

PERCENTILE CURVES IN BINARY LONGITUDINAL DATA

by

Pedro A. Torres Saavedra

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

MATHEMATICS (STATISTICS)

UNIVERSITY OF PUERTO RICO

MAYAGÜEZ CAMPUS

June, 2006

Approved by:

Edgar Acuña, Ph.D.
Member, Graduate Committee

Date

Edgardo Lorenzo, Ph.D.
Member, Graduate Committee

Date

Raúl Macchiavelli, Ph.D.
President, Graduate Committee

Date

José Fernando Vega, Ph.D.
Representative of Graduate Studies

Date

Luis F. Cáceres, Ph.D.
Chairperson of the Department

Date

Abstract of Dissertation Presented in Partial Fulfillment of the
Requirements for the Degree of Master of Science

PERCENTILE CURVES IN BINARY LONGITUDINAL DATA

By

Pedro A. Torres Saavedra

June 2006

Chair: Raúl Macchiavelli

Major Department: Mathematics Department

This work presents different alternatives for modeling binary longitudinal data for continuous and categorical outcomes. Using a logistic model with random intercept, we propose the *Percentile Curves* concept, which are conditional probability curves across time representing percentiles of distribution of curves generated by the random intercept. We also analyze the density and cumulative distribution of subject-specific probabilities across time induced by the logistic model with random intercept. We apply this concept to two binary longitudinal data sets (*Toenail* and *Garlic*). Also, we expand the percentile curves concept to a logistic model with a random intercept and slope, and we propose a methodology to compute them.

Resumen de Tesis Presentado como requisito parcial de los
Requerimientos para el grado de Maestría en Ciencias

**CURVAS PERCENTILES EN DATOS LONGITUDINALES
BINARIOS**

Por

Pedro A. Torres Saavedra

Junio 2006

Consejero: Raúl Macchiavelli

Departamento: Departamento de Matemáticas

En este trabajo se presentan diferentes alternativas de modelamiento de datos longitudinales para respuestas continuas y categóricas. Usando un modelo logístico con intercepto aleatorio, proponemos el concepto de *Curvas Percentiles*, que son curvas de probabilidades condicionales a través del tiempo que representan percentiles de una distribución de curvas generadas por el intercepto aleatorio. Así mismo, analizamos la función de densidad y la distribución acumulada de las proporciones sujeto-específicas a través del tiempo inducidas por el modelo logístico con intercepto aleatorio. Aplicamos este concepto a dos conjuntos de datos longitudinales binarios, *Uñas* y *Ajo*. Además, ampliamos el concepto de curvas percentiles a un modelo logístico con intercepto y pendiente aleatoria y proponemos una metodología para calcularlas.

Copyright © 2006

by

Pedro A. Torres Saavedra

To my dear wife, Marggie: thanks for feed my dreams. I love you.

ACKNOWLEDGMENTS

I would like to thank the Department of Mathematics. In special to Dr. Luis F. Cáceres, acting chairperson: thanks for the support and the opportunity you gave me to participate in investigation projects. I would also like to thank Dr. Raúl Macchiavelli: thanks for the support you provided me in this investigation. I also thank the members of the graduate committee.

TABLE OF CONTENTS

	<u>page</u>
ABSTRACT ENGLISH	ii
ABSTRACT SPANISH	iii
ACKNOWLEDGMENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	xi
1 Introduction	1
2 Approaches to Longitudinal Data	3
2.1 Introduction	3
2.2 A First Approach using a General Linear Model	5
2.2.1 General Notation	6
2.2.2 A General Linear Model	7
2.3 Linear Mixed Models for Normal Data	11
2.4 Generalized Linear Models (GLM)	17
2.5 Model Families in General	18
3 Modeling Longitudinal Data	21
3.1 Marginal Models	21
3.1.1 Generalized Estimating Equations (GEE)	22
3.2 Subject-specific Models	27
3.2.1 Generalized Linear Mixed Models (GLMM)	29
3.2.2 Inference in GLMM	31
3.2.3 Approximations to the Likelihood Function in GLMM	32
3.3 Transition Models	41
3.4 Marginalized Models	43
3.5 Non-linear Mixed Effects Model	45
4 Models for Binary Longitudinal Data	47
4.1 Introduction	47
4.2 Marginal and Random-Effects Models	47
4.2.1 Marginal Models	48
4.2.2 Random-Effects Models	51
4.2.3 Comparison between Marginal and Random-Effects Models	54

4.3	Marginalized Latent Variables Models	59
4.4	Random-Effects Models with Bridge Function	60
4.5	Available software to Fit Random-Effects Models	64
4.5.1	Functions Available in R and SAS Software Packages	64
5	Percentiles Curves in a Logistic Model with Random Effects	68
5.1	Heterogeneity in a Logistic Model with Random Intercept	69
5.1.1	Density Function and Cumulative Distribution Function of $\tilde{\mathbf{p}}$	70
5.1.2	Mean and Variance of $\tilde{\mathbf{p}}$	71
5.2	An Interpretation of Percentile Curves	78
5.3	Application to Real Datasets	81
5.3.1	Toenail Dataset	81
5.3.2	Garlic Dataset	87
5.4	Heterogeneity in a Logistic Model with Random Intercept and Slope	97
5.4.1	Density Function of $\tilde{\mathbf{p}}$ and Percentiles Curves	102
6	Conclusions	108
	APPENDICES	112
A	Codes of R Functions used in this Thesis	113
B	Codes to Fit a Logistic Model with Random Intercept for Toenail Dataset	115

LIST OF TABLES

Table	page
2-1 Expressions for a , $\psi(\theta)$, and $c(\mathbf{y})$ for one-parameter distributions . . .	17
2-2 Common link functions	18
3-1 Common choices of <i>working</i> correlations in standard <i>GEE</i> and moment-based estimators	23
3-2 Common link functions in GLMM	30
3-3 Four points of Gauss-Hermite quadrature	34
4-1 Relation between marginal and subject-specific estimates in a logistic model with random intercept, <i>Toenail Data</i>	56
4-2 Marginal and Subject-specific estimates in a logistic model with random intercept with <i>Bridge</i> distribution, <i>Toenail Data</i>	64
4-3 Available Functions in R to fit GLMMs	66
4-4 Available Procedures in SAS to fit GLMMs	67
5-1 Proportion of Infected Patients, <i>Toenail Dataset</i>	82
5-2 Dimensions and Fit Statistics for model, <i>Toenail Dataset</i>	83
5-3 Estimates for the model, <i>Toenail Dataset</i>	84
5-4 Induced Marginal Mean, <i>Toenail Dataset</i>	84
5-5 Dimensions and Fit Statistics (<i>Model 1</i>), <i>Garlic Dataset</i>	89
5-6 Estimates for <i>Model 1</i> , <i>Garlic Dataset</i>	90
5-7 Induced Marginal Mean Incidences (<i>Model 1</i>), <i>Garlic Dataset</i>	95
5-8 Dimensions and Fit Statistics (<i>Model 2</i>), <i>Garlic Dataset</i>	96
5-9 Estimates for <i>Model 2</i> , <i>Garlic Dataset</i>	97
5-10 Induced Marginal Mean Incidences (<i>Model 2</i>), <i>Garlic Dataset</i>	100
5-11 Fit Statistics for <i>Model 3</i> and <i>Model 4</i> , <i>Garlic Dataset</i>	102
5-12 Estimates and Standard Errors for <i>Model 4</i> , <i>Garlic Dataset</i>	105

5–13 Induced Marginal Mean Incidences (*Model 4*), *Garlic Dataset* 107

LIST OF FIGURES

Figure	page
2-1 Relationship between reading ability and age	4
2-2 Subject-specific and population-average profiles	12
3-1 Abscissas for Classic and Adaptive Quadrature	35
4-1 Random Effects Models for Longitudinal Binary Data Analysis	53
4-2 Relation between marginal and subject-specific estimates	55
4-3 Marginal and Subject-specific curves for logistic model with a random intercept	57
4-4 Normal, Logistic and Bridge Density Functions for Logistic-Normal Model	63
5-1 Density of \tilde{p} across time for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 0.15$	72
5-2 Density of \tilde{p} across time for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 0.6$	72
5-3 Density of \tilde{p} across time for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 1.5$	73
5-4 Cumulative Distribution of \tilde{p} for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 0.15$	73
5-5 Cumulative Distribution of \tilde{p} for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 0.6$	74
5-6 Cumulative Distribution of \tilde{p} for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 1.5$	74
5-7 Induced Marginal Mean for the model defined in Equation (5.6)	76
5-8 Variance of \tilde{p} for the model defined in Equation (5.6)	77
5-9 Percentiles Curves for the Model presented in Equation 5.6 with $\sigma = 0.5$ and $\sigma = 2$	80
5-10 Scheme of Percentiles Curves	80

5–11	Proportion of Infected Patients for each treatment, <i>Toenail Dataset</i>	83
5–12	Density Function of \tilde{p} Induced by the Random Intercept, <i>Toenail Dataset</i>	84
5–13	Cumulative Distribution of \tilde{p} Induced by the Random Intercept, <i>Toenail Dataset</i>	85
5–14	Percentiles Curves, <i>Toenail Dataset</i>	85
5–15	Infection Progress Curves for the five environments in the white rot experiment, <i>Garlic Dataset</i>	88
5–16	Density Functions of \tilde{p} (<i>Model 1</i>), <i>Garlic Dataset</i>	91
5–17	Cumulative Distribution of \tilde{p} (<i>Model 1</i>), <i>Garlic Dataset</i>	92
5–18	Histogram of Incidence in each Environment and Specific Time, <i>Garlic Dataset</i>	93
5–19	Percentile Curves (<i>Model 1</i>), <i>Garlic Dataset</i>	94
5–20	Density Functions of \tilde{p} (<i>Model 2</i>), <i>Garlic Dataset</i>	98
5–21	Cumulative Distribution of \tilde{p} (<i>Model 2</i>), <i>Garlic Dataset</i>	99
5–22	Density Function of \tilde{p} (<i>Model 2</i>), Environment <i>JM2</i> , <i>Garlic Dataset</i>	100
5–23	Percentile Curves (<i>Model 2</i>), <i>Garlic Dataset</i>	101
5–24	Conditional Probability in Function of Random Intercept and Slope	103
5–25	Density Functions of \tilde{p} (<i>Model 4</i>), <i>Garlic Dataset</i>	106
5–26	Percentile Curves (<i>Model 4</i>), <i>Garlic Dataset</i>	107

CHAPTER 1

INTRODUCTION

Longitudinal data analysis is a very important branch of statistical modeling, and it has many applications. In this work we present theoretical aspects, and applications of longitudinal data analysis. Chapters 2 to 4 present the literature review about longitudinal data analysis and the most recent advances. In Chapter 2, we also talk about the longitudinal data and its advantages in comparison with cross-sectional studies. We present a general linear model that includes several models, and we group them in three general families according to recent literature. We introduce a linear mixed model as the simplest case in the longitudinal data modeling and also because it provides a comparison point with the non-linear models. The different models used to analyze longitudinal data are discussed in Chapter 3. In this chapter, we also discuss the marginal, subject-specific, transitions and marginalized models, and its respective general properties. We emphasize in generalized linear mixed models (*GLMM*) as a special type of the subject-specific models used to analyze categorical longitudinal data. In Chapter 4, we study the available models to analyze binary longitudinal data with special emphasis on the logistic model. In this chapter, we also compare the marginal and subject-specific model estimates, and discuss the interpretations of each of them. In Chapter 5, we propose the *Percentile Curves* in a logistic model with random effects as an interesting concept to interpret the subject-specific probabilities induced by the random effects.

In a logistic model with random intercept, we propose to interpret the subject-specific probabilities across time as percentile curves of the distribution curves induced by the random effect. This curve indicates that 50% of the subject-specific probabilities are under, or above, this value across time. We formalize this result in Proposition 1, which it is proved. All these results are theoretically developed using classical inference and statistical model theory. Similarly, we analyze the densities and cumulative distributions of the subject-specific probabilities induced by the random effects across time, and how the random effect variance affects its shape.

We expand the *Percentile Curves* concept to a logistic model with random intercept and slope. In this case, we propose a methodology to find the density function of the subject-specific probabilities and the percentiles curves using the Monte Carlo method.

These results are applied to two data sets (*Toenail* and *Garlic*). The first one has a binary response, and the second one has a binomial response. The respective interpretations are discussed in Chapter 5.

R and SAS statistical packages were used in this thesis.

CHAPTER 2

APPROACHES TO LONGITUDINAL DATA

2.1 Introduction

The characteristic definition of a longitudinal study is that the subjects are measured in repeated occasions through time. By contrast, the cross-sectional studies measure a single time for each individual.

There are many advantages, or merits, that longitudinal studies have over cross-sectional studies [35]. First, in a longitudinal study an investigator can separate *age effects* (changes over time within individuals) from *cohort effects* (differences between subjects at baseline). In a cross-sectional study this is not possible. This is the prime advantage of the longitudinal studies. To understand this aspect better, Figure 2.1 shows a classical example presented by Diggle *et al.* [35]. The plots show the ability of reading against age for several subjects. The first plot (cross-sectional) indicate that the reading ability appears poorer among older subjects. This plot show only the relation between age and reading ability without consider the repeated measures. In the second plot (longitudinal), we suppose that same data were obtained in a longitudinal study in which each subject was measured twice. It is clear that while younger subjects began at a higher reading level, everyone improved with time. However, the third plot (cross-sectional and longitudinal) shows that reading ability deteriorates with age. In this plot, a pair of points are connected to show an unusual result. This reflects the importance of the age effects in the longitudinal studies.

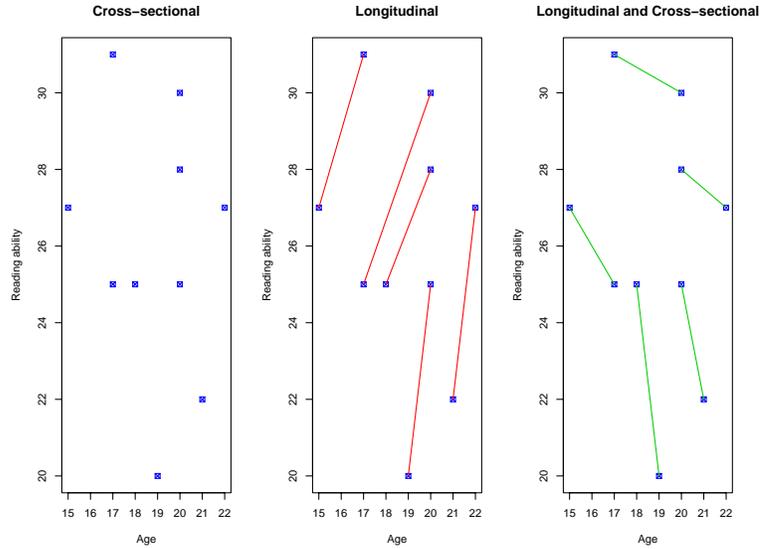


Figure 2-1: Relationship between reading ability and age

The second advantage is that, in a longitudinal study each subject can be its own control. For most outcomes, there exist a considerable variability across individuals due to the influence of unmeasured characteristics, such as environment exposure, personal habits, and others. This tend to persists over time. A third advantage is that, for a fixed number of subjects, the longitudinal studies are more powerful than the cross-sectional. This happens because the repeated observations from the same subject are rarely perfectly correlated. That is, repeated measurements from a single subject provide more information than a single measurement obtained from a single subject. Fourth, longitudinal data can provide information about individual change. Statistical estimates of individual trends can be used to better understand heterogeneity in the population, the determinants of growth, and change at the individual level.

On the other hand, for the analysis of longitudinal data it is necessary to consider that repeated observations on one subject are dependent, therefore, longitudinal data analysis requires special statistical methods. This correlation must be considered to draw valid scientific inferences, adding complexity to the models.

There are many aspects of longitudinal studies that must be considered in order to select an appropriate analysis [22]:

1. It is important to know whether the outcome is continuous or discrete. The selected model will depend on the type of outcome and its distribution. For continuous outcomes, there are several models, estimation techniques, and statistical software. However, for discrete outcomes, the models are more complex, and researchers still developing models and estimation techniques. Also, the statistical programs to fit models for discrete outcomes are limited, and the programs for parameter estimation can be computationally intensive due to the need of numerical methods.
2. The sample size (N), and the number of observations per subject (n_i) are also important aspects in a longitudinal study. In complex models such as generalized linear mixed models, small N and n_i may produce inappropriate estimates and large biases.
3. The number and type of covariates are important aspects in order to select and interpret a model for response mean. These can be baseline, or time-varying covariates. In any case, the model must consider these aspects.
4. In terms of the selection of a plausible covariance structure for $\text{Var}(Y)$. Different model specifications lead to homogeneous, or heterogeneous covariances of repeated measures over time.

