample size is large or the dispersion parameter σ^2 tends to be small (*small dispersion* asymptotic theory). These can be used to infer about the parameters of interest when the dispersion parameter is known. When $\lambda = 1/\sigma^2$ is not known, the estimate based on the modified profile likelihood

$$f(\boldsymbol{y}|\boldsymbol{\mu},\sigma^2) = \frac{1}{(2\pi\sigma^2)^{k/2}\,|V(\boldsymbol{y})|^{1/2}}\mathrm{e}^{-\lambda D(\boldsymbol{y};\boldsymbol{\mu})}$$

can be used jointly with the asymptotic properties to build appropriate F-tests.

A new dispersion model can be derived from two given dispersion models ED_1^* and ED_2^*. Suppose that the distributions of $\boldsymbol{Y}_i^{(\lambda)} = \boldsymbol{Y}_i/\lambda$ $(i = 1, 2)$ are given by

$$\begin{aligned}\boldsymbol{Y}_1^{(\lambda)}|\boldsymbol{Y}_2^{(\lambda)} &\sim ED_1^*(\lambda r + \boldsymbol{y}_2^{(\lambda)^T}\boldsymbol{q}, \boldsymbol{\theta}_1)\\ \boldsymbol{Y}_2^{(\lambda)} &\sim ED_2^*(\lambda, \boldsymbol{\theta}_2')\end{aligned}$$

Then it can be shown that the joint density of $(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$ is of the form

$$f(\boldsymbol{y}_1, \boldsymbol{y}_2|\lambda, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = a(\lambda, \boldsymbol{y}_1, \boldsymbol{y}_2)\exp\{\lambda[\boldsymbol{y}_1^T\boldsymbol{\theta}_1 + \boldsymbol{y}_2^T\boldsymbol{\theta}_2 - \kappa_{12}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]\} \tag{1.30}$$

where

$$\begin{aligned}\boldsymbol{\theta}_2 &= \boldsymbol{\theta}_2' - \boldsymbol{q}\kappa_1(\boldsymbol{\theta}_1)\\ \kappa_{12}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= r\kappa_1(\boldsymbol{\theta}_1) + \kappa_2\left(\boldsymbol{\theta}_2 + \boldsymbol{q}\kappa_1(\boldsymbol{\theta}_1)\right)\end{aligned}$$

The joint distribution in Equation (1.30), denoted by

$$(\boldsymbol{Y}_1, \boldsymbol{Y}_2) \sim ED_1 \times ED_2(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \sigma^2)$$

is still an exponential dispersion model called the *combination* of ED_1 and ED_2. The (marginal) resulting means for $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are respectively $\boldsymbol{\mu}_1 = \boldsymbol{\tau}_1(\boldsymbol{\theta}_1)(r + \boldsymbol{\mu}_2^T\boldsymbol{q})$ and $\boldsymbol{\mu}_2 = \boldsymbol{\tau}_2(\boldsymbol{\theta}_2')$. The marginal distribution or *mixture distribution* of $\boldsymbol{Y}_1$, which is still an exponential dispersion model, is denoted by

$$\boldsymbol{Y}_1 \sim ED_1 \bigwedge ED_2(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \sigma^2)$$

Similar results also hold with discrete exponential dispersion models.

Exponential dispersion models can be used together with this last technique to analyse longitudinal data. Consider, as a first example, an autoregressive model of order one with gamma distributions:

$$\begin{aligned} Y_1 &\sim ED^{(2)}(\mu_1, \sigma^2) \\ Y_i|Y_{i-1} &\sim ED^{(2)}(\mu_i' y_{i-1} q, \sigma^2 (y_{i-1} q)^{-1}) \quad i = 2 \ldots n \end{aligned}$$

Taking a log-linear model, $\log(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$, for the marginal means, one can show that the successive conditional means are given by

$$\begin{aligned} \log(\mu_1) &= \boldsymbol{x}_1^T \boldsymbol{\beta} \\ \log(\mu_i' y_{i-1} q) &= \log(y_{i-1}) + (\boldsymbol{x}_i - \boldsymbol{x}_{i-1})^T \boldsymbol{\beta} \quad i = 2 \ldots n \end{aligned}$$

Another example is provided by the *exponential dispersion linear growth model* in Jørgensen *et al.* (1994b),

$$\begin{aligned} X_0 &\sim ED_1^*(\tau_1^{-1}(\alpha/\eta), \eta) \\ \Delta X_i &\sim ED_1^*(\tau_1^{-1}(\beta/\lambda), \lambda \Delta t_i) \\ Y_i|X_i &\sim ED_2^*(\tau_2^{-1}(1/\delta), \delta X_i) \end{aligned}$$

from which we conclude that

$$\begin{aligned} &\mathrm{E}(X_0) = \alpha, && \mathrm{Var}(X_0) = \eta V_1(\alpha/\eta) \\ &\mathrm{E}(\Delta X_i) = \beta \Delta t_i, && \mathrm{Var}(\Delta X_i) = \lambda \Delta t_i V_1(\beta/\lambda) \\ &\mathrm{E}(Y_i|X_i) = X_i, && \mathrm{Var}(Y_i|X_i) = \delta X_i V_2(1/\delta) \end{aligned} \tag{1.31}$$

The response conditional mean is thus linear in time with intercept α and slope β.

These two examples illustrate the generality of the exponential dispersion family framework together with the availability of results in continuous time.

We now propose to treat heterogeneity more systematically, although it was already accounted for in several of the above models.

1.5 Random effects models

In this section, we propose to review the main ideas underlying random effects models. The need for random effects was already mentioned in Section 1.3 where heterogeneity was described as one of the main sources of variability in biological data. Neglecting this feature in a model can lead to inexact statements on the effect of covariates on the response. Inter-unit variability is usually easily detectable when the observed response is

plotted against time, where profiles related to different units have parallel evolutions.

Suppose that the location parameter under study is first modelled as a function of the covariates

$$\eta_{ij} = g(\mu_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i$$

The ideal solution to the heterogeneity problem would be to have a different intercept λ_i for each unit, thereby allowing for the parallelism between profiles. The number of parameters would thus be growing with the number of units under observation. Estimating these parameters is not a problem when the observed series are large, but directly maximizing the likelihood with short series can yield inconsistent estimates for the parameters of interest. Different alternatives have been proposed to deal with these intercepts. Of course the estimates of these parameters are usually not of direct interest to the modeller. They are just a convenient way to take the inter-unit variability into account. Hence eliminating these nuisance parameters is desirable.

When a sufficient statistic T_i for λ_i is available, conditioning on it provides a conditional likelihood with which inference on the remaining parameters can be done. Exponential family distributions, where λ_i appears linearly in a linear expression describing the canonical parameter, provide such a statistic. The Rasch model (Rasch, 1960) is an example.

Unfortunately, things are not always as simple. Using a probit model instead of a logistic means that the model does not provide a sufficient statistic for λ_i anymore. In these situations, alternative methods have to be used. One usually assumes that the units under observation form a random sample from a larger population that we want to study. Giving a distribution $h(.|\delta)$ to λ_i is then a possible way to model this. The traditional choice for the compounding distribution is the normal. When the response is also normally distributed, the marginal distribution

$$\int h(\lambda_i|\delta) \prod_j f(y_{ij}|\boldsymbol{\theta}_{ij}, \lambda_i) d\lambda_i \tag{1.32}$$

of the vector of responses on unit i, obtained by integrating out the random effect, is multivariate normal with a covariance matrix similar to the one in Equation (1.4). With a non-normal distribution for the response, an analytical form for the marginal distribution is usually not available. Numerical integration or alternative methods such as the EM algorithm (Anderson and Hinde, 1988) with generalized linear models, have to be used to compute the marginal density of the response.

Another convenient approach in the exponential family is to consider the conjugate distribution (Diaconis and Ylvisaker, 1979) for the random

effect instead of the normal. Closed analytic forms for Equation (1.32) can then be obtained under special circumstances. But the method fails with, for example, the logistic model, when explanatory variables are included in the systematic part.

As already mentioned, heterogeneity is the result of unobserved and thus uncontrolled characteristics distinguishing apparently identical units. The complexity of biological mechanisms, the cost of data collection, and the need for a simple model are some causes of the problem. Denote by x_j $(j = 0, 1, \ldots)$ the variables influencing the response and assume that only x_0 is observed. Let $c(x_0, x_1, \ldots)$ be the function relating the location parameter of the response distribution to the covariates. Locally, the location parameter can be approximated by a linear function of the observed explanatory variable x_0:

$$\begin{aligned}\eta = g(\mu) &= c(0, x_1, \ldots) + \left.\frac{\partial c(x_0, x_1, \ldots)}{\partial x_0}\right|_{x_0=0} x_0 \\ &= \beta_0(x_1, \ldots) + \beta_1(x_1, \ldots)x_0\end{aligned}$$

The regression parameters are function of the unobserved explanatory variables which therefore vary from one unit to the other. That mathematical derivation can easily be extended to several known covariates.

If the chosen sample is the result of a random allocation of patients to the different treatment branches in a clinical trial, then we expect the unknown covariates $x_1, \ldots$ to be 'fairly' distributed among the different strata. Extending the random intercept model to a more general framework where all the regression coefficients are given a distribution can then be justified, yielding the *random coefficient* model (Lindsey, 1993, pp. 52-53). This motivates the Laird and Ware (1982) model (see Section 1.4.3).

1.6 Marginal models

Although this work will focus on the development of conditional models, we now give a short review of marginal models for longitudinal data. Note that we shall not give a complete review and that the length of this section, compared to the one related to conditional models, might mislead the reader about the relative importance of these two subjects in the literature. However, because our approach will essentially be based on conditional arguments, we have preferred mainly to devote the introduction to related materials.

Another way of tackling longitudinal data is to focus attention on population averaged effects rather than subject specific effects. We have already given a word of caution on this in Section 1.2, mainly based on Simpson's paradox. Basically, quantities such as the marginal expectation of the response are directly modelled as a function of the explanatory variables,

and this separately from the association among the measurements made on the same unit. The popular tool for producing estimates of the regression parameters in marginal models is the generalized estimating equations (GEE) approach (Liang and Zeger, 1986; Zeger and Liang, 1986; Prentice, 1988; Zeger, Liang and Albert, 1988). The starting point of that methodology is that, in general, ignoring serial association and using a standard analysis assuming independence yields underestimated standard errors for time-stationary effects and overestimated s.e.'s for time-varying effects. For exponential family distributions, we know that the score equations for the regression parameters in an independence model are

$$\sum_{i=1}^{N} \frac{\partial l_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T \text{var}^{-1}(\boldsymbol{Y}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = 0 \tag{1.33}$$

where

$$\begin{aligned} \boldsymbol{Y}_i &= (y_{i1}, \ldots, y_{in_i})^T \\ \text{var}(\boldsymbol{Y}_i) &= \text{diag}\,[\text{var}(Y_{i1}), \ldots, \text{var}(Y_{in_i})] \end{aligned} \tag{1.34}$$

The idea would be to generalize these equations to deal with dependent data. Because the independence is expressed through the diagonal form of $\text{var}(\boldsymbol{Y}_i)$, one way to loosen this assumption would be to put non-zero off-diagonal elements in the covariance matrix in Equation (1.34). Unfortunately the resulting score Equation (1.33) cannot be integrated back to a proper likelihood function in most situations (McCullagh and Nelder, 1989, 2nd Ed., Section 9.3.2). Therefore no proper probability model really underlies the GEE. That means that comparing the performances of two such 'models' is impossible by likelihood or AIC.

In practice, the covariance matrix is replaced by a *working* covariance matrix of the form

$$\text{var}(\boldsymbol{Y}_i) = \boldsymbol{D}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{D}_i^{1/2}$$

where

$$\boldsymbol{D}_i = \text{diag}\,[\text{var}(Y_{i1}), \ldots, \text{var}(Y_{in_i})]$$

and $\boldsymbol{R}_i(\boldsymbol{\alpha})$ is a parameterized *working* correlation matrix. There are many different ways to define this matrix:

- $\boldsymbol{R}_i(\boldsymbol{\alpha}) = \boldsymbol{I}_{n_i}$ assumes that the data are independent.
- $[\boldsymbol{R}_i(\boldsymbol{\alpha})]_{jk} = \alpha$ for $j \neq k$, is the equivalent of a random intercept model assuming a constant correlation between observations made on the same unit.

- $[\boldsymbol{R}_i(\boldsymbol{\alpha})]_{jk} = \alpha^{|t_j - t_k|}$ is the equivalent of an autoregression model where the correlation between any two observations made on the same unit is (exponentially) decreasing with the time lag between them.

Of course using a correlation matrix puts constraints on the sample estimate of it. Mainly, the resulting estimate has to be positive definite, which might be an important technical problem. Some authors (Lipsitz *et al.*, 1991) use the more appropriate odds ratio as a measure of association between binary responses instead of the correlation matrix.

Further estimating equations are required to estimate the vector $\boldsymbol{\alpha}$ in the working correlation matrix. Let $r_{i,jk}$ be an estimate of the correlation $\rho_{i,jk}$ between the observations at time t_{ij} and t_{ik} on unit i. Then a possible set of extra estimating equations (one for each component of $\boldsymbol{\alpha}$) for the correlation matrix is

$$\mathrm{U}(\boldsymbol{\alpha}) = \sum_{i=1}^{I} \frac{\partial \boldsymbol{\rho}_i(\boldsymbol{\alpha})^T}{\partial \boldsymbol{\alpha}} \boldsymbol{R}_i(\boldsymbol{\alpha})^{-1} [\boldsymbol{r}_i - \boldsymbol{\rho}_i(\boldsymbol{\alpha})] \tag{1.35}$$

where $\boldsymbol{\rho}_i$ is a $n_i(n_i - 1)$ column vector containing all the elements of the upper right off-diagonal elements of the correlation matrix $\boldsymbol{R}_i(\boldsymbol{\alpha})$, $\boldsymbol{r}_i$ denoting the corresponding set of estimates. The consistent solution $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ to Equations (1.33) and (1.35) can be used to estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ by

$$\left[\sum_{i=1}^{I} \left(\frac{\widehat{\partial \mu_i}}{\partial \boldsymbol{\beta}^T}\right)^T \widehat{\mathrm{var}(\boldsymbol{Y}_i)}^{-1} \left(\frac{\widehat{\partial \mu_i}}{\partial \boldsymbol{\beta}^T}\right)\right]^{-1}$$

This estimate might be inconsistent if the working correlation matrix is misspecified. Liang and Zeger (1986) propose a robust estimate of $\mathrm{var}(\boldsymbol{\beta})$ which is consistent even when the correlation matrix is not the 'right' one. This is of course also true for the regression parameter estimator, a good specification of the systematic part $\boldsymbol{\mu}_i(\boldsymbol{\beta})$ being the only requirement. Wald tests can then be performed to assess the need for a given explanatory variable. These tests rely on the asymptotic normality of the estimates, which, with small sample sizes, might lead to unrealistic conclusions.

However, Crowder (1995) proposed a simple example where consistency does not hold, showing that a problem exists in the method. There could be 'no general asymptotic theory supporting existence and consistency of $(\widehat{\alpha}, \widehat{\boldsymbol{\beta}})$'.

The advantage of the GEE method is its extreme simplicity. It can be used to estimate any regression parameter in a marginal (exponential family) model. Moreover the association between observations on the same unit can be specified through a single correlation matrix. The form of that matrix is not really important because asymptotically consistent estimators

for both the regression parameters and their covariance matrix are still available when it is misspecified. It can easily cope with observations in continuous time.

From our point of view, simplicity is not necessarily a good argument to choose a method of analysis. Going from the general conclusion that, say, some treatment increases the proportion of recoveries, to the more subject specific claim (which, to a patient, is the only statement of interest) that there is an improvement in the health of a given patient under that treatment, is to us a dangerous step that probably misunderstands the complex links between the random variables in a marginal longitudinal model. Moreover, focusing the approach on the estimating step rather than on the modelling task, is rather artificial. Remember that the GEEs are generally not the score equations of a likelihood function. Finally, the associated tests rely on the normality of the estimators, which is not always a sensible assumption, as many examples from the log-linear methodology tend to show (Lindsey, 1995a).

However, there exist a few marginal methods for which a likelihood is available. An example of this is given in Molenberghs and Lesaffre (1994), where a full likelihood approach is used to analyse ordinal categorical responses based on an extension of the Plackett distribution. The chosen measure of association is the global cross-ratio. A second example is given by Becker and Balagtas (1993) who propose a marginal model to analyse two-period binary cross-over data.

Marginal and conditional methods are equivalent when the data are normal. The Laird and Ware (1982) model presented in Section 1.4.3 can be both analysed using conditional (the Kalman filter) and marginal arguments, leading to the same solution. Note that the GEE correspond to a proper likelihood function in the normal case. This is not surprising, because the normal distribution only requires the specification of the systematic part and of the (mean independent) covariance matrix to be completely defined.

1.7 Further reading

This section aims to review a selected subset of recent papers on longitudinal data analysis (LDA) that was used to write this chapter, and to give the reader some complementary information on the subject. This review is not meant to be comprehensive. An extended bibliography on repeated measurements up to 1993 can be found in Lindsey (1993). Finally note that the three books by Diggle *et al.* (1994), Fahrmeir and Tutz (1994) and Lindsey (1993) introduce the reader to multiple aspects of longitudinal data analysis.

1.7.1 AUTOREGRESSION MODELS

Continuous response Diggle and Zeger (1989) model the level of luteinizing hormone in the blood, which typically exhibits sudden increases named *pulses*. A normal autoregressive model of order one plus a gamma distributed pulse occurring with a probability defined in terms of the process history, is used to describe the hormone evolution measured in discrete time.

Schmid *et al.* (1994) allows for measurement error on both the covariates and the response in an autoregressive model. OLS and maximum likelihood estimates are compared in a simulation study.

Heitjan (1991b) relates an immunologic outcome variable to the level of drug received by patients in a clinical trial using a normal AR(1) model with a patient random effect. The mean of the non-linear response is modelled using the generalized logistic function which is defined in terms of biologically meaningful parameters.

Glasbey (1979) proposes a model for growth using a first-order normal autoregression with a non-linear mean. Different transformations of the response are considered. See Glasbey (1980) for a higher order model.

Discrete data Albert *et al.* (1994) model the number of (gadolinium enhanced) brain lesions detected using magnetic resonance imaging, to compare the activity of multiple sclerosis in treated and placebo patients. They first use an ('observation driven') Poisson autoregression model where, in addition to the usual seasonal components, serial association is modelled by conditioning on past residuals. They also propose an alternative ('parameter driven') approach where the probabilities of relapse (or increasing count) and remission (or decreasing count), defined by comparing the present count with the one at the previous observation time, are modelled using a latent Markov chain, which introduces autocorrelation. These two constructions are compared using the AIC.

Francis (1994) models longitudinal count data observed in continuous time using non-linear growth curves. Past transformed observations enter the systematic part through regression parameters multiplied by quantities exponentially decreasing to zero with the time elapsed since the corresponding datum was observed.

Korn and Whittemore (1979) analyse series of binary data indicating if the panelist under observation had respiratory problems on each day of an approximately eight-month study. They use logistic regression models with a patient random effect, where (in addition to covariates) conditioning is made on the previous day's outcome to take into account of the higher risk of asthma on the days following an attack. Note that the inference procedure used is approximate: the regression parameters are evaluated separately for each unit and some kind of average of these estimates is considered to yield the required parameter values. Cox and Snell (1970,

pp. 96-102), Lindsey (1993, Ch. 6), Muenz and Rubinstein (1985), Slud and Kedem (1994), Zeger, Liang and Self (1985) also use Markov chains to model dependence in series of binary outcomes.

Keenan (1982) relates series of binary data to an underlying normal model for continuous and correlated longitudinal data.

Albert (1994) proposes a Markov model to analyse sequences of ordinal data from a relapsing-remitting disease. The transition probabilities are decomposed as the product of two conditional probabilities with a biologically interpretable parameterization. The method is illustrated on a data set describing the effects of an experimental allergic encephalomyelitis on mice randomized to placebo or treatment. The observations were made daily on a forty day period and were recorded on a five level scale describing the severity of the disease symptoms.

Muenz *et al.* (1985) use Markov chains to study the psychological impact of breast cancer.

Gottschau (1992) models multivariate series of binary data using Markov chains and shows under what conditions analysing the aggregated number of individuals occupying a given state under the hypothesis of exchangeability can simplify the analysis. Methods for testing exchangeability and the conditional independence of the present responses given the past observations are developed.

Gottschau (1994) proposes other models for analysing the same kind of data. The case where each individual is exposed to a common risk factor between time $t-1$ and time t is considered. Conditional independence cannot be retained anymore and the Rasch model with a parameter common to all the individuals at time t is introduced. This parameter is then given a distribution and the likelihood obtained by conditioning on the total number of individuals occupying a given state, is used to infer about the within-group parameters. A separate analysis of the evolution of the total number of alike responses can then be made independently of the within-group analysis, thanks to the likelihood factorization. It is also shown how to introduce extra explanatory variables. Both the log-linear and logistic models can be used to estimate the parameters. A practical example studying the occurrence of bacteria in milk samples from each of the four teats of dairy cows is treated. The adequacy of the different models in Gottschau (1992 and 1994) is then assessed.

Exponential family Lindsey (1993, pp. 54-56) and Li (1994) show how to condition on past residuals in a GLM setting.

Shephard (1994) considers autoregression models in an exponential family setting where the a Taylor series expansion of the link transformed last response is used as regressor in the systematic part. That avoids problems such as the log of a zero observation in a Poisson model.

In a quasi-likelihood approach, Zeger and Qaqish (1988) consider a class of Markov models in which the conditional mean and variance given the past are explicit functions of past outcomes. It is illustrated with the analysis of interspike times for motor cortex neurons of a monkey using a gamma quasi-likelihood 'model'.

Miscellaneous Stanek *et al.* (1989) give a word of caution on misspecified forms for the functional relationship between time and the response, that can lead to false conclusions when assessing the need for some covariates in a model.

Some further references on autoregressive models include Rosner *et al.* (1985, 1988) in an epidemiological context. Their models can deal with missing data and continuous time.

1.7.2 DYNAMIC MODELS

Introductory papers to dynamic (generalized) linear models are proposed by Meinhold and Singpurwalla (1983), Diderrich (1985) and Bolstad (1986).

Discrete data Bolstad (1995) models count processes using a multiprocess dynamic Poisson model. The goal of the author is to build a model that quickly reacts to real parameter change while being insensitive to outliers. The mean parameter is assumed to be gamma distributed. The last posterior distribution for that mean is modified in three possible ways, to be used as a prior at the next observation time. First, if the observation to come is not an outlier, then the predictive distribution of the mean at time t is set equal to the last posterior at time $t-1$. Second, if the next observation indicates a real change in the data pattern, then the predictive variance is set to three times the last posterior variance while leaving the predictive mean equal to the posterior mean. Finally, if the observation is an outlier, then the predictive moments are transformed in the same way as in the second case, but the prior is not updated using Bayes theorem (as in the first two situations) to give a new posterior. The status of the observation to come is measured using a three-state probability distribution. That index distribution is updated as soon as a new observation is available. The conditional (on the index value at times t and $t-1$) posterior distribution of the mean is computed using Bayes theorem. The dependence on the index value at the previous time point can be eliminated by computing the corresponding marginal. This involves the weighted sum of three gamma distributed random variables. The resulting marginal can be approximated (or *condensed*) in an optimal way using the gamma distribution at the minimum Kullbeck-Liebler distance of the mixture. This way of estimating the new gamma parameters is compared with the traditional identification procedure of the first two moments of the desired gamma and of the mixture using simulations . The Kullbeck-Liebler distance method

seems to give substantial improvement. The paper does not consider the influence of explanatory variables on the response, though it seems easy to introduce.

Jørgensen *et al.* (1994a) use a Poisson model with a latent gamma process to model the number of emergency room visits due to four types of respiratory diseases. The latent process location parameter is expressed in terms of *long-term* covariates which are explanatory variables thought to have a cumulated effect on people's health. This inertia in the effects of these covariates is allowed precisely because they enter into the structure of the state vector. On the other hand, *short-term* covariates enter the model directly through the observation location parameter. Hence they have an immediate and temporary effect on the risk of respiratory disease. The model is able to cope with multivariate series of counts and to induce correlation among data in different series through the common latent process. Further details about the model setting and the estimation of the parameters are given in Jørgensen *et al.* (1995).

Ord *et al.* (1993) give more details on the models presented in Harvey and Fernandes (1989). An approximation to the posterior in the beta-binomial model involving covariates is also proposed.

Grunwald *et al.* (1993) propose the Dirichlet distribution to model series of 'continuous' (i.e. only known as percentages) proportions. This distribution is reparameterized to separate the effects of location and dispersion parameters. A conjugate prior distribution is then given to the location parameter vector. The loss of information between two time points is expressed by taking as prior distribution at time $t + 1$, the distribution obtained by setting the posterior at time t to some power in $[0, 1]$. The resulting mode is unchanged, but the dispersion is enlarged. Updating is done in the usual way using Bayes theorem. Grunwald *et al.* (1993) also show how to introduce covariates. The choice of the conjugate distribution as a prior provides a closed form for the likelihood that can be used to make inferences.

Gamerman (1992) uses the Kalman filter to make inference on point processes where the intensity rate is allowed to depend on covariates and on the history of the process.

Exponential family Schnatter (1992) approximates the posterior moments of the regression parameters in a non-Gaussian DGLM using Gauss-Hermite numerical integration together with the approximate formulae for the posterior mode and information in Fahrmeir (1992).

Singh and Roberts (1992) propose, in discrete time, a dynamic version of the generalized linear model for multiple series of possibly non-normal data. The underlying idea is to generalize the state space linear model for normal data in the same manner as the generalization of the linear (normal) model to the generalized linear model. In the independence case, the WLS algorithm

for normal data was applied iteratively to the linearized approximation of the linear predictor, a first-order Taylor expansion of $\boldsymbol{\eta} = g(\boldsymbol{\mu})$ about $\boldsymbol{\mu}$. In the dynamic case, Singh and Roberts (1992) iteratively use the normal Kalman filter on the approximation of the linear predictor. They make the assumption that the number of observations available at each time point is sufficiently large to support the linearization $\boldsymbol{z}_j$ of $g(\boldsymbol{\mu}_j)$. The variance of $\boldsymbol{z}_j$ is estimated by

$$\left(\frac{d\boldsymbol{\eta}_j}{d\boldsymbol{\mu}_j^T}\right) \text{var}(Y_j|\boldsymbol{\mu}_j) \left(\frac{d\boldsymbol{\eta}_j}{d\boldsymbol{\mu}_j^T}\right)^T$$

evaluated at its (independence) GLM estimate. This accounts for the relation existing between the mean and the variance in non-normal models. This last idea is similar to what Zehnwirth (1988) did. Finally, the methodology is illustrated on series of epidemiological count data.