2.2 A First Approach using a General Linear Model

In order to present a set of models to analyze longitudinal data, it is necessary to understand what the likely sources of random variation that occur in practice are, and which ones may be important to be included in the models [35]. These sources will permit to develop and study each model component. It is possible to distinguish between the following sources of random variation in a longitudinal study:

- *Random effects*: When the subjects are sampled at random from a population, various aspects of their behavior may show stochastic variation between subjects.

- *Serial correlation*: The observed measurement of the a subject may be a response to time-varying stochastic process operating within this subject. This type of stochastic variation results in a correlation between pairs of measurements on the same subject which depends on the time separation. Typically, the correlation becomes weaker as the time separation increases.
- *Measurement error*: Especially when the individual measurement involves some kind of sampling within units, the measurement may itself add a component of variation to the data.

This qualitative features must be introduced into the models by using appropriate terms, which will be discussed in the next section.

Following Diggle’s approach [35], we will begin talking about a general model for longitudinal data in Section 2.2.2. It will be possible to see that the models presented in most of the literature are special cases of this general model.

2.2.1 General Notation

Consider the following general notation. Suppose a longitudinal study with N subjects, each one measured in n_i , $i = 1, \dots, N$, occasions across time. Let y_{ij} be the outcome for the i – th subject in the j – th , $j = 1, \dots, n_i$, measurement.

- $Y_i = (Y_{i1}, \dots, Y_{in_i})$: $n_i \times 1$ vector of sequence of measurements on the i – th subject.
- $Y = (Y_1, \dots, Y_N)$: complete set of $N_T = \sum_{i=1}^N n_i$ measurements.
- X_i : $n_i \times p$ design matrix of p covariates for the i – th subject.
- X : $N_T \times p$ design matrix that contains the information of covariates for all subjects.
- β : p –dimensional parameter vector of fixed-effects associated with the covariates.
- $t_i = (t_{i1}, \dots, t_{in_i})$ the corresponding set of times at which the measurements were taken.
- $t = (t_1, \dots, t_N)$: $N_T \times 1$ vector with the time-points for all subjects.
- Z_i : $n_i \times q$ design matrix of q random effects for the i – th subject.
- \mathbf{u}_i : q –dimensional parameter vector of random-effects for the i – th subject.

- ϵ_i : $n_i \times 1$ error vector for the i -th subject.

2.2.2 A General Linear Model

In order to present the models used to analyze longitudinal data, first we present a general model that includes the different components associated to the random variation sources mentioned in Section 2.2.

Let us consider the following *general model*:

$$Y = X\beta + \mathbf{e} \quad (2.1)$$

where \mathbf{e} is a $N_T \times 1$ error component normally distributed with zero mean, and covariance matrix, $\Sigma = \Sigma(t, \alpha)$ that depends on a r -dimension parameter vector, α , and the time-points vector, t . In the previous formulation, we make an explicit separation between mean and variance structures. The mean of Y is given by $E(Y) = X\beta$, while the variance of \mathbf{e} , $\text{Var}(\mathbf{e})$, takes different forms depending on the assumptions on \mathbf{e} . Our interest is to study the variance components, α . Therefore, the additive decomposition of the ij -th element of \mathbf{e} , e_{ij} , in terms of *random effects*, *serial correlated variation* and *measurement error* can be expressed formally as [35]:

$$e_{ij} = z'_{ij}\mathbf{u}_i + w_i(t_{ij}) + \epsilon_{ij} \quad (2.2)$$

where ϵ_{ij} are a set of N_T mutually independent Gaussian random variables, each one with zero mean, and variance τ^2 , \mathbf{u}_i is a set of n_i mutually independent q -element Gaussian random vectors, each one with mean zero vector and covariance matrix D , and $w_i(t_{ij})$ are sampled from N independent copies of a stationary Gaussian process with zero mean, variance σ^2 , and correlation function $\rho(h)$, where h is the lag time. Note that u_i , $w_i(t_{ij})$, and ϵ_{ij} correspond to *random effects*, *serial correlation*, and *measurement error*, respectively. The variance of the error term for the i -th subject, e_i , is given by

$$\text{Var}(e_i) = Z_i D Z'_i + \sigma^2 H_i + \tau^2 I_i \quad (2.3)$$

where Z_i is the $n_i \times q$ design matrix with j -th row equal to z'_{ij} , H_i a $n_i \times n_i$ matrix with the jk -th element given by $h_{ijk} = \rho(|t_{ij} - t_{ik}|)$, and I_i the $n_i \times n_i$ identity matrix. When H_i is the identity matrix, that is, the measurements taken over the same subject are independent, we have the *classical linear regression model*. Diggle calls this formulation as a *Parametric Model for Covariance Structure* [35].

Let us consider some special cases of the variance structure of \mathbf{e} presented in Equation 2.3 to derive some well-known models used to analyze longitudinal data:

- *Covariance Patterns Models (CPM)*: If the decomposition of the error term does not consider neither random effects nor measurement error, then

$$e_{ij} = w_i(t_{ij})$$

and, therefore, the resulting model implies that

$$\text{Var}(Y_i) = \sigma^2 H_i$$

In this case we are considering that the measurements taken over the same subject have a correlation described by the function $\rho(h)$. Two classical alternatives for H_i are [35]:

- *Uniform correlation*: This structure of H_i considers that there is a positive correlation, ρ , between any two measurements on the same subject. In matrix terms,

$$H_i = (1 - \rho)I_i + \rho J_i \tag{2.4}$$

where I_i denotes the $n_i \times n_i$ identity matrix, and J_i the $n_i \times n_i$ matrix all of whose elements are 1. A *uniform correlation* could be the result of considering

$$Y_{ij} = \mu_{ij} + \mathbf{u}_i + \gamma_{ij}$$

where $\mu_{ij} = E(Y_{ij})$, \mathbf{u}_i are mutually independent $N(0, \lambda^2)$ random variables, the γ_{ij} are mutually independent $N(0, \sigma^2)$ random variables, and \mathbf{u}_i and γ_{ij} are independent.

- *Exponential correlation*: In contrast to the uniform correlation, the correlation between a pair of measurements on the same unit decays towards zero when the distance between units increases. In this case, H_i has the jk -th element, ν_{jk} , of the form:

$$\nu_{jk} = \exp(-\phi |t_j - t_k|) \quad (2.5)$$

with ϕ a unknown constant. If the time-points are equally-spaced, $t_{j+1} - t_j = d$, for all j , then ν_{jk} could be defined as

$$\nu_{jk} = \rho^{|j-k|} \quad (2.6)$$

where $\rho = \exp(-\phi d)$.

Similarly to the uniform correlation, an *exponential correlation* could be the result of considering

$$Y_{ij} = \mu_{ij} + W_{ij} \quad (2.7)$$

where

$$W_{ij} = \rho W_{ij-1} + \gamma_{ij} \quad (2.8)$$

and, the γ_{ij} are mutually independent $N[0, \sigma^2(1 - \rho^2)]$ random variables, to give $\text{Var}(Y_{ij}) = \text{Var}(W_{ij}) = \sigma^2$ as required. In view of Equations 2.7 and 2.8, the exponential correlation model is sometimes called the *First-order Autoregressive Model AR(1)*. This concept can be generalized to *p*-th Order Autoregressive Models.

- *Mixed Models*: If we eliminate the serially correlated component altogether, Equation 2.2 reduces to

$$e_{ij} = z'_{ij} \mathbf{u}_i + \epsilon_{ij} \quad (2.9)$$

and therefore,

$$\text{Var}(e_i) = Z_i D Z_i' + \tau^2 I_i$$

Equation 2.9 corresponds to a mixed model with the *conditional independence* assumption, that is, given \mathbf{u}_i the repeated measurements of a subject are independent. The *mixed* word is because the model includes *fixed effects* which are represented by β , and *random effects* included in \mathbf{u}_i . The simplest model of this kind is obtained considering a scalar random intercept, \mathbf{u}_{i0} , with covariance matrix equal to $\lambda^2 I$. Therefore,

$$\text{Var}(e_i) = \lambda^2 J_i + \tau^2 I_i$$

The previous result implies that the correlation between any two measurements on the same subject is given by:

$$\rho = \frac{\lambda^2}{\lambda^2 + \tau^2}$$

This is the well-known *intra-class correlation coefficient* cited usually in cluster analysis. Note that this result is like the one obtained with the uniform correlation structure presented in the covariance patterns models. This result shows that the uniform correlation between observations of the same subject can be induced using a model with random intercept.

- *Mixed Model with Covariance Structures*: Finally, if the model has no measurement error term, that is, it considers only random effects and serial correlation, then its formulation is given by:

$$e_{ij} = z'_{ij} \mathbf{u}_i + w_i(t_{ij})$$

and,

$$\text{Var}(e_i) = Z_i D Z_i' + \sigma^2 H_i$$

This model is more complex because it includes random effects, and serial correlation. Generally, it requires special estimation techniques, and specialized software tools to be fitted [34].

2.3 Linear Mixed Models for Normal Data

In the previous section, we talked about random variation sources in longitudinal studies, and how these are represented in a general model. In this section, we present a linear mixed model for response variables with normal distribution. First, let us consider the following linear model, similar to Equation 2.1:

$$Y_i = X_i\beta + \varepsilon_i \quad (2.10)$$

where ε_i is a $n_i \times 1$ error vector normally distributed with zero mean. The model is completed by specifying an appropriate covariance matrix Σ_i for ε_i , leading to the multivariate model

$$Y_i \sim N(X_i\beta, \Sigma_i) \quad (2.11)$$

When $\Sigma = \sigma^2 I_i$, this model corresponds to the linear regression model, which assumes all repeated measurements are independent. That is, it ignores the fact that repeated measurement within subjects may be correlated. Depending on the context, it is possible to select other structures for Σ_i , such as uniform correlation, or first-order autoregressive, which were showed in Section 2.2.2.

The random effects approach extends the univariate linear regression model, but it assumes that the response can be modeled by a linear regression model with *subject-specific* regression coefficients. In many situations, one could assume that each subject responds differently across time, and consequently, each one must have an intercept and a slope. Therefore, one can assume that the outcome y_{ij} , measured at time t_{ij} satisfies

$$Y_{ij} = \tilde{b}_{i0} + \tilde{b}_{i1}t_{ij} + \varepsilon_{ij} \quad (2.12)$$

where $\tilde{b}_{i0} = \tilde{\beta}_{i0} + u_{i0}$ and $\tilde{b}_{i1} = \tilde{\beta}_{i1} + u_{i1}$. Because subjects are randomly selected from a population, it is reasonable to consider that the subject-specific regression coefficients are randomly sampled from a population of coefficients. Assuming $\tilde{b}_i = (\tilde{b}_{i0}, \tilde{b}_{i1})$ to be bivariate normal with mean $(\tilde{\beta}_{i0}, \tilde{\beta}_{i1})'$ and 2×2 covariance matrix D ,

we can rewrite the model in Equation 2.12 as:

$$Y_{ij} = (\tilde{\beta}_{0i} + u_{i0}) + (\tilde{\beta}_{i1} + u_{i1})t_{ij} + \varepsilon_{ij} \quad (2.13)$$

where the new random effects, u_{i0} and u_{i1} , are normally distributed with zero mean and covariance matrix D . Under this model, it is reasonable to think in terms of the *population-average* profile which represents the linear regression model for the average subject. The previous statement is valid only in the linear mixed model context. In other types of models, such as generalized linear mixed models, it is not true. Taken expectations to both sides in Equation 2.13, we find that the *population-averaged* regression model is given by:

$$E(Y_{ij}) = \tilde{\beta}_{i0} + \tilde{\beta}_{i1}t_{ij} + \varepsilon_{ij} \quad (2.14)$$

Figure 2–2 shows the subject-specific and population-average profiles for a hypothetical example of longitudinal data that can be well described using a linear mixed model with random intercepts and slopes.

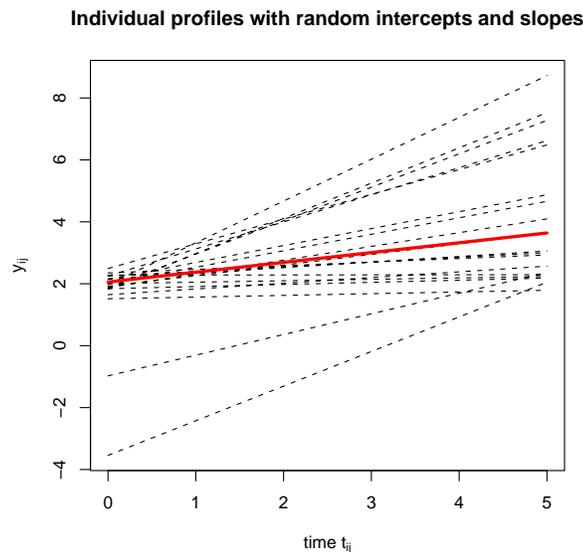


Figure 2–2: Subject-specific and population-average profiles. The continue line represents population-average evolution and the dash lines indicates the subject-specific evolution

The model presented in Equation 2.12 is a special case of the general *linear mixed model* which assumes that the vector Y_i of repeated measurements for the i -th subject satisfies

$$\begin{aligned} Y_i | \mathbf{u}_i &\sim N(X_i\tilde{\beta} + Z_i\mathbf{u}_i, \Sigma_i) \\ \mathbf{u}_i &\sim N(0, D) \end{aligned} \tag{2.15}$$

All implicit terms in the model were defined in Section 2.2.1. The components in $\tilde{\beta}$ are called *fixed effects*, and the components in \mathbf{u}_i are called *random effects*. The fact that the model contains fixed as well as random effects generate the term *mixed model*. The formulation of this model is also called *hierarchical* formulation of the linear mixed model.

Inference in Linear Mixed Models

Let $f_i(\mathbf{y}_i | \mathbf{u}_i)$ and $f(\mathbf{u}_i)$ be the density functions corresponding to the random variables defined in Equation 2.15, respectively. Then, the marginal (induced) density function of Y_i is

$$f_i(\mathbf{y}_i) = \int_{\mathfrak{R}} f_i(\mathbf{y}_i | \mathbf{u}_i) f(\mathbf{u}_i) d\mathbf{u}_i \tag{2.16}$$

which has a n_i -dimensional normal density function with mean and variance given by:

$$E(Y_i) = X_i\beta \tag{2.17}$$

and,

$$\text{Var}(Y_i) = V_i = Z_i D Z_i' + \Sigma_i \tag{2.18}$$

This result shows that the linear mixed model presented in Equation 2.15 implies a *marginal model* as presented in Equation 2.10, but with a specific form for the marginal covariance matrix V_i , which depends on an unknown vector α of parameters in the covariance matrices usually called *variance components*. The corresponding

marginal normal distribution with mean $X_i\beta$ and covariance V_i is called the *marginal formulation* of the model. A marginal model follows from a hierarchical one, however, different random-effects models can produce the same marginal model [34]. The model presented in Equation 2.15 is called by some authors *conditional model* because of the conditional formulation with respect to the random effects, or is also called *subject-specific model* because it models the response for each subject based on the respective random effect value. Precisely, the mean of the model presented in Equation 2.15, $E(Y_i | \mathbf{u}_i)$, is called the *conditional mean*, which is denoted by $\tilde{\mu}_i$, and $E(Y_i)$ is called the *marginal mean*, which is denoted by μ_i .

An important property of linear mixed models is that the marginal mean of Y_i coincide with the conditional mean given $\mathbf{u}_i = 0$, that is

$$E(Y_i) = E(Y_i | \mathbf{u}_i = 0) = X_i\beta$$

The previous equation indicates that the *population-average profile* obtained from integrating the random effects in the Equation 2.16 is equal to the *typical profile*, or *conditional mean* given a zero random effect, $\mathbf{u}_i = 0$. In the linear mixed models with normal distribution, it is due to the interesting properties of multivariate normal distribution, and the linear mean structure. Consequently, this property is not necessary true in other models, such as generalized linear mixed models, which will be presented later.

The classical inference for linear mixed models is based on maximum likelihood (*ML*). Assuming independence across subjects, the likelihood takes the form

$$L(\theta) = \prod_{i=1}^N \left\{ (2\pi)^{-n_i/2} |V_i(\alpha)|^{-1/2} \exp \left[-\frac{1}{2} (Y_i - X_i\beta)' V_i^{-1}(\alpha) (Y_i - X_i\beta) \right] \right\} \quad (2.19)$$

Estimation of $\theta' = (\beta', \alpha)'$ requires joint maximization of 2.19 with respect to all elements of θ . Conditionally on α , the maximum likelihood estimator (*MLE*) for β

is given by:

$$\beta(\hat{\alpha}) = \left(\sum_{i=1}^N X_i' W_i X_i \right)^{-1} \sum_{i=1}^N X_i' W_i Y_i \quad (2.20)$$

where $W_i = V_i^{-1}$, and α can be replaced by the Restricted Maximum Likelihood (*REML*) estimator, which is obtained without first having to estimate β . $\hat{\beta}(\alpha)$ has a normal distribution with mean and variance given by:

$$E[\hat{\beta}(\alpha)] = \beta \quad (2.21)$$

and,

$$\text{Var}[\hat{\beta}(\alpha)] = \left(\sum_{i=1}^N X_i' W_i X_i \right)^{-1} \quad (2.22)$$

Note that standard errors based on Equation 2.22 are valid, only if the mean and covariance were correctly specified; however, in practice this can hardly be reached.

Thus, it is recommended to calculate the standard errors based on the formula:

$$\text{Var}[\hat{\beta}(\alpha)] = \left(\sum_{i=1}^N X_i' W_i X_i \right)^{-1} \left(\sum_{i=1}^N X_i' W_i \text{Var}(Y_i) W_i X_i \right) \left(\sum_{i=1}^N X_i' W_i X_i \right) \quad (2.23)$$

where $\text{Var}(Y_i)$ is estimated using

$$(\mathbf{y}_i - X_i \hat{\beta})(\mathbf{y}_i - X_i \hat{\beta})' \quad (2.24)$$

rather than \hat{V}_i . Standard errors of $\hat{\beta}$ based on Equation 2.24 are the so-called *robust* or *empirical* standard errors, which will be discussed with more detail in Section 3.1.1.

Random Effects Predictions

Although in practice the primary interest is to estimate the parameters in the marginal distribution for Y_i , β and D , it is often useful to calculate the estimates for the random effects \mathbf{u}_i as well. They reflect the between-subject variability, which could have interesting interpretations in terms of clustering of individuals, or special profiles. The assumptions about the random and fixed effects are different. A

fixed effect is considered to be a constant to be estimated, while a random effect is considered an effect sampled from a population, and therefore, we prefer to use the term *prediction* to refer to random effects instead *estimate*. Nevertheless, some authors do not distinguish between the two terms. They use *estimate* for both fixed and random effects inference [2, 34].

In general, the prediction of the random effects, \mathbf{u}_i , is based on the conditional *posterior* distribution of \mathbf{u}_i given all the information of the marginal model, that is [34]:

$$f_{\mathbf{u}_i}(u_i | \mathbf{y}_i, \beta, D) = \frac{f_i(\mathbf{y}_i | \mathbf{u}_i, \beta) f(\mathbf{u}_i | D)}{\int_{\mathcal{R}} f_i(\mathbf{y}_i | \mathbf{u}_i, \beta) f(\mathbf{u}_i | D) d\mathbf{u}_i} \quad (2.25)$$

The resulting predictions from this method are called *Bayes Estimates* because they use the *prior* information on the marginal model. For a linear mixed model, Equation 2.25 takes the form of a normal distribution with mean given by:

$$\hat{\mathbf{u}}_i(\theta) = DZ_i'V_i^{-1}(\alpha)(\mathbf{y}_i - X_i\beta) \quad (2.26)$$

which is used in practice as predictor of \mathbf{u}_i . Its covariance is equal to

$$\text{Var}[\hat{\mathbf{u}}_i(\theta)] = \left\{ V_i^{-1} - V_i X_i \left(\sum_{i=1}^N X_i' V_i^{-1} X_i \right)^{-1} X_i' V_i^{-1} \right\} Z_i D \quad (2.27)$$

If the unknown parameters α and β in θ are replaced by their respective estimators, then $\hat{\mathbf{u}}_i$ is called the *Empirical Bayes Estimate* for the random effects in the linear mixed model. Using a classical result of inference, it is important to note that

$$\text{Var}(\mathbf{u}_i) = \text{Var}[E(\mathbf{u}_i | \mathbf{y})] + E[\text{Var}(\mathbf{u}_i | \mathbf{y})] \quad (2.28)$$

where $E(\mathbf{u}_i | \mathbf{y}) = \hat{\mathbf{u}}_i$, and the second term in Equation 2.28 is positive. Therefore, the Equation 2.27 underestimates the variability of the random effects, \mathbf{u}_i . This phenomenon is usually referred to as *shrinkage*, and $\hat{\mathbf{u}}_i$ is called a *shrinkage estimator* [33].

2.4 Generalized Linear Models (GLM)

Before talking about other models to analyze longitudinal data, it is important to define the *generalized linear models* (*GLM*), which are used to analyze categorical data, such as binary and count data [2].

Suppose that Y is a response variable that has a density distribution that belongs to the exponential family:

$$f(\mathbf{y}; \theta, \phi) = \exp \left\{ \frac{[\mathbf{y}\theta - \psi(\theta)]}{a\phi} + c(\mathbf{y}, \phi) \right\} \quad (2.29)$$

where θ is a location parameter (not necessarily the mean), ϕ is a dispersion parameter, $\psi(\cdot)$, and $c(\cdot)$ are known functions. If the density distribution has one-parameter θ , Equation 2.29 can be simplified to:

$$f(\mathbf{y}; \theta, \phi) = \exp \left\{ \frac{[\mathbf{y}\theta - \psi(\theta)]}{a} + c(\mathbf{y}) \right\} \quad (2.30)$$

The expressions for a , $\psi(\theta)$, and $c(\mathbf{y})$ for some well-known distributions with sample size equal to n are shown in Table 2–1. The mean and variance of the distribution, $f(\mathbf{y}; \theta, \phi)$, can be written in terms of the functions a , and ψ [6, 33].

Table 2–1: Expressions for a , $\psi(\theta)$, and $c(\mathbf{y})$ for one-parameter distributions

Distribution	a	$\psi(\theta)$	$c(\mathbf{y})$
Bernoulli	1	$\log(1 + \exp(\theta))$	1
Binomial	$\frac{1}{n}$	$\log(1 + \exp(\theta))$	$\log \left[\frac{n!}{(ny)!(n-by)!} \right]$
Poisson	1	$\exp(\theta)$	$-\log(y!)$
Poisson with offset t	$\frac{1}{t}$	$y \log(yt) - \log(yt!)$	1

In general, the systematic component of a *GLM* is defined using the following equation:

$$g(\mu_i) = g\{E(Y_i)\} = X_i\beta \quad (2.31)$$

where Y_i is the response variable for the i – *th* subject, $g(\cdot)$ is a known monotone differentiable function, called *link function*, and β is a parameter vector related to

the covariates. Table 2–2 shows the link functions commonly used for different types of outcomes with their respective equation models.

Table 2–2: Common link functions

Outcome	Link Function	Equation Model
Normal	Identity	$\mu_i = X_i\beta$
Binary	Logit	$\log\left(\frac{\mu_i}{1-\mu_i}\right) = X_i\beta$
Binary	Probit	$\Phi^{-1}(\mu_i) = X_i\beta$
Binary	Complement log – log	$\ln(-\ln(1 - \mu_i)) = X_i\beta$
Count	Log	$\log(\mu_i) = X_i\beta$

The model presented in Equation 2.31 can be rewritten as

$$\mu_i = E(Y_i) = g^{-1}(X_i\beta)$$

The previous models are used if the observations are not correlated; however, in the context of longitudinal studies this assumption is no valid, and it is necessary to extend these models in order to consider correlated observations.

2.5 Model Families in General

In this section, we consider a classification of the possible models to analyze longitudinal data in three general families. First, let us consider the following results for models with a linear mean structure. A *marginal model* is characterized by a marginal mean function given by

$$E(Y_{ij} | x_{ij}) = x'_{ij}\beta$$

where x'_{ij} is a vector of covariates for the i – th subject at the j – th occasion. Although the mean is conditionally expressed in terms of the covariates, this formulation is considered marginal because it does not depend on other variables or previous results. In *random-effects models*, we focus on the conditioning expectation on a random effect

$$E(Y_{ij} | \mathbf{u}_i, x_{ij}) = x'_{ij}\beta + z'_{ij}\mathbf{u}_i$$

Clearly, the mean response, depends on the covariates and the random effects. Finally, there is a third type of models which conditions a particular outcome on the previous response or a subset thereof. A specific case, a simple first-order stationary *transition model* takes the form

$$E(Y_{ij} | Y_{ij-1}, \dots, Y_{i1}, x'_{ij}) = x'_{ij}\beta + \alpha Y_{ij-1}$$

The previous type of models have a clear relation in Gaussian responses. However, there is not a close connection between them when the outcomes are discrete, such as binary, or categorical. In order to study other alternatives to model longitudinal data, different to the normal responses, we consider three general model families which are described briefly [34, 35]:

- ***Marginal models***: Responses are modeled marginalized over all other responses or results, and the association structure is then typically captured using a set of association parameters, such as correlations, odds ratios, etc. These models are also called *population-averaged models* because the parameters characterize the marginal expectation.

The marginal models include an extensive range of possibilities, especially, those that use methods based on likelihood such as the *Bahadur model*, or log-linear models [34]. We will focus on marginal models in which the dependence between observations within each individual is modeled using correlation structure matrices, and generalized estimation equations (*GEE*) to obtain estimates. These themes will be presented in Section 3.1.

- ***Subject-specific models***: Subject-specific models are differentiated from marginal, or population-averaged models, by the inclusion of specific parameters for each subject. Unlike the normal responses, the parameters of subject-specific and population-averaged models describe different types of effects of the covariates. The

subject-specific term is usually equated to *random-effects*; however, the subject-specific parameters can be treated as (1) fixed-effects, (2) as random-effects, and (3) eliminated by conditioning. The first case is very simple but presents many problems when the number of parameters is large. We will emphasize the second approach, the *generalized linear mixed models*. These models will be discussed in more detail in Section 3.2.1. Finally, the third approach is well-known in epidemiology to analyze case-control studies. In particular, the *conditional logistic regression* is often used in these studies. This method conditions on a sufficient statistics for each subject.

- ***Transition models***: These are models in which any response within the sequence of repeated measures is modeled conditional upon the other outcomes. Diggle *et al.* criticize the transition approach because the interpretation of a fixed effect parameter of one response is conditional on other responses for the same subject, outcomes of others subjects, and the number of repeated measures. We will discuss briefly these models in Section 3.3.

CHAPTER 3

MODELING LONGITUDINAL DATA

In the previous chapter, we introduced the three model families used to analyze longitudinal data. They can be used for both continuous and discrete outcomes. Next, we will study some theoretical aspects, and characteristics of the models that belong to each family.

3.1 Marginal Models

When inferences based on mean parameters, or population-average are of primary interest it is unnecessary to specify the full joint distribution, and many times the complete likelihood can be complicated to evaluate, except with small n_i . In these situations, a *marginal model* is an adequate form to analyze the data originated by a longitudinal study. A *marginal model* specifies the mean of the response variable, or marginal expectation, and the correlation of the repeated measures separately.

First, consider Y_{ij} , the outcome for the i -th subject in the j -th measurement. Define a general form to relate the response and its mean:

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij}$$

Then, a marginal model establishes the following assumptions:

1. The marginal expectation of the response variable and the covariates are related by means of:

$$g(\mu_{ij}) = g[E(Y_{ij})] = x'_{ij}\beta \tag{3.1}$$

where $g(\cdot)$ is the *link function*, and x'_{ij} is the covariates vector.

2. The variance of Y_{ij} is a function of the μ_{ij} , and a parameter ϕ called *dispersion parameter*. Thus, $\text{Var}(Y_{ij}) = \phi v(\mu_{ij})$, where $v(\cdot)$ is a known *variance function*. For example, for Poisson response variable, the variance function is the identity, and ϕ is a dispersion parameter, thus $\text{Var}(Y_{ij}) = \phi \mu_{ij}$.
3. Finally, the correlation between observations on the same subject, $\text{Corr}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha)$, depends on the marginal means, and on a parameter vector α . This correlation is modeled using a correlation structure matrix called *working correlation matrix*. It is assumed that the correlation matrix $\mathbf{R}(\alpha)$ depends on a vector of association parameters α , and is equal for all subjects.

If $g(\cdot)$ is the identity function, then the model in Equation 3.1 reduces to the classical regression model: $E(Y_{ij}) = x'_{ij}\beta$. The specification of a covariance structure for the errors establishes a marginal model to continuous responses, such as we discussed in Chapter 2. Therefore, we can see that the marginal models can be used for both continuous and discrete responses.

A method used to calculate the estimates in the marginal models are *generalized estimating equations (GEE)*. This method uses *quasi-likelihood*, which requires the mean response to be expressed as a parametric function of covariates (*first moment*), and the variance is assumed to be a function of the mean up to possibly unknown scale parameters (*second moment*). The estimates obtained with this model are *population average estimates* because they represent the average effect of the covariates over the marginal mean of the response for the subjects in the population who share the same values in the covariates.

3.1.1 Generalized Estimating Equations (GEE)

The generalized estimating equations method was developed in the 1980's to extend the generalized linear models to correlated data [22, 29].

Based on the marginal model, the working covariance matrix of Y_i is specified as

$$V_i = \text{Var}(Y_i) = \phi A_i^{1/2} \mathbf{R}_i(\alpha) A_i^{1/2} \quad (3.2)$$

where ϕ is a dispersion parameter, A_i is a $n_i \times n_i$ diagonal matrix with $a_{i,jj} = v(\mu_{ij})$ as the j -th diagonal element, and $\mathbf{R}_i(\alpha)$ the $n_i \times n_i$ working correlation for i -th subject. For example, in Poisson responses, $a_{i,jj} = \mu_{ij}$, and for binomial responses, $a_{i,jj} = \mu_{ij}(1 - \mu_{ij})$. $\mathbf{R}_i(\alpha)$ must be selected according the observed correlation. Nevertheless, the *GEE* method yields consistent estimates of the regression coefficients, and their standard errors, even with misspecification of the correlational structure. This is a relevant property of the *GEE* estimates.

Table 3–1: Common choices of *working* correlations in standard *GEE* and moment-based estimators

Structure	$\text{corr}(y_{ij}, y_{ik})$	Estimator
Independence	0	–
Exchangeable (CS)	α	$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N \frac{1}{n_i(n_i-1)} \sum_{j \neq k} e_{ij} e_{ik}$
AR(1)	$\alpha^{ j-k }$	$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N \frac{1}{(n_i-1)} \sum_{j \leq n_i-1} e_{ij} e_{i,j+1}$
Unstructured	α_{jk}	$\hat{\alpha}_{jk} = \frac{1}{N} \sum_{i=1}^N e_{ij} e_{ik}$

We present the parameters for some common choices for the *working correlation matrix* in Table 3–1. It can be used either for continuous or discrete data. However, in practice, it is possible that some correlation structures work better for some type of data [34, 35]. The term $e_{ij} = \frac{y_{ij} - \mu_{ij}}{v\sqrt{\mu_{ij}}}$ represents the standardized residuals.

For example, let us suppose a study with four measurements ($n_i = 4$) taken through time. Under each working correlation structure, the respective matrices $R_i(\alpha)$ for each subject i are given by:

Independence:

$$R_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The independence structure assumes that repeated measurements for a subject are not correlated. In this case, *GEE* simplifies to the *GLM* estimating equations.

Exchangeable or Compound Symmetry:

$$R_i = \begin{pmatrix} 1 & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & 1 \end{pmatrix}$$

Such as shown in the previous chapter, exchangeable is the correlation structure induced by the model with random intercept, but only for the identity link function. This structure may not be justified in a longitudinal study, but it is often reasonable in situations in which repeated measures are not obtained over time. This structure is appropriate when cluster sampling is involved.

Auto-regressive $AR(1)$:

$$R_i = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{pmatrix}$$

Unstructured:

$$R_i = \begin{pmatrix} 1 & \alpha_{21} & \alpha_{31} & \alpha_{41} \\ \alpha_{21} & 1 & \alpha_{32} & \alpha_{42} \\ \alpha_{31} & \alpha_{32} & 1 & \alpha_{43} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & 1 \end{pmatrix}$$

The unstructured correlation structure implies that the correlation matrix has $\frac{n_i(n_i-1)}{2}$ parameters to estimate. The estimation can present problems when there are missing data [31].

The *GEE* estimator of β for the model described in Equation 3.1 is the solution of:

$$\sum_{i=1}^N D_i' [V_i(\hat{\alpha})]^{-1} (y_i - \mu_i) = 0 \quad (3.3)$$

where $\hat{\alpha}$ is a consistent estimate of α , and $D_i' = \left(\partial \mu_i / \partial \beta \right)'$. Equation 3.3 is called the *Score Equation*. These are similar to the score equations in the *GLM*. However, in this context, the variance of Y_i , depends on the parameters of the working structure correlation. The estimates are called *quasi-likelihood estimates* because Equation 3.3 depends only of the mean and variance. Notice that for the normal case, Equation 3.3 reduced to the normal estimation equation for the linear regression using weighted least-squares (*WLS*) with weight matrix $[R_i(\alpha)]^{-1}$. When $R_i(\alpha) = I$, the solution to this equation is reduced to the normal estimation equation for the classic linear regression problem.