Refined Kalman filtering Zehnwirth (1988) shows how to adapt the normal Kalman filter to deal with a state dependent observation variance.

Gordon and Smith (1993) robustify the updating equation in the linear Kalman filter by preventing the updated location parameter from going too quickly to the likelihood mode, but instead staying at a reasonable distance from the prior mode when an extreme observation arises.

Meinhold and Singpurwalla (1989) show how to robustify the normal Kalman filter. They adopt the point of view that observations deviating markedly from their predictions should be given less weight to compute the posterior. The underlying belief is that the prior distribution has been carefully chosen and is not simply a technical requirement to make Bayesian inference. Therefore they really want the predictive prior distribution to be close to the computed posterior. Hence one must avoid having the posterior defined as a compromise between the prior and the likelihood. This can be done by replacing the normal distribution by one with a wider variety of possible tails. Meinhold and Singpurwalla (1989) have chosen to work with a Student-t distribution.

Kitagawa (1987) has developed a model for data that occasionally show jumps after a period of smooth and gradual change. In these situations, a linear Gaussian model with a small variance cannot cope with jumps, whereas one with a large variance can produce unjustified peaks. Distributions with thicker tails, such as *stable* distributions (with the Cauchy density as a well-known member), would obviously be more adapted. This kind of distributions can be used in an empirical Bayesian approach similar to what has been done in the Kalman filter derivation. The basic idea is to approximate probability densities by piecewise linear functions and to make inferences using these approximations. Of course that method is

numerically very demanding. The examples presented in the original paper, which are based on the Pearson family of distributions, only involve one parameter for which filtering is of interest. Extension to parameter vectors is theoretically easy, but the computation burden quickly becomes problematic.

Komaki (1993) describes pulses arising in endocrinological data in a different way than Diggle and Zeger (1989). The dynamics of the luteinizing hormone is decomposed into two parts: a linearly (exponentially) rising and an exponentially decreasing phases. Each part is modelled in continuous time using a stochastic differential equation. Actually the sampling procedure works in discrete time, but the resulting fifteen minute intervals are further (artificially) divided into ten parts. This allows smooth model transition between two hormone production phases. Two different logistic models are used to predict the start and the end of the hormone rising. Modelling the termination probability of the rising mode separately allows variability in the heights of the pulses, as observed in practice. The likelihood is then evaluated using the usual factorization in conditional settings. The procedure is essentially numerical and relies on the Kitagawa (1987) approximation to densities using piecewise linear functions.

Miscellaneous Jones and Ackerson (1990) model serial correlation in normal longitudinal data using continuous time autoregressive moving averages. In addition to the usual marginal approach, they propose to use the Kalman filter to compute the likelihood.

Recursive residuals and model diagnostic tools for normal and non-normal state-space models are developed in Schnatter (1994). The recursive residuals are defined as the predictive probability of observing a value for the response smaller or equal to its actual value, and this conditional on past outcomes. If the data were approximately generated by the model considered, then these residuals, called the *P-scores*, should be uniformly distributed on $[0, 1]$. The corresponding normal P-score*th* quantile, called the *transformed P-score*, can be used instead for model checking. It coincides with the recursive residual (which is the Pearson residual computed from the first two moments of the predictive distribution) in the normal case. An approximation to the P-score is derived using a Gauss-Hermite integration where the mixing distribution of the state vector is assumed to be normally distributed. Indices for bias, dispersion, skewness and tail are used to check the distribution of the transformed P-scores. Note that the predictive distribution used to compute the P-scores, is invariant under invertible data transformations.

Fahrmeir and Tutz (1994b) suggest a dynamic model for analysing paired comparisons made over time. The model is also suitable for ordinal responses

Gordon and Smith (1990) develop a framework for monitoring medical times series. Their goal is to detect dynamically any level or trend change in the observed time series. Four possible states are considered: the series can be stable, its level may change, the slope can also be modified, and finally a transient unusual value could be observed. Distinguishing among these four situations is important in practice, where, for example, clinicians have to anticipate possible physiological changes. Conditionally on this state, a dynamic linear model is used to model the data evolution. The variance of the state vector can further be given a gamma distribution. The different parameter distributions are updated using Bayes theorem. Integrating out the state of the observation yields a mixture. The resulting form is approximated to avoid complex numerical work. It is shown that the proposed approximation minimizes the Kullback-Liebler distances for the normal and gamma mixtures.

Smith and Miller (1986) propose a non-Gaussian state-space model in order to predict records. A record is a maxima (or minima) of some observed random variable. Some time points do not necessarily give a record. In such situations we only know that the best performance at that time is lower (or greater) that the last record, thereby yielding a censored observation.

1.7.3 RANDOM EFFECTS MODELS

Anderson and Hinde (1988) propose to use the EM algorithm to estimate the parameters in GLMs with random effects.

Fahrmeir and Tutz (1994, Ch. 7) estimate the parameters in a random effects model using different methods, among which a technique based on posterior modes.

Schall (1991) gives a simple algorithm to compute the estimates of the fixed effects, the random effects and the dispersion parameters in a generalized linear model with random effects. This algorithm approximately yields the MLEs or restricted MLEs. The distribution of the random effects does not need to be multivariate normal: it only has to be in the exponential family.

Stiratelli *et al.* (1984) propose an empirical Bayes method for estimating normal random effects in serially correlated binary responses. An approximation based on the mode of the random effects posterior distribution is used, because of the intractability of exact methods.

Davies (1987) shows that misspecification of the compounding distribution in random effects models can lead to seriously misleading estimates for the observed covariate effects. This problem only seems to be serious when the assumed underlying compound distribution highly differs from the true one. Therefore, when nothing can be said about the unknown mixing component, Davies (1987) proposes to use a non-parametric mass point distribution which assigns a discrete set of unknown possible values for the

unobserved covariate with a relative weight to be estimated. Such mass points are added to the 'naive' non-random model until the corresponding deviance stabilizes. Simulations tend to support the idea that the flexibility of the mixing distribution does not yield conservative tests for the effect of the observed covariates. Note also that the resulting marginal distribution has an easy to handle analytic form which is no longer the result of restrictive assumptions on the model structure.

With binary data, Conaway (1990) suggests using the log-log link together with a log-gamma distribution for the random intercept, instead of the traditional logistic model with a normal latent distribution. This avoids the evaluation of integrals by numerical methods because an analytic marginal is available. The log-gamma distribution can take a wide variety of shapes, including the normal bell as a limiting case. Serial association extra to the random effect can be defined using autoregression. Explanatory variables can also be introduced. The latent distribution can be chosen to be some transform of a beta, while keeping a tractable form for the marginal.

McDonald (1994) uses an additive random effect on the logit scale to model heterogeneity between series of binary outcomes. The random effect is assumed to take only a fixed number of values that are estimated from the data.

Thall (1988) models the number of events occurring in time intervals delimited by the observations times, using a Poisson distribution with a gamma mixing component to allow for heterogeneity. Alternative nonparametric methods are used in Thall and Lachin (1988).

1.7.4 MARGINAL MODELS

GEE based methods Prentice (1988) discusses a method that allows the analysis of correlated binary data with unit specific covariates. A logistic regression, with an intercept allowed to depend on the responses of the other units in the same block, is used to derive the joint distribution of the responses in the block. Estimation of the parameters by maximization of the joint likelihood is possible for a pair or triplet of responses, but the problem quickly becomes computationally infeasible. In these situations, Prentice (1988) recommends the use of the GEE.

Fitzmaurice and Lipsitz (1995) analyse series of binary responses measured in continuous time using serial odds ratios for measuring association, instead of the less adapted (and constrained by marginals) correlation matrix. Basically, the odds ratio related to two binary responses observed at consecutive time points, is modelled as a quantity decaying exponentially with time. The correlation matrix related to the vector of observations on any given unit, can then easily be derived from that model because an expression relating correlations and odds ratios in binary data is available. Estimation of the regression parameters can then be made using the GEE. They recommend the use of the GEEs proposed by Carey *et al.* (1993).

Zeger *et al.* (1985) model binary longitudinal data using a marginal logistic regression where a constant dependence between any two observations is specified using correlation. Serial association can also be modelled using Markov chains.

Thall and Vail (1990) describes a family of covariance models for series of count data modelled using quasi-likelihood arguments. Generalized estimating equations are used to evaluate both the covariate and variance-covariance parameters.

Zeger (1988) presents a marginal regression model for analysing time series of counts. Correlation between observations from the same series is introduced using an unobservable process added to the linear predictor in a log-linear model. The parameters are estimated using a procedure similar to the GEE. Campbell (1994) uses this model to investigate the relationship between the sudden infant death syndrome and the environmental temperature.

Zeger *et al.* (1988) compare subject specific and population averaged approaches in the estimation of regression parameters. This is done in a longitudinal study of the incidence of mothers smoking status on the risk of respiratory disease for their children. The parameters are estimated in both approaches using the GEE.

Paik (1992) models variance heterogeneity using an extension of the GEE procedure (Liang and Zeger, 1986).

Other methods Azzalini (1994) explains how to build a Markov chain model based on transitional probabilities, where the covariates only enter the process mean, independently of the serial association parameter.

Lipsitz *et al.* (1992) describe a three stage estimator for the regression parameters in a marginal logistic model for series of binary outcomes. The first stage makes the assumption that the data are independent. Given these estimates, residuals are computed and used to estimate the correlation matrix for the vector of observations made on the same unit. In the third and last stage, generalized least squares estimating equations based on the second step correlation matrix are used to yield the final parameter estimates.

Zhao and Prentice (1990) model serial association in series of binary data in terms of correlations and conditional odd-ratios.

Fitzmaurice and Laird (1993) model this type of data using likelihood methods based on a multivariate binary model with unknown normalizing constant. A logistic marginal regression is used to model the influence of explanatory variables on the response. Association between any two responses on the same unit is expressed in terms of a conditional odds ratios. The estimation of the parameters is made with an *iterative proportional fitting procedure* together with the Fisher scoring algorithm, despite the

unknown normalizing constant required to estimate joint probabilities for each unit. Higher-way association can be incorporated.

In the same context, Stiratelli, Laird and Ware (1984) consider a general logistic linear mixed model.

Lipsitz *et al.* (1995) use a joint marginal distribution to derive the probability of a least one 'success' in a set of repeated binary events. The parameters in this model are estimated by the MLEs when the number of observations per unit is moderate. When this number is large, they propose to use either a 'one-step' MLE, or a moment-based method close to the GEE approach.

In a review paper, Fitzmaurice *et al.* (1993) compare likelihood based approaches with non-likelihood approaches. They essentially focus on the analysis of series of binary outcomes.

Wei and Stram (1988) use quasi-likelihood methods to study the marginal distribution of response variables observed repeatedly.

Stukel (1993) compares the Zeger and Liang (1986) model with the gamma-Poisson model in the analysis of the longitudinal count data. Simulations with an underlying process consistent with a Poisson, show that the gamma-Poisson model is more efficient than the Zeger and Liang model, probably thanks to its ability to deal with heterogeneity. When the process is non-Poisson, and under non-negligible heterogeneity, the gamma-Poisson process is more efficient with fixed covariates. The situation is reversed under mild heterogeneity.

An early paper on multivariate categorical data obtained from repeated measurements is Koch *et al.* (1977) who propose different hypotheses of possible interest with the corresponding test statistics.

Agresti (1989) surveys models for repeated ordered categorical data. These models essentially rely on classical log-linear or logistic techniques where the dependence on time is simply introduced through a factor variate taking a different value at each occasion. This is done using the cumulative or adjacent-category logits.

Molenberghs and Lesaffre (1994) develop a full likelihood approach to the analysis of ordinal categorical responses based on an extension of the Plackett distribution. The chosen measure of association is the global cross-ratio.

Stram *et al.* (1988) propose a two-step procedure to analyse repeated ordered categorical data with possibly missing observations and time dependent covariates. In the first step, the data are analysed separately at each time point using a GLM. In the second step, the asymptotic distribution of the parameter estimates is derived to make inferences. The technique is illustrated by analysing the effects of indoor and outdoor air pollution on respiratory health.

1.7.5 MISSING DATA

Key references in the missing data literature are Rubin (1976), who introduces the concept of *missing at random* data where the probability for one observation to be missing is independent of the unobserved values; Little and Rubin (1987), who make a distinction between data missing at random and *completely at random* where a further independence of the missing data mechanism from the other observations is assumed.

Laird (1988), Diggle and Kenward (1994), Diggle *et al.* (1994), Robins *et al.* (1995) and Fitzmaurice *et al.* (1995) consider the problem in a longitudinal data setting.

There are numerous other papers on the subject, but the proposed ones should provide a reasonable introduction to the problem.

1.7.6 OTHER MODELS

Diggle (1988) proposes to model a (possibly transformed) continuous response using a linear normal model. The stochastic element is decomposed into three parts: a term modelling subsampling variation within units, another one modelling unit heterogeneity, and a final term modelling within-unit or measurement error autocorrelation. The variance structure of the latter element is parameterized using a reduced number of elements to avoid inefficient estimation (Altham, 1984), as with an unstructured covariance matrix. This model simplification enables one to treat series observed in continuous time. The empirical semi-variogram is used to analyse the different sources of variation in the data. An overparameterized model for the systematic part is first considered to avoid inducing spurious autocorrelation due to model misspecification. The resulting residuals are used to compute the empirical semi-variogram which helps to choose a model for the covariance structure. As soon as the different parameters in the model have been estimated, the theoretical semi-variogram can be compared with the empirical one to check the adequacy of the model.

Jennrich and Schluchter (1986), Muñoz *et al.* (1992), Núñez-Antón and Woodworth (1994) propose different structures for the normal covariance matrix to define heterogeneity and autocorrelation.

Mauger *et al.* (1995) compare different models for pulsatile series using simulation.

Ware *et al.* (1988) discusses some existing marginal and conditional methods for use in the analysis of repeated categorical outcomes.

2
Positive longitudinal data modelling

In this chapter, we propose to develop a general methodology for modelling series of non-negative data observed at unequally spaced times. This subject is a major concern in biomedical sciences where the follow-up of patients or animals often gives rise to sequences of positive observations that might have a skewed distribution. Moreover, the sampling is not necessarily done at equi-spaced time points. This can be the result of a protocol prescription where, for example, important changes are thought to happen in a limited time range, with a more stable or less interesting evolution for other time values. Or the irregular sampling is simply due to missing values in an experiment where the data should have been observed after fixed and constant time intervals. The model developed below will assume that the data are *missing (completely) at random*. Situations where the missing data mechanism is informative can arise; we refer to Section 1.7.5 for such problems.

Thus, both distributional and longitudinal aspects are the source of technical problems. In the coming lines, we present a flexible model for longitudinal data that includes a wide variety of distributions such as the normal, exponential, gamma, Weibull and log-normal. The parameterization enables both the importance of serial association, as well the 'order' of this dependence to be expressed. The theory will be illustrated by an example where the effects of three fiber based diets on dog triglyceride profiles are compared.

This chapter is an extended version of Lambert (1996c).

2.1 Data set of interest

The data were collected by a veterinary team of the nutrition department of the University of Liège (Belgium).

Two groups of four young adult Beagle dogs were used in two separate but similar experiments. A 4 by 4 Latin-square design, described in Table 2.1 where the numbers stand for the identification numbers of the dogs, was adopted. Two groups {1,7,9,10} and {2,5,8,11} of four dogs were considered in 1992 and 1993 respectively. Dogs 5 and 7 were castrated males, whereas the others were females. Each dog was given four different diets

		Diet			
	Period	C	G	I	FB
Exp 1	11.03.92	10	1	7	9
	10.04.92	9	10	1	7
	27.05.92	7	9	10	1
	22.06.92	1	7	9	10
Exp 2	16.12.92	5	11	2	8
	13.01.93	8	5	11	2
	18.02.93	2	8	5	11
	23.03.93	11	2	8	5

Table 2.1. 4 by 4 Latin-square design used to collect blood parameter values in adult Beagle dogs. The numbers in the Table are the identification numbers of the dogs.

at four different calendar times. All the diets were based on minced meat, cooked rice, maize oil and a mineral mixture (= control diet C). Diets (G), (I) and (FB) were supplemented with respectively guar gum, inuline and beet fibers. The time between two experiments on the same dog was long enough to avoid carry-over effects. Note that each of the four evaluations (of a given treatment) was performed at a different calendar time to avoid confounding with period effect. Moreover, with this cross-over design, a dog was not restricted to a given diet, avoiding a possibly unfair comparison of the treatments. Indeed it might happen that a given dog is systematically responding at a lower or higher level for a variable of interest, yielding an apparent treatment effect. However, one weakness of this particular design is that each treatment always follows the same previous one.

At a given period, for a given dog and diet, e.g. dog 10 on diet (C) on the 11 March, 1992, seven profiles related to different plasma components, namely α-amino-N (mg/l), urea (mg/l), glucose (mg/l), insulin (g/l), cholesterol (g/l), triglycerides (g/l) and non esterified fatty acid (NEFA) (μEq/l), were recorded. Blood was taken before feeding and then serially over a six hour period at times 20, 40, 60, 90, 120, 180, 240, 300 and 360 minutes. The corresponding profiles for the triglycerides are plotted on Figures 2.1, 2.2, 2.3 and 2.4, each plot corresponding to one treatment. The sampling was successfully completed for all the blood variables, except for the triglyceride and NEFA profiles where more irregular sampling was adopted. But the missing data mechanism was not informative, missing data appearing because of technical problems in the analysis of some blood samples.

We shall focus our attention on the triglyceride profiles. In the next section, we develop both the rationale and the technical aspects related to the *generalized autoregression model.*

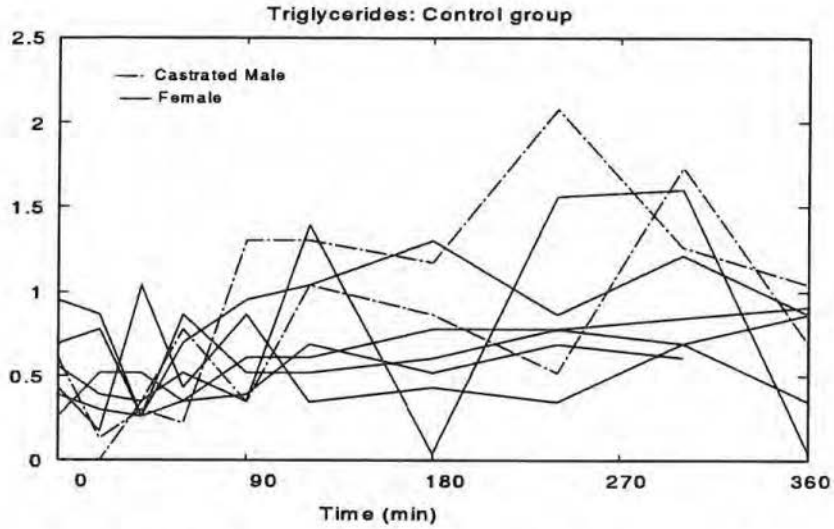


Fig. 2.1. Observed triglyceride (g/l) profiles for the control diet.

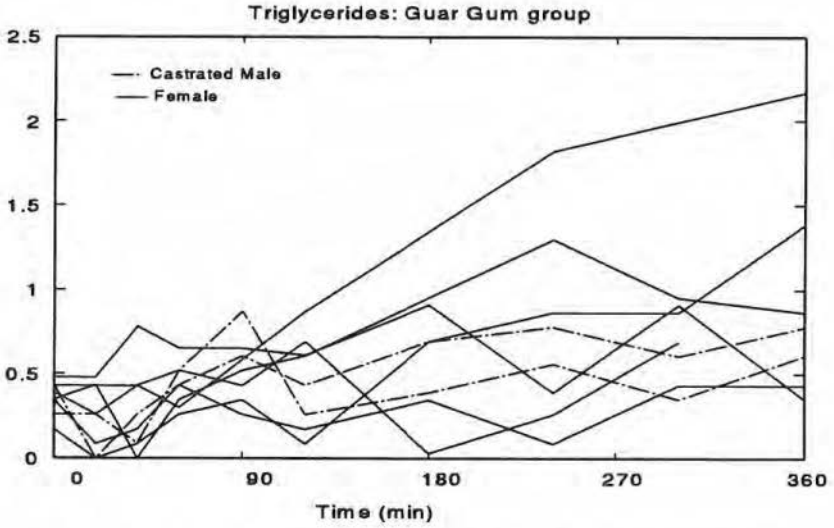


Fig. 2.2. Observed triglyceride (g/l) profiles for the guar gum diet.

2.2 Generalized autoregression models

Assuming normality when analysing the triglyceride data set might not be sensible. Indeed a quick inspection of the data set shows the possible skewness of the distribution. This is mainly due to the positive character

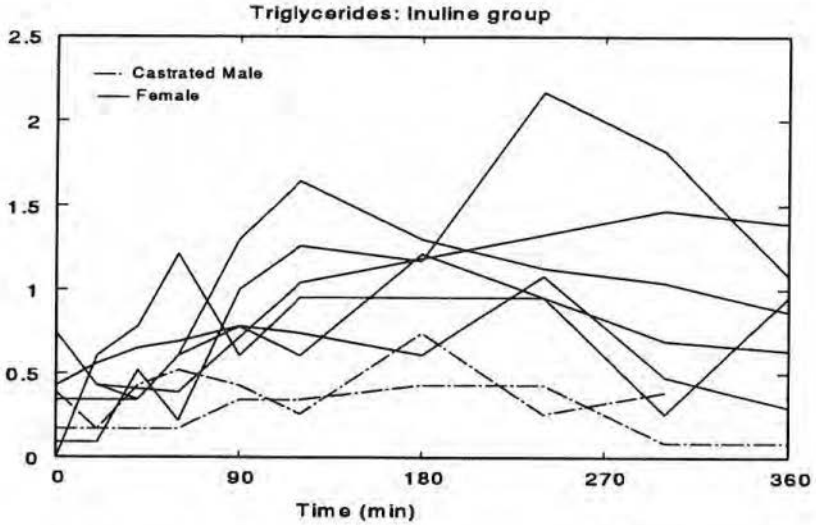


Fig. 2.3. Observed triglyceride (g/l) profiles for the inuline diet.

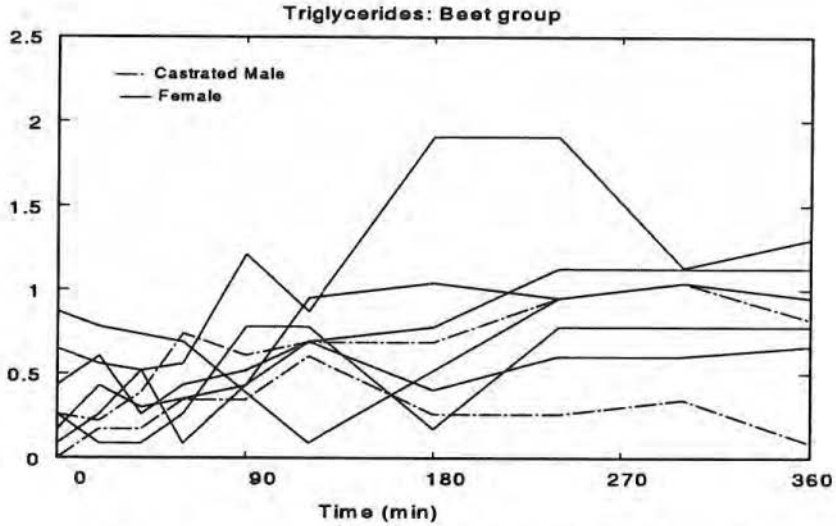


Fig. 2.4. Observed triglyceride (g/l) profiles for the beet diet.

of the response. The normal distribution does not take this into account, meaning that it allows a negative fit on some time range. However normal models have been used and assessed below on this data set. Their poor performance (measured using the AIC) in this context confirms the need for non-normal longitudinal models.

Of course one could analyse the log of the response to reduce the danger of a symmetric error term assumption, but this choice would ignore the many other sensible candidate distributions for non-negative data, such as the exponential, gamma, Weibull, power transformed normal densities, and so on, which possibly would provide a better description of the data pattern.

Specification of serial association in a non-normal setting is technically more difficult than in the normal case, because the covariance matrix does not completely specify the dependence structure.

The method of analysis that we propose is an attempt to solve these theoretical problems. It is a generalization of the autoregression models presented in Lindsey (1993, p. 55) to unequally spaced time sampling.

Consider a distribution $f(y_{ij}|\mu_{ij}, \boldsymbol{\theta}_{ij})$ (not necessarily a member of the exponential family) for the response where y_{ij} and μ_{ij} respectively denote the response and the mean response for unit i $(i = 1, \ldots, I)$ at the $j^{th}(j = 1, \ldots, n_i)$ sampling time t_{ij}. The vector $\boldsymbol{\theta}_{ij}$ stands for nuisance parameters such as scale or shape parameters.

Different families of distributions such as the normal, log-normal, exponential, gamma and Weibull will be considered below. Note that many of these densities are special cases of the generalized gamma family. Its distribution is given by

$$f(y|\kappa, \mu, \delta) = \frac{\delta \kappa^{\kappa} y^{\delta\kappa - 1}}{\mu^{\kappa\delta}\Gamma(\kappa)} \mathrm{e}^{-\kappa\left(\frac{y}{\mu}\right)^{\delta}}$$

It includes well-known members such as the exponential $(\kappa = 1, \delta = 1)$, Weibull $(\kappa = 1)$ and gamma densities $(\delta = 1)$, the log-normal being $(\kappa \to \infty)$ a limiting case.

Denote by

- $g(\cdot)$ the desired link function. One could for example take the log-link for a gamma distribution, or any other one such as the inverse (canonical) link.
- r_{ij}, the residual for unit i at time t_{ij} on the g-scale, defined by

$$r_{ij} = g(y_{ij}) - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij}$$
$$r_{i0} = 0$$

  where $\boldsymbol{\beta}_{ij}$ stands for regression parameters and $\boldsymbol{x}_{ij}$ for the vector of regressors at time t_{ij}. This definition for the residual is rather arbitrary, as pointed out by Lindsey (1993, p. 56). Note also that non-linear forms can be considered for the systematic part of the model. But, in this chapter, we shall restrict our attention to the usual linear regression setting.
- r_{ij}^c, the *cumulated residual* for unit i at time t_{ij} on the g-scale, defined by

$$r^c_{ij} = \mathrm{e}^{-\phi \Delta t_{ij}} r^c_{i,j-1} + r_{ij}$$
$$r^c_{ij} = 0$$

where $\Delta t_{ij} = t_{ij} - t_{i,j-1}$ and $0 < \phi$. Note that ϕ will be modelling the relative importance of former residuals to explain the last observation.