Solving the *GEE* involves iterating between the quasi-likelihood solution to estimate β , and a robust method to estimate α as a function of β . Specifically, it involves the following steps:

1. Given estimates of $R_i(\alpha)$, and ϕ , calculate estimates of β using iteratively re-weighted least squares where iterative estimates of α are used to yield new estimates of β .
2. Given estimates of β , obtain estimates of α , and ϕ . To do this, it is necessary to calculate Pearson (or standardized) residuals

$$r_{ij} = \frac{(y_{ij} - \hat{\mu}_{ij})}{\sqrt{[V(\hat{\alpha})]_{jj}}}$$

and then, use these residuals to estimate α , and ϕ .

These steps are repeated until convergence is reached. Finally, the standard errors for $\hat{\beta}$ are obtained using robust, or empirical estimates (also called *sandwich* estimates)

$$V(\hat{\beta}) = M_0^{-1}M_1M_0^{-1}$$

where

$$M_0 = \sum_{i=1}^N D_i' \hat{V}_i^{-1} D_i$$

$$M_1 = \sum_{i=1}^N D_i' \hat{V}_i^{-1} (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)' \hat{V}_i^{-1} D_i$$

Finally, the following are some important properties of the *GEE* method [31]:

- *GEE* reduces to *GLM* estimating equations for $t_i = 1$. In this case, Equation 3.2 reduces to:

$$V_i = \phi v(\mu_i) \tag{3.4}$$

and,

$$D_i' = \partial \mu_i / \partial \beta = \left[\frac{\partial g(\mu_i)}{\partial \mu_i} \right]^{-1} x_i \tag{3.5}$$

Then, combining Equation 3.4 and 3.5, the Equation 3.3 reduce to

$$\sum_{i=1}^N \left[\frac{\partial g(\mu_i)}{\partial \mu_i} \right]^{-1} x_i [\phi v(\mu_i)]^{-1} (y_i - \mu_i) = 0 \tag{3.6}$$

which is equivalent to the score equation of a *GLM* [33].

- *GEE* are the maximum likelihood score equations for multivariate Gaussian data when unstructured correlation is specified.
- The regression parameter estimates are consistent if the number of subjects increase, even if the working correlation matrix is misspecified, as long as the model for the mean is correct.

- The *empirical sandwich estimator* of the covariance matrix of $\hat{\beta}$ is also consistent if the number of subjects is large, even if the working correlation matrix is misspecified, as long the model for the mean is correct.

Currently, there are many statistical software packages for fitting models for longitudinal data with categorical outcomes using *GEE*, such as SAS PROC GENMOD [31], *gee*, and *geepack* R libraries, Stata, SPSS, and others.

3.2 Subject-specific Models

As presented in the previous chapter, subject-specified models can be treated using three approaches: fixed effects, random effects, and conditioning on a sufficient statistic for the subjects. In this section, we emphasize the *random effects models* as a way to model discrete longitudinal data. When we refer to subject-specific model we will always talk about the random-effects models.

In contrast with the marginal models, which model the correlation between repeated measures over the same subject using a working correlation matrix, the *subject-specific models*, induce the correlation between observations by including random effects for the subjects. As was mentioned by Diggle *et al* [35] “*the correlation among repeated observations arises because we cannot observe the underlying growth curve, that is, the true regression coefficients*”. The importance of including random effects in the models has been studied by many authors, which consider that omitting them could generate a loss of efficiency and, therefore, an increase in the standard errors of parameter estimates [7]. These models are also referred by some authors as *conditional models* [27, 35], *multilevel models* [15, 44] or, *latent variable models*, depending on the model formulation.

A subject-specific model uses the following assumptions:

1. The conditional expectation of the response variable is related to the covariates and the random effects by:

$$g(\tilde{\mu}_{ij}) = g[\mathbb{E}(Y_{ij} \mid \mathbf{u}_i)] = x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i \quad (3.7)$$

where $g(\cdot)$ is the link function, $\tilde{\mu}_{ij}$ is the conditional mean, x'_{ij} is the covariate vector, $\tilde{\beta}$ is the fixed effects parameter, z'_{ij} is the random effects covariate vector, and \mathbf{u}_i represents the random effects.

2. Given \mathbf{u}_i , the responses $y_{i1} \dots y_{in_i}$ are mutually independent, and follow a *GLM*. \mathbf{y} has a $f(\mathbf{y} \mid \beta, D, \sigma^2)$ density function in the exponential family.
3. The random effects, \mathbf{u}_i , are mutually independent with a common underlying multivariate distribution $f(\mathbf{u})$.

Similarly to the marginal models, if $g(\cdot)$ is the identity function, then, the model in Equation 3.7 reduces to the linear mixed model: $\mathbb{E}(Y_{ij} \mid \mathbf{u}_i) = x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i$.

In these models, \mathbf{u}_i is a sample of independent unobservable variables from a random effects distribution. Qualitatively, this assumption implies that we can learn about one individual's coefficients by understanding the variability in coefficients across the population. When there is little variability, we should rely on the population average coefficients to estimate those for an individual. When the variation is considerable, we must rely more heavily on the data from each individual in order to estimate their own coefficients. The random effects models is most useful when the objective is to make inference about individuals rather than the population average. The regression coefficients, $\tilde{\beta}$ represent the effects of the covariances on an particular subject. This is a *subject-specific interpretation*, unlike the marginal models where the β 's have a *population-average interpretation*.

The maximum likelihood estimation in Equation 3.7 is based on the likelihood function of \mathbf{y} :

$$f(\mathbf{y}) = \prod_{i=1}^N \int \prod_{j=1}^{n_i} f(y_{ij} \mid \beta, D, \sigma) f(\mathbf{u}) d\mathbf{u} \quad (3.8)$$

This is a marginal distribution of \mathbf{y} obtained by integrating the joint distribution of (\mathbf{y}, \mathbf{u}) with respect to \mathbf{u} . In general, this integral does not have a close-form expression, so different approximations have been proposed, for example, the Gauss-Hermite quadrature, linear mixed effects approach (LME), and the Laplacian approximation [36], Monte Carlo algorithms [14, 32] and others [12]. We will discuss further some of these methods.

There exist some well-known subject-specific models: beta-binomial, probit-normal, and the *generalized linear mixed models* (GLMM). GLMMs will be our point of interest in this thesis because of their wide use in discrete longitudinal data analysis.

3.2.1 Generalized Linear Mixed Models (GLMM)

GLMMs are an extension of the generalized linear model (GLM), adding random effects [12]. In a GLMM the response vector, \mathbf{y} , is assumed to have conditionally independent elements given the random effects \mathbf{u} , each one with density function in the exponential family [33]. Let us suppose that $\tilde{\mu}_{ij}$, the conditional mean of y_{ij} , is related with the systematic, and random components by the following equation:

$$g(\tilde{\mu}_{ij}) = x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i \quad (3.9)$$

where x'_{ij} , and z'_{ij} are the covariates vector of the fixed, and random effects, respectively. The mean-to-variance relationship is given by

$$\text{Var}(Y_i | \mathbf{u}_i) = \phi a_i v(\tilde{\mu}_{ij})$$

where $v(\cdot)$ is the variance function that relates the conditional means and variances, ϕ is a scale factor (i.e. ϕ is assumed equal to one for standard binomial, and Poisson models), a_i is a prior weight, such as the reciprocal of a binomial denominator. Finally, \mathbf{u}_i has a q -dimensional known distribution function $f_{\mathbf{u}_i}(u_i)$, such as the normal distribution $N(0, D)$, which is used in many cases.

Because of the properties of function g , then it is possible to apply g^{-1} to both sides in Equation 3.9:

$$E(y_{ij} | \mathbf{u}_i) = \tilde{\mu}_{ij} = g^{-1}(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)$$

Table 3–2 shows common link functions with their respective models. In this case, $\Phi(\cdot)$ is the standard normal cumulative distribution.

Table 3–2: Common link functions in GLMM

Outcome	Link Function	Equation Model
Binary	Logit	$\log\left(\frac{\tilde{\mu}_{ij}}{1-\tilde{\mu}_{ij}}\right) = x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i$
Binary	Probit	$\Phi^{-1}(\tilde{\mu}_{ij}) = x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i$
Binary	Complementary log – log	$\ln(-\ln(1 - \tilde{\mu}_{ij})) = x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i$
Count	Log	$\log(\tilde{\mu}_{ij}) = x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i$

The consequences of including random effects in the models, can be analyzed by studying the first two moments of the marginal distribution of Y_{ij} . Note that the marginal mean of Y_{ij} induced by the random effects, denoted by $E_u(Y_{ij})$, is given by:

$$E_u(Y_{ij}) = E_{\mathbf{u}_i}[g^{-1}(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)] \quad (3.10)$$

which, in general, cannot be simplified, due to the nonlinear function g^{-1} . A particular case is the Poisson regression with a random intercept distributed normally with zero mean and variance $\sigma_{\mathbf{u}}^2$. In this case, the induced marginal mean of Y_{ij} is equal to:

$$E_u(Y_{ij}) = x'_{ij}\tilde{\beta} + \frac{\sigma_{\mathbf{u}}^2}{2}$$

The marginal (induced) mean obtained by integrating out the individual's heterogeneities, will be a population-average if and only if the subjects in the study can be regarded as a random sample from a population. According to some authors, the *population-average* concept is true only if we are analyzing a random sample [27].

The marginal variance of Y_{ij} induced by the random effects, denoted by $\text{Var}_u(Y_{ij})$, has the following expression:

$$\begin{aligned}\text{Var}_u(Y_{ij}) &= \text{Var}(\mathbb{E}[Y_{ij} \mid \mathbf{u}_i]) + \mathbb{E}(\text{Var}[Y_{ij} \mid \mathbf{u}_i]) \\ &= \text{Var}(\mu_{ij}) + \mathbb{E}[\phi a_i v(\mu_{ij})] \\ &= \text{Var}(g^{-1}[x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i]) + \mathbb{E}\{\phi a_i v[g^{-1}(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)]\}\end{aligned}\tag{3.11}$$

which again cannot be simplified in most cases. In a linear mixed model, the induced marginal mean in Equation 3.10 is reduced to: $\mathbb{E}_u(Y_{ij}) = x'_{ij}\tilde{\beta}$. Likewise, the induced marginal variance in Equation 3.11 can be reduced to: $\text{Var}_u(Y_{ij}) = Z_i D Z'_i + \Sigma_i$, similar to Equation 2.18.

Finally, assuming conditional independence of the elements of Y_i , we have that

$$\begin{aligned}\text{Cov}_u(Y_{ij}, Y_{ik}) &= \text{Cov}(\mathbb{E}[Y_{ij} \mid \mathbf{u}_i], \mathbb{E}[Y_{ik} \mid \mathbf{u}_i]) + \mathbb{E}[\text{Cov}(Y_{ij}, Y_{ik} \mid \mathbf{u}_i)] \\ &= \text{Cov}[g^{-1}(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i), g^{-1}(x'_{ik}\tilde{\beta} + z'_{ik}\mathbf{u}_i)]\end{aligned}$$

Similarly, in a linear mixed model, the previous equation is reduced to $\text{Cov}_u(Y_{ij}, Y_{ik}) = D_i z_{ij} z'_{ik}$. If the linear mixed model has only a random intercept, $\text{Cov}_u(Y_{ij}, Y_{ik}) = \sigma_u^2$, where σ_u^2 is the variance of the random intercepts.

3.2.2 Inference in GLMM

To fit a *GLMM* it is possible to use two alternatives: *Bayesian approach* and *Maximum Likelihood Estimation (MLE)*. In the Bayesian approach, it is necessary to specify the prior densities for β , D , and ϕ , denoted by $f(\beta)$, $f(D)$, and $f(\phi)$, respectively. Once priors have been specified, the posterior distribution can be found [34].

On the other hand, the complete likelihood function for *GLMM* can be written from the product of the known distributions of $\mathbf{y} \mid \mathbf{u}$, and \mathbf{u} . The true likelihood

function can be written as:

$$L(\mathbf{y}, \mathbf{u}) = L(\mathbf{y} | \mathbf{u})L(\mathbf{u}) \quad (3.12)$$

If we assumed that \mathbf{u} has a multivariate normal distribution $N(0, D)$, then

$$L(\mathbf{y}, \mathbf{u}) \propto L(\mathbf{y} | \mathbf{u}) |D|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{u}'D^{-1}\mathbf{u}\right) \quad (3.13)$$

Based on Equation 3.13, the marginal likelihood for \mathbf{y} could be obtained integrating over the random effects \mathbf{u} . However, this is not always possible because the resultant integral of the right side in Equation 3.13 does not have an analytical solution, that is, it must be solved using numerical methods.

3.2.3 Approximations to the Likelihood Function in GLMM

GLMM can be fitted maximizing the marginal likelihood, obtained by integrating over the random effects. The contribution of the i – *th* subject to the likelihood is given by:

$$L_i = f_i(y_i, \beta, D) = \int_{\mathfrak{R}} \prod_{j=1}^{n_i} f_{\mathbf{y}_{ij}|\mathbf{u}}(y_{ij} | u) f_{\mathbf{u}}(u) du \quad (3.14)$$

Thus, the likelihood function L can be written as:

$$L = \prod_{i=1}^N L_i = \prod_{i=1}^N \int_{\mathfrak{R}} \prod_{j=1}^{n_i} f_{\mathbf{y}_{ij}|\mathbf{u}}(y_{ij} | u) f_{\mathbf{u}}(u) du \quad (3.15)$$

where \mathbf{y}_i is the n_i –dimensional vector containing all the measurements available for the i – *th* subject. Notice that Equation 3.15 has the form of Equation 3.12. This integration is done over the q -dimensional distribution of \mathbf{u} . In some cases, Equation 3.15 can be worked out analytically, for example probit-normal model [34]. In other cases, such as the logistic mixed model, the integral in Equation 3.15 cannot be evaluated in closed form. Numerical approximation is necessary in order to evaluate the likelihood, and the derivatives needed in the optimization algorithm [9, 14]. Numerical approximations for the integral of the marginal likelihood in Equation

3.15 can be divided into those based on the *approximation of the integral itself*, such as the Gauss-Hermite quadrature [34], those based on the *approximation of the integrand*, such as the Laplace approximation, and those based on *an approximation of the data*, such as the penalized quasi-likelihood [34].

Approximation of the Integral: Gauss-Hermite Quadrature

In many applications, it is convenient to consider an integral of the form:

$$\int_{\mathfrak{R}} W(x)f(x)dx$$

where $W(x)$ is a non-negative function over the integration interval. $W(x)$ is called the *weighted* or *kernel* function. The integral can be interpreted as a weighted average of $f(x)$.

Gauss-Hermite quadrature approximate integrals of the form:

$$\int_{\mathfrak{R}} f(x) \exp(-x^2)dx \tag{3.16}$$

where the kernel function is $W(x) = \exp(-x^2)$. Gauss-Hermite quadrature is often used for numerical integration in statistics because of the direct relationship it has with Gaussian densities [26, 30]. In many statistical applications, a Gaussian density is an explicit factor of the integrand. Using a linear transformation this factor takes the form $\exp(-x^2)$. When a Gaussian density is not a factor of the integrand, the original integrand must be divided and multiplied by $\exp(-x^2)$, or some other Gaussian density, in order to put it into the Gauss-Hermite quadrature form [30].

We want to apply Gauss-Hermite quadrature to approximate integrals of the form:

$$\int_{\mathfrak{R}} g(x)dx \tag{3.17}$$

Table 3–3: Four points of Gauss-Hermite quadrature

Abscissa	Weight
-1.650668	0.081312
-0.524647	0.804914
0.524647	0.804914
1.650668	0.081312

where $g(t) > 0$, and it is a smooth and unimodal function. Gauss-Hermite quadrature can be rewritten in terms of Gaussian density as:

$$\int_{\mathbb{R}} f(x)\phi(x; \mu, \sigma)dx \approx \sum_{j=1}^K w_j f(z_j) \quad (3.18)$$

where $\phi(x; \mu, \sigma)$ is an arbitrary Gaussian density. The abscissas are $z_j = \mu + \sqrt{2}\sigma x_j$, and the weights are modified from w_j to $w_j/\sqrt{\pi}$, $j = 1, 2, \dots, K$ points of quadrature. This approximation is called *Classical Quadrature*.

We implemented the **phermite** function in **R** to calculate the abscissas, and weights for a classic Gauss-Hermite quadrature (see *Appendix A*). These values can be found in books on numerical analysis [11, 41]. Table 3–3 shows the abscissas, and weights using four points of quadrature. These points are symmetric about zero. Sometimes it is necessary to take many points in order to approximate an integral because the abscissas are sampled in an inappropriate region. *Adaptive Quadrature* samples the abscissas in an appropriate region where $\hat{\mu}$ will be the mode of the integrand $g(x)$, and $\hat{\sigma} = \frac{1}{\sqrt{\hat{j}}}$, where:

$$\hat{j} = -\frac{\partial^2}{\partial x^2} \log g(x) \Big|_{x=\hat{\mu}} \quad (3.19)$$

Figure 3–1 clearly shows the differences in the abscissas between classic, and adaptive quadrature for a specific function $f(x)$. Using $\phi(x; \hat{\mu}, \hat{\sigma})$, the approximation is:

$$\int g(x)dx \approx \sqrt{2}\hat{\sigma} \sum_{j=1}^K w_j^* g(\hat{\mu} + \sqrt{2}\hat{\sigma}x_j) \quad (3.20)$$

where $w_j^* = w_j \exp x_j^2$.