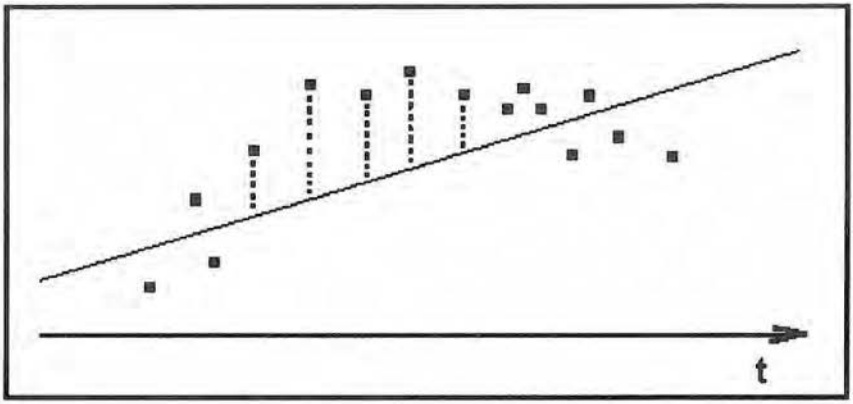
- w^c_{ij}, the weight for the cumulated residual r^c_{ij}, where

$$w^c_{ij} = \mathrm{e}^{-\phi \Delta t_{ij}} w^c_{i,j-1} + 1$$
$$w^c_{i0} = 0$$

The idea is to model the g-transform of the mean as

$$g(\mu_{ij}) = \boldsymbol{x}^T_{ij}\boldsymbol{\beta}_{ij} + \mathrm{e}^{-\rho \Delta t_{ij}} \frac{r^c_{i,j-1}}{w^c_{i,j-1}} \tag{2.1}$$

where $0 < \rho$. An illustration of this is given on the figure below. The dots



represent the data on the g-scale plotted against time for a given unit. The line is the systematic part $\boldsymbol{x}^T_{ij}\boldsymbol{\beta}_{ij}$ of the g-mean estimated by the model. We can see that a positive residual tends to be followed by a residual of the same sign and order. This is a typical situation with longitudinal data. Equation (2.1) is precisely made to account for this feature. The first term is the systematic part given by the line on the figure, while the second term is a weighted average of past residuals. The importance of this correction is decreasing with ρ and Δt_{ij}. A large value for ρ is an indication of independence. The relative importance of former residuals is decreasing with ϕ. In other words a large residual value can still induce a large residual several steps further if ϕ is close to zero.

Note also that full confidence is given to the model (and its systematic part) at the start of the series when no previous observation occurred

because, then, no past residual is available.

Estimation of ϕ, ρ, the regression parameters and $\boldsymbol{\theta}_{ij}$ can be made by maximizing the likelihood

$$\prod_i \prod_j f(y_{ij}|\mu_{ij}, \boldsymbol{\theta}_{ij})$$

where μ_{ij} is a function of ϕ, ρ and the regressors.

No random effect has been included in the model just presented. A number of technical problems arise as soon we relax the normality hypothesis for the response distribution. One could choose the conjugate distribution for a random effect for the intercept to keep a closed analytic form for the likelihood. More traditionally people tend to work with normal random effects to model the effect of unrecorded covariates. Estimation in such a context can be made using the EM algorithm (Anderson and Hinde, 1988).

Another possibility is to use a nonparametric random effect (Davies, 1987; Oskrochi, 1994). Technically speaking, one could modify Equation (2.1) to yield

$$g(\mu_{ij}) = \lambda_i + \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij} + \mathrm{e}^{-\rho \Delta t_{ij}} \frac{r_{i,j-1}^c}{w_{i,j-1}^c}$$

It is assumed that λ_i can only take a finite number k of values on the real line. Take the example of a symmetric random effect with $k = 5$. The random effect will be completely specified by the set $\{(r_0, \pi_0), (r_1, \pi_1), (r_2, \pi_2)\}$ with

$$\begin{aligned} r_{-s} &= -r_s \\ \pi_{-s} &= \pi_s \\ r_{|s|} < r_{|u|} &\quad \text{if} \quad |s| > |u| \\ \Pr(\lambda_i = r_s) &= \pi_s \\ \sum_{s=-2}^{s=2} \pi_s &= 1 \end{aligned}$$

and $0 < \pi_s < 1$ for $s \in \{0, 1, 2\}$. Hence the expression for the likelihood is

$$\prod_i \sum_s \pi_s \prod_j f(y_{ij}|\lambda_i = r_s, \mu_{ij}, \theta_{ij})$$

A non-linear optimizer can be used to maximize this likelihood.

2.3 Model selection

Different types of models have been presented so far. Some of them are reasonable candidates for modelling the dog data set:

- the Jones model (Jones and Boadi-Boateng, 1991) as a natural extension of the traditional ANOVA.
- the generalized autoregression family of models with normal, gamma, log-normal, generalized gamma, power-transformed normal density functions for $f(\cdot)$.

Comparison of these models can be made using the AIC

$$-2\sum_i \sum_j \log f(y_{ij}) + 2p$$

where $f(\cdot)$ is the density function and p the number of parameters in the model. Arguments for using the AIC instead of the traditional likelihood ratio test in a model selection context are given in Lindsey (1994).

2.4 Analysis of the triglyceride data set

Five explanatory variables are candidates for explaining the variability in the data:

- **Period** An 8-level factor variable: it takes different values for any two different calendar times and thus accounts for any possible seasonal effect on the response.
- **Sex** A 2-level factor variable: it takes values 1 for females and 2 otherwise.
- **Treatment** A 4-level factor variable: it takes values 1 for control, 2 for guar gum, 3 for inuline and 4 for beet fibers.
- $\mathbf{Y_{i0}}$ A continuous explanatory variable: it contains the value of the baseline response just before feeding.
- $\mathbf{t_{ij}}$ A continuous (non-randomized) explanatory variable: it measures the time (in minutes) elapsed since the last meal.

The transformation $g(\mu_{ij})$ of the mean for all the models described below is of the form

$$g(\mu_{ij}) = \boldsymbol{x}_i^T\boldsymbol{\beta} \; + \mathrm{e}^{-\rho\Delta t_{ij}} \frac{r^c_{i,j-1}}{w^c_{i,j-1}} \tag{2.2}$$

where

- $\boldsymbol{\beta}^T = (\beta_0 \; \boldsymbol{\beta}_i^M \; \boldsymbol{\beta}_i^{P^T} \; \boldsymbol{\beta}_i^{TrF^T} \; \boldsymbol{\beta}_i^{TrM^T} \; \alpha \; \beta_1 \; \beta_2)$. The size of these parameter vectors should be clear as soon as the row $\boldsymbol{x}_i^T$ of the design matrix has been described.

- $\boldsymbol{x}_i^T = (1\ x_i^M\ \boldsymbol{x}_i^{PT}\ \boldsymbol{x}_i^{TrF^T}\ \boldsymbol{x}_i^{TrM^T}\ y_{i0}\ t_{ij}\ t_{ij}^2)$ where
  - x_i^M is 1 if the observation was made on a male dog, and 0 otherwise.
  - $\boldsymbol{x}_{ik}^P$ is 1 if the observation took place at the k^{th} period and 0 otherwise ($k = 2 \ldots 8$).
  - $\boldsymbol{x}_{ik}^{TrF}$ is 1 if the observation took place on a female dog under treatment $Tr = k$ ($k = 1 \ldots 4$), and 0 otherwise. Similarly for males with $\boldsymbol{x}_{ik}^{TrM}$.
- $\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2$ models a quadratic profile in time.
- the last term in Equation (2.2) is responsible for the modelling of serial association.

Of course, β_2 can be set to 0 if the profile is thought to be linear.

The main results of the analysis are presented in Tables 2.2 and 2.3 where a star in the column spanned by an explanatory variable indicates that the variable appears linearly in the systematic part of the model. In the special case of the treatment column, a sequence of eight symbols (four for each sex) describes how treatment effects are modelled: a star at the i^{th} position in one of the quadruple indicates that treatment i effect is contrasted with the treatment(s) with a zero symbol. For example, the sequence '$0 * * * \mid * 000$' indicates that, for females, treatments 2, 3 and 4 are contrasted with the control diet, whereas, for males, the control diet is compared with the other 3 treatments (assumed to have the same effect on the response). This is equivalent to setting $\boldsymbol{x}_{i1}^{TrF} = 0$ and $\boldsymbol{x}_{ik}^{TrM} = 0$ ($k = 2, 3, 4$) in Equation (2.2). Note that SA and Rd stand respectively for serial association and a random effect for the intercept. The selected model is indicated by a bold AIC value.

From the AIC values, we see that the log-normal distribution is discarded. It is even worse than the normal. This result alone is very important for applied statisticians who, very often, just take the log of positive and possibly skewed data to use commercialized software dealing with series of normal observations.

The Jones (Jones and Boadi-Boateng, 1991) and normal generalized autoregression models (AIC=147.2 and 147.0 respectively) perform better, but still lie behind the three remaining families of distributions. This might be due to the non-zero probability assigned to negative responses.

Next comes the gamma family of models which markedly improves the fit (AIC=114.0). However the square-root transformed normal and Weibull generalized autoregression models have even lower AIC values (104.2 and 93.4 respectively). Note that the choice of the square root transform was suggested by the MLE 0.534 for the power coefficient in the power-transformed normal generalized autoregression model. Because of this, the corresponding number of parameters was increased by one.

Per	Sex	$\text{Tr}_{female}\|\text{Tr}_{male}$	Y_{i0}	t_{ij}	t_{ij}^2	SA	Rd	Par	AIC
Jones									
*	*	0***|0***	*	*	*	*	*	21	157.2
*	*	0***|*000	*	*	*	*	*	19	153.6
*	*	0*00|*000	*	*	*	*	*	17	150.8
	*	0*00|*000	*	*	*	*	*	10	173.5
*	*	0*00|*000		*	*	*	*	16	149.2
*	*	0000|*000		*	*	*	*	15	149.7
*	*	0*00|0000		*	*	*	*	15	149.6
*	*	0*00|*000		*		*	*	15	167.8
*	*	0*00|*000		*	*	*		15	**147.2**
Normal generalized autoregression with identity-link									
*	*	0***|0***	*	*	*	*		22	156.0
*	*	0***|*000	*	*	*	*		20	152.2
*	*	0*00|*000	*	*	*	*		18	149.0
	*	0*00|*000	*	*	*	*		10	167.4
*	*	0*00|*000		*	*	*		17	147.1
*	*	0000|*000		*	*	*		16	**147.0**
*	*	0*00|0000		*	*	*		16	147.8
*	*	0000|*000		*		*		15	196.3
Log-normal generalized autoregression with identity-link									
*	*	0***|0***	*	*	*	*		22	182.1
*	*	0***|*000	*	*	*	*		20	178.3
*	*	0*00|*000	*	*	*	*		18	175.6
	*	0*00|*000	*	*	*	*		10	191.3
*	*	0*00|*000		*	*	*		17	174.4
*	*	0000|*000		*	*	*		16	176.4
*	*	0*00|0000		*	*	*		16	**174.2**
*		0*00|0000		*	*	*		15	183.2
*	*	0*00|0000		*		*		15	191.1

Table 2.2. AICs for the triglyceride data set: normal and log-normal distributions (with a bold AIC value for the best model in the family of densities considered).

The parameter estimates in the Weibull model together with the deviance changes when they are eliminated from the best model are displayed in Table 2.4. This table shows that conditioning on the baseline response is not fundamental and that one might eliminate it. Note that some parameters (such as the intercept β_0) cannot be suppressed, and that others such as β^M have to stay in the model if e.g. a treatment-sex interaction is present. Finally the elimination of the period effect was not considered with isolated periods, but rather for the whole parameter vector β^P. This explains the blanks in the last two columns of Table 2.4.

Per	Sex	Tr_{female}|Tr_{male}	Y_{i0}	t_{ij}	t_{ij}^2	SA	Rd	Par	AIC
Gamma generalized autoregression with log-link									
*	*	0***|0***	*	*	*	*		22	121.8
*	*	0***|*000	*	*	*	*		20	118.0
*	*	0*00|*000	*	*	*	*		18	114.7
	*	0*00|*000	*	*	*	*		10	134.0
*	*	0*00|*000		*	*	*		17	**114.0**
*	*	0000|*000		*	*	*		16	117.8
*	*	0*00|0000		*	*	*		16	116.8
*	*	0*00|*000		*		*		16	131.4
Power-transformed normal gen. auto. with identity-link									
*	*	0***|0***	*	*	*	*		23	112.4
Square root-transformed normal gen. auto. with identity-link									
*	*	0***|0***	*	*	*	*		23	112.7
*	*	0***|*000	*	*	*	*		21	108.8
*	*	0*00|*000	*	*	*	*		19	105.9
	*	0*00|*000	*	*	*	*		11	125.0
*	*	0*00|*000		*	*	*		18	**104.2**
*	*	0000|*000		*	*	*		17	105.4
*	*	0*00|0000		*	*	*		17	105.0
*	*	0*00|*000		*	*	*		17	125.2
Generalized gamma generalized autoregression with log-link									
*	*	0***|0***	*	*	*	*		23	100.3
Weibull generalized autoregression with log-link									
*	*	0***|0***	*	*	*	*		22	99.5
*	*	0***|*000	*	*	*	*		20	96.3
*	*	0*00|*000	*	*	*	*		18	**93.4**
*	*	0*00|*000	*	*	*	*	*	20	97.4
	*	0*00|*000	*	*	*	*		11	120.8
*	*	0*00|*000		*	*	*		17	94.3
*	*	0000|*000	*	*	*	*		17	98.7
*	*	0*00|0000	*	*	*	*		17	99.7
*	*	0*00|*000	*	*		*		17	110.7

Table 2.3. AICs for the triglyceride data set: gamma, power-transformed normal, generalized gamma and Weibull distributions (with a bold AIC value for the best model in the family of densities considered).

The gamma, square-root and Weibull models essentially provide the same type of conclusion:

- there is a period effect.
- female profile is lower under the guar gum diet than under the other treatments.

Par.	Est.	Δdev.	P-value
β_0	-0.0979	–	–
β_1	0.00219	–	–
β_2	-0.0000128	19.3	0.000
β_i^M	-0.447	–	–
$\boldsymbol{\beta}^P$		41.4	0.000
β_2^P	0.280		
β_3^P	0.561		
β_4^P	0.143		
β_5^P	0.125		
β_6^P	0.220		
β_7^P	-0.165		
β_8^P	0.189		
β_2^{TrF}	-0.230	7.3	0.007
β_1^{TrM}	0.415	8.3	0.004
α	0.0928	2.9	0.089

Table 2.4. Parameter estimates and deviance changes when the corresponding explanatory variable is withdrawn (if it is sensible) from the selected Weibull model.

- male profile is higher under the control diet than under the other treatments.
- a degree 2 (at least) profile is required to model the triglyceride evolution.

In addition, the Weibull model indicates that the triglyceride level before diet is related to its subsequent values.

Note that the non-selected candidates are more discordant on the treatment effects, but this is not worrying in the light of their AIC values.

From a technical point of view, the zero observations caused certain problems. These zeros appeared not because the triglyceride level was zero but because the measurement process was not sensitive enough for values of the response close to zero. This was not a problem with normal models, where y_{ij} only appears in a positive power form in the deviance, but with models such as the log-normal generalized autoregression where one must compute the logarithm of a zero value. In these situations the zeros were replaced by a parameter whose MLE was evaluated.

Note also that the Weibull is a good reduction of the generalized gamma as indicated by the AIC values.

No random effect was detected as shown by the Jones and one of the Weibull models.

In terms of goodness of fit we can classify the merits of the different models. We also give, as an illustration, some estimated profile equations for females under respectively non-guar gum and guar gum diets at period

1, with $\tilde{t}_{ij}$ standing for $(t_{ij} - 180)$:

(1) Weibull generalized autoregression with log-link (18 par.; $\widehat{\phi} = 0.529$, $\widehat{\rho} = 0.101$; $\widehat{\delta} = 2.28$):

$$\mu_{ij} = \exp[-0.0979 + 0.0928\ \log(y_{i0}) + 0.00219\ \tilde{t}_{ij} - 0.0000128\ \tilde{t}_{ij}^2]$$
$$\mu_{ij} = \exp[-0.328 + 0.0928\ \log(y_{i0}) + 0.00219\ \tilde{t}_{ij} - 0.0000128\ \tilde{t}_{ij}^2]$$

(2) Square root-transformed normal generalized autoregression with identity link (17 par.; $\widehat{\phi} = 0.0253$, $\widehat{\rho} = 0.0376$):

$$\mu_{ij} = \left[0.819 + 0.000885\ \tilde{t}_{ij} - 0.00000596\ \tilde{t}_{ij}^2\right]^2$$
$$\mu_{ij} = \left[0.742 + 0.000885\ \tilde{t}_{ij} - 0.00000596\ \tilde{t}_{ij}^2\right]^2$$

(3) Gamma generalized autoregression with log-link (17 par.; $\widehat{\phi} = 0.049$, $\widehat{\rho} = 0.063$):

$$\mu_{ij} = \exp(-0.268 + 0.00224\ \tilde{t}_{ij} - 0.000015\ \tilde{t}_{ij}^2)$$
$$\mu_{ij} = \exp(-0.518 + 0.00224\ \tilde{t}_{ij} - 0.000015\ \tilde{t}_{ij}^2)$$

(4) Jones model (15 par.) and Normal generalized autoregression with identity-link (16 par.; $\widehat{\phi} = 0.0302$, $\widehat{\rho} = 0.0280$)

(5) Log-normal generalized autoregression with identity-link (16 par.; $\widehat{\phi} = 0.0200$, $\widehat{\rho} = 0.0492$)

These female profiles for the guar gum diet are plotted on Figure 2.5. One can see that all the models, except the log-normal (which is too skewed), provide similar curves. This is not a surprise: the log-normal model has the largest AIC value. The normal generalized autoregression and the Jones models behave differently from the others for small values of t_{ij}. They even provide negative values for times close to zero. This can happen (with the identity link) because data were only included in the fit after time 20, the observations at time 0 being used as covariates.

Note that we have omitted the autoregression term in the fitted profile equations because its contribution is specific to the animal under consideration. Hence the departure between e.g. the observed profiles for females under guar gum and the fitted ones in Figure 2.5 is overestimated. To conclude the analysis we mention that we have computed the deviance residuals (not displayed) to detect possible 'outliers'. Some stand out, such as the top profile on Figure 2.2. In this example the fit tends to be pushed upwards yielding a few series of negative residuals. However these data could

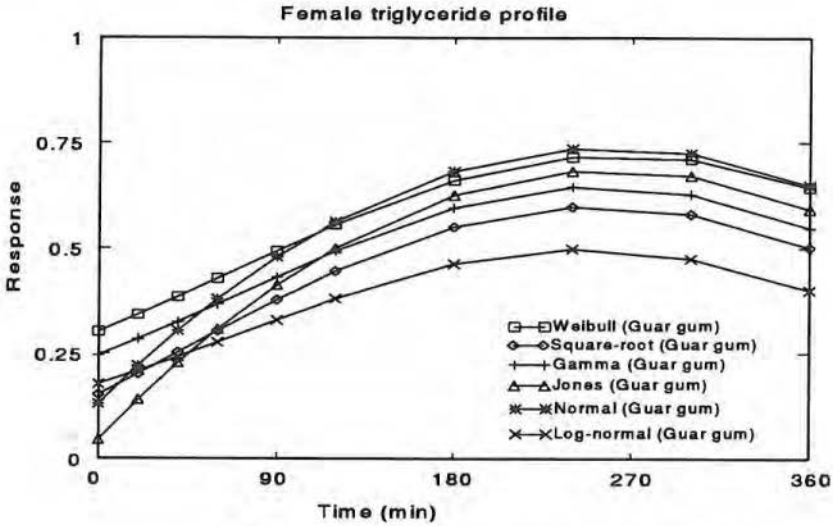


Fig. 2.5. Fitted female triglyceride profiles.

not be discarded, as pointed out by the veterinarians, because no protocol violation was detected.

3
Longitudinal count data in continuous time

Our second contribution is devoted to modelling series of count data measured at irregularly spaced time points. This has been the subject of two papers, Lambert (1996a and b), that we propose to present in a single chapter.

Models in discrete time are available. We have mentioned several methods in Section 1.7, most of them based on the Kalman filter. Related references include West *et al.* (1985), Harvey and Fernandes (1989), Ord *et al.* (1993) (See Sections 1.4.4 and 1.4.5), Lindsey (1993, pp. 58–60, 206–209), Lindsey and Lambert (1995) and Fahrmeir and Tutz (1994, Sections 8.2, 8.3 and 8.4). Our goal in this chapter, will be to extend the ideas in these references to be able to cope with count data in continuous time. Other approaches include Jørgensen *et al.* (1994) (see Section 1.4.7) who use exponential dispersion models to model linear growth in count data series, and Francis (1994) who analyses the growth of plants measured by leaf counts using an autoregression model.

3.1 Problem setting with two examples

In order to determine precisely the nature of the challenge, we now present two data sets that will be used throughout this chapter to illustrate the techniques that we have developed.

The first example concerns bovine respiratory diseases which are a major concern to workers involved in meat production. Even if efficient prophylactic measures and therapeutic drugs are available, respiratory diseases are reported to be the most frequent causes of death. In the U.K., such losses exceed 50 million pounds annually (Génicot *et al.*, 1993). For this reason veterinarians need to be able to induce acute respiratory distress syndrome artificially in healthy calves to test the effectiveness of the drugs that they are developing. This is often done by intramuscular injection of 5-hydroxytryptamine receptor blocker. The data set of interest (shown in Table 3.1 and plotted in Figures 3.1 and 3.2) consists of six series of the respiratory rates of Belgian White and Blue calves. The animals were submitted to a continuous injection of 5-hydroxytryptamine receptor blocker, at different doses over 30 minutes. The goal of the study was to find out

Time	0	5	15	25	35	40	45	50	55	60
$y_1(t_{1j})$	48	126	120	126	-	56	-	46	-	56
Dose	0	20	20	20	0	0	0	0	0	0
$y_2(t_{2j})$	33	49	46	54	-	32	-	-	-	-
Dose	0	10	10	10	0	0	-	-	-	-
$y_3(t_{3j})$	32	73	82	108	-	27	-	-	-	-
Dose	0	15	15	15	0	0	-	-	-	-
$y_4(t_{4j})$	39	58	58	88	68	-	58	-	40	-
Dose	0	15	15	15	0	0	0	0	0	-
$y_5(t_{5j})$	19	21	39	45	31	-	28	-	26	-
Dose	0	15	15	15	0	0	0	0	0	-
$y_6(t_{6j})$	19	52	43	52	26	-	27	-	22	-
Dose	0	17.5	17.5	17.5	0	0	0	0	0	-

Table 3.1. Measurements of the respiratory rate (min^{-1}) of calves submitted to a continuous injection of 5-hydroxytryptamine receptor blocker at different doses (in μgr kg^{-1} min^{-1}) during 30 minutes.

the drug doses that the calves could tolerate. A second, but not essential, objective was to model the respiratory rate profiles as a function of the dose. These profiles were observed at irregularly spaced time points during and after the drug injection. In addition to serial association, several technical difficulties appear in this data set. Firstly, the data will not be normal simply because they are counts. Therefore traditional methods, such as the autoregression model of Jones and Boadi-Boateng (1991), have to be discarded. Secondly, the observations were made at irregularly spaced time points. Moreover the same timing was not necessarily used for two different calves. Hence a discrete time setting is not acceptable any longer. And finally, fixed and time varying covariates (the dose received by the animal) have to be included in the model.

The second example concerns the growth of three closed colonies of *Paramecium aurelium* in a nutritive medium. The observed counts (in Table 3.2) are plotted in Figure 3.3. Details concerning the experiment can be found in Diggle (1990, p. 8). Our goal is to build a model for the profile of the mean number of individuals. One can easily see from Figure 3.3 that the colony growth cannot be modelled using polynomial or spline based methods. Indeed, the colony sizes seem to stabilize around day 10, suggesting a model with an asymptotic behaviour. Actually, biological models would be more appropriate than any artificial mathematical construct in this situation.

The next sections in this chapter will aim to build a flexible tool for longitudinal count data that can be used to analyse the data sets in the above two examples.

Day	0	2	3	4	5	6	7	8	9	10
Sample 1	2	17	29	39	63	185	258	267	392	510
Day	11	12	13	14	15	16	17	18	19	
Sample 1	570	650	560	575	650	550	480	520	500	
Day	0	2	3	4	5	6	7	8	9	10
Sample 2	2	15	36	62	84	156	234	348	370	480
Day	11	12	13	14	15	16	17	18	19	
Sample 2	520	575	400	545	560	480	510	650	500	
Day	0	2	3	4	5	6	7	8	9	10
Sample 3	2	11	37	67	134	226	306	376	485	530
Day	11	12	13	14	15	16	17	18	19	
Sample 3	650	605	580	660	460	650	575	525	550	

Table 3.2. Growth of three closed colonies of paramecium aurelium in a nutritive medium (Gause, 1934).

3.2 Gamma-Poisson model

We shall keep consistent with the notation used in the preceding chapters. This will facilitate the comparison of the existing models with ours.

Suppose that we observe independent series of counts $\{y_{i1}, \ldots, y_{in_i}\}$ on I units ($i = 1, \ldots, I$) at unequally spaced times $\{t_{i1}, \ldots, t_{in_i}\}$, together with a set of covariates $\{\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i}\}$. One possible model for such data is the gamma-Poisson model

$$(Y_{ij}|\lambda_{ij}, \boldsymbol{\beta}_{ij}, \mathcal{F}_{i,j|j-1}) \sim \text{Poisson}\{\lambda_{ij} \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij})\}$$
$$(\lambda_{ij}|\mathcal{F}_{i,j|j-1}) \sim \text{gamma}(\kappa_{i,j|j-1}, \upsilon_{i,j|j-1})$$

where $\mathcal{F}_{i,j|j-1}$ denotes the *filtration* or history of the responses for unit i up to, but not including time t_{ij} and gamma(κ, υ) the gamma distribution with observed Fisher information υ/κ at the mode κ. Here the value y_{ij} is seen as generated by a Poisson process with a time dependent mean $\lambda_{ij} \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij})$.

The non-random part $\exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij})$ of the mean, modelling the influence of covariates, can be used in several ways. It can be defined to model the influence of continuous explanatory variables such as the dose of a given medicine (cfr. dose of 5-Hydroxytryptamine in example 1, a function of time defining the shape of the profile, etc.), or the influence of an indicator variable (e.g. a variable that might distinguish two breeds of calves in example 1). Note that one can release the linearity assumption for the part of the mean involving the covariates.