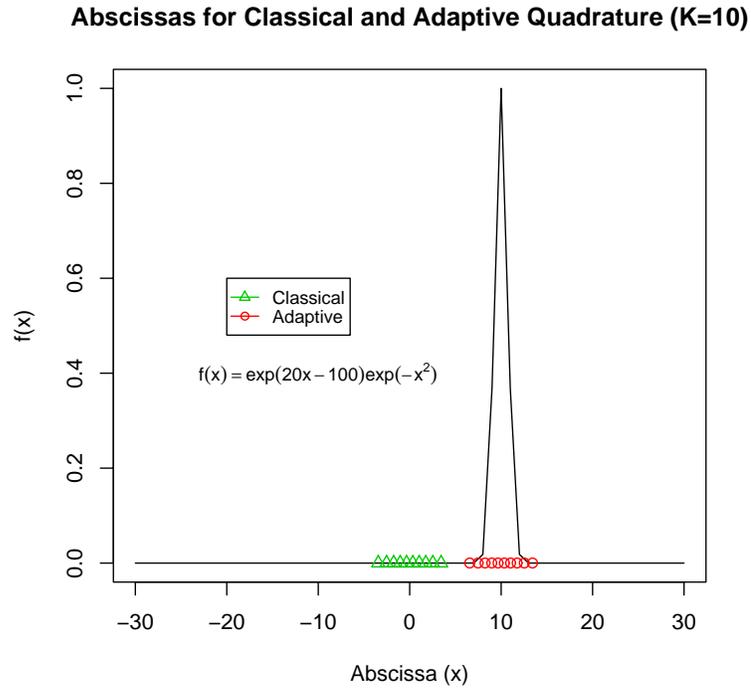


Figure 3–1: Abscissas for Classic and Adaptive Quadrature

When Equation 3.20 is applied with only one abscissa the result is the Laplace approximation to the integral [30]:

$$\int_{\mathbb{R}} g(x) dx \approx \sqrt{2\pi} \hat{\sigma} g(\hat{\mu}) \quad (3.21)$$

Therefore, the m -order Gauss-Hermite quadrature is, alternatively, the form of the m -order Laplace approximation. Some simulation results suggest that in the classic quadrature a larger number of abscissas are necessary to obtain high accuracy (100 points or more), while the adaptive quadrature provides good accuracy with 20 points or fewer [12, 35]. For the random intercept logistic model, we have confirmed that these values are reasonable in practice [28, 39].

The previous approach show the case of a one-dimensional random effect. The generalization of adaptive Gaussian quadrature, with K points of quadrature, is

given by [23]

$$\int_{\mathfrak{R}} L_i(y_i | \beta, \mathbf{u}_i) L(\mathbf{u}_i) d\mathbf{u}_i \approx 2^{q/2} |Q(\theta)|^{-1/2} \sum_{j_1=1}^K \cdots \sum_{j_q=1}^K [L(y_i | \beta, a_{j_1}, \dots, a_{j_q})] L(u_{j_1}, \dots, u_{j_q}) \prod_{k=1}^q w_{j_k} \exp(z_{j_k}^2)$$

where $Q(\theta)$ is the Hessian matrix of the Empirical Bayes Minimization, and $a_{j_1}, \dots, a_{j_q} = \hat{u}_i + 2^{1/2} Q(\theta)^{-1/2} z_{j_1, \dots, j_q}$, where z_{j_1, \dots, j_q} is a vector with elements $(z_{j_1}, \dots, z_{j_q})$. \hat{u}_i minimizes $-\log [L_i(y_i | \beta, u_i) L(u_i)]$, and (z_j, w_j) are the abscissas, and weights of the classic quadrature, respectively.

Numerical integration to calculate the likelihood is straightforward, and hence, numerical maximization of the likelihood can be evaluated accurately. This approximation works relatively well in simple situations, such as single random effect, two, or three nested random effects [35]. However, it fails for more complicated structures, such as crossed random effects. In addition, it could be heavy computationally for two, or more random effects. In some cases *Monte Carlo Markov-Chain (MCMC)* is a good alternative to estimate parameters in these models. Some of the best known algorithms are *Monte Carlo EM (MCEM)*, *Monte Carlo Newton Raphson (MCNR)*, and *Simulated Maximum Likelihood* [32].

Many statistical software packages use the Gauss-Hermite approximation in their procedures, such as SAS PROC NL MIXED [24], several R functions [37], and MIXNO [22].

Approximation of the Integrand: Laplace Approximation

To illustrate how Laplace approximation works, let us suppose that we want to approximate $I = \int_{\mathfrak{R}} e^{-Q(\mathbf{u})} d\mathbf{u}$, and that $\tilde{\mathbf{u}}$ is the value of \mathbf{u} for which Q is minimized.

Then, the second-order Taylor expansion of $Q(\mathbf{u})$ around $\tilde{\mathbf{u}}$ is of the form

$$Q(\mathbf{u}) \approx Q(\tilde{\mathbf{u}}) + \frac{1}{2}(\mathbf{u} - \tilde{\mathbf{u}})Q''(\tilde{\mathbf{u}})(\mathbf{u} - \tilde{\mathbf{u}}) \quad (3.22)$$

where $Q''(\tilde{\mathbf{u}})$ is equal to the Hessian of Q evaluated at $\tilde{\mathbf{u}}$. The integral I can be approximated using Equation 3.22, thus

$$I \approx (2\pi)^{q/2} |Q''(\tilde{\mathbf{u}})|^{-1/2} e^{-Q(\tilde{\mathbf{u}})} \quad (3.23)$$

In this case, it is considered that $Q(\cdot)$ is unimodal. When Q is bimodal, it is necessary to use an improved Laplace approximation [12]. In this method, the approximation to the integral uses as many estimates of \mathbf{u} as necessary according to the different modes of the Q function. Clearly, the integral in Equation 3.15 is proportional to an integral I , for a $Q(\mathbf{u})$ function given by:

$$Q(\mathbf{u}) = (a_i\phi)^{-1} \sum_{j=1}^{n_i} [y_{ij}(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}) - \psi(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u})] - \frac{1}{2}\mathbf{u}'D^{-1}\mathbf{u} \quad (3.24)$$

such that Laplace's method can be applied here. In the previous equation, $\psi(\cdot)$ are the functions presented in Table 2-1.

Approximation of the Data: Penalized and Marginal Quasi-Likelihood

For *GLM*, maximum quasi-likelihood is an attractive method because of its ability to generate highly efficient estimators without making precise distributional assumptions. The *quasi-likelihood* does not specify a distribution, only the mean and the variance [2]. For *GLMM*, the objective function to maximize is modified by a term of penalty due to the random effects, and therefore, it is called the *Penalized Quasi-likelihood (PQL)*. *PQL* could be viewed as a method of approximated inference in *GLMM*. In the statistical literature, different versions of *GLMM* can be found [3, 34]. We will describe the initial proposal presented by Breslow and Clayton [4].

In a *GLM* context, and based on Equation 3.12, for any type of *GLMM*, the quasi-likelihood form is given by

$$QL(\mathbf{y}, \beta, \mathbf{u}) = QL(\mathbf{y} \mid \mathbf{u})L(\mathbf{u}) \quad (3.25)$$

Following the same principles used in the likelihood functions, the integrated quasi-likelihood function (*PQL*) is given by:

$$PQL = (2\pi)^{-q/2} |D|^{-1/2} \int_{\mathfrak{R}} \exp \left[-\frac{1}{2\phi} \sum_{i=1}^N d_i(y_i, \tilde{\mu}_i) - \frac{1}{2} \mathbf{u}' D^{-1} \mathbf{u} \right] d\mathbf{u} \quad (3.26)$$

where

$$d_i(y_i, \tilde{\mu}_i) = -2 \int_{y_i}^{\tilde{\mu}_i} \frac{y_i - t}{a_i v(t)} dt \quad (3.27)$$

denotes the weighted deviance. If y is Gaussian, and g^{-1} is the identity, the integral in Equation 3.27 is normal, and may be evaluated in closed form. Otherwise, this expression contains integrals that must be solved numerically with methods such as Laplace approximation.

Equation 3.26 has the form $c |D|^{-1/2} \int e^{-Q(\mathbf{u})} d\mathbf{u}$, with c a constant term, and Q a function of \mathbf{u} . Therefore, Laplace approximation can be applied to approximate this integral. Using the result of Equation 3.23, the log-quasilikelihood is given by

$$pql \approx -\frac{1}{2} \log |D| - \frac{1}{2} \log |Q''(\tilde{\mathbf{u}})| - Q(\tilde{\mathbf{u}}) \quad (3.28)$$

where $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}(\beta, \theta)$ is the solution to

$$Q'(\mathbf{u}) = - \sum_{i=1}^N \frac{(y_i - \tilde{\mu}_i) z_i}{\phi a_i v(\tilde{\mu}_i) g'(\tilde{\mu}_i)} + D^{-1} \mathbf{u} = 0$$

that minimizes $Q(\mathbf{u})$. In addition, the approximation of the second derivative of Q is given by

$$Q''(\mathbf{u}) = \sum_{i=1}^N \frac{z_i z_i'}{\phi a_i v(\tilde{\mu}_i) [g'(\tilde{\mu}_i)]^2} + D^{-1} + R \approx Z' W Z + D^{-1}$$

Therefore, the log quasi-likelihood is given by

$$pql \approx -\frac{1}{2} \log |D| - \frac{1}{2} \log |I + Z'WZD| - \frac{1}{2\phi} \sum_i d_i(y_i, \hat{\mu}_i) - \frac{1}{2} \tilde{\mathbf{u}}' D^{-1} \tilde{\mathbf{u}} \quad (3.29)$$

where $\tilde{\mathbf{u}}$ is chosen to maximize the last two terms, I is the identity matrix, and W is the $N \times N$ diagonal matrix with diagonal terms $w_i = \{\phi a_i v(\hat{\mu}_i) [g'(\tilde{\mu}_i)]^2\}^{-1}$. This expression leads to the *PQL* algorithm to estimate β , and θ in the model presented in Equation 3.9. Differentiating with respect to β and u , the score equations for the mean parameters are:

$$\sum_{i=1}^N \frac{(y_i - \tilde{\mu}_i) x_i}{\phi a_i v(\tilde{\mu}_i) g'(\tilde{\mu}_i)} = 0 \quad (3.30)$$

and

$$\sum_{i=1}^N \frac{(y_i - \tilde{\mu}_i) z_i}{\phi a_i v(\tilde{\mu}_i) g'(\tilde{\mu}_i)} = D^{-1} \mathbf{u} \quad (3.31)$$

The solutions of Equations 3.30, and 3.31 are based on normal theory defining the working vector Y^* with components $Y_i^* = (x_i' \hat{\beta} + z_i' \hat{\mathbf{u}}) + (y_i - \hat{\mu}_i) g'(\hat{\mu}_i)$. Then, the Fisher scoring algorithm is applied repeatedly in order to determine $(\hat{\beta}, \hat{\mathbf{u}})$ as a function of θ . This procedure leads to the familiar mixed model equations for joint estimation of fixed, and random effects, but involving Y^* and weights w_i .

An alternative derivation of the *PQL* algorithm was developed by Schall [3] in the context of longitudinal data. This method proposes to use a linearization of the conditional mean as a function of fixed and random effects. This method is based on a decomposition of the data into the conditional mean, and an appropriate error term, with a Taylor series expansion of the mean that is a non-linear function of the linear predictor [34]. Consider the decomposition

$$Y_{ij} = \tilde{\mu}_{ij} + \varepsilon_{ij} = h(x'_{ij} \tilde{\beta} + z'_{ij} \mathbf{u}_i) + \varepsilon_{ij} \quad (3.32)$$

where $\text{Var}(Y_{ij}) = \phi v(\mu_{ij})$ for $v(\cdot)$ the usual variance function, and $h = g^{-1}$. For example, with the logistic model

$$\mu_{ij} = P(Y_{ij} = 1) = \pi_{ij} = \frac{\exp(x'_{ij}\tilde{\beta} + z'_{ij}u_i)}{1 + \exp(x'_{ij}\tilde{\beta} + z'_{ij}u_i)}$$

ε_{ij} is equal to $1 - \pi_{ij}$ with probability π_{ij} , and is equal to $-\pi_{ij}$ with probability $1 - \pi_{ij}$.

Consider a linear Taylor expansion of Equation 3.32, around current estimates $\hat{\beta}$, and \hat{u}_i of the fixed effects, and random effects, respectively. Thus,

$$\begin{aligned} Y_{ij} &\approx h(x'_{ij}\hat{\beta} + z'_{ij}\hat{u}_i) & (3.33) \\ &+ h'(x'_{ij}\hat{\beta} + z'_{ij}\hat{u}_i)x'_{ij}(\tilde{\beta} - \hat{\beta}) \\ &+ h'(x'_{ij}\hat{\beta} + z'_{ij}\hat{u}_i)z'_{ij}(u_i - \hat{u}_i) + \varepsilon_{ij} \\ &= \hat{\mu}_{ij} + v(\hat{\mu}_{ij})x'_{ij}(\tilde{\beta} - \hat{\beta}) + v(\hat{\mu}_{ij})z'_{ij}(u_i - \hat{u}_i) + \varepsilon_{ij} \end{aligned}$$

where $\hat{\mu}$ equals the current predictor $h(x'_{ij}\hat{\beta} + z'_{ij}\hat{u}_i)$ for the conditional mean $E(Y_{ij} | u_i)$. Rewriting Equation 3.33 in vector notation, and reordering the terms yields:

$$Y_i^* \equiv \hat{V}_i^{-1}(Y_i - \hat{\mu}) + X_i\hat{\beta} + Z_i\hat{u}_i \approx X_i\tilde{\beta} + Z_i\hat{u}_i + \varepsilon_i^* \quad (3.34)$$

where $\varepsilon_i^* = \hat{V}_i^{-1}\varepsilon_i$, with zero mean. Equation 3.34 can be viewed as a linear mixed model for the pseudo data Y_i^* , with fixed effects $\tilde{\beta}$, random effects u_i , and error terms ε_i^* . This result yields estimates for a *GLMM*. Given the starting values for parameters β and D in the marginal likelihood, empirical Bayes estimates for u_i using the posterior density function and pseudo data Y_i^* are computed. Then, the approximate linear mixed model presented in Equation 3.34 is fitted, and then, the estimates of the parameters are updated. This process is iterated until convergence is reached. The resulting estimates are called *penalized quasi-likelihood* estimates because they are obtained from optimizing a quasi-likelihood function using approximations of first

and second order. Other approximation can be found in Wolfinger and O'Connell [43].

An alternative approximation is very similar to the *PQL* method, but is based on a linear Taylor expansion of the mean $\tilde{\mu}_{ij}$ around current estimates $\hat{\beta}$ for the fixed effects, and around $u_i = 0$ for the random effects. This produces similar expressions to *PQL*, but $\tilde{\mu}_{ij}$ takes the form $h(x'_{ij}\hat{\beta})$ rather than $h(x'_{ij}\hat{\beta} + z'_{ij}\hat{u}_i)$. The pseudo data are now of the form $Y_i^* \equiv \hat{V}_i^{-1}(Y_i - \hat{\mu}) + X_i\hat{\beta}$, and satisfy the approximate linear mixed model

$$Y_i^* \approx X_i\hat{\beta} + Z_i u_i + \varepsilon_i^*$$

Following the same procedure than in *PQL*, the resulting estimates are called *Marginal Quasi-likelihood (MQL)* estimates.

3.3 Transition Models

Under a transition model, correlation among Y_{i1}, \dots, Y_{in_i} exists because the past values Y_{i1}, \dots, Y_{ij-1} explicitly influence the present observation, Y_{ij} . The past outcomes are treated as additional predictor variables. Denote, $\mathbf{H}_{ij} = \{y_{ik}, k = 1, \dots, j-1\}$, the history for i -th subject in the j -th visit. The most useful transition models are *Markov* chains for which the conditional distribution of Y_{ij} given H_{ij} depends only on the q prior observations Y_{i1}, \dots, Y_{ij-q} . The integer q is called the *order* of the model [35].

A transition model establishes the following assumptions:

1. The conditional expectation, denoted by μ_{ij}^* , is specified by:

$$g(\mu_{ij}^*) = g(\mathbb{E}(y_{ij} \mid \mathbf{H}_{ij})) = x'_{ij}\beta^* + \sum_{r=1}^s f_r(\mathbf{H}_{ij}; \alpha) \quad (3.35)$$

for suitable functions $f_r(\cdot)$ and $g(\cdot)$, the link function.