The residual at time t_{ij} is $\log \lambda_{ij}$ on the log-scale for unit i. If the data are serially associated, values of two residuals observed at close time points should be closely related. One way to allow for this is to give a gamma prior to λ_{ij} and to use Bayes theorem

$$p(\lambda_{ij}|\mathcal{F}_{ij}) \propto \Pr(y_{ij}|\lambda_{ij}, \boldsymbol{\beta}_{ij}, \mathcal{F}_{i,j|j-1})\ p(\lambda_{ij}|\mathcal{F}_{i,j|j-1})$$

to construct the posterior distribution (with $\mathcal{F}_{ij}$ denoting the filtration for unit i until and including time t_{ij}) from the residual prior and the likelihood for the previous observation. The choice of the conjugate gamma distribution enables us to derive a closed analytic form for the posterior. The prior distribution is built by taking into account the time elapsed since the last observation. The further back in time the previous observation, the less weight the last posterior is given in building the new prior distribution. This can be done by taking as prior at time t_{ij} a distribution with the same mode as the (last) posterior distribution at time $t_{i,j-1}$, but with a smaller Fisher information. We suggest taking

$$\begin{aligned}\kappa_{i,j|j-1} &= \kappa_{i,j-1}\\ \upsilon_{i,j|j-1} &= \zeta(\Delta t_{ij})\upsilon_{i,j-1}\end{aligned}$$

where $\Delta t_{ij} = t_{ij} - t_{i,j-1}$ and $\zeta(\cdot)$ is a monotonally decreasing real function with values on $[0, 1]$ such that $\zeta(0) = 1$. We shall restrict our attention to $\zeta(\Delta t) = \exp(-\phi\Delta t)$. Notice that the above specification of the prior enables us to cope with continuous observation times.

Updating the prior distribution of $\log \lambda_{ij}$ (using Bayes' theorem) to take into account that it has generated the observed residual, we get a gamma posterior:

$$(\lambda_{ij}|\mathcal{F}_{ij}) \sim \text{gamma}(\kappa_{ij}, \upsilon_{ij})$$

where

$$\kappa_{ij} = \kappa_{i,j|j-1} + \frac{y_{ij} - \kappa_{i,j|j-1}\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{ij})}{\upsilon_{i,j|j-1} + \exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{ij})} \tag{3.1}$$

$$\upsilon_{ij} = \upsilon_{i,j|j-1} + \exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{ij}) \tag{3.2}$$

This recursive procedure can be started by taking a vague prior for the first residual, e.g. $\kappa_{i,1|0} = 1$ and $\upsilon_{i,1|0} = 0$. Considering a distribution for λ_{ij} different for each unit avoids problems related to heterogeneity, because an evolution for the residuals different for each unit is allowed.

The parameter used for modelling the prior weight, a generalization of the discount parameter used by Harvey and Fernandes (1989), West *et al.* (1985), Grunwald *et al.* (1993a, b) and Smith (1979) in a discrete time setting, behaves somewhat like an autocorrelation coefficient. Indeed, remember that ϕ is used to model the loss of information on λ_{ij} between two observation times. Basically the prior information on λ_{ij} at time t_{ij} is obtained by multiplying the posterior information at time $t_{i,j-1}$ by a quantity that decreases from 1 with the time elapsed since the last observation. This

loss of information is even more marked for large values of ϕ. Hence we expect serial association to be a decreasing function of ϕ.

The likelihood function can easily be written down if it is decomposed into a product of conditional probabilities. The contribution of the i^{th} series to the likelihood is

$$\prod_{j=2}^{n_i} \frac{v_{i,j|j-1}^{\kappa_{i,j|j-1}v_{i,j|j-1}}}{y_{ij}\{v_{i,j|j-1}+\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{ij})\}^{\kappa_{i,j|j-1}v_{i,j|j-1}+y_{ij}}} \frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{ij}y_{ij})}{B(\kappa_{i,j|j-1}v_{i,j|j-1},y_{ij})}$$

where $B(.,.)$ is the beta function.

This likelihood must be maximized with respect to the regression parameters $\boldsymbol{\beta}_{ij}$ and ϕ.

3.3 Analysis of data set 1

We now have all that is required to analyse the respiratory rate data set. As seen in Figures 3.1 and 3.2 and Table 3.1, the measures were made at irregular times between 0 and 60 minutes. The observation times are common during the administration of the drug. This is no longer true afterwards. Note that the observations at time 0 were obtained before the drug injection. They will be used as baseline covariates in the following models, our purpose being to decrease the uncontrolled heterogeneity among calves. The gamma-Poisson model was considered with different linear forms $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{ij}$ for the regression part. These forms are summarized as follows.

(a) $\eta_{ij} = \alpha \log y_{i0} + \left[I_{(t_{ij}<30)}\beta_1 + I_{(t_{ij}\geq 30)}\ \beta_2\right](t_{ij}-30)$;

(b) $\eta_{ij} = \alpha \log y_{i0} + \left[I_{(t_{ij}<30)}\beta_1 + I_{(t_{ij}\geq 30)}\ \beta_2\right](t_{ij}-30) + \delta_{\text{dose}_i}$;

(c) $\eta_{ij} = \alpha \log y_{i0} + \left[I_{(t_{ij}<30)}\beta_1 + I_{(t_{ij}\geq 30)}\ \beta_2\right](t_{ij}-30) + \gamma \log(\text{dose}_i)$;

(d) $\eta_{ij} = \alpha \log y_{i0} + \left[I_{(t_{ij}<30)}\beta_1 + I_{(t_{ij}\geq 30)}\ \beta_2\right](t_{ij}-30) + \gamma \text{dose}_i$;

(e) $\eta_{ij} = \alpha \log y_{i0} + I_{(t_{ij}\geq 30)}\ \beta_2(t_{ij}-30) + \gamma \text{dose}_i$

The deviances (with the deviance of the most complex model arbitrarily set to 0), the AIC values and the corresponding numbers of parameters are displayed in Table 3.3.

Form (a) considers a slope β_1 to model the evolution of the respiratory rate during the 30 minutes of the drug injection. Indeed the calves might grow accustomed to the drug as time passes, or the drug might have a cumulative effect on the respiratory rate. A slope β_2 modelling the decrease in respiratory rate is proposed for times greater than 30 minutes (the drug injection stopping at $t = 30$). We also condition on the control value y_{i0} (at rest) of the respiratory rate. Note that the regression parameters are assumed to be identical for all the calves, i.e. we consider parallel profiles.

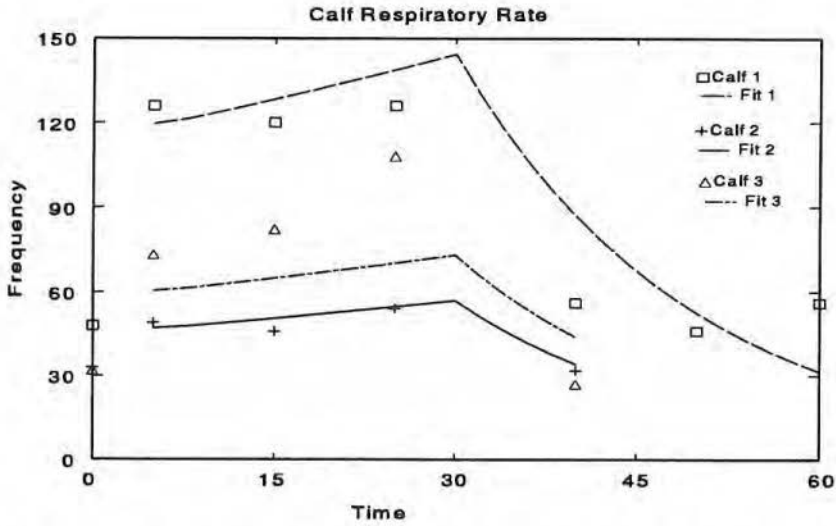


Fig. 3.1. Measurements of the respiratory rate (min^{-1}) of calves submitted to a continuous injection of 5-hydroxytryptamine receptor blocker at different doses (in μgr kg^{-1} min^{-1}) over 30 minutes: data and fitted profile for calves 1, 2 and 3.

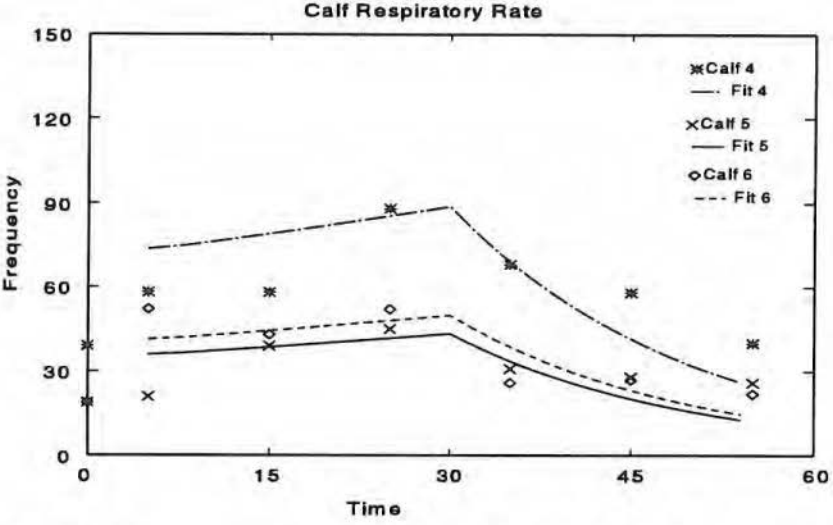


Fig. 3.2. Measurements of the respiratory rate (min^{-1}) of calves submitted to a continuous injection of 5-hydroxytryptamine receptor blocker at different doses (in μgr kg^{-1} min^{-1}) over 30 minutes: data and fitted profile for calves 4, 5 and 6.

Gamma-Poisson model			
η_{ij}	Par.	Dev.	AIC
1	5	37.0	47.0
2	8	0.0	16.0
3	6	6.7	18.7
4	6	**0.6**	**12.6**
5	5	5.8	15.8

Table 3.3. Deviance and AIC table for the models (selected model in boldface) considered in Section 3.

Parameter	Estimate	Standard error
α	0.9962	0.0079
β_1	0.0079	0.0014
β_2	-0.0507	0.0095
γ	0.0558	0.0017

Table 3.4. Estimates and standard errors for the regression parameters of the model selected.

Forms (b)-(d) are equivalent to form (a) except that the level of the time series is allowed to vary with the dose. Form (b) considers a non-parametric function of the dose (using a 4-level factor variable), whereas forms (c) and (d) assume that the respiratory rate varies linearly with respectively log-dose and the dose. From the AIC values in Table 3.3, we conclude that there is a dose effect and that a linear function of the dose seems to account for the drug effect.

At this stage, we might wonder whether the respiratory rate can be assumed to be stable during the 30 minutes of the drug injection, as in form (e), but this form yields a larger AIC, thereby rejecting the hypothesis.

We finally select form (d):

$$\widehat{\eta}_{ij} = 0.996 \log(y_{i0}) + 0.00787\ (t_{ij} - 30) + 0.0558\ \text{dose}_i\ \ (t_{ij} \leq 30)$$
$$\widehat{\eta}_{ij} = 0.996 \log(y_{i0}) - 0.0507\ (t_{ij} - 30) + 0.0558\ \text{dose}_i\ \ (t_{ij} > 30)$$

where the respiratory rate increases during the drug injection and then decreases afterwards. The estimates and the standard errors corresponding to the regression parameters of this model are given in Table 3.4. Note that one should be cautious with standard errors in a non-normal context where non-symmetric profile likelihoods are likely to arise. A 10% likelihood interval for ϕ is $[0.217, 0.333]$ with $\widehat{\phi} = 0.278$. The fitted profiles are plotted together with the data on Figures 3.1 and 3.2. Note that no observation is available at time 30; the fitted value is a projection. Some calf profiles are not well modelled. This is due to the assumption that profiles for all calves

are parallel. Relaxing this assumption would not be realistic (because there would be too many parameters for the amount of data) or interesting in a situation where we want to make statements valid for our (whole) population of calves.

3.4 Robustification of the gamma-Poisson model

In this section, we extend the gamma-Poisson model to include the Poisson distribution as a special case and to reduce its sensitivity to extreme observations. The first point is desirable to ensure that the gamma-Poisson model performs at least as well as the Poisson, the 'performance' of each model being measured using the AIC. We have known several data sets in discrete time where the Harvey and Fernandes (1989) gamma-Poisson model had a larger AIC than the negative binomial which only accounts for heterogeneity. But we should remember that the last two authors were working in a forecasting context. The robust gamma-Poisson model turned out to be more satisfactory than the negative binomial in all these examples.

The second point concerns the sensitivity of the model to extreme observations. As can be seen from Equation (3.1), the updated mean of the residual distribution can markedly differ from its predicted value if an unexpected observation arises. We shall add a new parameter to our model to limit the impact of such outlying observations.

As mentioned in the last section, we can release the linearity assumption for the response mean. Using the same notation as above, we denote the resulting mean by $\lambda_{ij}\mu(\boldsymbol{x}_{ij})$ where $\mu(\boldsymbol{x}_{ij})$ models the influence of covariates. Suitable forms for the non-random and possibly non-linear part $\mu(\boldsymbol{x}_{ij})$ in a growth curve context are described in Section 3.6.

3.4.1 THE POISSON DISTRIBUTION AS A LIMITING CASE

In the preamble, we have found it desirable to have the Poisson distribution as a limiting case of the gamma-Poisson model. This would ensure that our model does perform al least as well as the Poisson. This can be achieved by considering $\lambda_{ij}^{'}$ instead of λ_{ij} where

$$p(\lambda_{ij}^{'}|\mathcal{F}_{i,j|j-1}) \propto p(\lambda_{ij}|\mathcal{F}_{i,j|j-1})^{\delta}$$

The gamma mixing distribution will reduce to a point if δ tends to zero and to a vague distribution if δ tends to infinity. The recursive procedure will be similar to above, but with Equations (3.1) and (3.2) replaced by

$$\kappa_{ij}^{'} = \kappa_{i,j|j-1}^{'} + \frac{y_{ij} - \kappa_{i,j|j-1}^{'}\mu_{ij}}{v_{i,j|j-1}^{'}\delta + \mu_{ij}} \tag{3.3}$$

$$v_{ij}^{'} = v_{i,j|j-1}^{'} + \frac{\mu_{ij}}{\delta} \tag{3.4}$$

3.4.2 ROBUSTIFICATION

Extreme observations can have undesirable effects on inferential procedures. This is particularly true with dynamic models where an extreme observation can completely change the profile of the model. Different methods have been proposed in the literature to deal with such problems. Bolstad (1995) distinguishes three types of count observations in a gamma-Poisson setting: outliers, unsurprising observations (i.e. consistent with what has been observed before), and observations indicating a real change in the data pattern. For each type of observation, a different gamma mixing distribution for the mean is chosen.

Gordon and Smith (1993) robustify the updating equation in the Kalman filter to ensure that the posterior distribution of the mean stays close to the prior when an extreme observation is observed. Meinhold and Singpurwalla (1989) robustify the Kalman filter using the Student-t as a mixing distribution instead of the normal in a normal data setting. We refer to Section 1.7.2 for more details on these papers.

In our setting the problem arises in Equation (3.3) where a large difference (measured by the second term in the equation) between the observed data and its predicted value based on past information, can radically change the residual distribution. Hence reducing the impact of outlying values would be desirable in the procedure. A further parameter $\alpha \in [0, 1]$ in Equations (3.3) and (3.4) such that

$$\begin{aligned} \kappa'_{ij} &= \kappa'_{i,j|j-1} + \alpha \frac{y_{ij} - \kappa'_{i,j|j-1}\mu_{ij}}{v'_{i,j|j-1}\delta + \mu_{ij}} \\ v'_{ij} &= v'_{i,j|j-1} + \alpha \frac{\mu_{ij}}{\delta} \end{aligned} \tag{3.5}$$

could reduce the sensitivity of the model to extreme observations. Values of α close to zero avoid having too quick a correction (as implied by Equation (3.1)) of the model towards an outlier, i.e. an observation for which $y_{ij} - \kappa'_{i,j|j-1}\mu_{ij}$ is large.

The two new parameters δ and α can be estimated by maximizing the likelihood using a non-linear optimizer. Section 3.6 will illustrate this on data set 2 (related to the growth of colonies of *Paramecium aurelium*).

3.5 Modelling the growth curve

One striking feature of data set 2 is the stabilization of the colony sizes after about ten days. Therefore the systematic part of the model should tend to an asymptote as time passes. The model considered by Diggle (1990, p. 155) does not allow for this. He just considered a quartic polynomial in time for the mean number of individuals in each colony. The fit may be reasonable within the observed time-span, but is not realistic for larger

time values.

Nelder (1961 and 1962) considers a generalization of the logistic growth curve further developed by Heitjan (1991a and b) and including the Mitscherlich, Gompertz, logistic and exponential forms as well-known members. Heitjan (1991b) uses this family of models to assess the effect of three multiple sclerosis treatments on the evolution of ACFR, a measure of autoimmunity. Inclusion of explanatory variables in the systematic part is illustrated.

The equation of these profiles is given by

$$\mu_{ij} = e^{\kappa_2}\left[1 + (e^{(\kappa_2-\kappa_1)\kappa_4} - 1)e^{-\kappa_3 t_{ij} e^{\kappa_2 \kappa_4}}\right]^{-\frac{1}{\kappa_4}} \quad \kappa_4 \neq 0 \qquad (3.6)$$

$$= e^{\kappa_2 + (\kappa_1-\kappa_2)e^{-\kappa_3 t_{ij}}} \quad \kappa_4 = 0 \qquad (3.7)$$

It is the solution of the differential equation

$$\frac{d\mu_{ij}}{dt_{ij}} = \kappa_3 \mu_{ij}\left[d(e^{\kappa_2}, \kappa_4) - d(\mu_{ij}, \kappa_4)\right]$$

where

$$d(\mu_{ij}, \kappa_4) = \frac{\mu_{ij}^{\kappa_4} - 1}{\kappa_4} \quad \kappa_4 \neq 0$$
$$= \log(\mu_{ij})\kappa_4 \quad \kappa_4 = 0$$

Note that $\kappa_1 = \log(\mu_{i0})$ is the initial condition and $\kappa_2 = \lim_{t_{ij}\to\infty} \log(\mu_{ij})$, the asymptote. The parameters κ_3 and κ_4 model the rate of growth. The parameter κ_4 also determines the type of the curve, varying from the Mitscherlich ($\kappa_4 = -1$) through the Gompertz ($\kappa_4 = 0$), and the logistic ($\kappa_4 = 1$) to the exponential ($\kappa_4 \to \infty$ and $d(e^{\kappa_2}, \kappa_4) \to$ constant) (Heitjan, 1991b).

3.6 Analysis of data set 2

We now propose to use the robust gamma model to analyse the three series of data (in Table 3.2) plotted in Figure 3.3, giving the daily counts of *Paramecium aurelium* over a period of twenty days in a nutritive medium.

Two families of models are considered. The first one, ignoring the longitudinal aspect of the data set, and thus ignoring serial association between data on the same unit, assumes that the counts are distributed as a negative binomial. The second family is the gamma-Poisson model where serial association is modelled using an empirical Bayes approach for the residuals on the log-scale, this choice being suggested by the canonical link for Poisson data.

	Ind. Neg. Bin.			gamma-Poisson		
Syst. Part	Dev.	Par.	AIC	Dev.	Par.	AIC
Polynom	553.7	6	565.7	546.2	8	562.2
Gen.Log.	556.5	5	566.5	548.8	7	562.8

Table 3.5. Deviance and AIC table.

For each family of models, two types of systematic parts for the mean were fitted to the data:

- a fourth degree polynomial in time as suggested by Diggle (1990, p. 155).
- the generalized logistic growth curve given by Equations (3.6) and (3.7) (Nelder, 1961 and 1962).

The parameter estimation of the gamma-Poisson model was performed in three steps. Firstly, the MLEs for the regression parameters in the independence negative binomial model were computed. Then the serial association parameters ϕ, δ and α of the gamma-Poisson model were estimated, the regression parameters being held fixed at their first step values. Finally the gamma-Poisson likelihood was maximized over both the serial association and regression parameters, starting values being given by the first two steps of the algorithm.

The deviance, number of parameters and AIC for the above four models are displayed in Table 3.5. Whatever the chosen systematic part, we see that the gamma-Poisson model performs better than the independence negative-binomial model, thereby showing the need for modelling serial association within unit.

The AIC provides no clear-cut choice for the systematic part. Therefore it seems advisable to select the gamma-Poisson model with a generalized logistic growth curve for the systematic part, because it is simpler and is properly modelling the asymptotic behavior of the colony development for large time values.

The equations for the profiles of the gamma-Poisson model are respectively given by

$$\widehat{\mu}_{ij} = 0.6850 + 1097.4\, t_{ij} - 66.02\, t_{ij}^2 + 0.9843\, t_{ij}^3 + 0.0139\, t_{ij}^4$$

and

$$\widehat{\mu}_{ij} = \mathrm{e}^{6.304}\left[1 + (\mathrm{e}^{5.791} - 1)\mathrm{e}^{-0.0004 t_{ij} \mathrm{e}^{7.600}}\right]^{-0.830}$$

for the polynomial and generalized logistic growth curves. They are both plotted on Figure 3.3. One can see that the polynomial model will predict an explosive number of individuals for large time values, whereas this number is estimated by $\mathrm{e}^{\widehat{\kappa_2}} \doteq 547$ for the logistic growth curve model.

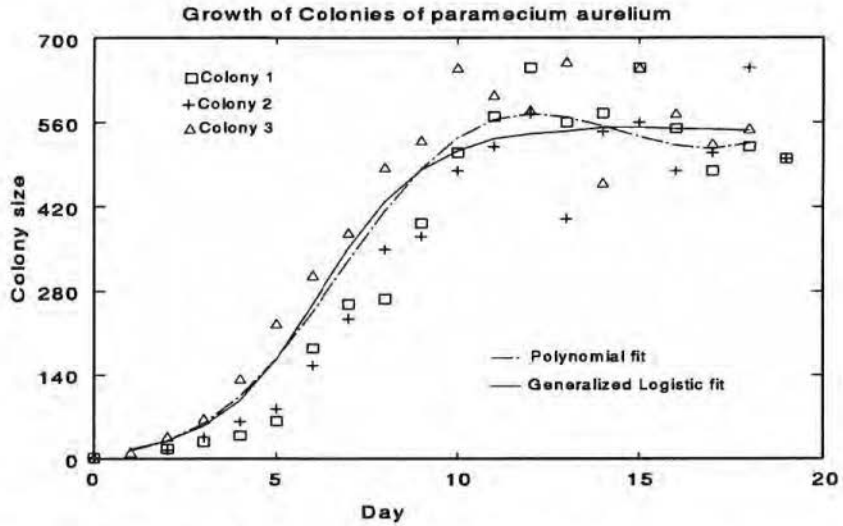


Fig. 3.3. Growth of three closed colonies of *Paramecium aurelium* in a nutritive medium (Gause, 1934): fitted profiles and data.

The serial association parameters for the above models are respectively estimated as $\widehat{\phi} = 0.2245$, $\widehat{\delta} = 0.000084$, $\widehat{\alpha} = 0.05937$ and $\widehat{\phi} = 0.1944$, $\widehat{\delta} = 0.000077$, $\widehat{\alpha} = 0.04837$ for the two best models.

Note that the model described in the above sections can easily cope with sampling at irregularly spaced time points. Such situations might be caused by the design. For example, a regular sampling is very important at the beginning of the colony development (which undergo quick changes at the early stage), whereas its asymptotic behavior allows for more sparse observations after day 10. But coping with unequally spaced observation time may also be needed when missing (completely) at random data appear in an equally spaced sampling design.

Logistic growth curves (Nelder 1961 and 1962) are more realistic than a polynomial or any spline based systematic part, because each of the parameters has an interpretation as explained in Section 3.5. Moreover it is more sensible than polynomials to model biological mechanisms of growth.

Covariates such as an indicator of the growth condition can easily be included as shown by Lindsey (1993, p. 133). They might affect the level of the asymptote, or the growth rate of the colony at an early stage, and so on.

The parameters used by the full likelihood approach were estimated using the procedure OPTMUM in GAUSS, thereby enabling the use of non-linear expressions for the mean.

4
Other Applications of the GARM

In Chapter 2, we have presented the generalized autoregression model or GARM, to model series of positive data in continuous time. The basic idea was to express some transformation of a location parameter (such as the mean response) as a function of past covariates plus a correction enabling inertia in the sign of residuals. The setting was very general and the model defined was not even restricted to exponential family distributions. For example, in this same chapter, we have used the generalized gamma distribution, which includes well-known nonexponential members such as the Weibull density, to model series of positive data.

Hence nothing prevents us from modelling other types of data such as series of binary and multinomial data (Lambert, 1995d). We shall consider this problem in Section 4.1. The use of the resulting tool will be illustrated using two data sets from the literature. The performances of the binary model will be compared with Markov chain models.

The next theoretical point will be devoted to rewriting the GARM as a GLM when the covariates enter linearly in the systematic part (Section 4.2). In this particular setting, the different parameters in the GARM will be computable using GLM software such as GLIM or S-PLUS. This will enable us to use the powerful IWLS algorithm to compute the regression parameters, the other two or three autoregression parameters being estimated using a non-linear optimizer or by considering a grid of sensible values.

Finally, predictions using an approximate predictive likelihood that corrects for the randomness in the parameter estimates of the basic model are computed (Lambert, 1995e; Section 4.3). An example is considered where the basic model is the negative binomial GARM for overdispersed count data. The rewriting of the GARM based on a GLM significantly improves the time required to compute likelihood prediction envelopes.

4.1 Series of binary and multinomial data

In this section, we show how the GARM can be extended to deal with series of binary and multinomial data in continuous time. Three extra parameters compared to the independence model setting enable it to compete with low order Markov chain models even when the series are short or the serial as-

Profile		Larval origin		Profile		Larval origin	
Apple	Hawth.	Apple	Hawth.	Apple	Hawth.	Apple	Hawth.
1111	1111	2	1	0001	1011	1	0
1110	1111	1	0	0001	1111	1	0
1101	1111	1	0	0000	1111	6	7
1011	1010	1	0	0000	0010	1	0
1000	1111	1	1	0000	0111	3	5
0111	1111	1	0	0000	1011	2	1
0110	1111	1	0	0000	1100	1	0
0100	1111	1	2	0000	1101	2	1
0100	0010	1	0	0000	0011	2	2
0011	0111	2	1	0000	0001	1	6
0011	1011	1	0	0000	0000	0	2
0010	1111	1	0	0000	0101	0	1
0010	1100	1	0	0000	0100	0	1
0010	1011	1	0	0000	1000	0	1
0001	0011	1	0	0000	1110	0	1

Table 4.1. Frequency of observed profiles by fly origin for eight binary responses (1: Success; 0: Failure).

sociation order is small. But the superiority of generalized autoregression models is more obvious with longer series of data where a larger autoregression order might arise. The technique is illustrated on two data sets from the literature.

The first example that we shall consider studies acceptance for oviposition by *Rhagoletis pomonella* adult female flies (Stanek and Diehl, 1988). Female flies, which grew as larvae either in apple or hawthorn fruit, were first placed on apple and then on hawthorn at four different ages: 8-9, 11-12, 15-16 and 18-19 days after adult eclosion. At each occasion they were observed for five minutes for possible oviposition, the success or failure of the procedure being recorded (see Table 4.1). The whole process was successfully repeated with 70 flies (37 of apple and 33 of hawthorn origin). Our goal is to model the response profiles. The response of a fly at a given day should be allowed to depend on its past history, as well as on the fly origin and the type of fruit support.

The second example is a series of multinomial data giving the number of *Pinus*, *Abies*, *Quercus* and *Alnus* pollen grains in samples of size 100 (Mosimann, 1962). These samples coming from a single core were extracted from the soil at increasing depths that we shall assume to be equi-spaced. Space will now be used instead of time as a qualitative tool for ordering the observations and defining a measure of serial association. The corresponding data are plotted on Figure 4.1. Our goal is to model and compare the different pollen profiles to enable the study of past changes in vegetation and

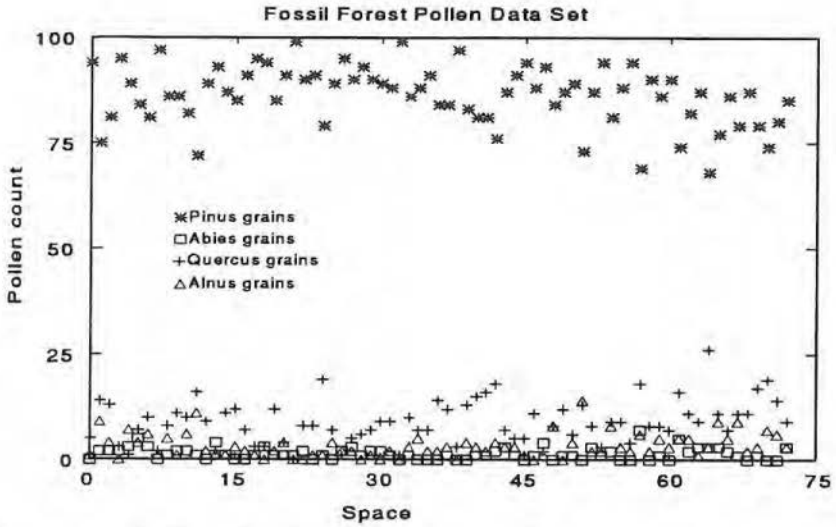


Fig. 4.1. Number of pollen grains of *Pinus*, *Abies*, *Quercus* and *Alnus* fossil forests in samples of 100 grains (Mosimann, 1962) extracted at increasing depths in the soil.

hence in climate. We refer the reader to Section 1.7 for a review of the key papers on longitudinal models for binary and multinomial data.

4.1.1 BINOMIAL SERIES

We now explain how to adapt the GARM to deal with non-stationary series of binomial data observed at unequally spaced times.

Suppose that we observe a series $\{y_{i1}/n_{i1}, \ldots, y_{im_i}/n_{im_i}\}$ $(i = 1, \ldots, I)$ of m_i proportions on unit i at unequally spaced times $\{t_{i1}, \ldots, t_{im_i}\}$ together with covariates $\{\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im_i}\}$.

Denote by

- r_{ij}, the residual for unit i at time t_{ij} on the logit-scale, defined by

$$r_{ij} = \alpha \log\left(\frac{y_{ij} + 0.5}{n_{ij} - y_{ij} + 0.5}\right) - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij} \qquad (4.1)$$
$$r_{i0} = 0$$

  where $\boldsymbol{\beta}_{ij}$ stands for regression parameters. This is simply the difference between the empirical and the theoretical logits. The double subscript for the regression parameters indicates that we do not put a constraint on the parameter structure. But very often it will just be assumed that $\boldsymbol{\beta}_{ij} = \boldsymbol{\beta}$.

- r_{ij}^c, the *cumulated residual* for unit i at time t_{ij} on the logit-scale, defined by

$$r^c_{ij} = e^{-\phi \Delta t_{ij}} r^c_{i,j-1} + n_{ij} r_{ij} \quad (j > 0)$$
$$r^c_{i0} = 0$$

where $\Delta t_{ij} = t_{ij} - t_{i,j-1}$ and $0 < \phi$.

- w^c_{ij}, the weight for the cumulated residual r^c_{ij}, where

$$w^c_{ij} = e^{-\phi \Delta t_{ij}} w^c_{i,j-1} + n_{ij} \quad (j > 0)$$
$$w^c_{i0} = 0$$

The idea is to model the logit of π_{ij}, the proportion of successes on unit i at time t_{ij}, as

$$\text{logit}(\pi_{ij}) = \boldsymbol{x}^T_{ij} \boldsymbol{\beta}_{ij} + e^{-\rho \Delta t_{ij}} \frac{r^c_{i,j-1}}{w^c_{i,j-1}} \tag{4.2}$$

where $0 < \rho$. Note that ϕ and ρ are respectively modelling the order and importance of the serial association.

The introduction of n_{ij} in the definition of the cumulated residual and of the cumulated weight expresses that a deviation by the data from the systematic part of the model is more likely to be 'serious' (and thus to be followed by a deviation of the same order) if it is based on a large number of observations.

Note that, because $r_{i0} = 0$, full confidence is given to the systematic part when no previous observation (and hence no previous cumulated residual) is available.

The above model specification can also cope with binary observations (where $n_{ij} = 1$) thanks to the use of the empirical logit in Equation (4.1).

Estimates for the parameters can be computed by maximizing the likelihood

$$L(\boldsymbol{\beta}_{ij} s, \rho, \phi) = \prod_i \prod_j \pi_{ij}(\boldsymbol{\beta}_{ij}, \rho, \phi)^{y_{ij}} [1 - \pi_{ij}(\boldsymbol{\beta}_{ij}, \rho, \phi)]^{m_{ij} - y_{ij}}$$

(with $\pi_{ij}(\boldsymbol{\beta}_{ij}, \rho, \phi)$ defined in Equation (4.2)) using a non-linear optimization routine such as Proc OPTMUM in GAUSS.

4.1.2 ANALYSIS OF THE FLY DATA SET

We have applied the technique described in the last section to the fly data set. Several factors were liable to influence the presence or absence of oviposition by any given insect, among which the fly origin (ORI, indexed by k and parameterized using α_k), the support for oviposition (SUP, indexed by l and parameterized using γ_l) and age (AGE, which stands for time t_j). We also expect that the probability of 'success' ('failure') is larger

if one or several successes (failures) were recorded in the fly history. A summary of the models considered, including several of these explanatory variables, is given in Table 4.2 together with the corresponding number of parameters and AIC values. The Wilkinson and Rogers (1973) notation is used to describe the systematic part of the model. For example, the fourth generalized autoregression model in Table 4.2 is

$$\text{logit}(\pi_{ij}) = \beta_0 + \alpha_k + \gamma_l + \beta_{1l} t_j + e^{-\rho \Delta t_{ij}} \frac{r^c_{i,j-1}}{w^c_{i,j-1}}$$

which assumes that the probability for a fly to lay its eggs depends on both its origin and the support. In addition, the effect of AGE is allowed to vary with the last covariate.

For comparison, the independence model together with the first and second order Markov chain models with the same systematic part were computed. In these three settings, the logit of the probability of success is respectively modelled as

- $\text{logit}(\pi_{ij}) = \beta_0 + \alpha_k + \gamma_l + \beta_{1l} t_j$ (independence)
- $\text{logit}(\pi_{ij}) = \beta_0 + \alpha_k + \gamma_l + \beta_{1l} t_j + \rho(2y_{i,j-1} - 1)$ (1st order Markov chain model)
- $\text{logit}(\pi_{ij}) = \beta_0 + \alpha_k + \gamma_l + \beta_{1l} t_j + \rho_1(2y_{i,j-1} - 1) + \rho_2(2y_{i,j-2} - 1)$ (2nd order Markov chain model)

In Markov chain models, conditioning on past events is impossible and thus ignored for the first observation. In the 2nd order Markov chain model, full conditioning is only considered after the second observation time. This procedure is equivalent to assigning a response of one half for missing past events.

The results are summarized in Table 4.2. Note that the Markov chain models were computed using all the observations. The covariates were solely used to define the systematic part when no previous observation is available. This permits a sensible comparison of the AIC values with these for the generalized binomial autoregression model. We now propose to explain the meaning of the systematic parts considered in the table. The first line assumes that the probability for a fly to lay its eggs depends on its origin and on the support. Moreover, the effect of each of these two variables is allowed to depend on the value of the other one. The effect of AGE can differ for each combination of origin and support. The second systematic part makes the same assumptions as the first one, except that there is no origin-support interaction. The third line further assumes that the difference in AGE effects between flies of two different origins does not depend one the type of the support. In the fourth model, the incidence of AGE is assumed to be independent of the fly origin. The fifth systematic part supposes that there is no AGE effect for flies of apple origin. In the

Systematic Part	Par	AIC	AIC		
	IND		M1	M2	GA
(ORI*SUP).(1+AGE)	8	559.2	516.9	513.6	513.3
(ORI+SUP)+(ORI*SUP).AGE	7	557.5	515.9	511.8	512.0
(ORI+SUP).(1+AGE)	6	559.2	516.3	513.5	513.3
(ORI+SUP)+SUP.AGE	5	557.5	**514.3**	**511.6**	511.7
(ORI+SUP)+(SUP=2).AGE	4	**557.2**	516.1	515.5	**510.0**
ORI+SUP	3	570.5	524.2	523.2	524.6
SUP+(SUP=2).AGE	3	576.6	527.0	525.5	520.0

Table 4.2. AIC table for the generalized binary autoregression (GA), independence (IND), first (M1) and second order (M2) Markov chain models applied to the fly data set (with a bold AIC value for the selected model). Each successive column has one more parameter.

sixth model, the probability of success is stationary. Finally, the seventh systematic part is the same as the fifth one, except that there is no origin effect.

By comparing the AICs in the last four columns of Table 4.2, we see that we cannot assume independence. The generalized autoregression (GA) and second order Markov chain (M2) models are clearly superior to the other two. The 'best' GA model performs slightly better (with an AIC value of 510.0) than the 'best' M2 model (AIC=511.6). The corresponding systematic parts are not in perfect agreement even if the conclusions are quite similar.

In order to compare the estimates of the two models, we shall compute the regression parameters using the systematic part of the best GA model.

The GA and the M2 models both indicate an effect of larval origin. When the fly is observed for the first time, the conditional odds for oviposition for a fly from apple origin is estimated by the GA (M2) model to be 3.290 (2.203) larger than the conditional odds for a fly from hawthorn origin.

The need for autoregression indicates that we can say more when the fly has performed the test earlier. The conditional odds for success at the second observation time is (estimated to be) more than five times larger if the last recorded observation is a success than if it was a failure just before. Clearly there might be situations in practice where the information carried by covariates is negligible once the history of a given unit is available. Then marginal modelling would be completely unadapted.

We also detect a support effect. The conditional odds for success on an apple support is estimated to be 2.546 (3.357) larger than the corresponding odds on a hawthorn support.

Note that no origin-support interaction was detected.

There seems to be an evolution of the support effect with AGE, even

when conditioning on the fly history. The older the fly the more likely the success. The selected GA model suggests that there might be no evolution of the odds of success when the fly is from apple origin. This is in contradiction with the M2 model conclusions. Hence caution is needed on this point of the analysis.

4.1.3 MULTINOMIAL SERIES

Suppose that we observe I series $\{y_{i1}, \ldots, y_{im}\}$ of m data at common and possibly unequally spaced times $\{t_1, \ldots, t_m\}$ where $y_{.j}$ is fixed by the design for any $j \in \{1, \ldots, m\}$. In other words, we assume that we have a multinomial sample at each time point t_j. Our goal is to model the evolution of the probability of being in category i, while taking into account the multinomial constraint and the possible serial association between proportions observed at close time points.

Denote by $\{\pi_{1j}, \ldots, \pi_{Ij}\}$ the theoretical multinomial probabilities at time t_j with $\pi_{.j} = 1$ for all j. Consider the log-linear model

$$\log(\mu_{ij}) = \log(y_{.j}\pi_{ij}) = \alpha_j + \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij} \tag{4.3}$$

where μ_{ij} and $\boldsymbol{x}_{ij}$ are respectively the mean response and a vector of covariates for unit i at time t_j.

Serial association can be modelled by conditioning on cumulated residuals as in binary generalized autoregression models. Cumulated residuals r_{ij}^c can be defined using

$$\begin{aligned}
r_{ij} &= \log\left(\frac{y_{ij}}{y_{.j}}\right) - (\alpha_j + \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij}) \qquad (4.4)\\
r_{i0} &= 0\\
r_{ij}^c &= \mathrm{e}^{-\phi \Delta t_j} r_{i,j-1}^c + y_{.j} r_{ij}\\
r_{ij}^c &= 0\\
w_{ij}^c &= \mathrm{e}^{-\phi \Delta t_j} w_{i,j-1}^c + y_{.j}\\
w_{i0}^c &= 0
\end{aligned}$$

where again a departure of the theoretical model from the observed data is more completely accounted for when the related multinomial total $y_{.j}$ is large. The meaning of this notation and the rationale for the above procedure are the same as in Section 4.1.1 and need not be repeated. As in the binary case, one could add one half to a zero observation and the related multinomial total to avoid trouble with Equation (4.4).

Equation (4.3) can then be modified to yield

$$\log(\mu_{ij}) = \alpha_j + \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij} + \mathrm{e}^{-\rho \Delta t_j} \frac{r_{i,j-1}^c}{w_{i,j-1}^c}$$

which now accounts for serial association.

The next step is to maximize the Poisson likelihood

$$L(\boldsymbol{\beta}_{ij} s, \rho, \phi) = \prod_j \prod_i \mathrm{e}^{-\mu_{ij}(\boldsymbol{\beta}_{ij}, \rho, \phi)} \mu_{ij}(\boldsymbol{\beta}_{ij}, \rho, \phi)^{y_{ij}} \tag{4.5}$$

$$= \prod_j \exp\left[\alpha_j y_{.j} + \sum_i y_{ij}\left(\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij} + \mathrm{e}^{-\rho \Delta t_j} \frac{r^c_{i,j-1}}{w^c_{i,j-1}}\right)\right] \prod_i \mathrm{e}^{-\mu_{ij}(\boldsymbol{\beta}_{ij}, \rho, \phi)}$$

Thus $Y_{.j}$ is a sufficient statistic for α_j. Now, in exponential family models, we know that $E[\widehat{T(Y)}] = T(Y)$ for any sufficient statistic $T(Y)$. Hence $\widehat{\mu}_{.j} = y_{.j}$, meaning that the estimated multinomial frequencies $\widehat{\mu}_{ij}$ $(j = 1 \ldots J)$ maintain the observed marginal (and multinomial) total $y_{.j}$. This is the usual log-linear model 'trick' where estimation of the regression parameters is done by maximizing the Poisson likelihood. By conditioning on $Y_{.j}$, the sufficient statistic for α_j, we recover the multinomial likelihood

$$\prod_j \prod_i \left(\frac{\mu_{ij}}{\mu_{.j}}\right)^{y_{ij}} = \prod_j \prod_i \left\{ \frac{\exp\left[\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{ij} + \mathrm{e}^{-\rho \Delta t_j} \frac{r^c_{i,j-1}}{w^c_{i,j-1}}\right]}{\sum_k \exp\left[\boldsymbol{x}_{kj}^T \boldsymbol{\beta}_{kj} + \mathrm{e}^{-\rho \Delta t_j} \frac{r^c_{k,j-1}}{w^c_{k,j-1}}\right]} \right\}^{y_{ij}} \tag{4.6}$$

which does not contain α_j anymore. Maximizing the likelihood in Equation (4.5) or the conditional likelihood in Equation (4.6) gives the same estimates for the regression parameters. The probability of observing a datum in the i^{th} category at time t_{ij} is then estimated by

$$\widehat{\pi}_{ij} = \frac{\exp\left[\boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}}_{ij} + \mathrm{e}^{-\widehat{\rho} \Delta t_j} \frac{\widehat{r}^c_{i,j-1}}{w^c_{i,j-1}}\right]}{\sum_k \exp\left[\boldsymbol{x}_{kj}^T \widehat{\boldsymbol{\beta}}_{kj} + \mathrm{e}^{-\widehat{\rho} \Delta t_j} \frac{\widehat{r}^c_{k,j-1}}{w^c_{k,j-1}}\right]}$$

4.1.4 ANALYSIS OF THE POLLEN DATA SET

We are now in a position to model and compare the profiles of the four types of fossil forest pollens presented in the introductory section. Space will now be used instead of time to classify the observations on a given unit.

Different linear forms were considered for the profiles. Unfortunately we were not able to simplify the most complex model considered, which assumes a different quadratic profile for each type of pollen. Note that it is only necessary to specify three out of the four profiles, the other one being computed using the multinomial constraint. The final model is plotted in Figure 4.2. The large values for the parameter estimates $\widehat{\rho}$ (=12.21) and $\widehat{\phi}$ (=0.055) suggest that no serial association is present in the data, the specification of the systematic part using polynomials in space being sufficiently precise to model the data evolution.

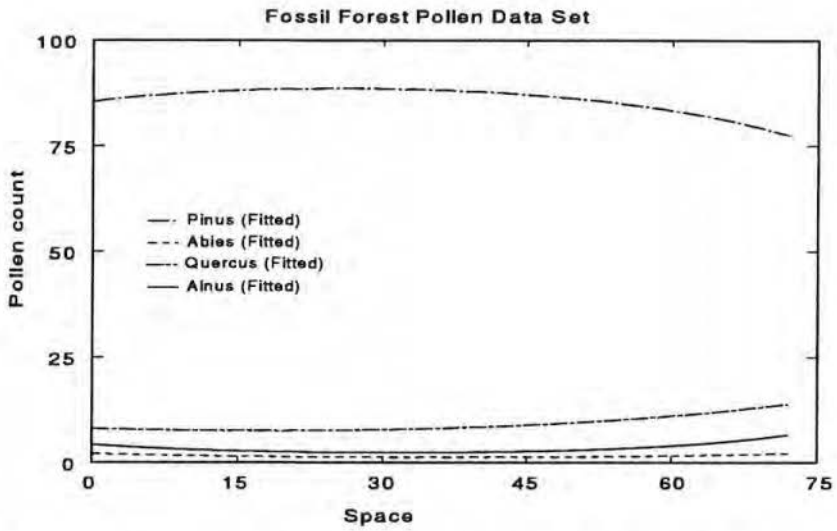


Fig. 4.2. Number of pollen grains of *Pinus*, *Abies*, *Quercus* and *Alnus* fossil forests in samples of 100 grains (Mosimann, 1962) extracted at increasing depths in the soil: fitted profiles.

4.2 The GARM and generalized linear models

In particular situations, the GARM can be expressed as a generalized linear model. As already explained in the preamble, this is particularly interesting in practice because the IWLS algorithm, which is implemented in many statistical software packages, can then be used to compute the regression parameters. We now give the details and conditions under which that approach is feasible.

Consider the special case of a GARM with a linear model for the systematic part and an exponential family distribution as stochastic element. From Equation (2.1), one can easily show that

$$\begin{aligned} g(\mu_{i1}) &= \boldsymbol{x}_{i1}^T\boldsymbol{\beta} \\ g(\mu_{i2}) &= \boldsymbol{x}_{i2}^T\boldsymbol{\beta} + \mathrm{e}^{-\rho\Delta t_{i2}}[g(y_{i1}) - \boldsymbol{x}_{i1}^T\boldsymbol{\beta}] \\ &= [\boldsymbol{x}_{i2} - \mathrm{e}^{-\rho\Delta t_{i2}}\boldsymbol{x}_{i1}]^T\boldsymbol{\beta} + \mathrm{e}^{-\rho\Delta t_{i2}}g(y_{i1}) \end{aligned}$$

$$g(\mu_{i3}) = \boldsymbol{x}_{i3}^T\boldsymbol{\beta} + \mathrm{e}^{-\rho\Delta t_{i3}}\frac{[g(y_{i2}) - \boldsymbol{x}_{i2}^T\boldsymbol{\beta}] + \mathrm{e}^{-\phi\Delta t_{i2}}[g(y_{i1}) - \boldsymbol{x}_{i1}^T\boldsymbol{\beta}]}{1 + \mathrm{e}^{-\phi\Delta t_{i2}}}$$

$$= \left[\boldsymbol{x}_{i3} - \mathrm{e}^{-\phi\Delta t_{i3}}\frac{\boldsymbol{x}_{i2} + \mathrm{e}^{-\phi\Delta t_{i2}}\boldsymbol{x}_{i1}}{1 + \mathrm{e}^{-\phi\Delta t_{i2}}}\right]^T\boldsymbol{\beta} + \mathrm{e}^{-\rho\Delta t_{i3}}\frac{g(y_{i2}) + \mathrm{e}^{-\phi\Delta t_{i2}}g(y_{i1})}{1 + \mathrm{e}^{-\phi\Delta t_{i2}}}$$

Similarly, one can show that

$$g(\mu_{i4}) = \left[\boldsymbol{x}_{i4} - \mathrm{e}^{-\rho\Delta t_{i4}} \frac{\boldsymbol{x}_{i3} + \mathrm{e}^{-\phi\Delta t_{i3}}\boldsymbol{x}_{i2} + \mathrm{e}^{-\phi(\Delta t_{i3}+\Delta t_{i2})}\boldsymbol{x}_{i1}}{1 + \mathrm{e}^{-\phi\Delta t_{i3}} + \mathrm{e}^{-\phi(\Delta t_{i3}+\Delta t_{i2})}}\right]^T \boldsymbol{\beta}$$
$$+\mathrm{e}^{-\phi\Delta t_{i4}} \frac{g(y_{i3}) + \mathrm{e}^{-\phi\Delta t_{i3}} g(y_{i2}) + \mathrm{e}^{-\phi(\Delta t_{i3}+\Delta t_{i2})} g(y_{i1})}{1 + \mathrm{e}^{-\phi\Delta t_{i2}} + \mathrm{e}^{-\phi(\Delta t_{i3}+\Delta t_{i2})}}$$

More generally, we have

$$g(\mu_{ij}) = \left[\boldsymbol{x}_{ij} - \mathrm{e}^{-\rho\Delta t_{ij}} \frac{\boldsymbol{x}^c_{i,j-1}}{w^c_{i,j-1}}\right]^T \boldsymbol{\beta} + \mathrm{e}^{-\rho\Delta t_{ij}} \frac{\gamma^c_{i,j-1}}{w^c_{i,j-1}} \tag{4.7}$$
$$= \boldsymbol{x}^{d\,T}_{ij} \boldsymbol{\beta} + \mathrm{offset}_{ij} \tag{4.8}$$

where

$$\begin{aligned}
\boldsymbol{x}^c_{ij} &= \boldsymbol{x}_{ij} + \mathrm{e}^{-\rho\Delta t_{ij}} \boldsymbol{x}^c_{i,j-1} \\
\boldsymbol{x}^c_{i1} &= \boldsymbol{x}_{i1} \\
\gamma^c_{ij} &= g(y_{ij}) + \mathrm{e}^{-\rho\Delta t_{ij}} \gamma^c_{i,j-1} \\
\gamma^c_{i1} &= g(y_{i1})
\end{aligned}$$

From Equation (4.7) we thus conclude that, for fixed values of ρ and ϕ, the GARM can be expressed as a GLM where, for each unit, and some observation time t_{ij},

- any design matrix row $\boldsymbol{x}^{d\,T}_{ij}$ can be expressed as the difference between the original design matrix row $\boldsymbol{x}^T_{ij}$ and some kind of weighted average of design matrix rows from previous observation times.
- an offset γ_{ij}, a weighted average of the last observation on the g-scale and the previous g-observations on this unit, is introduced.

This formulation will be particularly useful below to compute modified or approximate predictive likelihoods.

4.3 An approximate predictive likelihood

The goal of this section is to apply existing methods to make predictions in time series of non-normal data. We shall focus on the particular case of overdispersed count data, although the presentation is sufficiently general to be applied in other contexts.

The tool developed will be applied to the second example of Section 3.1 which was further analysed in Section 3.6. The data set of interest concerns the growth of three closed colonies of *Paramecium aurelium* in a nutritive medium on a twenty day period. One of the series (Colony 1) has been (artificially) truncated at day 10 (see Figure 4.3). We propose to construct a likelihood prediction envelope based on a model built from the three series

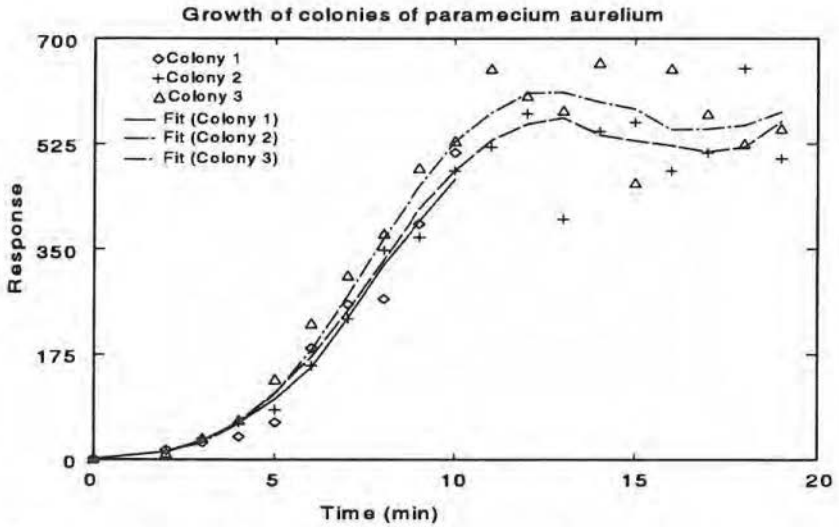


Fig. 4.3. Growth of three closed colonies of paramecium aurelium in a nutritive medium (Gause, 1934): fitted profiles and data.

at the preceding time points, and to check that the actually observed values fall into these intervals.

4.3.1 LIKELIHOOD PREDICTION ENVELOPES

The first thing to do to make predictions is to build a model suitable for the observed data. Various techniques have been proposed in the literature to model series of count data. We refer to Chapter 3 for a review of these methods. Here we shall adapt the generalized autoregression model to deal with our discrete data.