2. The conditional variance satisfies the equation:

$$\text{Var}(Y_{ij} \mid \mathbf{H}_{ij}) = \phi v(\mu_{ij}^*) \quad (3.36)$$

where $v(\cdot)$ is a known variance function.

This model assumes that the past affects the present through the sum of s terms, each of which may depend on the q prior values.

As an example, a linear regression model with autoregressive errors for Gaussian data is a Markov model. It has the form:

$$Y_{ij} = x'_{ij}\beta^* + \sum_{r=1}^q \alpha_r(Y_{ij-r} - x'_{ij-r}\beta^*) + Z_{ij}$$

where the Z_{ij} are independent, and zero-mean Gaussian errors. This is a transition model with $g(\cdot)$ equal to the identity function, $v(\mu_{ij}^*) = 1$, and $f_r(\mathbf{H}_{ij}; \alpha) = \alpha_r(Y_{ij-r} - x'_{ij-r}\beta^*)$. Note that the present observation, Y_{ij} , is a linear function of x'_{ij} and of the earlier deviations $Y_{ij-r} - x'_{ij-r}\beta^*$, $r = 1, \dots, q$.

An alternative to estimate β and α in Equation 3.35 is maximizing the conditional likelihood:

$$\prod_{i=1}^N f(y_{iq+1}, \dots, y_{in_i} \mid y_{i1}, \dots, y_{iq}) = \prod_{i=1}^N \prod_{j=q+1}^{n_i} f(y_{ij} \mid \mathbf{H}_{ij}) \quad (3.37)$$

When maximizing Equation 3.37 there are two distinct cases to consider. In the first, $f_r(\mathbf{H}_{ij}; \alpha) = \alpha_r f_r(\mathbf{H}_{ij})$. Here, $g(\mu_{ij}^*)$ is a linear function of both β^* and $\alpha = (\alpha_1, \dots, \alpha_s)$, therefore, the estimation proceeds as in *GLMs* for independent data. The second case occurs when the functions of past responses include both α and β^* . Here, iterative weighted least squares are necessary as estimation procedure.

The parameters in a transition model, β^* , have a *population-average* interpretation conditioned on the past responses. That is, the effect of a covariate over the response variable depends on the value of the past responses.

These models were discussed here for comparison purpose, and will not be used in the thesis.

3.4 Marginalized Models

Marginalized models were originally proposed by Heagerty [20, 21]. These models establish a *regression structure* for the marginal mean μ_{ij} :

$$g(\mu_{ij}) = g\{E(Y_{ij})\} = x'_{ij}\beta \quad (3.38)$$

The second component of the model describes the *dependence* among measurements within a cluster conditioning on a latent variable rather than on other response:

$$g(\mu_{ij} \mid x'_{ij}, \mathbf{A}_{ij}) = \Delta_{ij}(x'_{ij}) + \gamma'_{ij}\mathbf{A}_{ij} \quad (3.39)$$

where $\Delta_{ij}(x'_{ij})$ represents a function of the marginal mean and dependence parameters, and the variables \mathbf{A}_{ij} are used to structure dependence among the repeated measurements. \mathbf{A}_{ij} can take different forms. For example, $\mathbf{A}_{ij} = \{Y_{ik} : k \neq j\}$. In this case, γ'_{ij} indicates how strongly all other response variables, Y_{ik} , predict the current response, Y_{ij} . Alternatively, $\mathbf{A}_{ij} = \mathbf{u}_i$, a collection of random effects. Here, γ'_{ij} represents the variance components that characterize the magnitude of unobservable or random variation which induces all within-subject correlation. These models are called *marginalized latent variables models*. Finally, when $\mathbf{A}_{ij} = \{Y_{ik} : k < j\} = \mathbf{H}_{ij}$, these models take the form of the so-called *marginalized transition models*.

When the conditional mean in Equation 3.39 is averaged over the distribution of \mathbf{A}_{ij} , the value $\Delta_{ij}(X_i)$ is chosen such that the resulting marginal mean structure is properly induced:

$$\mu_{ij} = E_{\mathbf{A}_{ij}}[E\{g^{-1}(\Delta_{ij}(x'_{ij}) + \mathbf{u}_i)\}]$$

We will discuss with more details the *marginalized latent variables models*. Consider the following model:

$$g(\mu_{ij}) = \Delta_{ij}(x'_{ij}) + \mathbf{u}_i \quad (3.40)$$

This model assumes that the elements of the response vector Y_i are conditionally independent given \mathbf{u}_i and the distribution of \mathbf{u}_i is completely specified, for example, $\mathbf{u}_i \sim N(0, D)$. Common models for longitudinal data include a subject-level random effects model such that $u_{ij} = u_{i0}$ yielding $\text{cov}(u_{ij}, u_{ik}) = D_i(j, k) = \sigma^2$, or a temporal association model, where u_{ij} is assumed to have an autoregressive covariance structure and $\text{cov}(u_{ij}, u_{ik}) = D_i(j, k) = \sigma^2 \exp(-\gamma |t_{ij} - t_{ik}|^\theta)$, where $1 \leq \theta \leq 2$ is fixed.

This formulation is an alternative to the generalized linear mixed model which parameterizes the conditional mean function $\Delta_{ij}(x'_{ij}) = x'_{ij}\tilde{\beta}$. The marginalized model in 3.38 and 3.39 also permits conditional statements via implicitly defined $\Delta_{ij}(x'_{ij})$, recognizing their dependence on model assumptions. The parameter $\Delta_{ij}(x'_{ij})$ is a function of both the marginal linear predictor $x'_{ij}\beta$, and the random effects, \mathbf{u}_i . $\Delta_{ij}(x'_{ij})$ is defined as the solution to the integral equation that links the marginal and conditional means:

$$h(x'_{ij}\beta) = \int_{\mathfrak{R}} h[\Delta_{ij}(x'_{ij}) + u_i] f_{\mathbf{u}_i}(u_i) d\mathbf{u}_i \quad (3.41)$$

where $h = g^{-1}$ is the inverse of the link function.

For the simplest case where $u_i \sim N(0, \sigma^2)$, we can rewrite $u_i = \sigma z_i$, where $z \sim N(0, 1)$, and the Equation 3.41 is now

$$h(x'_{ij}\beta) = \int_{\mathfrak{R}} h[\Delta_{ij}(x'_{ij}) + \sigma z] \phi(z) dz \quad (3.42)$$

where ϕ is the standard normal density function. Given $x'_{ij}\beta$, and σ the integral equation can be numerically solved for $\Delta_{ij}(x'_{ij})$ using numerical integration and Newton-Raphson iteration [26, 39]. For some special cases, such as the probit link function and Gaussian random effects, $u = \sigma z$, it yields the relationship

$$\Phi(x_{ij}\beta) = E[\Phi\{\Delta_{ij}(x'_{ij}) + \sigma z\}] = \Phi\left[\frac{\Delta_{ij}(x'_{ij})}{\sqrt{1 + \sigma^2}}\right]$$

showing that the marginal linear predictor $x'_{ij}\beta$ is a rescaling of the conditional linear predictor $\Delta_{ij}(x'_{ij})$. Also it happens with the Poisson-normal regression, where:

$$\Phi(x_{ij}\beta) = \exp \left[\Delta_{ij}(x'_{ij}) + \frac{1}{2}\sigma^2 \right]$$

The estimation of parameters in marginalized models is based on maximum likelihood, quadratic estimating equations and empirical Bayes for individual-level estimation [20, 35]. All these steps require computationally challenging numerical methods [19]. The parameters in a marginalized model, β , have a *population-average* interpretation.

Although we consider that marginalized models are theoretically interesting, in this work they will not be used because they are not yet implemented in commercial software packages and hence have very little use in practice [18].

Comparisons between some of the models presented in the previous sections will be shown in Chapter 4 for the logistic model.

3.5 Non-linear Mixed Effects Model

This type of models have a distinct formulation in comparison with the previous ones. The first stage of a non-linear mixed model (*NLME*) consists of N nonlinear regression models with some random parameters (subject-specific),

$$Y_i = \mathbf{f}_i(\gamma, \delta_i) + \varepsilon_i, \quad i = 1, \dots, N$$

where \mathbf{f}_i is $n_i \times 1$ vector function,

$$\mathbf{f}_i(\gamma, \delta_i) = [f(\gamma, \delta_i; x'_{i1}), \dots, f(\gamma, \delta_i; x'_{in_i})]'$$

γ is a $k \times 1$ vector of fixed parameters. In addition, $\text{Var}(\varepsilon_i) = \sigma^2 I$ and $\text{Var}(\delta_i) = \sigma^2 D$.

The second stage of the *NLME* has the form:

$$\delta_i = X_i\beta + \mathbf{u}_i$$

Random effects \mathbf{u}_i and error terms ε_i are assumed mutually independent and independent across subjects [10, 12].

This formulation implies that generally the marginal expectation of the response variable, $E(Y_{ij})$, cannot be expressed in terms of population-average parameters in close terms, namely,

$$E(Y_{ij}) = E_{\mathbf{u}_i}\{f_i(\gamma, X_i\beta + \mathbf{u}_i)\} \neq f_i(\gamma, X_i\beta)$$

such as with the *GLMM*. Note that the fixed effects parameters, γ , do not enter as linear terms in the model. The parameter estimation in these are calculated using maximum likelihood and techniques to approximate it [36].

CHAPTER 4

MODELS FOR BINARY LONGITUDINAL DATA

4.1 Introduction

Binary data can be specified either as a series of zeros and ones (*Bernoulli form*), or aggregated as frequencies of successes out of a certain number of trials (*Binomial form*).

In many longitudinal studies, the researcher is interested in a dichotomous variable as response, for example, presence, or absence of a disease in patients of a clinic treated with different medications, effectiveness of a particular health care service, or presence of symptoms in plants treated with different fungicides. In all these cases, the binary responses are clustered within observational units, and hence, the classical model fails in its independence assumption (i.e. data are taken at several occasions on the same unit). In this case, it is possible to analyze this type of data using the models described in the previous chapter with appropriate link functions. We will concentrate on binary longitudinal data analysis with the logistic model because of its wide use.

4.2 Marginal and Random-Effects Models

In this section, we discuss the use of marginal and random-effects models to model binary longitudinal data. Also, we compare the parameter interpretation of both models for the logistic case.

4.2.1 Marginal Models

First, let us consider a longitudinal binary response variables vector $Y_i = (y_{i1}, \dots, y_{in_i})$ measured at recorded times vector $t_i = (t_{i1}, \dots, t_{in_i})$ along with possibly time-dependent covariates matrix $X_i = (X_{i1}, \dots, X_{in_i})$, where in each observation the covariate is a p -dimensional vector x'_{ij} , for $i = 1, 2, \dots, N$ subjects. The assumptions of the marginal model are:

1. The marginal expectation is given by:

$$g(\mu_{ij}) = x'_{ij}\beta \quad (4.1)$$

where $Y_{ij} \sim \text{Bernoulli}(p_{ij})$, or $Y_{ij} \sim \text{Binomial}(m_{ij}, p_{ij})$, m_{ij} is the number of trials for the i -th subject in the j -th measurement, and $g(\cdot)$ is the *link function*, such as logit, probit or complementary log-log. Specifically, for a marginal model with logit link function, Equation 4.1 can be rewritten as:

$$\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = x'_{ij}\beta \quad (4.2)$$

If we solve for p_{ij} in Equation 4.2, the result is

$$p_{ij} = \frac{\exp(x'_{ij}\beta)}{1 + \exp(x'_{ij}\beta)}$$

which indicates the marginal mean as a function of the covariates. The function $\frac{\exp(\cdot)}{1+\exp(\cdot)}$ is the inverse link function, that is, the inverse logit.

2. The variance of Y_{ij} is a function of the p_{ij} , and a parameter ϕ called *dispersion parameter*. Thus,

$$\text{Var}(Y_{ij}) = \phi v(p_{ij}) \quad (4.3)$$

If $Y_{ij} \sim \text{Bernoulli}(p_{ij})$ or $Y_{ij} \sim \text{Binomial}(m_{ij}, p_{ij})$, $\text{Var}(Y_{ij})$ can be rewritten as, respectively:

$$\text{Var}(Y_{ij}) = \phi p_{ij}(1 - p_{ij})$$

and,

$$\text{Var}(Y_{ij}) = \phi m_{ij} p_{ij} (1 - p_{ij})$$

When the data have no overdispersion, the parameter ϕ in the previous equations is equal to 1.

3. The correlation between observations over the same subject is:

$$\text{Corr}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha) \quad (4.4)$$

Some authors suggest to use $AR(1)$ correlation structures for this type of data [1, 34]. However, the choice of the working correlation matrix, as always, depends on the observed data patterns and, the underlying assumptions.

To interpret the fixed coefficients in a logistic model, let us consider a longitudinal study with a binary covariate x ($x = 0, 1$), and a response variable $Y_{ij} \sim \text{Bernoulli}(p_{ij})$. In addition, let us suppose that the data are analyzed with a logistic model given by the following equation:

$$\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_0 + \beta_1 x + \beta_2 t \quad (4.5)$$

Furthermore, assume that the residuals have a correlation structure given by a matrix R_i , where R_i is any structured matrix such as the ones presented in Section 3.1.1. Based on Equation 4.5, we obtain

$$\beta_1 = \text{logit}(p_{ij} | x = 1) - \text{logit}(p_{ij} | x = 0)$$

which measures the variation in success between groups determined by the values of x . This is the *log odds ratio* ($\log OR$). This is an *population-average* interpretation.

The use of these models for binary longitudinal data have been questioned because it does not accomplish the necessary assumptions. Some authors have studied the efficiency of *GEE* in binary data [1]. Specifically, if the response variable $y = (y_1, \dots, y_p)$ is a Bernoulli random vector with marginal probabilities p_j , and

$q_j = 1 - p_j$ for $j = 1, \dots, p$, and constant correlation ρ between any two pairs, then

$$\max_{j \neq k} \left\{ -\sqrt{\left(\frac{p_j p_k}{q_j q_k}\right)}, -\sqrt{\left(\frac{q_j q_k}{p_j p_k}\right)} \right\} \leq \rho \leq \min_{j \neq k} \left\{ -\sqrt{\left(\frac{p_j q_k}{q_j p_k}\right)}, -\sqrt{\left(\frac{q_j p_k}{p_j q_k}\right)} \right\}$$

However, if $p_i(\mathbf{x})$ is a function of a covariate vector \mathbf{x} , then a constant correlation matrix over all \mathbf{x} would imply that the constant correlation must lie in the interval

$$\max_{\mathbf{x}} \left(\max_{j \neq k} \left\{ -\sqrt{\left(\frac{p_j p_k}{q_j q_k}\right)}, -\sqrt{\left(\frac{q_j q_k}{p_j p_k}\right)} \right\} \right) \leq \rho \leq \min_{\mathbf{x}} \left(\min_{j \neq k} \left\{ -\sqrt{\left(\frac{p_j q_k}{q_j p_k}\right)}, -\sqrt{\left(\frac{q_j p_k}{p_j q_k}\right)} \right\} \right) \quad (4.6)$$

Inequality 4.6 shows that the constraint on ρ will depend on the covariates, but current *GEE* software ignores them. When the range of \mathbf{x} is wide, the interval in the inequality can be quite narrow. Chaganty and Joe [1] suggested an initial data analysis with tabulations and odds ratios to assess the strength of the dependence. If the covariate vectors have just a few values, then an initial analysis consist of tabulating the frequencies of the p -dimensional binary vectors for each case of the covariate vector, and computing empirical odds ratios and correlations for each bivariate margin. If all or majority of the p -dimensional binary vectors occur, then the dependence is not strong; the strongest dependence is indicated if the frequencies concentrate near the vectors of all 0's and all 1's. Otherwise the odds ratios and correlations will suggest whether the dependence is weak or moderate, and whether and exchangeable or *AR*(1) structure is better for the working correlation matrix. They also suggest that generally one could use an exchangeable matrix $R_i(\alpha)$ for cluster-type samples, and an *AR*(1) matrix for longitudinal data analysis.

In addition to *GEE*, there exist others methods to calculate the estimates in marginal models such as *Prentice's GEE method*, *Second-Order Generalized Estimating (GEE2)*, and *Alternative Logistic Regression (ALR)* [34]. *GEE* estimates can be calculated using SAS PROC GENMOD, *gee* and *geepack* R libraries, and other procedures are available in commercial software packages.

4.2.2 Random-Effects Models

The assumptions of the random-effects model with logit link function and normal distribution are:

1. The conditional expectation is given by:

$$\text{logit}(\tilde{\mu}) = \log\left(\frac{\tilde{p}_{ij}}{1 - \tilde{p}_{ij}}\right) = x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i \quad (4.7)$$

If we solve for \tilde{p}_{ij} in Equation 4.7, the result is:

$$E(Y_{ij} | \mathbf{u}_i) = \tilde{p}_{ij} = \frac{\exp(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)}{1 + \exp(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)}$$

The previous equation shows clearly that the conditional mean is a function of the random effects, that is, $\tilde{p}_{ij} = f(\mathbf{u}_i)$.