Predictions can then be made using the final model built from the observed data. One naive but simple way to make predictions is to consider the *profile predictive likelihood* (Fisher, 1959, pp. 128-133; Lejeune and Faulkenberry, 1982)

$$L(z|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_I) = f(z|\boldsymbol{y}_k;\widehat{\mu}^{(z)}_{k,n_k+1},\widehat{\boldsymbol{\nu}}^{(z)}_k)\prod_i\prod_j f(y_{ij}|\mathcal{F}_{i,j-1};\widehat{\mu}^{(z)}_{ij},\widehat{\boldsymbol{\nu}}^{(z)}_i)$$

where z stands for the observation to come on unit k (which now plays the role of the parameter of interest in the likelihood), $\boldsymbol{y}_i$ denotes the set of observations on unit i, and $\widehat{\mu}^{(z)}_{ij}$ and $\widehat{\boldsymbol{\nu}}^{(z)}_i$ respectively stand for the MLEs of μ_{ij} and $\boldsymbol{\nu}_i$ given $\{y_{i1},\ldots,y_{in_i},z\}$ and the observations on the other series. The symbol (z) as superscript is used to make a distinction between the MLEs computed from the likelihood including the contribution of the unobserved z datum, and the 'classical' MLEs computed only using the observed data

contributions to the likelihood. More precision is achieved by estimating the model parameters using a likelihood based on the three series instead of using the sole likelihood contribution from the series of interest. Note that the first element in the profile predictive likelihood is related to the observation to come, whereas the others are the contribution of the observed data.

One major criticism of the profile predictive likelihood is that it does not take into account the uncertainty attached to each of the estimated parameters. The profile likelihood just assumes that the estimates of the nuisance parameters are the true (or 'population') values. A consequence of this in our setting, is an underestimated width for prediction intervals. One way to release this unrealistic hypothesis (Hinkley, 1979) is to condition on the maximum likelihood estimators which have a given distribution. Deriving such quantities is not an easy task, and except in very special circumstances, one has to approximate the conditional distribution. There is a considerable literature on the subject which often relies on known analytic forms for the MLEs or on a substantial reduction of the data to sufficient statistics (Butler, 1986 and 1989; Kalbfleisch and Sprott, 1970; Hinkley, 1979; Bjornstad, 1990; and, indirectly, Barndorff-Nielsen, 1983, 1993). For a review of prediction techniques based on the likelihood, see Bjornstad (1990). For example, the resulting conditional density can be approximated using the modified profile likelihood. The so-called Barndorff-Nielsen (1983) p^*-formula provides an approximation to the distribution of MLEs given an ancillary statistic. The required (approximate) conditional distribution can then be derived, yielding

$$L^*(z|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_I) = L(z|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_I)\ \left|J^{(z)}(\widehat{\boldsymbol{\theta}}^{(z)})\right|^{-\frac{1}{2}}\ \left|\frac{\partial\widehat{\boldsymbol{\theta}}}{\partial\widehat{\boldsymbol{\theta}}^{(z)}}\right| \qquad (4.9)$$

where $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\nu}_1^T, \ldots, \boldsymbol{\nu}_I^T,)$ stands for all the parameters in the model such as the regression parameters $\boldsymbol{\beta}$ defining μ_{ij} and the nuisance parameters; $J^{(z)}(\widehat{\boldsymbol{\theta}}^{(z)})$ is the observed information matrix about $\boldsymbol{\theta}$ computed at the MLEs for a given value of the unobserved data z, i.e.

$$J^{(z)}(\widehat{\boldsymbol{\theta}}^{(z)}) = \left.\frac{\partial^2 \log f(z|\boldsymbol{y}_k;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(z)}} + \sum_i \sum_j \left.\frac{\partial^2 \log f(y_{ij}|\mathcal{F}_{i,j-1};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(z)}}$$

Unfortunately the last factor in Equation (4.9) is particularly tedious to compute, because it requires knowledge of the analytical forms for the MLEs, which, outside the exponential family, are usually impossible to determine. In such settings, Butler (1986, Rejoinder) proposes an approximate

predictive likelihood which can be evaluated with any type of distribution, because it only requires the maximum likelihood estimates of the nuisance parameters based on $\{\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_I^T, z\}$:

$$L^{(B)}(z|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_I) = L(z|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_I)\ |J^{(z)}(\widehat{\boldsymbol{\theta}}^{(z)})|^{\frac{1}{2}}\ \left|H(\widehat{\boldsymbol{\theta}}^{(z)})H^T(\widehat{\boldsymbol{\theta}}^{(z)})\right| \tag{4.10}$$

with

$$H^{(z)}(\widehat{\boldsymbol{\theta}}^{(z)}) = \left.\frac{\partial^2 \log f(z|\boldsymbol{y}_k;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\ \partial\{\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_I^T\}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(z)}} + \sum_i \sum_j \left.\frac{\partial^2 \log f(y_{ij}|\mathcal{F}_{i,j-1};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\ \partial\{\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_I^T\}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(z)}}$$

As briefly explained in the original paper, this last formula for the conditional likelihood can be derived by using a Taylor series expansion for the joint density of the observed and unobserved data about the nuisance parameter MLEs, and by dividing the whole by a normal approximation to the nuisance parameter joint distribution. Note that using the inverse of the Fisher information evaluated at $\widehat{\boldsymbol{\theta}}$ as an approximation to the covariance matrix of the above normal distribution simplifies the formula, finally yielding Equation (4.10).

The goal of this section is to construct likelihood prediction envelopes. By a 100 p% likelihood prediction envelope, we mean a succession of 100 p% (Butler approximate) predictive likelihood intervals computed at the time points of interest. These likelihood intervals are obtained in the same way as with a traditional parameter likelihood (Kalbfleisch, 1985), the role of the parameter here being played by the unobserved quantity.

Note that we have not tried to compute *simultaneous* prediction intervals, to make an analogy with the frequentist simultaneous confidence intervals. In our view, one is more interested by what is 'likely' to be observed in the future at one given time point independently of what the other predictions are. Plotting an envelope is more a way to summarize graphically a series of independent results than giving artificially related statements. However a *simultaneous* approach is feasible, but this is technically far more difficult, particularly in a non-normal setting, because it requires the computation of N_p- (and thus possibly large) dimension normed likelihood regions (if one wants to predict N_p unobserved data). Finally note that a method for modelling series of data observed at unequally spaced times is required, because the time at which the prediction is made is (in practice) totally arbitrary. A second (but not rigorous) choice would be to proceed step by step by using a conditioning argument. The observation at time $t_{n_k} + 1$ could be predicted conditional on the last observation; the one at

time $t_{n_k}+2$ could be derived by conditioning on the first step prediction; and so on until the time of interest has been reached. This sounds very useful when the point forecast is really what interests us, but this becomes far more complicated when prediction intervals are required.

4.3.2 APPLICATION

Applying the theory of Sections 4.2 and 4.3.1 to construct a likelihood prediction envelope for the truncated part of the introductory data set is the subject of this section. The full data set has been studied by Diggle (1990, p. 155) who proposed a quartic polynomial in time to model the observed growth curve. Lambert (1996b and Ch. 3 above) uses a generalized form of the logistic growth curve (Nelder, 1961 & 1962) to take into account the asymptotic behaviour of the colony sizes, and compares it with the quartic polynomial fit. It was noticed that both solutions are realistic in the observed time range, but that the generalized logistic form should be prefered because it is more sensible than polynomials to model biological mechanisms of growth.

However, in order to illustrate the expression of the GARM as a GLM, we have decided to present the construction of the likelihood prediction envelopes with the quartic polynomial (and thus linear) model. Note that the same approach can be used with the generalized logistic form, but this will require the use of FORTRAN or GAUSS (in our case) codes to compute MLEs.

As was already pointed out in Ch. 3, a negative binomial distribution seems to be more adapted than the Poisson alternative. Hence we shall focus on the building of the Butler approximate predictive likelihood using the stochastic and systematic elements derived in a traditional approach to modelling the observed data.

We shall denote by $\mathrm{NB}(\upsilon, \pi_{ij})$, the negative binomial distribution

$$\frac{\Gamma(\upsilon + y_{ij})}{y_{ij}!\,\Gamma(\upsilon)}(1-\pi_{ij})^{\upsilon}\pi_{ij}^{y_{ij}}$$

with mean $\upsilon\frac{\pi_{ij}}{1-\pi_{ij}}$ for the random variable $\mathrm{Y_{ij}}$.

The influence of time as well as of any other explanatory variable on the mean response μ_{ij} on unit i at time t_{ij} can be modelled using a logistic regression

$$\eta_{ij} = \log\frac{\pi_{ij}}{1-\pi_{ij}} = {x_{ij}^{d}}^{T}\beta + \mathrm{offset}_{ij}$$

jointly with a GARM to take the serial association into account. Note that the GARM appears in this last equation through the offset and the transformed design matrix $\boldsymbol{X}^{d}$ computed using Equation (4.7).

Par.	Est.
υ	63.38
Autoregression Par.	
ϕ	0.2786
ρ	0.7685
Regression Par.	
β_0	0.6804
β_1	1.104
β_2	-0.06493
β_3	0.000755
β_4	0.0000226

Table 4.3. Parameter MLEs computed on the triglyceride data set using a logistic regression jointly with a GARM.

If the negative binomial parameter υ and the autoregression parameters ρ and ϕ were known, then one could simply compute the regression parameter MLEs using the IWLS algorithm. Taking a grid of values for the three unknown parameters might be one solution to determine the MLEs. In this example we have used the non-linear optimizing procedure PROC OPTMUM in GAUSS. The corresponding MLEs and fit are respectively displayed in Table 4.3 and on Figure 4.3. The fitted profiles displayed on this figure are not smooth curves because corrections due to serial association were added to the polynomial contribution.

The next step is the computation of the Butler approximate predictive likelihood for various values of z at the time point of interest. This can be done in three steps:

(1) Given the likelihood based on the observed data and z, compute the MLEs of the eight parameters in the model (see Table 4.3).

(2) Compute the Jacobian $J^{(z)}(\widehat{\boldsymbol{\theta}}^{(z)})$ and the matrix $H(\widehat{\boldsymbol{\theta}}^{(z)})$ at the MLEs from step 1.

(3) Compute the approximate predictive likelihood using Equation (4.10) together with the results of the first two steps.

In our example, the procedure has been simplified by only conditioning on the regression parameter estimates. This reduces the dimension of the Jacobian and of the H matrices from eight to five. Of course it is not necessary to use numerical methods to compute the J and H matrices: analytic forms are easy to derive. This point is essential to ensure a reasonable rapidity to the procedure. Indeed one has to repeat the three above described steps for different values of z to determine first the forecast $\widehat{z}$ (which maximizes the predictive likelihood) at the time point of interest. Once $\widehat{z}$ is known, the predictive likelihood is rescaled by dividing it by its maximum value. The 10% (say) predictive likelihood interval can then be determined. This

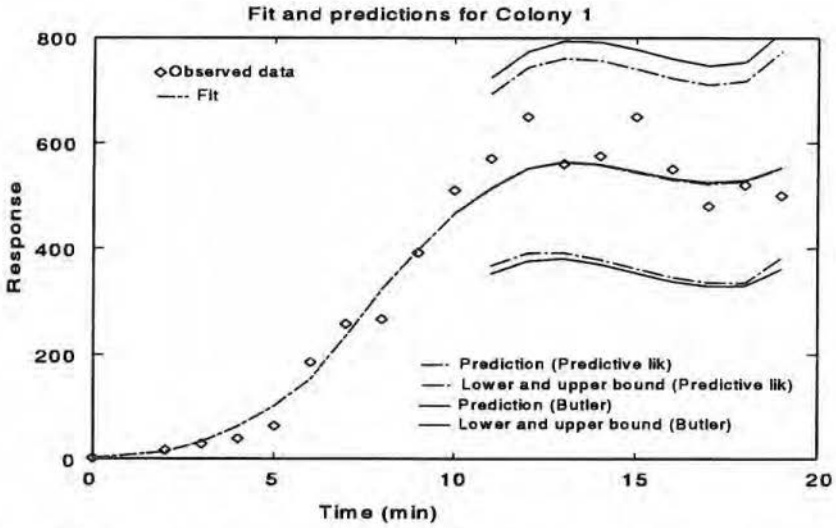


Fig. 4.4. Growth of three closed colonies of *Paramecium aurelium* in a nutritive medium (Gause, 1934): predictions and likelihood prediction envelope for the artificially truncated series.

can be done using e.g. the secant method for determining the zeros of a function. This method has the advantage of not requiring the derivative of the predictive likelihood (at the cost of a function evaluation).

The corresponding results for the *Paramecium aurelium* data set are displayed on Figure 4.4. Two approaches were used to compute the 10% likelihood prediction envelope. The first one was based on the 'naive' Lejeune and Faulkenberry (1982) predictive likelihood which assumes that the parameters are known (and equal to their MLEs). The second method is the one described in Section 4.3.1. As can be seen from Figure 4.4, there is a small bias correction and the likelihood prediction envelope derived using the Butler approximation is slightly wider, reflecting the extra uncertainty in the regression parameters values. It would probably be even wider if all eight parameters were used when conditioning. The correction with respect to the 'naive' method is not very important in this setting. We would expect larger corrections with smaller counts and shorter series, or if all series had been truncated. Finally note that the actually observed data fall well within the 10% likelihood prediction envelope (whatever the chosen method).

5
Conclusion

In this chapter, we propose to review and discuss some key points in the writing of the preceding lines. Finally, in the last section, we suggest some extra topics that could be the subject of further research to provide the user with a complete and flexible tool to analyse series of non-normal data in continuous time.

5.1 Conditional models

Conditional models were presented as the natural approach to analyse longitudinal data. The key underlying argument was based on Simpson's paradox: with longitudinal data sets, the conclusions drawn using conditional and marginal models can be in complete disagreement, although a careful inspection of these results simply reveals that the data generating mechanism was more complex than what the marginal point of view was ready to assume (cfr. 'treat or not to treat'). Hence, it was concluded that a model for longitudinal data has to describe, as far as possible, the relationship between the response and the causal mechanism. Then, if marginal conclusions are required, they should be derived from such an appropriately constructed conditional model.

In practice, and more particularly in biological settings, it is desirable to explain the observed 'data generator' using simple models. This can usually be done easily using conditional models, where both covariates and past history of the process contribute to explain what has been observed. However, the marginal correspondent of such a model is usually very complex (when normality cannot assumed), and often requires the use of numerical methods to be derived. Hence, one might question the reality underlying simple marginal models, which, in turn, have complex conditional equivalents. In our view, this disconnection with reality is the most worrying problem in marginal models.

We have also briefly discussed technical problems arising in these settings. Many of the proposed approaches rely on the GEE, which cannot be integrated back to a proper likelihood, except in very special cases. Hence, these methods of analysis do not correspond to a proper model, because the probability of the observed data cannot be calculated. Therefore, it is impossible to compare competing 'models' based on GEE.

5.2 Models in continuous time

As shown in the literature review of Chapter 1, most of the models for longitudinal data assume that time is discrete. This hypothesis is a real handicap in practice for several reasons. Experimenters, such as veterinarians, often do not want to be restricted by a discrete time schedule when they collect observations. An example of this was given in Section 3.6, where the size of colonies of *Paramecium aurelium* undergoes quick changes at the start of the experiment before reaching stability. There, it would be more interesting to observe the colonies intensively at the beginning of the process, and to reduce the effort after day 10 (say).

In some situations, veterinarians simply cannot collect data regularly, giving rise to 'missing data' in a discrete time design. Sometimes, observations cannot be gathered on week-ends because the technical staff is limited on these days; or a suddenly ill nonexperimented animal might distract the doctor from the trial for a short period of time, giving rise to a missing at random observation; or the observed animal might be moving at some stage of the data collection, making the process unobservable at this time.

Therefore, from Chapter 2 on, we have restricted attention to longitudinal data models which are not constrained by the artificial assumption that time is discrete.

5.3 Nonlinear regression

The technical phase in the modelling of data is made in two steps. Firstly, a family of distributions, likely to approximate the stochastic process which gave rise to the data, is chosen. Secondly, some location parameter (and less often, shape or scale parameters) is described in terms of covariates. For technical reasons, and due to the availability of GLM packages such as GLIM, modellers very often restrict attention to linear functions of the explanatory variables; unfortunately, these popular polynomial and spline based methods do not model the underlying reality. This is somehow regrettable when the biological or economical mechanism is well understood and could be transposed in the modelling process. This kind of approach, although desirable, was impossible until recently. But the wide availability of fast computers and the existence of powerful nonlinear optimizer no longer justify the nearly exclusive use of (generalized) linear models in many settings. Of course, these (locally) linear approximations to the reality still remain the only alternative with large database, repetitive analyses and recursive models, but the data analyst should always wonder if a more realistic approach is feasible.

5.4 Non-normal models

In the last section, the first step in the modelling process was mentioned to be the choice of a distribution for the data. A large part of the longitudinal data literature in continuous time restricts attention to normal models. A few exceptions to this rule include the exponential dispersion models and the GEE 'models'. The same arguments used to introduce generalized linear models, can justify the need for non-normal longitudinal models. This has been one of the main underlying themes of our work: Chapter 2 was devoted to the modelling of series of positive longitudinal data, for which skewed distributions are necessary alternatives to normality; Chapter 3 focused attention on series of overdispersed counts; finally, Chapter 4 proposed models for binary, binomial and multinomial longitudinal data. The analysis of the triglyceride data set in Section 2.4 shows that the conclusions can be very sensitive to the choice of the distribution. Therefore, the consideration of wide families of distributions, such as the generalized gamma for positive data, to model the data generating mechanism, is an important step in any modelling strategy. The use of these 'highly' parameterized densities, as recommended by Sprott (1982), can point out unexpected candidates, such as the Weibull (in our example) which is more traditionally used in a survival context. Again, fast computers and efficient nonlinear optimizers enable to fit atypical distributions outside the restrictive exponential family.

5.5 Model selection

Throughout this work, the Akaike information criterion (AIC) has been used to select models. Other authors, such as Jones (1993) and Lindsey (1995a and b), have used the AIC as a guideline in model selection. As explained in Lindsey (1994), likelihood ratio tests based on the Chi-squared distribution, are not well adapted in this context, because they do not treat the models that we wish to compare symmetrically. There has been an important debate on this problem, from which the AIC of Akaike (1973) and the BIC of Schwarz (1978) arose.

Another problem appears in the modelling of stochastic processes (see Lindsey, 1995a, pp. 174–178) where (when using likelihood ratio tests based on the Chi-square) models with few parameters are systematically rejected in favour of very complex ones. This phenomenon seems to be particularly worrisome when the models compared differ by a large number of parameters. This goes in the opposite direction to what any scientist would desire: a simple and easily interpretable model.

Finally note that the AIC can be modified to yield models smoothing the data more. Collet (1991) suggests to add $3p$ instead of $2p$ in Equation (2.2), thereby penalizing complex models. This kind of generalization of the Akaike's criterion was originally proposed by Atkinson (1980) and Bhansali and Downham (1977).

5.6 Further research

In this section, we propose to review some aspects that might be the subject of future research. As mentioned in Section 1.1, this work voluntarily ignores some important steps in the development of a longitudinal study. Problems related to the design of such experiments are very challenging and important in practice. The available methods to compute sample sizes in a longitudinal context assume that the response is normally distributed, which, with large data sets, often turns out to be an unrealistic hypothesis. As already mentioned, the formula proposed in Lindsey (1995b, pp. 88–91) in an independence context could be adapted in a non-normal longitudinal context if the likelihood is decomposed as the product of independent contributions, as in Equation (1.1). Of course, this supposes that the model likely to be satisfactory, is already known.

A refinement of the existing exploratory methods (Diggle *et al.*, 1994, Ch. 3) in a longitudinal context, would also be welcome to deal with non-normal and irregularly spaced observations. The definition of appropriate residuals in the context of the above models is also important, although, again, the independence of the conditional contributions in Equation (1.1) allows the use of most of the existing methods. The convenience of deviance residuals in this setting was already pointed out at the end of Section 2.4. Schnatter (1994; see also Section 1.7.2) also suggests using the P-scores in a similar context.

Another important problem concerns the choice of the initial conditions in conditional models containing random effects. The paper by Heckman (1981) is certainly an interesting starting point.

The modelling of missing data patterns in the models of the preceding chapters, could also be investigated. Diggle and Kenward (1994) is a possible first reference.

Finally, a few technical improvements could also be made to the models that we have presented. Even if efficient nonlinear optimizers are available, it would be interesting to develop efficient algorithms which are more adapted to the problem at hand, thereby enabling the treatment of large data sets.

Similarly to the gamma-Poisson, the beta-binomial model might be an interesting alternative to the GARM for binary and binomial data. But, as already mentioned in Section 1.1, important difficulties arise when covariates are present, particularly with the logit link. One could avoid these by using a log-log link with a log-gamma distribution for the residual part, although this idea was originally proposed to introduce a random intercept (Conaway, 1990). But the nonsymmetric treatment of 'success' and 'failure' by the log-log link restricts the appeal of such a model.

One could also inquire on possibly more appropriate forms for the $\zeta(\cdot)$ function, which models the relation between time and serial association in

Chapter 3. The same type of question arises in the GARM, where functions of the type $\exp(-\phi\Delta t_{ij})$ were systematically used to model the decaying influence of past residuals on the response. Of course, one could wonder if, in practice, there is enough information in the data to discriminate between such (parametric) functions.

It would also be interesting to consider wider families of distributions to model discrete data, in the same way as the generalized gamma in Chapter 2 with positive continuous data.

6
Summary

We shall begin from the general fact that most of the methods proposed in the literature to analyse non-normal longitudinal data make the assumption that the observations are equally-spaced in time. Very often, the authors of such papers mention that extension to continuous time is not a problem, although without explicitly explaining how to do so. However, such a generalization usually turns out to be theoretically difficult, if not impossible, the original model making a large use of the discrete structure of time.

A second worrying limitation is the wide use of marginal models when irregularly sampled longitudinal data have to be analysed, few flexible conditional model being available. This is not a problem in observational population studies, such as epidemiology, where marginal tools exist and are well adapted to answer the usual type of questions attached to such settings. However, when the data generating mechanism is the central question of the study, when the way that the observations evolve over time is of interest, such as in clinical trials or biological experiments, conditional models are desired (Sheiner *et al.*, 1989; Davidian and Giltinan, 1995, p. 122).

For all these reasons, building conditional models for non-normal longitudinal data turns out to be both an interesting and challenging problem to be considered. An extra motivating problem was to concentrate on likelihood based approaches to enable different models to be compared, either directly through their likelihood functions, which indicate how probable they each make the observed data, or by more sophisticated model selection procedures, such as the AIC (Akaike, 1973), which penalizes for the complexity of the competing models. Note that this is not possible with most of the marginal models in the literature (with a few remarkable exceptions such as in Molenberghs and Lesaffre, 1994), because they essentially rely on some kind of score equations, the generalized estimating equations (Liang and Zeger, 1986), which, most often, cannot be integrated back to obtain a likelihood function (McCullagh and Nelder, 1989, 2nd Ed., Section 9.3.2).

After these few motivating remarks, we now describe the structure of our presentation.

Chapter 1: Introduction

This chapter first presents a few practical situations, in different scientific areas, where longitudinal data arise. Some of these examples will be analysed in details in the subsequent chapters, as suitable theoretical tools are developed. The different steps, required to setup and to analyse data from a longitudinal study, are then presented. The points around which this work will be built up, are indicated and referenced. Arguments for using conditional likelihood based approaches instead of marginal methods are developed. The extra difficulties when analysing longitudinal data instead of the usual independent observations in cross-sectional studies, are also discussed. The different ways with which these extra features — namely heterogeneity and serial association — are handled in the literature, are the subject of the following sections.

Normal autoregression models, which are equivalent in conditional and marginal settings, are then reviewed. The possible combinations of random effects — modelling heterogeneity — and of autoregression terms — modelling serial association — and their consequences on the normal covariance matrix structure, are considered. The non-normal equivalent of autoregression models is also discussed. It is shown how, by suitably conditioning on past responses, we can compare autoregression models of different order.

State-space models, where the likelihood function is derived using the Kalman filter, are then presented as an alternative in a normal setting (Jones and Ackerson, 1990; Jones, 1993) giving the dynamic linear model. The Kalman filter, which enables a dynamic derivation of the likelihood function in complex settings, is shown to have numerical advantages and to be very flexible, particularly when time is assumed to be continuous.

Dynamic generalized linear models (West *et al.*, 1985) are then proposed as a generalization of the last approach in a non-normal context, but in a discrete time setting. These are derived by assigning a conjugate prior distribution to the linear part (or *linear predictor*) of the usual static generalized linear model. This prior distribution is then updated using Bayes theorem. The choice of the conjugate as a prior yields a closed analytic form for both the posterior and the likelihood, thereby avoiding the computation of numerical integrals, which would have to be repeatedly evaluated to estimate other integrals, as observations accumulate.

Other approaches to modelling series of count data are also presented (Harvey and Fernandes, 1989; Ord *et al.*, 1993), again in a discrete time setting. Here, the intercept, in the canonical regression, is given a conjugate prior distribution, which, again, is updated using Bayes theorem. This technique can be used with count, binomial and multinomial data. Note that approximations are required for the logistic models when explanatory variables are present.

These dynamic generalized linear models can be further generalized to

deal with a vector of observations (Fahrmeir, 1992), while still assuming time to be discrete. Here, the dynamic regression parameters are assumed to evolve according to transitions equations fixing their first two moments, and to be normally distributed. The use of the posterior mode instead of the traditional posterior mean to estimate these parameters, avoids the computation of numerical multidimensional integrals.

Heterogeneity, which was accounted for in several of the above models, is then more systematically treated in the next section on random effects. It is introduced in conditional models by giving a distribution to the intercept in regression models. Random effects are further generalized to the other parameters in the regression, yielding the random coefficient model.

Despite the fact that we have decided to concentrate on conditional models, we give a short review of marginal ones in longitudinal data analysis. The generalized estimating equations are presented, although no proper model usually corresponds to these 'score' equations. Marginal likelihood methods are also mentioned.

Finally, a review of the literature on subjects related to the themes developed above, is proposed.

Chapter 2: Positive longitudinal data modelling

This second chapter is really the start of our contribution to the modelling of non-normal longitudinal data, although the discussion on conditional models in Ch. 1 already brought some new material to the debate. As its title shows, Ch. 2 is dedicated to the modelling of series of positive data. This is a subject of particular importance in several sciences such as human and veterinary medicine, economics, etc. As pointed out in the first chapter, we focus our attention on the modelling of biomedical data. More particularly, we develop a model to analyse the measures of the triglyceride profiles for two sets of four Beagle dogs under four types of fiber based diets. These data are typically positive and bounded below by zero, meaning that their distribution might be skewed. A second technical problem is that the responses were not measured at equally spaced times, meaning that a continuous time setting would be more appropriate than a discrete time model assuming that some data are missing at random. The *generalized autoregression model* (GARM) is then developed as a tool for modelling series of irregularly sampled observations. Serial association is modelled using two extra parameters compared to the traditional model assuming independence. Different ways to model heterogeneity are suggested, including a non-parametric random effect that will be used to analyse the triglyceride data set. No particular hypothesis is made on the type of the distribution, meaning that the GARM could be used in other settings (see Ch. 4). The generalized gamma family of distributions, which includes well-known members such as the exponential, Weibull, gamma and log-

normal as a limiting case, is proposed to analyse the positive longitudinal data. These different models are compared using the Akaike information criterion (AIC), which enables one to assess and compare the goodness of fit on non-nested models. The traditional normal and log-normal distributions, which are widely used in practice to analyse this kind of data, are shown to be completely unadapted, the Weibull being our final choice.