2. The random effects are normally distributed: $\mathbf{u}_i \sim N(0, D)$.

If we consider a model with random intercept, $\mathbf{u}_i = u_{i0}$, the model presented in Equation 4.7 with a time covariate is reduced to

$$\text{logit}(\tilde{\mu}) = \log\left(\frac{\tilde{p}_{ij}}{1 - \tilde{p}_{ij}}\right) = \beta_0 + \beta_1 t_{ij} + u_{i0} \quad (4.8)$$

and, therefore,

$$\tilde{p}_{ij} = \frac{\exp(\beta_0 + \beta_1 t_{ij} + u_{i0})}{1 + \exp(\beta_0 + \beta_1 t_{ij} + u_{i0})}$$

where $u_{i0} \sim N(0, \sigma^2)$. On the other hand, u_{i0} could depend on the covariates. In such case, the random-effect variance, σ^2 , may vary according to the levels of the covariates. This random effect, u_{i0} , also denoted as u_i , indicates that there exists a between-subject variability that must be explained by a random-effect associated with each subject. This random-effect is a consequence of the characteristics of each subject, and therefore, to the distinct forms in that they could respond across time. The model in Equation 4.8 is called *logistic model with random intercept*, or *logistic-normal model* [2, 32].

The log-likelihood function for a logistic model with random intercept takes the form [12]

$$\log L = l(\tilde{\beta}, \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) + r'\tilde{\beta} + \sum_{i=1}^N \ln \int_{\mathcal{R}} \exp[h_i(\tilde{\beta}, u)] du \quad (4.9)$$

where

$$h_i(\tilde{\beta}, u) = k_i u - \frac{u^2}{2\sigma^2} - \sum_{j=1}^{n_i} \ln[1 + \exp(x'_{ij}\tilde{\beta} + u)] \quad (4.10)$$

with $k_i = \sum_{j=1}^{n_i} y_{ij}$, and

$$r = \sum_{i=1}^N \sum_{j=1}^{n_i} y_{ij} x'_{ij}$$

It is assumed that the first component of x'_{ij} is 1. $\tilde{\beta}$ indicates the subject-specific parameters.

The previous integrals could be solved using Gauss-Hermite quadrature. The maximal error for Gauss-Hermite quadrature is increasing as the variance of the random errors increases. For $K = 20$ abscissas, some authors have found that the maximal possible error in the approximation using Gauss-Hermite quadrature is approximately 10^{-3} [35]. For a logistic-normal model, several approximations of these integrals and comparisons between them can be found in Demidenko [12].

Now, if we assume a logistic model with random intercepts and slopes, then the model equation is given by

$$\text{logit}(\tilde{\mu}_{ij}) = \log\left(\frac{\tilde{p}_{ij}}{1 - \tilde{p}_{ij}}\right) = \tilde{\beta}_0 + (\tilde{\beta}_1 + u_{i1})t_{ij} + u_{i0} \quad (4.11)$$

where $\mathbf{u}_i = (u_{i0}, u_{i1}) \sim N[(0, 0), D]$. The matrix D is the variance-covariance of \mathbf{u} . Figure 4-1 shows the difference between a logistic model with random intercept only and one with both random intercept and slope. In the model with only random intercept, the subject-specific curves differ by a small value due to the random intercept. In contrast, the subject-specific curves in the random intercept and slope model, vary considerably across time due to both random quantities, intercept and

slope. In both cases, the differences between the curves depend on the magnitude of the entries in the D matrix.

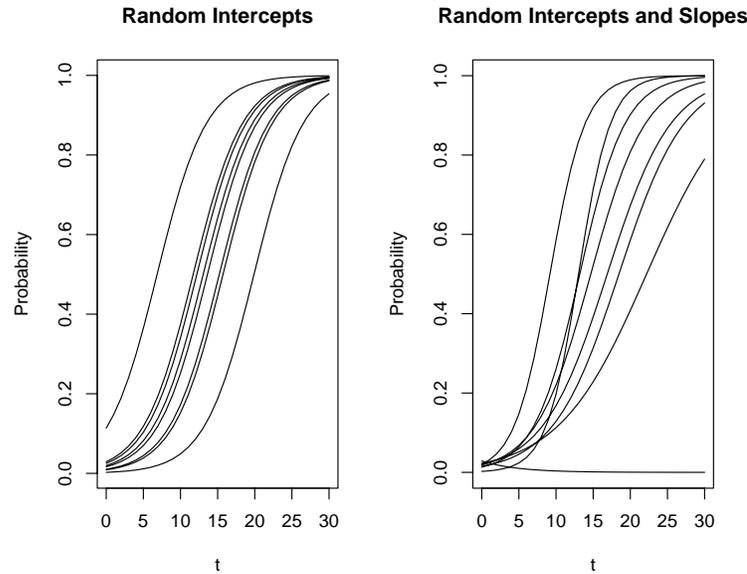


Figure 4–1: Alternative Random Effects Models for Longitudinal Binary Data Analysis

In a logistic model with random intercept, the standard deviation of the random effects, σ has an interesting interpretation. We redefine the model presented in Equation 4.5 in conditional terms as

$$\text{logit}(p_{ij} | u_i) = \log \left(\frac{\tilde{p}_{ij}}{1 - \tilde{p}_{ij}} \right) = \tilde{\beta}_0 + \tilde{\beta}_1 x + \tilde{\beta}_2 t_{ij} + z_i \sigma \quad (4.12)$$

where $z_i \sim N(0, 1)$. Based on the Equation 4.12, we have that

$$\tilde{\beta}_1 = \text{logit}(\tilde{p}_{ij} | x = 1) - \text{logit}(\tilde{p}_{ij} | x = 0)$$

measures the variation in success log odds between groups determined by the values of x for the i -th subject. This is also a subject-specific *log odds ratio* ($\log OR$). This is a *subject-specific* interpretation. Likewise, σ could be interpreted as a regression coefficient for a standardized omitted covariate, with σ contrasting individuals with

equal covariate x and time t_{ij} , whose z_i 's differ by one unit:

$$\sigma = \text{logit}(\tilde{p}_{ij} \mid z_i = c) - \text{logit}(\tilde{p}_{ij} \mid z_i = c + 1)$$

where c is a constant. That is, σ measures the magnitude of variation in log odds between subjects within a group defined by the covariates [19].

4.2.3 Comparison between Marginal and Random-Effects Models

Distinguishing random-effects or conditional models from marginal models is straightforward probabilistically, but it is often difficult in practice. We will present the main theoretical differences between the two model families for the logistic model. Based on the model presented in Equation 4.7, we can compare the two approaches by considering three aspects, which will be analyzed next.

Marginal and Conditional Mean

The relation between population-average and subject-specific estimates have been discussed by many authors. Neuhaus *et al.* [35] shows that if $\text{Var}(\mathbf{u}_i) > 0$, then the elements of the marginal β and random effects $\tilde{\beta}$ regression vector satisfy

1. $|\beta_k| \leq |\tilde{\beta}_k|$, for all k . β and $\tilde{\beta}$ indicate the marginal and conditional estimates, respectively.
2. equality holds if and only if $\tilde{\beta} = 0$, or $\sigma = 0$;
3. the discrepancy between β and $\tilde{\beta}$ increases with $\text{Var}(\mathbf{u}_i)$.

In particular, if \mathbf{u}_i is assumed to follow a Gaussian distribution with mean zero and variance σ^2 in a logistic model with random intercept, Zeger *et al.* showed that

$$\beta \approx \frac{\tilde{\beta}}{\sqrt{c^2\sigma^2 + 1}} \quad (4.13)$$

where $c = \frac{16\sqrt{3}}{15\pi}$, that is, $c^2 \approx 0.346$. Clearly, $\tilde{\beta}$ is always greater than β . We can write $\beta = k\tilde{\beta}$, where k is a constant that depends on σ . Figure 4-2 shows the k values for different values of σ . For example, when $\sigma = 2$ the marginal estimate, β_k , will be approximately 75% smaller than the conditional one, $\tilde{\beta}_k$. Note that $\beta_k = \tilde{\beta}_k$

when $\sigma = 0$. The plot illustrates clearly that the discrepancy between β and $\tilde{\beta}$ increases with σ . Intuitively, this relation between curves is due to the fact that the variance of the random effect is large and all curves are averaged. The resulting unconditional mean has a smaller slope and value β [33].

To illustrate these relations in practice, we show the marginal and subject-specific estimates for a fitted logistic model with random intercept for the Toenail dataset [34], which will be described in Chapter 5.

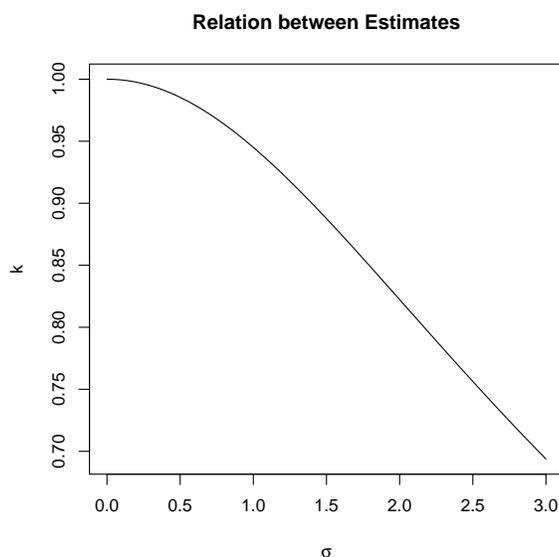


Figure 4-2: Constant k versus σ values for the relation between marginal and subject-specific estimates: $\beta = k\tilde{\beta}$

The marginal and random effect models are given by the following equations, respectively:

$$\text{logit}(p_{ij}) = \tilde{\beta}_0 + \tilde{\beta}_1 \text{treat} + \tilde{\beta}_2 t_{ij}$$

$$\text{logit}(\tilde{p}_{ij}) = \tilde{\beta}_0 + \tilde{\beta}_1 \text{treat} + \tilde{\beta}_2 t_{ij} + u_{i0}$$

In the marginal model, we consider an exchangeable correlation structure. The estimated standard deviation of the random intercepts is $\hat{\sigma}=4.0082$. Table 4-1 shows

the relation between the estimates mentioned in Equation 4.13. The marginal estimates obtained using the relationship 4.13 are very similar to the ones obtained using a marginal model.

Table 4–1: Relation between marginal and subject-specific estimates in a logistic model with random intercept, *Toenail Data*

Effect	Marginal	Subject-specific	Marginal using (4.13)
Intercept	−0.5819	−1.6190	−0.6321
Treatment	0.0072	−0.1609	−0.0628
Time	−0.1713	−0.3911	−0.1528
Interaction	−0.0777	−0.1368	−0.0534

In other cases, such as the probit model this relation is direct. Considering

$$\tilde{p}_{ij} = \Phi(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)$$

where $\mathbf{u} \sim N(0, D)$, then

$$\beta = \frac{\tilde{\beta}}{\sqrt{z'_{ij}Dz'_{ij}}}$$

For *GLMM* models with logit link and *Bridge function*, which will be discussed in Section 4–4, these relations can also be found analytically. This type of relations are important to establish the impact that produces the inclusion of a random effect in the model and to show the difference between population-average and subject-specific estimates. In addition, these equations allow to compare the curves underlying the marginal and random-effects models, such the ones shown in Figure 4–3. The solid line indicates the population-average curve and the dashed lines indicate the subject-specific curves for a logistic model with random intercept and a continuous covariate x .

Induced marginal mean

The model presented in Equation 4.7 generates two expectations: (1) the conditional mean, $E(Y_{ij} | \mathbf{u}_i)$, and (2) the marginal mean induced by the random effects,

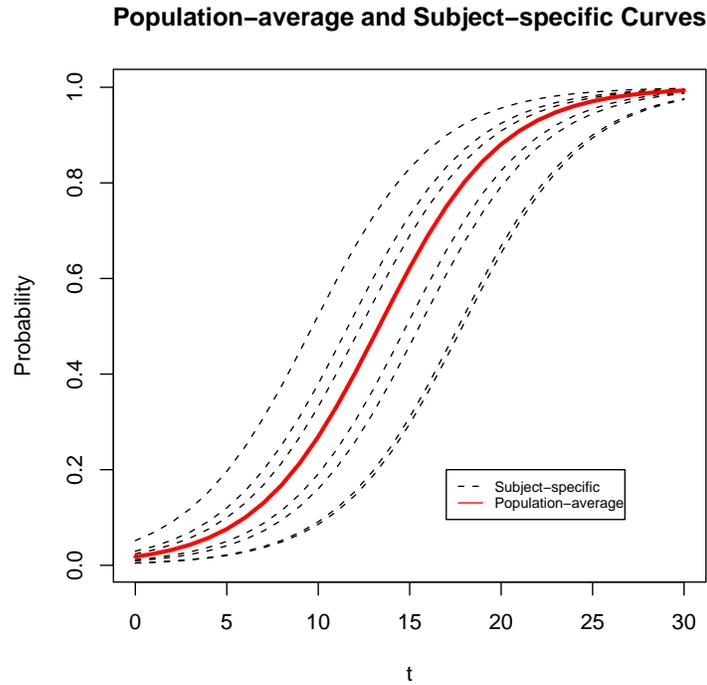


Figure 4–3: Marginal and Subject-specific curves for logistic model with a random intercept

denoted by $E_u(Y_{ij})$. The induced marginal mean is given by:

$$\begin{aligned}
 E_u(Y_{ij}) &= E_{\mathbf{u}}[E(Y_{ij} \mid \mathbf{u}_i)] \\
 &= E_{\mathbf{u}} \left[\frac{\exp(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)}{1 + \exp(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)} \right] \\
 &= \int_{\mathfrak{R}} \frac{\exp(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)}{1 + \exp(x'_{ij}\tilde{\beta} + z'_{ij}\mathbf{u}_i)} f(\mathbf{u}) d\mathbf{u}
 \end{aligned}$$

This result is one of the principal advantages of the conditional models with the marginal models. In a conditional model, the induced marginal mean could be obtained by integrating the random effects, while in a marginal model it is only possible to specify a marginal mean. Several authors consider this advantage as the ability of the random effects to produce marginal inference [30].

In general, the induced marginal mean and the marginal mean are different,

$$E_u(Y_{ij}) \neq \frac{\exp(x'_{ij}\beta)}{1 + \exp(x'_{ij}\beta)}$$

where the β 's are the population-average parameters presented in Section 3.1. Intuitively, this difference is due to the fact that the random-effect model considers the between-subject variability, while in a marginal model, this aspect is omitted.

Finally, unlike the linear mixed model, the marginal mean (*population-average profile*) and the conditional mean given $\mathbf{u}_i = 0$ (*typical profile*) are also different:

$$E(Y_{ij}) \neq E(Y_{ij} \mid \mathbf{u}_i = 0) = \frac{\exp(x'_{ij}\tilde{\beta})}{1 + \exp(x'_{ij}\tilde{\beta})}$$

Functional shape of the distributions

This result is related with the functional shape of the conditional and induced marginal distributions. In the conditional model, the functional shape has a logistic form, as shown in Equation 4.7. This implies that the conditional parameters $\tilde{\beta}$ can be interpreted in terms of log odds ratios given the random effects for each subject. However, the induced marginal distribution obtained integrating the random effects, \mathbf{u} , in the model 4.7, does not have not a logistic form. This result is due to the fact that the integral

$$\int_{\mathfrak{R}} \frac{\exp(x'_{ij}\tilde{\beta} + \mathbf{u}_i)}{1 + \exp(x'_{ij}\tilde{\beta} + \mathbf{u}_i)} f_{\mathbf{u}}(u) d\mathbf{u}$$

is not a logistic form when $\mathbf{u} \sim N(0, D)$. However, Wang and Louis [42] proposed a special distribution function for \mathbf{u} where the marginal (induced) functional shape is still logistic and parameters have an explicit marginal interpretation. This models will be discussed in Section 4-4. For Poisson and Probit regressions with normal random intercepts, the functional shape in the induced marginal distribution is retained, as we showed previously.

4.3 Marginalized Latent Variables Models

Let us consider the simplest marginally specified logistic-normal. The model for random individual variation is given by the marginal expectation:

$$\text{logit } E(Y_i) = x'_{ij}\beta \quad (4.14)$$

The second model is used to describe the dependence among longitudinal measurements and is based on a conditional model where instead of conditioning on other response variables, we condition on a latent variable:

$$\text{logit } E(Y_i | \mathbf{u}_i) = \Delta_{ij} + \mathbf{u}_i \quad (4.15)$$

The response Y_i is conditionally independent given \mathbf{u}_i , and $\mathbf{u}_i \sim N(0, D_i)$. The covariance matrix D_i can be obtained as a function of the observation times t_i and a parameter vector α . In the marginally specified logistic-normal model, the parameter Δ_{ij} is a function of both the marginal linear predictor $\eta_{ij} = x'_{ij}\beta$ and the random effects standard deviation $\sigma_{ij} = \sqrt{\text{Var}(u_{ij})}$. Δ_{ij} can be obtained as the solution to the convolution equation

$$h(\eta_{ij}) = \int_{\Re} h(\Delta_{ij} + \mathbf{u}_i)\phi(z)dz$$

where $h = \text{logit}^{-1}$ and ϕ is the standard normal density function. Given $(\Delta_{ij}, \sigma_{ij})$, this equation can be solved for Δ_{ij} using numerical integration and Newton-Raphson iteration. σ_{ij} is allowed to depend on covariates, like in *GLMM*'s. For example, $\log(\sigma_i) = Z_i\alpha$, where Z_i is a subset of X_i . The marginally specified logistic-normal model is related to the conditionally specified logistic-normal model, where the conditional log odds Δ_{ij} are directly modeled as a function of the covariates.