Chapter 3: Longitudinal count data in continuous time

This chapter is devoted to the modelling of discrete longitudinal data observed at irregularly spaced time points. The initial idea was to generalize the dynamic generalized linear models of West *et al.* (1985) and Harvey and Fernandes (1989) to model count data in continuous time. Basically, the intercept in a log-linear model, is given a gamma prior (conjugate) distribution. This prior is used to predict the state of the process at the next observation time, simply by considering a gamma predictive distribution with the same mode as the original prior, but with a smaller Fisher information at this point. As soon a one observation is available, the distribution of the intercept, which is a kind of residual in a model assuming independence, is updated using Bayes theorem. The resulting (unconditional) likelihood is then the product over observation times of negative binomial densities.

This technique, based on the Kalman filter, can be modified to reduce its sensitivity to extreme observations and to include the Poisson distribution as a special case. This last improvement is desirable in practice, because the Harvey and Fernandes (1989) model was found, in some examples, to have a larger AIC than the negative binomial model, which only accounts for heterogeneity.

Two examples are then presented and analysed using the above new technique. The first one studies the respiratory rate profile of calves submitted to a thirty minutes injection of a receptor blocker, which simulates acute respiratory distress syndrome in healthy animals. The data, collected before, during and after the drug injection, were irregularly sampled.

The second example studies the evolution of three colonies of *Paramecium aurelium* in identical nutritive media over a twenty day period. The goal is to build a profile equation giving the mean number of individuals in each colony at any time point. One interesting feature of the data set, is the apparent stabilization of the colony size after day 10. Thus polynomial or the related spline based methods cannot take this behaviour into account. Biological models would be more appropriate than any artificial mathematical construct in this situation. Therefore, a generalization of the logistic growth curve (Nelder, 1961 and 1962), further developed by Heitjan (1991a and b), is proposed to model the profile.

Chapter 4: Other applications of the GARM

This chapter considers the generalized autoregression model in another context than in Chapter 2. There, no particular assumption was made on the form of the density function involved in the GARM definition.

Hence nothing prevents us from using the GARM to model series of binary, binomial and multinomial data. This idea is first used to analyse series of binary data on the acceptance for oviposition by *Rhagoletis pomonella* adult female flies (Stanek and Diehl, 1988) on two types of support. Our goal is simply to model the response profiles by allowing the response of a fly on given day to depend on its past history, in addition to the fly origin and the type of fruit support.

The second example consists of a series of multinomial data giving the number of *Pinus*, *Abies*, *Quercus* and *Alnus* pollen grains in samples of size 100 (Mosimann, 1962). These samples coming from a single core were extracted from the soil at increasing depths that we assume to be equispaced. Thus, space is now used instead of time as a qualitative tool for ordering the observations and for defining a measure of serial association.

The next theoretical point is to rewrite the GARM as a GLM, when the covariates enter linearly in the systematic part. Under this assumption, the regression parameters in the GARM are computable using GLM software such as GLIM or S-PLUS. The other two or three autoregression parameters have to be estimated either by using a non-linear optimizer, or by considering the likelihood at a grid of sensible values.

Finally, the last section is devoted to the application of prediction methods to series of non-normal data. From the particular case of overdispersed count data, it is shown how the GARM can be used jointly with an approximate predictive likelihood (Butler, 1986), to build prediction envelopes. This approximate predictive likelihood, where the observation to come now plays the role of a parameter in the likelihood function, is used instead of the better known profile predictive likelihood (Fisher, 1959, pp. 128–133), because it does not assume that the (regression) parameters are fixed quantities equal to their MLEs as traditionally. Hence the imprecision due to the parameter estimates is directly taken into account when predictions are drawn from the estimated model.

An example is then considered, where the missing data from an artificially truncated series of the *Paramecium aurelium* data set of Ch. 3 are predicted by using the information present in the observations remaining in the three series. The resulting prediction envelopes are shown, as expected, to be wider than the predictive likelihood ones.

Finally, a short concluding chapter discusses the advantages of the models presented in the preceding pages.

7
Samenvatting

Ons uitgangspunt is dat de meeste methoden voor de analyse van niet-normale longitudinale data de onderstelling maken dat de observaties op equidistante tijdstippen gebeuren. De auteurs van deze artikels stippen meestal aan dat de uitbreiding naar continue tijd geen probleem vormt. Er wordt echter nooit expliciet aangegeven hoe dit moet gebeuren. Nochtans is het doorgaans zo dat het oorspronkelijk model uitvoerig gebruik maakt van de discrete tijdsstructuur, hetgeen de uitbreiding theoretisch zeer moeilijk, zoniet onmogelijk maakt.

Een tweede zorgwekkende beperking is het algemeen gebruik van marginale modellen bij onregelmatig gespatieerde longitudinale data daar weinig flexibele conditionele modellen beschikbaar zijn. Dit is geen probleem in observationele populatie studies, zoals in epidemiologie, waar marginale modellen bestaan en goed geschikt zijn om de typische vragen in voornoemde context te beantwoorden. Echter, indien het data genererend mechanisme de centrale vraag is in de studie, indien het tijdsverloop van de gegevens belangrijk is, zoals in klinische studies of biologische experimenten, dan zijn conditionele modellen gewenst (Sheiner *et al.*, 1989; Davidian en Giltinan, 1995, p. 122).

Om al deze redenen is het ontwerpen van conditionele modellen voor niet-normale longitudinale data een interessante uitdaging. Een extra motiverend probleem was ons te concentreren op likelihood-gebaseerde benaderingen om toe te laten verschillende modellen te vergelijken. Dit gebeurt, ofwel direct via de likelihoodfuncties welke aangeven hoe waarschijnlijk zij de geobserveerde data maken, ofwel door meer gesofisticeerde model selectie zoals het AIC (Akaike, 1973) die de complexiteit van de modellen penaliseert. We merken op dat dit niet mogelijk is voor de meeste marginale modellen in de literatuur (met enkele merkwaardige uitzonderingen zoals in Molenberghs en Lesaffre, 1994). Dit komt omdat zij in hoofdzaak steunen op een soort van score vergelijkingen, de veralgemeende schattingsvergelijkingen (Liang en Zeger, 1986), die doorgaans niet terug kunnen geïntegreerd worden om de likelihoodfunctie te verkrijgen (McCullagh en Nelder, 1989, 2nd Ed., Sectie 9.3.2).

Na deze motiverende opmerkingen beschrijven we nu de structuur van ons werk.

Hoofdstuk 1: Inleiding

In dit hoofdstuk worden enkele praktische situaties uit verschillende wetenschappelijke gebieden voorgesteld, waarin longitudinale data opduiken. Sommige van deze voorbeelden zullen in de hiernavolgende hoofdstukken in detail geanalyseerd worden, eens de geschikte theoretische instrumenten ontwikkeld zijn. Daarna geven we de verschillende stappen in het opzet en de analyse van een longitudinale studie aan. De themata in de opbouw van dit werk worden aangebracht en van referenties voorzien. Ook geven we argumentatie voor de ontwikkeling van conditionele likelihood modellen versus marginale modellen. De bijkomende moeilijkheden bij de analyse van longitudinale data i.p.v. onafhankelijke data in cross-sectionele studies, worden besproken. De verschillende wijzen waarmee deze extra kenmerken (namelijk heterogeniteit en seriële associatie) aangepakt worden in de literatuur vormt het onderwerp van de volgende secties.

Vervolgens geven we een overzicht van normale autoregressie modellen. Hier zijn marginale en conditionele aanpak equivalent. We beschouwen de mogelijke combinaties van random effecten (die de heterogeniteit modelleren) en van autoregressie termen (die de seriële associatie modelleren) en hun gevolgen op de normale covariantiematrix structuur. Het niet-normale equivalent van autoregressie modellen wordt ook besproken. Er wordt aangetoond hoe, door gepast conditioneren op respons uit het verleden, het mogelijk wordt autoregressie modellen van verschillende orden te vergelijken. Toestandsruimte modellen waarbij de likelihoodfunctie afgeleid wordt door gebruik te maken van de Kalman filter, worden voorgesteld als een alternatief in een normale set up (Jones en Ackerson, 1990; Jones, 1993), hetgeen leidt tot het dynamisch lineair model. Van de Kalman filter, die een dynamische afleiding van de likelihoodfunctie in complexe set up toelaat, wordt aangetoond dat hij numerische voordelen biedt en bijzonder flexibel is, in het bijzonder bij continue tijd.

Dynamische veralgemeende lineaire modellen (West *et al.*, 1985) worden voorgesteld als een veralgemening naar een niet-normale context, maar in een discrete tijdssituatie. Deze worden afgeleid door een toegevoegde a priori verdeling toe te kennen aan het lineair deel (of lineaire predictor) van het gebruikelijk statische veralgemeend lineair model. Deze a priori verdeling wordt dan geactualiseerd met de stelling van Bayes. De keuze van de toegevoegde als a priori verdeling leidt tot een gesloten analytische vorm voor zowel de a posteriori verdeling als van de likelihood. Aldus vermijden we het herhaaldelijk berekenen van numerische integralen wanneer het aantal observaties toeneemt.

Andere benaderingen voor het modelleren van reeksen van aantallen worden ook gegeven (Harvey en Fernandes, 1989; Ord *et al.*, 1993), opnieuw in discrete tijd. Hier wordt het intercept in het regressiemodel voorzien van een toegevoegde a priori verdeling die dan geactualiseerd wordt met de

stelling van Bayes. Deze techniek kan gebruikt worden voor aantallen, binomiale of multinomiale gegevens. Merk op dat benaderingen vereist zijn voor de logistische modellen indien verklarende variabelen aanwezig zijn. Deze dynamische veralgemeende lineaire modellen kunnen verder veralgemeend worden naar vectoren van observaties (Fahrmeir, 1992) in discrete tijd. Hier wordt ondersteld dat de dynamische regressieparameters evolueren volgens de transitievergelijkingen die hun eerste twee momenten vastleggen en daarbij normaal verdeeld zijn. Het gebruik van de a posteriori modus in plaats van het traditionele a posteriori gemiddelde bij het schatten van parameters, vermijdt het berekenen van numerische multidimensionale integralen.

Heterogeniteit, waarmee rekening gehouden werd in verschillende van hoger genoemde modellen, wordt dan meer systematisch besproken in de volgende sectie over random effecten. Het begrip wordt ingevoerd in conditionele modellen door een verdeling toe te kennen aan het intercept in regressiemodellen. Random effecten worden verder veralgemeend naar de andere parameters in de regressie, hetgeen leidt tot het random coëfficiënt model. Niettegenstaande we besloten hebben ons te concentreren op de conditionele modellen, geven we toch een kort overzicht van marginale modellen in de analyse van longitudinale data. De veralgemeende schattingsvergelijkingen worden gegeven, alhoewel meestal geen echt model correspondeert met deze 'score' vergelijkingen. We vermelden ook marginale likelihood methoden. Tenslotte geven we een overzicht van de literatuur rond de hogervermelde onderwerpen.

Hoofdstuk 2: Modelleren van positieve longitudinale data

In dit tweede hoofdstuk wordt een aanvang genomen met onze eigen bijdrage tot het modelleren van niet-normale longitudinale data. Het dient echter gezegd dat de discussie in Hoofdstuk 1 over conditionele modellen reeds enig nieuw materiaal aanbracht in de discussie. Zoals de titel aangeeft is Hoofdstuk 2 gewijd aan het modelleren van reeksen van positieve gegevens. Dit onderwerp is van bijzonder belang in verschillende wetenschappelijke domeinen zoals menselijke en dierlijke geneeskunde, economie, enz. Zoals aangegeven in het eerste hoofdstuk zal onze aandacht vooral gaan naar het modelleren van biomedische data. Meer in het bijzonder ontwikkelen we een model voor het analyseren van maten van de triglyceride profielen voor twee groepen van vier Beagle honden onder vier types van vezelrijk dieet. De data zijn typisch positief en naar onder begrensd door nul, hetgeen maakt dat de verdeling scheef kan zijn.

Een tweede technisch probleem is dat de respons gemeten werd op niet equidistante tijdstippen, hetgeen betekent dat een continue tijd opzet meer geschikt zal zijn dan een discrete tijd model waarbij sommige data worden verondersteld op willekeurige wijze te ontbreken (missing at random). Het gegeneraliseerd autoregressie model (GARM) wordt dan ontwikkeld als een

instrument om reeksen te modelleren die op onregelmatige tijdstippen gemeten werden. Seriële associatie wordt gemodelleerd met twee extra parameters in vergelijking met het traditionele model dat onafhankelijkheid onderstelt. Er worden verschillende manieren gesuggereerd om heterogeniteit te modelleren, zoals onder andere een niet-parametrisch random effect model dat zal gebruikt worden om de triglyceride data te analyseren. Geen speciale onderstelling wordt gemaakt over het type verdeling, hetgeen betekent dat het GARM kan gebruikt worden in andere situaties (zie Hoofdstuk 4).

De familie van veralgemeende gamma verdelingen die de exponentiële, de Weibull, de gamma en de lognormale als limietgevallen heeft, wordt voorgesteld om positieve longitudinale data te analyseren. Deze verschillende modellen worden vergeleken met het Akaike informatie criterium (AIC) hetgeen toelaat de aanpassing te berekenen en te vergelijken in niet-geneste modellen. De traditionele normale en lognormale verdelingen, die in de praktijk zeer veel gebruikt worden om dit soort data te analyseren, blijken hier totaal ongeschikt te zijn. Onze finale keuze is de Weibull verdeling.

Hoofdstuk 3: Longitudinale aantallen in continue tijd

Dit hoofdstuk is gewijd aan het modelleren van discrete longitudinale gegevens op onregelmatig gespatieerde tijdspunten. De oorspronkelijke idee was dan de dynamische veralgemeende lineaire modellen van West *et al.* (1985) en Harvey en Fernandes (1989) te veralgemenen voor aantallen in continue tijd. In essentie wordt aan het intercept in het log-lineair model een gamma a priori (toegevoegde) verdeling toegekend. Deze a priori verdeling wordt gebruikt om de toestand van het proces te beschrijven bij de volgende observatietijd door eenvoudigweg een gamma predictieve verdeling te beschouwen met dezelfde modus als de oorspronkelijke a priori verdeling, maar met een kleiner Fisher informatie in dit punt.

Van zodra een observatie beschikbaar is kan de verdeling van het intercept (dat een soort residu is in een model met onafhankelijkheid) geactualiseerd worden met de stelling van Bayes. De resulterende (niet-conditionele) likelihood is dan het produkt van de negatief binomiale dichtheden over de observatietijden. Deze techniek, gebaseerd op de Kalman filter, kan aangepast worden om de gevoeligheid voor extreme observaties te reduceren en om de Poisson verdeling meer op te nemen als bijzonder geval. Deze laatste verbetering is zeer gewenst in de praktijk, aangezien gebleken is dat het Harvey en Fernandes (1989) model in sommige voorbeelden een hoger AIC heeft dan het negatief binomiaal model, dat enkel heterogeniteit in rekening brengt. Vervolgens worden twee voorbeelden gegeven en geanalyseerd met deze nieuwe techniek. Het eerste bestudeert het ademhalingstempo profiel van kalveren die onderworpen werden aan een dertig minuten in-

jectie van een receptor blocker die een acuut ademhalingsnood syndroom simuleert bij gezonde dieren. De data voor, tijdens en na de injectie met het geneesmiddel, werden verzameld op onregelmatig gespatieerde tijdstippen. In het tweede voorbeeld bestuderen we de evolutie van drie kolonies van *Paramecium aurelium* in hetzelfde voedingsmidden over een periode van twintig dagen. De bedoeling is een profielvergelijking te construeren die het gemiddeld aantal individuen geeft op elk tijdstip. Een interessant kenmerk van deze data is de uitgesproken stabilisatie van de grootte van de kolonie na de tiende dag. Vandaar dat de methodes gebaseerd op veeltermen of splines niet in staat zijn met dit gedrag rekening te houden. Biologische modellen zouden meer geschikt zijn in deze situatie dan om het even welk wiskundig artefact. Vandaar dat een veralgemening van de logistische groeikromme (Nelder, 1961 en 1962), verder ontwikkeld door Heitjan (1991a en b), voorgesteld wordt om het profiel te modelleren.

Hoofdstuk 4: Andere toepassingen van het GARM

In dit hoofdstuk wordt het gegeneraliseerd autoregressie model beschouwd in een andere context dan in Hoofdstuk 2. Daar werd geen speciale onderstelling gemaakt over de vorm van de dichtheidsfunctie in de GARM definitie. Bijgevolg is er niets dat ons belet om het GARM te gebruiken voor het modelleren van reeksen van binaire, binomiale en multinomiale data. Dat idee wordt eerst gebruikt om reeksen van binaire data te analyseren over de geschiktheid tot eileggen bij de volwassen *Rhagoletis pomonella* vrouwelijke vliegen (Stanek en Diehl, 1988) op twee soorten substraat. Onze bedoeling is eenvoudig het modelleren van de respons profielen door toe te laten dat de respons van een vlieg op een gegeven dag afhangt van het verleden, het soort vlieg en het soort substraat.

Het tweede voorbeeld bestaat uit een reeks multinomiale data over het aantal *Pinus*, *Abies*, *Quercus* en *Alnus* pollen in steekproeven van grootte 100 (Mosimann, 1962). Deze steekproeven werden genomen op dezelfde kern en op verschillende diepten in de bodem. Deze bodemdiepten worden equidistant ondersteld. Op die manier wordt ruimte gebruikt in plaats van tijd als een kwalitatief instrument om de observaties te ordenen en een maat voor seriële associatie te definiëren. Het volgende theoretisch punt is het GARM te herschrijven als een GLM waarbij de covariabelen lineair voorkomen in het systematisch gedeelte. Bij deze onderstelling zijn de regressieparameters in het GARM te berekenen met GLM software zoals GLIM of S-PLUS. De andere twee of drie regressieparameters moeten geschat worden ofwel via niet-lineaire optimizatie of door de likelihood te beschouwen in een rooster van zinvolle waarden.

Tenslotte wordt de laatste sectie gewijd aan het toepassen van voorspellingsmethoden op reeksen van niet-normale data. Vanuit het bijzonder geval van aantallen met overdispersie wordt aangetoond hoe het GARM kan

gebruikt worden samen met een benaderde predictieve likelihood (Butler, 1986). Deze benaderde predictieve likelihood, waarbij de toekomstige observatie nu de rol speelt van parameter in de likelihoodfunctie, wordt gebruikt in plaats van de beter bekende profiel predictieve likelihood (Fisher, 1959, p. 128-133). Dit gebeurt omdat er niet ondersteld wordt dat de (regressie) parameters vaste grootheden zijn die gelijk zijn aan hun maximum likelihood schattingen. Bijgevolg wordt rechtstreeks rekening gehouden met de onnauwkeurigheid te wijten aan parameterschatting door predicties te maken vanuit het geschatte model.

Er wordt een voorbeeld gegeven, waarin de ontbrekende gegevens in een kunstmatig afgebroken reeks van de *Paramecium aurelium* data in Hoofdstuk 3 voorspeld worden door gebruik te maken van de overblijvende informatie in de drie reeksen. De resulterende predictie omhullende blijken, zoals verwacht, breder te zijn dan deze verkregen met predictieve likelihood.

Tenslotte bespreken we in een kort hoofdstuk de voordelen van de hierboven voorgestelde modellen.

8
Résumé

Modèles pour données longitudinales non-normales en temps continu, basés sur la fonction de vraisemblance

Ce travail tire son origine du constat que la plupart des méthodes d'analyse de données longitudinales non-normales font l'hypothèse que les données ont été observées en des temps équidistants. Les auteurs de tels articles mentionnent fréquemment qu'une extension en temps continu ne pose pas de problèmes, mais ils ne détaillent pratiquement jamais une telle démarche. Pourtant, de telles généralisations s'avèrent habituellement difficiles, et même souvent impossibles d'un point de vue théorique car le modèle de départ fait un large usage de la structure discrète du temps.

Un deuxième constat préoccupant en analyse longitudinale en temps continu, est le monopole des modèles marginaux, peu de modèles conditionnels étant disponibles dans ce contexte. Ce n'est pas un problème dans des études observationnelles, telles que l'épidémiologie, pour lesquelles les méthodes marginales existent et sont bien adaptées pour répondre aux questions surgissant dans de tels contextes. Cependant, lorsque le mécanisme générant les données est la préoccupation majeure de l'étude, quand la façon dont les données évoluent au cours du temps est au centre des questions, comme dans les essais cliniques et autres expériences biologiques, des modèles conditionnels sont requis (Sheiner *et al.*, 1989; Davidian et Giltinan, 1995, p. 122).

Pour toutes ces raisons, le développement de modèles conditionnels pour des données longitudinales non-normales en temps continu s'est avéré être un défi intéressant.

Une autre source de motivation était de se limiter à des méthodes basées sur la fonction de vraisemblance, cette dernière permettant de comparer les mérites de modèles de nature différente. Ces comparaisons peuvent se faire, soit directement en calculant la probabilité d'observer les données étudiées sous les hypothèses considérées, soit en utilisant l'AIC (Akaike, 1973) qui pénalise les modèles complexes. Cette façon de procéder est impossible avec la plupart des modèles marginaux (avec quelques exceptions notables telles que Molenberghs et Lesaffre, 1994), parce qu'ils sont essentiellement basés sur les équations généralisées d'estimation (Liang et Zeger, 1986) qui,

dans la majorité des cas, ne peuvent pas être réintégrées pour obtenir la fonction de vraisemblance (McCullagh et Nelder, 1989, 2nd Ed., section 9.3.2).

Après ces quelques remarques motivantes, nous décrivons la structure de notre travail.

Chapitre 1: Introduction

Dans un premier temps, nous présentons quelques situations concrètes où les données longitudinales surviennent. Quelques exemples du même type seront analysés en détail afin d'illustrer les propos théoriques des chapitres suivants. Les différentes étapes requises pour collecter et analyser des données longitudinales sont alors décrites. Les thèmes autour desquels ce travail est construit, sont énoncés avec les renvois correspondants dans le texte. Nous exposons aussi des arguments militant en faveur de l'usage de modèles conditionnels (par opposition aux modèles marginaux) ayant une fonction de vraisemblance sous-jacente. Les difficultés supplémentaires survenant dans un contexte longitudinal (l'hypothèse d'indépendance usuelle des études transversales étant la plupart du temps irréaliste) sont également discutées. Les différentes méthodes utilisées dans la littérature pour modéliser ces nouvelles caractéristiques, nommément l'hétérogénéité et la dépendance sérielle, sont le sujet des sections suivantes.

Les modèles autorégressifs normaux, qui sont équivalents dans un contexte conditionnel et marginal, sont alors présentés. Par la suite, nous énumérons les différentes façons de combiner les effets aléatoires, modélisant l'hétérogénéité, avec les termes autorégressifs, modélisant la dépendance sérielle. Nous considérons l'équivalent de ces modèles autorégressifs dans un contexte non-normal. Nous montrons comment il est possible, par un conditionnement adéquat, de comparer des modèles d'ordres différents.

Un modèle basé sur le filtre de Kalman (Jones et Ackerson, 1990; Jones, 1993), le modèle linéaire dynamique, est alors mis en avant comme une alternative dans un contexte normal. Nous montrons alors la flexibilité et les avantages numériques du filtre de Kalman, qui permet de calculer la fonction de vraisemblance de manière itérative (dynamiquement) dans des situations compliquées, même lorsque le temps est supposé continu.

Le modèle linéaire généralisé dynamique (West *et al.*, 1985) élargit le domaine d'application de cette dernière approche à des données non-normales observées en des temps équidistants. Celui-ci est construit en donnant une distribution conjuguée à la partie linéaire du désormais bien connu modèle linéaire généralisé (statique). Cette distribution conjuguée est alors remise à jour en utilisant le théorème de Bayes. Le choix de la distribution conjuguée comme distribution a priori, résulte en des formes analytiques explicites pour la distribution postérieure et la vraisemblance. Ceci permet d'éviter l'évaluation numérique d'intégrales, qui, au fur et à mesure que les

observations s'accumulent, devraient être calculées de façon répétée pour estimer d'autres intégrales.

D'autres approches permettant de modéliser des séries de comptages sont aussi présentées (Harvey et Fernandes, 1989; Ord *et al.*, 1993), de nouveau en temps discret. On y donne une distribution a priori conjuguée au terme indépendant de la régression canonique, qui est mise à jour à l'aide du théorème de Bayes. Cette technique peut être utilisée avec des données de comptage, binomiales et multinomiales. Remarquons que des approximations sont requises avec des modèles logistiques comportant des variables explicatives.

Ces modèles linéaires généralisés dynamiques peuvent être étendus pour modéliser des vecteurs d'observations (Fahrmeir, 1992), tout en continuant de supposer que le temps est discret. Les paramètres de régression dynamiques, en plus d'avoir une distribution normale, sont supposés évoluer conformément à des équations de transition fixant leurs deux premiers moments. L'usage du mode a posteriori plutôt que de la moyenne a posteriori, évite de devoir calculer numériquement des intégrales multidimensionnelles.

L'hétérogénéité, dont nous avons tenu compte dans certains modèles parmi les précédents, est traitée de façon systématique dans la section relative aux effets aléatoires. Cette notion est introduite dans les modèles conditionnels en donnant une distribution au terme indépendant des modèles de régression. L'usage de cette technique peut être étendu aux autres paramètres de la régression, donnant un modèle à coefficients aléatoires.

Bien que nous ayons décidé de concentrer nos efforts sur le développement de modèles conditionnels, nous proposons de revoir brièvement quelques modèles marginaux fondamentaux en analyse longitudinale. Les équations d'estimation généralisées sont présentées, bien qu'elles ne puissent être dérivées d'aucune fonction de vraisemblance.

Enfin, nous proposons un survol de la littérature relative aux thèmes mentionnés précédemment.

Chapitre 2: Modélisation de données longitudinales positives

Ce deuxième chapitre marque réellement le début de notre contribution à la modélisation de données longitudinales non-normales, bien que la discussion des modèles conditionnels dans le premier chapitre ait déjà apporté quelque chose de nouveau au débat. Comme son titre l'indique, le deuxième chapitre est consacré à la modélisation de données positives. Ce sujet est particulièrement important dans des domaines tels que la médecine humaine et vétérinaire, l'économie, etc. Ainsi que nous l'avons mentionné dans le premier chapitre, nous nous consacrons essentiellement à la modélisation de données biomédicales. Plus particulièrement, nous développons un modèle permettant d'analyser les profils de triglycérides de deux groupes de quatre chiens de race Beagle nourris à l'aide d'un régime choisi parmi quatre ali-

mentations à base de fibres. Ces données sont typiquement positives et bornées à gauche par zéro, ce qui pourrait donner lieu à une distribution asymétrique. Un deuxième problème technique survient parce que les données n'ont pas été observées en des temps équidistants. Un modèle en temps continu serait plus approprié qu'un modèle en temps discret supposant que certaines données sont aléatoirement manquantes. Pour cette raison, nous avons développé le modèle autorégressif généralisé (MAG), qui permet de modéliser des séries de données échantillonnées de façon irrégulière. La dépendance sérielle est modélisée avec deux paramètres additionnels par rapport au modèle d'indépendance. Nous suggérons plusieurs méthodes pour rendre compte de l'hétérogénéité. En particulier, nous considérons un effet aléatoire non-paramétrique pour analyser les données relatives aux chiens. Aucune hypothèse particulière n'a été émise quant à la distribution des données. Dès lors, rien ne s'oppose à l'utilisation du MAG dans d'autres contextes (voir chapitre 4). La famille de distributions gamma généralisées, qui inclut comme membres connus, les distributions exponentielle, Weibull, gamma et log-normale à la limite, est proposée pour analyser des données positives longitudinales. Ces différents modèles sont comparés en utilisant le critère d'information de Akaike (AIC), qui permet d'évaluer et de comparer l'ajustement de modèles non-imbriqués. Nous montrons que, dans notre exemple, les traditionnels modèles normal et log-normal, communément utilisés pour analyser ce genre de données, sont complètement inadaptés, la Weibull étant notre choix final.

Chapitre 3: Modélisation de comptages longitudinaux en temps continu

Nous consacrons ce chapitre à la modélisation de données longitudinales discrètes observées en des temps irrégulièrement espacés. L'idée de départ était de généraliser le modèle linéaire généralisé dynamique de West *et al.* (1985) et de Harvey et Fernandes (1989) pour modéliser des séries de comptage en temps continu. Fondamentalement, on donne au terme indépendant du modèle log-linéaire, une distribution a priori gamma (conjuguée). Cette dernière est utilisée pour prédire l'état du processus au temps d'observation suivant, en considérant une distribution gamma (prédictive) avec le même mode que la distribution a priori en ce point, mais avec une information de Fisher plus petite. Dès qu'une observation est disponible, la distribution a priori du terme indépendant, qui est en quelque sorte un résidu dans un modèle faisant l'hypothèse d'indépendance, est mise à jour à l'aide du théorème de Bayes. La vraisemblance (inconditionnelle) résultante est alors le produit de densités binomiales négatives, toute nouvelle observation amenant un facteur supplémentaire dans ce produit.

Cette technique, basée sur le filtre de Kalman, peut être modifiée afin de réduire la sensibilité du processus aux observations extrêmes et d'assurer

que la distribution de Poisson en soit un cas particulier. Cette dernière amélioration est particulièrement désirable en pratique, car le modèle de Harvey et Fernandes (1989) présente, dans quelques exemples, un plus grand AIC que le modèle binomial négatif, qui tient seulement compte de l'hétérogénéité.

Nous donnons ensuite deux exemples que nous nous proposons d'analyser à l'aide de cette technique. Le premier étudie l'effet d'une perfusion de trente minutes de 5-hydroxytryptamine sur le profil respiratoire de veaux. Cette substance est utilisée pour simuler des syndromes de détresse respiratoire chez des bovins sains. Les données ont été collectées avant, pendant, et après la perfusion, et ce, de façon irrégulière dans le temps.

Dans le deuxième exemple, nous étudions, sur une période de vingt jours, l'évolution de trois colonies de *Paramecium aurelium* observées dans des milieux nutritivement équivalents. Notre but est de construire un modèle estimant le nombre moyen d'individus présents à chaque instant dans chacune des colonies. Il est intéressant de remarquer que la taille des colonies semble se stabiliser après environ dix jours. Il est évident que toute méthode basée sur des polynômes ou des splines ne peut pas rendre compte de ce comportement. Dans ce genre de situation, un modèle biologique serait plus approprié qu'un quelconque artifice mathématique. Pour cette raison, nous proposons d'utiliser une généralisation de la courbe de croissance logistique (Nelder, 1961 et 1962), que Heitjan (1991a et b) a plus amplement développée, pour modéliser ce genre de profil.

Chapitre 4: Autres applications du MAG

Ce chapitre considère le modèle autorégressif généralisé dans un autre contexte que dans le deuxième chapitre. Aucune hypothèse sur la forme de la densité n'a été faite lors de la définition du MAG.

Par conséquent, rien ne nous empêche d'utiliser le MAG pour modéliser des séries de données binaires, binomiales ou multinomiales. Dans un premier temps, nous exploitons cette idée pour analyser des séries de données binaires (Stanek et Diehl, 1988) décrivant l'aptitude de mouches femelles *Rhagoletis pomonella* à pondre leurs oeufs sur deux types de support. Notre but est simplement de modéliser le profil de la réponse, en tenant compte que le comportement d'une mouche un jour donné peut non seulement être expliqué par l'origine de l'insecte et le type du support (fruitier) proposé, mais aussi par ses comportements antérieurs.

Le second exemple, une série de données multinomiales, donne le nombre de grains de pollen de type *Pinus*, *Abies*, *Quercus* et *Alnus* dans des échantillons de cent unités (Mosimann, 1962). Ces échantillons proviennent de carottes prélevées dans le sol à des profondeurs croissantes que nous supposons régulièrement espacées. Par conséquent, nous utilisons maintenant l'espace au lieu du temps comme outil qualitatif permettant de classer les

observations et de définir une mesure de dépendance sérielle.

L'objectif théorique suivant est de réécrire le MAG sous forme d'un modèle linéaire généralisé, lorsque les variables explicatives apparaissent linéairement dans la partie systématique du modèle. Sous cette hypothèse, il est possible d'estimer les paramètres de régression du MAG en utilisant des logiciels tels que GLIM ou S-PLUS. Les deux ou trois paramètres restants, relatifs à l'autorégression, doivent, quant à eux, être estimés à l'aide d'une routine d'optimisation non-linéaire, ou en évaluant la fonction de vraisemblance sur une grille de valeurs raisonnables pour ces paramètres.

Enfin, nous consacrons la dernière partie de ce travail à une méthode de prédiction dans un contexte non-normal. A partir du cas particulier de séries de comptages surdispersés, nous montrons comment le MAG peut être utilisé conjointement avec une vraisemblance prédictive approximative (Butler, 1986) pour construire des 'enveloppes de confiance'. Cette vraisemblance prédictive approximative, dans laquelle l'observation à venir joue le rôle d'un paramètre, est utilisée au lieu de la classique vraisemblance prédictive (Fisher, 1959, p. 128-133), parce que cette dernière suppose que les paramètres (de régression) sont des quantités fixes égales à leur maximum de vraisemblance. L'imprécision présente dans l'estimation des paramètres est directement prise en compte lorsque des prédictions sont dérivées du modèle estimé.

Nous considérons alors un exemple où les données manquantes provenant de la troncature d'une des trois séries relatives aux colonies de *Paramecium aurelium* présentées dans le troisième chapitre, sont prédites à partir des observations restant dans les trois séries. Nous montrons que les enveloppes de prédiction résultantes sont, sans surprise, plus larges que celles obtenues avec la vraisemblance prédictive traditionnelle.

Pour conclure, nous présentons un court chapitre discutant des avantages des modèles présentés précédemment.

Bibliography

1. Agresti, A. (1989) A survey of models for repeated ordered categorical response data. *Statistics in Medicine*, **8**: 1209–1224.
2. Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In Petrov, B.N. and Csáki, F., *Second International Symposium on Inference Theory*, Budapest: Akadémia Kiadó, pp. 267–281.
3. Albert, P.S. (1994) A Markov model for sequences of ordinal data from a relapsing-remitting disease. *Biometrics*, **50**: 51–60.
4. Albert, P.S., McFarland, H.F., Smith, M.E., Frank, J.A. (1994) Time series for modelling counts from a relapsing-remitting disease: application to modelling disease activity in multiple sclerosis. *Statistics in Medicine*, **13**: 453–466.
5. Altham, P.M. (1984) Improving the precision of estimation by fitting a model. *Journal of the Royal Statistical Association*, **B46**
6. Anderson, D. and Hinde, J. (1988) Random effects in generalized linear models and the EM algorithm. *Communications in Statistics*, **A17**: 3847-56.
7. Atkinson, A.C. (1980) A note on the generalized information criterion for choice of a model. *Biometrika*, **67**: 413–418.
8. Azzalini, A. (1994) Logistic regression for autocorrelated data with application to repeated measures. *Biometrika*, **81**: 767–775.
9. Barndorff-Nielsen, O.E. (1983) On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.
10. Barndorff-Nielsen, O.E. (1993) Adjusted likelihood inference about interest parameters. *Theory Probab. Appl.*, **38**, 179–193.
11. Becker, M. and Balagtas, C.C. (1993) Marginal modelling of binary crossover data. *Biometrics*, **49**, 997–1009.
12. Bhansali, R.J. and Downham, D.Y. (1977) Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPF criterion. *Biometrika*, **67**: 547–551.
13. Bjornstad, J.F. (1990) Predictive likelihood: a review. *Statistical Science*, **5**, 242–265.
14. Bolstad, W.M. (1986) Harrison-Stevens forecasting and the multiprocess dynamic linear model. *The American Statistician*, **40**: 129–135.
15. Bolstad, W.M. (1995) The multiprocess dynamic Poisson model. *Journal of the American Statistical Association*, **90**: 227–232.
16. Butler, R.W. (1986) Predictive likelihood inference with applications (with discussion). *Journal of the Royal Statistical Society*, **B48**, 1–38.

17. Butler, R.W. (1989) Approximate predictive pivots and densities. *Biometrika*, **76**, 489–501.
18. Carey, V., Zeger, S.L. and Diggle, P.J. (1993) Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, **80**: 517–526.
19. Collet, D. (1991) *Modelling of Binary Data*. London: Chapman and Hall.
20. Conaway, M.R. (1990) A random effects model for binary data. *Biometrics*, **46**: 317–328.
21. Cox, D.R. and Snell (1970) *Analysis of Binary Data*. Chapman and Hall.
22. Crowder, M. (1995) On the use of a working correlation matrix in using generalised linear models for repeated mesures. *Biometrika*, **82**: 407–10.
23. Davidian, M. and Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurements*. London: Chapman and Hall.
24. Davies, R.B. (1987) Mass point methods for dealing with nuisance parameters in longitudinal studies. In *Longitudinal Data Analysis, Surrey Conferences on Sociological Theory and Method*. Averbury, Ed. Robert Crouchley.
25. Davison, A.C. and Gigli, A. (1989) Deviance residuals and normal scores plots. *Biometrika*, **76**: 211–221.
26. Davison, A.C. and Tsai, C.-L. (1992) Regression models and diagnostics. *International Statistical Review*, **60**: 337–353.
27. Diaconis, P. and Ylvisaker, D. (1979) Conjugate priors for exponential families. *The Annals of Statistics*, **7**: 269–281.
28. Diderrich, G.T. (1985) The Kalman filter from the perspective of Goldberger-Theil estimators. *The American Statistician*, **39**: 193–198.
29. Diggle, P.J. (1988) An approach to the analysis of repeated measurements. *Biometrics*, **44**: 959–71.
30. Diggle, P.J. (1990) *Times Series. A biostatistical Introduction*. Oxford: Oxford University Press.
31. Diggle, P.J. and Kenward, M.G. (1994) Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, **43**: 49–93.
32. Diggle, P.J., Liang, K-Y. and Zeger, S.L. (1994) *Analysis of Longitudinal Data*. Oxford: Oxford Statistical Science Series.
33. Diggle, P.J. and Zeger, S.L. (1989) A non-Gaussian model for time series with pulses. *Journal of the American Statistical Association*, **84**: 354–359.
34. Fahrmeir, L. (1992) Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, **87**: 501–509.
35. Fahrmeir, L. and Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag.
36. Fahrmeir, L. and Tutz, G. (1994b) Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association*, **89**: 1438–1449.
37. Fisher R.A. (1959) *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.
38. Fitzmaurice, G.M. and Laird, N.M. (1993) A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**: 141–151.

39. Fitzmaurice, G.M., Laird, N.M. and Rotnitzky (1993) Regression models for discrete longitudinal responses. *Statistical Science*, *8*: 284–309.

40. Fitzmaurice, G.M. and Lipsitz, S.R. (1995) A model for binary time series data with serial odds ratio patterns. *Applied Statistics*, **44**: 51–61.

41. Fitzmaurice, G.M., Molenberghs, G. and Lipsitz, S.R. (1995) Regression models for longitudinal binary responses with informative drop-outs. *Journal of the Royal Statistical Society*, **B57**: 691–704.

42. Francis, B. (1994) Poisson growth curves, temporal dependence and over-dispersion - assessing treatment effects through repeated leaf counts. *Proceedings of the IX International Workshop on Statistical Modelling, Exeter, U.K.*

43. Gamerman, D. (1992) A dynamic approach to the statistical analysis of point processes. *Biometrika*, **79**: 39–50.

44. Gause, G.F. (1934) *The struggle for existence*. Williams and Williams: Baltimore.

45. Génicot, B., Mouligneau, F., Lindsey, J.K., Lambert, P., Close, R. and Lekeux, P. (1993) Induction of a serotonin-S_2 receptor blockade during early or late stage of acute respiratory distress syndrome in double-muscled calves: a comparative study. *Journal of Veterinary Medicine*, **40**: 241–248.

46. Glasbey, C.A. (1979) Correlated residuals in non-linear regression applied to growth data. *Applied Statistics*, **28**: 251–259.

47. Glasbey, C.A. (1980) Non-linear regression with autoregressive time series errors. *Biometrics*, **36**: 135–140.

48. Gordon, K. and Smith, A.F. (1990) Modeling and monitoring biomedical time series. *Journal of the American Statistical Association*, **85**: 328–337.

49. Gordon, N.J. and Smith, A.F. (1993) Approximate non-Gaussian Bayesian estimation and modal consistency. *Journal of the Royal Statistical Society*, **B55**: 913–918.

50. Gottschau, A. (1992) Exchangeability in multivariate Markov chain models. *Biometrics*, **48**: 751–763.

51. Gottschau, A. (1994) Markov chain models for multivariate binary panel data. *Scandinavian Journal of Statistics*, **21**: 57–71.

52. Grunwald, G.K., Raftery, A.E. and Guttorp, P. (1993) Time series of continuous proportions. *Journal of the Royal Statistical Society*, **B55**: 103–116.

53. Harvey, A.C. and Fernandes, C. (1989) Time series models for count or qualitative observations. *Journal of Business and Economic Statistics*, **7**: 407–423.

54. Heckman, J.J. (1981) The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Menski, C.F. and McFadden, D., *Structural Analysis of discrete data with economic applications.*

55. Heitjan, D.F. (1991a) Generalized Norton-Simon models of tumor growth. *Statistics in Medicine*, **10**: 1075–1088.

56. Heitjan, D.F. (1991b) Nonlinear modeling of serial immunologic data: a case study. *Journal of the American Statistical Association*, **86**: 891–898.

57. Hinkley, D. (1979) Predictive likelihood. *The Annals of Statistics*, **79**, 718–728.
58. Jennrich, R.I. and Schluchter, M.D. (1986) Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, **42**: 805–820.
59. Jones, R.H. (1993) *Longitudinal Data with Serial Correlation : a State-space Approach*. London: Chapman and Hall.
60. Jones, R.H. and Ackerson, L.M. (1990) Serial correlation in unequally spaced longitudinal data. *Biometrika* **77**: 721–731.
61. Jones, R.H. and Boadi-Boateng, F. (1991) Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics*, **47**: 161–175.
62. Jørgensen, B. (1987) Exponential dispersion models. *Journal of the Royal Statistical Society*, **B49**: 127–162.
63. Jørgensen, B. (1992) Exponential dispersion models and extensions: a review. *International Statistical Review*, **60**: 5–20.
64. Jørgensen, B., Lundbye-Christensen, S., Song, X. and Sun, L. (1994a) A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Technical Report*, University of British Columbia: Vancouver.
65. Jørgensen, B., Labouriau, R. and Lundbye-Christensen, S. (1994b) Linear growth curve analysis based on exponential dispersion models. *Technical Report 140*, University of British Columbia: Vancouver.
66. Jørgensen, B., Lundbye-Christensen, S., Song, X. and Sun, L. (1995) A state space model for multivariate longitudinal data. *Technical Report 148*, University of British Columbia: Vancouver.
67. Kalbfleisch, J.D. (1985) *Probability and Statistical Inference*. Volume 2: *Inference*. New York: Springer Verlag.
68. Kalbfleisch, J.D. and Sprott, D.A. (1970) Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society*, **B32**, 175–208
69. Keenan, D.M. (1982) A time series analysis of binary data. *Journal of the American Statistical Association*, **77**: 816–821.
70. Kitagawa, G. (1987) Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association* **82**: 1032–1063.
71. Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, J.R. and Lehnen, R.G. (1977) A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, **33**: 133–158.
72. Komaki, F. (1993) State-space modelling of time series sampled from continuous processes with pulses. *Biometrika*, **80**: 417–29.
73. Korn, E.L. and Whittemore, A.S. (1979) Methods for analysing panel studies of acute health effects of air pollution. *Biometrics*, **35**: 795–802.
74. Laird, N.M. (1988) Missing data in longitudinal studies. *Statistics in Medicine*, **7**: 305–15.
75. Laird, N.M. and Ware, J.H. (1982) Random effects models for longitudinal data. *Biometrics*, **38**: 963–974.
76. Lambert P. (1996a) Modelling of repeated series of count data measured at unequally spaced times. *Applied Statistics*, **45** (in press).

77. Lambert, P. (1996b) Modelling of non-linear growth curve on series of correlated count data measured at unequally spaced times: a full likelihood based approach. *Biometrics*, **51** (in press).
78. Lambert, P. (1996c) Modelling irregularly sampled profiles of nonnegative dog triglyceride responses under different distributional assumptions. *Statistics in Medicine* (in press).
79. Lambert, P. (1995d) The generalized autoregression model as a tool for modelling series of irregularly sampled binomial and multinomial data. Submitted to *Biometrics*.
80. Lambert, P. (1995e) Predictions in overdispersed series of counts using an approximate predictive likelihood. Submitted to *Applied Statistics*.
81. Lejeune, M. and Faulkenberry, G.D. (1982) A simple predictive density function. *Journal of the American Statistical Association*, **77**, 654–659.
82. Li, W.K. (1994) Time series models based on generalized linear models: some further results. *Biometrics*, **50**: 506–511.
83. Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**: 13–22.
84. Lindsey, J.K. (1993) *Models for Repeated Measurements*. Oxford: Oxford Statistical Science Series.
85. Lindsey, J.K. (1994) A direct likelihood inference derivation of the AIC or why likelihood ratio tests can be misleading. *Publications du Laboratoire de Statistique et Probabilités, Toulouse*, **10**: 56–62.
86. Lindsey, J.K. (1995a) *Modelling Frequency and Count Data*. Oxford: Oxford Science Series.
87. Lindsey, J.K. (1995b) *Introductory Statistics*. Oxford: Oxford Science Series.
88. Lindsey, J.K. and Jones, B. (1995) Choosing among generalized linear models applied to medical data. Submitted to *Statistics in Medicine*.
89. Lindsey, J.K. and Lambert, P. (1995) Dynamic generalized linear models and repeated measurements. *Journal of Statistical Planning and Inference*, **47**: 129–139.
90. Lipsitz, S.R., Fitzmaurice, G.M., Sleeper, L. and Zhao, L.P. (1995) Estimation methods for the joint distribution of repeated binary observations. Biometrics, **51**: 562–570.
91. Lipsitz, S.R., Laird, N.M. and Harrington, D.P. (1991) Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, **78**: 153–160.
92. Lipsitz, S.R., Laird, N.M. and Harrington, D.P. (1992) A three-stage estimator for studies with repeated and possibly missing binary outcomes. *Applied Statistics*, **41**: 203–213.
93. Little, R.J. and Rubin, D.B. (1987) *Statistical Analysis with Missing Data*. Wiley: New-York.
94. Mauger, D.T., Brown, M.B. and Kushler, R.H. (1995) A comparison of methods that characterize pulses in time series. *Statistics in Medicine*, **14**: 311–325.
95. McCullagh, P. and Nelder, J.A. (1989, 2nd ed.) *Generalized Linear Models*. London: Chapman and Hall.

96. McDonald, B.W. (1994) Two random effects models for multivariate binary data. *Biometrics*, **50**: 164–172.

97. Meinhold, R.J. and Singpurwalla, N.D. (1983) Understanding the Kalman filter. *The American Statistician*, **37**: 123–127.

98. Meinhold, R.J. and Singpurwalla, N.D. (1989) Robustification of the Kalman filter models. *Journal of the American Statistical Association*, **84**: 479–486.

99. Molenberghs, G. and Lesaffre, E. (1994) Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, **89**: 633–644.

100. Mosimann ,J.E. (1962) On the compound multinomial distribution, the multivariate beta-distribution and correlations among distributions. *Biometrika*, **49**: 65–82.

101. Muenz, L.R. and Rubinstein, L.V. (1985) Markov models for covariate dependence of binary sequences. *Biometrics*, **41**: 91–101.

102. Muñoz, A., Carey, V., Schouten, J.P., Segal, M. and Rosner, B. (1992) A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*, **48**: 733–742.

103. Nelder, J.A. (1961) The fitting of a generalization of the logistic curve. *Biometrika*, **17**: 89–100.

104. Nelder, J.A. (1962) An alternative form of a generalized logistic equation. *Biometrics*, **18**: 614–616.

105. Núñez-Antón, V. and Woodworth, G.G. (1994) Analysis of longitudinal data with unequally spaced observations and time-dependent correlated errors. *Biometrics*, **50**: 445–456.

106. Ord, K., Fernandes, C. and Harvey, A.C. (1993) Time series models for multivariate series of count data. *Developments in Times Series Analysis. In honor of Maurice B. Priestley. Ed. Rao, S.* Chapman and Hall.

107. Oskrochi, GH.R. (1994) *Analysis of Censored Correlated Observations.* K.U. Leuven: unpublished PhD thesis.

108. Paik, M.C. (1992) Parametric variance function estimation for nonnormal repeated measurement data. *Biometrics*, **48**: 19–30.

109. Potthoff, R.R. and Roy, S.N. (1984) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**: 313–26.

110. Prentice, R.L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**: 1033–1048.

111. Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danmarks Paedagogiske Institut.

112. Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**: 106–121.

113. Rosner, B., Muñoz, A. (1988) Autoregressive modelling for the analysis of longitudinal data with unequally spaced examinations. *Statistics in Medicine*, **7**: 59–71.

114. Rosner, B., Muñoz, A., Tager, I., Speizer, F. and Weiss, S. (1985) The use of an autoregressive model for the analysis of longitudinal data in epidemiologic studies. *Statistics in Medicine*, **4**: 457–467.

115. Rubin, D.B. (1976) Inference and missing data. *Biometrika*, **63**: 581–92.

116. Schall, R. (1991) Estimation in generalized linear models with random effects. *Biometrika*, **78**: 719–727.

117. Schmid, C.H., Segal, M.R. and Rosner, B. (1994) Incorporating measurement error in the estimation of autoregressive models for longitudinal data. *Journal of Statistical Planning and Inference*, **42**: 1–18.

118. Schnatter, S. (1992) Integration-based Kalman filtering for a dynamic generalized linear trend model. *Computational Statistics and Data Analysis*, **13**: 447–459.

119. Schnatter, S. (1994) Recursive residuals and model diagnostics for normal and non-normal state space models. *Research Memorandum* 40: Institut für Statistik, Wirtschaftsuniversität, Wien.

120. Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**: 461–464.

121. Sheiner, L.B., Beal, S.L., and Sambol, N.C. (1989) Study designs for dose-ranging. *Clinical Pharmacology and Therapeutics*, **46**, 63–77.

122. Shephard, N. (1994) Autoregressive time series models based on the exponential family. *Working paper.*

123. Singh, A.C. and Roberts, G.R. (1992) State space modelling of cross-classified time series of counts. *International Statistical Review*, **60**: 321–335.

124. Slud, E. and Kedem, B. (1994) Partial likelihood analysis of logistic regression and autoregression. *Statistical Sinica*, **4**: 89–106.

125. Smith, R.L. and Miller, J.E. (1986) A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society*, **B48**: 79–88.

126. Sprott, D.A. (1982) Robustness and maximum likelihood estimation. *Communications in Statistics*, **A11**: 2513–2529.

127. Stanek, E.J. and Diehl, S.R. (1988) Growth curve models of repeated binary response. *Biometrics*, **44**: 973–983.

128. Stanek, E.J., Shetterley, S.S., Allen, L.H., Pelto, G.H. and Chavez, A. (1989) A cautionary note on the use of autoregressive models in the analysis of longitudinal data. *Statistics in Medicine*, **8**: 1523–1528.

129. Stiratelli, R., Laird, N. and Ware, J.H. (1984) Random-effects models for serial observations with binary response. *Biometrics*, **40**: 961–971.

130. Stram, D.O., Wei, L.J. and Ware, J.H. (1988) Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates. *Journal of the American Statistical Association*, **83**: 631–637.

131. Stukel, T.A. (1993) Comparison of methods for the analysis of longitudinal interval count data. *Statistics in Medicine*, **12**: 1339–1351.

132. Thall, P.F. (1988) Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics*, **44**: 197–209.

133. Thall, P.F. and Lachin, M. (1988) Analysis of recurrent events: nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, **83**: 339–347.

134. Thall, P.F. and Vail, S.C. (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**: 657–671.

135. Ware, J.H., Lipsitz, S. and Speizer, F.E. (1988) Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, **7**: 95–107.

136. Wei, L.J. and Stram, D.O. (1988) Analysing repeated measurements with possibly missing observations by modelling marginal distributions. *Statistics in Medicine*, **7**: 139–148.

137. West, M. Harrison, P.J., & Migon, H.S. (1985) Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, **80**: 73–97.

138. Wilkinson, G.N. and Rogers, C.E. (1973) Symbolic description of factorial models for analysis of variance. *Applied Statistics*, **22**: 392–399.

139. Zeger, S.L. (1988) A regression model for time series of counts. *Biometrika*, **75**: 621–629.

140. Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**: 121–130.

141. Zeger, S.L., Liang, K.Y. and Albert, P.S. (1988) Models for longitudinal data: a generalized estimating approach. *Biometrics*, **44**: 1049–1060.

142. Zeger, S.L., Liang, K.Y. and Self, S.G. (1985) The analysis of binary longitudinal data with time-independent covariates. *Biometrika*, **72**: 31–38.

143. Zeger, S.L. and Qagish, B. (1988) Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, **44**: 1019–1031.

144. Zehnwirth, B. (1988) A generalization of the Kalman filter for models with state-dependent observation variance. *Journal of the American Statistical Association*, **83**: 164–167.

145. Zhao, L.P. and Prentice, R.L. (1990) Correlated binary regression using a quadratic exponential model. *Biometrika*, **77**: 642–648.