Let us assume a logistic model with random intercept, $\mathbf{u}_i = u_{i0}$, where $u_{i0} \sim N(0, \sigma^2)$. We can substitute $u_{ij} = \sigma z_i$, where $z_i \sim N(0, 1)$, and express the model

for random individual variation by the conditional expectation:

$$\text{logit } E(Y_i | \mathbf{u}_i) = \Delta_{ij} + \sigma z_i$$

This representation shows that the variance component σ may be interpreted as a regression coefficient for a standardized covariate, with σ contrasting subjects with equal Δ_{ij} , whose z_i 's differ by one unit. Because Δ_{ij} is determined by x'_{ij} and the parameters (β, α) , subgroups defined by the covariates x'_{ij} are the same subgroups defined by Δ_{ij} [19].

These models can be defined for nested random effects. Let Y_{ijk} denote the binary response for the i -th cluster, j -th subject and k -th measurement. Consider the following regression model for the marginal log odds $\eta(x'_{ijk})$ [21]:

$$\text{logit } E(Y_{ijk}) = \eta(x'_{ijk})$$

The multilevel model is completed with a second conditional assumption:

$$\text{logit } E(Y_{ijk} | z_{2,ij}, z_{3,i}) = \Delta(x'_{ijk}) + \sigma_2 z_{2,ij} + \sigma z_{3,i}$$

for $z_{2,ij} \sim N(0, 1)$ and $z_{3,i} \sim N(0, 1)$ mutually independent. Each one measures the random variation in the second and third level, respectively.

4.4 Random-Effects Models with Bridge Function

These models, originally proposed by Wang and Louis, are a *GLMM* [42]. The fundamental feature is that the random effects \mathbf{u}_i do not necessarily have a normal distribution. Let us consider the random-intercept logistic regression model commonly used in binary longitudinal data. The model for the conditional expectation has the following form:

$$\text{logit}(\mu_{ij} | u_i) = x'_{ij} \tilde{\beta} + u_i \tag{4.16}$$

The parameter $\tilde{\beta}$ measures the change in the conditional logit of the mean with covariance x'_{ij} for each subject with random effect u_i . Integrating out the random effects in the model 4.16, the induced marginal mean is given by

$$\mu_{ij} = \int_{\mathfrak{R}} \frac{\exp(x'_{ij}\tilde{\beta} + u_i)}{1 + \exp(x'_{ij}\tilde{\beta} + u_i)} f_{\mathbf{u}_i}(u_i) du_i \quad (4.17)$$

which is usually not of logistic form if u_i is normally distributed. That is, the model for the conditional expectation has a logistic form but the marginal expectation induced by the conditional model does not have a logistic form. This result motivates a new distribution for the random effects. Let $H(\cdot)$ be the inverse link function, and let $G(u_i)$ be the probability distribution function for random effects. It is of interest whether or not there exists a distribution $G(u)$ such that the conditional functional shape is retained in the marginal scale, thus,

$$\int_{\mathfrak{R}} H(x'_{ij}\tilde{\beta} + u_i) dG(u) = H(k + \phi x'_{ij}\tilde{\beta})$$

where k and ϕ are unknown parameters. For $\eta = x'_{ij}\tilde{\beta}$, and non-degenerate $G(u)$ (neither null nor uniform distribution), differentiation with respect to η yields

$$\int_{\mathfrak{R}} h(\eta + u_i) dG(u) = \phi h(k + \phi\eta) \quad (4.18)$$

where $h = H'$. Equation 4.18 can be written as

$$h * g_{-b}(\eta) = \phi h(k + \phi\eta) \quad (4.19)$$

where g_{-b} is the probability density function of $-b$ and the symbol $*$ indicates the convolution operation. Taking the Fourier transform F of both sides of Equation 4.19, applying some properties, and the Fourier Inversion Theorem, the probability density function of u is given by

$$g_u(x) = \frac{1}{2\pi} \int_{\mathfrak{R}} \exp[i(\frac{k}{\phi} - x)\zeta] \frac{Fh(\zeta/\phi)}{Fh(\zeta)} d\zeta \quad (4.20)$$

This function is called the *Bridge density function* [42]. This function has the characteristic that it retains the functional form in the marginal mean of the conditional model.

For the logit link, $H(\eta) = \exp(\eta)/[1 + \exp(\eta)]$, $h(\eta) = \exp(\eta)[1 + \exp(\eta)]^{-2}$, and the bridge density function is given by

$$g_u(x, \phi) = \frac{1}{2\pi} \frac{\sin \phi\pi}{\cosh(\phi x) + \cos(\phi\pi)} \quad (4.21)$$

for $0 < \phi < 1$, and $-\infty < x < \infty$. The Bridge distribution presented in Equation 4.21 has zero mean, and variance $\sigma_u^2 = \pi^2(\phi^{-2} - 1)/3$. Figure 4-4 compares the *Bridge*, Normal and Logistic density functions with zero mean and variance equal to 1. The *Bridge* density is more concentrate around the zero. It also has slightly heavier tails than the Gaussian and lighter tails than the logistic.

Under the Bridge distributional assumption, the regression parameters in the marginal and conditional logistic models are related by

$$\beta = \frac{\tilde{\beta}}{\sqrt{1 + 3\sigma_u^2/\pi^2}} \quad (4.22)$$

This relation is attractive because it permits to interpret the parameters as population-average and subject-specific as well, fitting only a random effects model. In the logistic model, it is possible to talk about *log odds* in both the induced marginal and conditional forms.

In addition, $\beta = \tilde{\beta}(1 - \rho_y)$ where $\rho_y = \text{corr}(Y_{ij}, Y_{ij'} \mid \tilde{\beta} = 0)$ is the intra-cluster correlation in the binary response scale. When the random effects are nested, for example of the form $u_i + u_{ij}$, the Bridge distribution can still be used to preserve marginalization. The marginal mean is given by:

$$\int H(x'_{ij}\tilde{\beta} + u_i + u_{ij})dG(u_{ij} \mid u_i)dG(u_i) = \int H(\phi_1 u_i + \phi_1 \tilde{\beta} x'_{ij})dG(u_i)$$

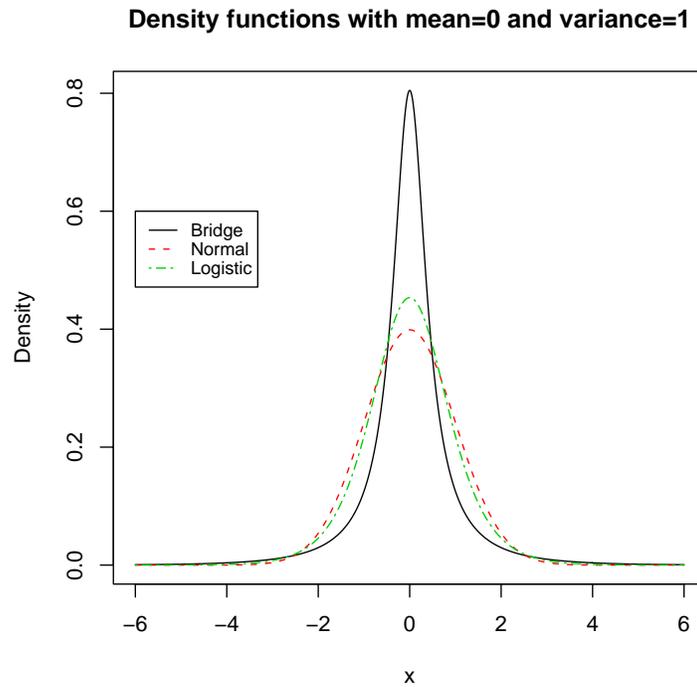


Figure 4-4: Normal, Logistic and Bridge Density Functions for Logistic-Normal Model

If we assume that $u_i \sim \phi^{-1}B(0, \phi_2)$, not $B_l(0, \phi_2)$, then the marginal mean is given by $H(\phi_1\phi_2X_i\tilde{\beta})$.

A generalized model with random intercept with Bridge distribution can be easily fitted using PROC NLMIXED SAS and applying the *Probability Integral Transformation Theorem* [8] such as was presented by Wang and Louis [42]. Table 4-2 shows the estimates obtained for a fitted logistic model with random intercepts with a Bridge distribution, $\mathbf{u}_i \sim \text{Bridge}(0, \sigma_{\mathbf{u}}^2)$. The estimate for $\hat{\phi} = 0.3481$, and therefore, $\hat{\sigma}_{\mathbf{u}} = 4.3597$. Note that the estimates are similar to the ones obtained with the model with normal random effects.

Table 4–2: Marginal and Subject-specific estimates in a logistic model with random intercept with *Bridge* distribution, *Toenail Data*

Effect	Subject-specific	Marginal using (4.22)
Intercept	−1.4966	−0.8773
Treatment	−0.1255	−0.0736
Time	−0.4010	−0.02351
Interaction	−0.1410	−0.0826

4.5 Available software to Fit Random-Effects Models

Currently there exist many softwares packages to fit *GLMMs* such as SAS, R, S-PLUS, EGRET, LIMDEP, MIXOR and MIXREG, MLWiN, STATA, OSWALD, WINBUGS (it uses *MCMC*), and others [17]. All these programs offer wide possibilities to analyze binary longitudinal data, however, each one has its advantages and limitations.

We will briefly describe the functions and procedures that exists in R v2.2.1 and SAS v9.1.3 packages [24, 37].

4.5.1 Functions Available in R and SAS Software Packages

Table 4–3 shows the best known functions of R to fit generalized linear mixed models for binary longitudinal data. We have summarized the more important topics for each function.

Similarly, Table 4–4 shows the two SAS procedures to fit longitudinal binary data. This table includes the GLIMMIX procedure, the last procedure released by SAS Institute to fit these type of models.

In general, the SAS routines have better consistency and the processing time is less than the ones obtained using R. In some R functions it is necessary to consider too many points of quadrature ($Q \geq 30$) to obtain good estimates while SAS produces similar estimates with fewer points [39]. *Appendix B* shows codes and outputs for an example with each R function and SAS procedures for the Toenail dataset, which

will be described in Section 5.3. The estimates of parameter regression differ for R functions and SAS procedures.

Table 4–3: Available Functions in R to fit GLMMs

Function	Library	Brief Description
glmmML (Bröstom 2001)	glmmML [5]	<ul style="list-style-type: none"> • Fits <i>GLMMs</i> with random intercept by Maximum Likelihood and numerical integration via classical Gauss-Hermite quadrature. • Accepts binomial and Poisson distributions. • Logit and Complement log-log are link functions permitted to binary data. • It invokes the <code>optim</code> function of the MASS library. • Newton-Raphson procedure finishes the maximization.
glmmPQL (Venables and Ripley 2002)	MASS [40]	<ul style="list-style-type: none"> • Fits <i>GLMMs</i> with multivariate normal random effects, using Penalized Quasi-Likelihood. • Accepts different distributions in the response and link functions. • Invokes the <code>lme</code> function of the nlme library. • It is possible to fit binary longitudinal data with nested random structure. • Permits correlation structures (R matrix). • It also produces the empirical Bayes estimates for random effects.
lmer (Bates and Maechler 2006)	Matrix	<ul style="list-style-type: none"> • This generic function fits a generalized linear mixed-effects model with nested or crossed grouping factors for the random effects. This function was called GLMM in the lme4 library. • Uses the approximation penalized quasi-likelihood (<i>PQL</i>). • Laplacian and adaptive Gaussian quadrature are used to approximate the integrals. • The <i>PQL</i> method is the fastest but least accurate. The Laplace method is intermediate in speed and accuracy. The Adaptive Quadrature method is the most accurate but can be considerably slower than the others.
glmm (Lindsey 2005)	Repeated	<ul style="list-style-type: none"> • Requires <code>rmutil</code> library. • Fits a generalized linear mixed model with a random intercept using a normal mixing distribution computed by Gauss-Hermite integration. • It permits several response distributions and link functions. • This library also has more functions to fit generalized non-linear mixed models and others (<code>gnlmm</code>, <code>gnlmm3</code>, <code>binnest</code>, and <code>gnlmix</code> functions).

Table 4–4: Available Procedures in SAS to fit GLMMs

Function	Brief Description
NLMIXED Procedure [24]	<ul style="list-style-type: none"> • Fits nonlinear mixed models using Maximum Likelihood Estimation. • Permits normal, binomial, Poisson or a general distribution for the response. • Different integral approximations are available, the principal ones being adaptive Gaussian quadrature and a first-order Taylor series approximation. • A variety of alternative optimization techniques are available to carry out the maximization; the default is a dual quasi-Newton algorithm. • Does not permit correlation structures (\mathbf{R} matrix). • Standard errors are estimated using Delta method. • Produces the empirical Bayes estimates of the random effects. • It is possible to fit a logistic model with random intercept with a Bridge distribution. • It permits programming code to incorporate heterogeneous variances, etc.
GLIMMIX Macro	<ul style="list-style-type: none"> • Fits <i>GLMM</i> using pseudo-likelihood and PROC MIXED. • This allows multiple random effects, nested and crossed random effects, and multiple cluster types. • Permits covariance structures for random effects and correlated errors. • Linearization and Taylor-series techniques are used to construct Wald-type test statistics and confidence intervals. • This is a production version Procedure in SAS v9.2.1.

CHAPTER 5

PERCENTILES CURVES IN A LOGISTIC MODEL WITH RANDOM EFFECTS

Several books and papers about binary data longitudinal modeling have interpreted and compared the estimates in marginal and random effects models [27, 35]. These works analyzed the relation between subject-specific and population-average estimation, as shown in Chapter 4. They have discussed the convenience of using the marginal, or random effects (or conditional) models in different situations. As an example, Lee and Nelder [27] considered that the principal distinction between marginal and conditional models has often been asserted to depend on whether the regression coefficients are going to describe an individual's response or the marginal response to changing covariates. In the same article, they state that they prefer the random-effect models over marginal models: "*the use of marginal models can be dangerous, even when marginal inferences are of interest*". The reasons to prefer the random-effects models are because they permit a simple marginal interpretation, and because ignoring important random effects may render invalid many traditional techniques of statistical analysis.

In this chapter we are not interested in comparing the estimates. Our interest is to propose a methodology to interpret subject-specific curves in a logistic model with random intercept in a longitudinal study based on a recent paper about heterogeneity in *GLMMs* [13]. Our proposal is to interpret the subject-specific curves across time as percentiles of a curve distribution that depend on the values of the random effects variance and the the covariates parameters. We will apply this result on two datasets.

5.1 Heterogeneity in a Logistic Model with Random Intercept

Consider a model for a binary response $(0, 1)$ in a longitudinal study with N subjects. The measurements were taken across time during n_i occasions, $i = 1, \dots, N$. Using a logistic model with random intercept, u_i , the conditional mean of the response variable Y_{ij} can be modeled as:

$$\text{logit} [P(Y_{ij} | \mathbf{u}_i)] = \text{logit} (\tilde{p}_{ij}) = x'_{ij}\tilde{\beta} + u_i \quad (5.1)$$

where x'_{ij} is the covariates vector for the i -th subject measured in the j -th occasion ($j = 1, 2, \dots, n_i$), and \mathbf{u}_i is a random variable normally distributed with zero mean and variance σ^2 . The probability of success depends on an unobserved, or latent variable related to each subject u_i , and on the respective values of the covariates x'_{ij} . Some models consider σ , the standard deviation of the random intercept, as a function of the covariates (i.e. treatment). This assumption implies that the variance of u_i is different for each level of the treatment in the study.

We will study the effect of the variability between subjects over the subject-specific probabilities of success across time. Applying to longitudinal data the idea presented by Duchateau and Janssen [13] about heterogeneity in *GLMMs* for clustered data, Equation 5.1 implies that:

$$\tilde{p}_{ij} = P(Y_{ij} | \mathbf{u}_i) = \varphi(\mathbf{u}_i) \quad (5.2)$$

where $\varphi(\cdot)$ is a known function obtained by solving for \tilde{p}_{ij} in Equation 5.1. $\varphi(\cdot)$ corresponds to the inverse link function. In the logistic case, $\varphi(\cdot)$ is the inverse logit function. Because the random effects are unobservable in practice, we can suppress the dependence on the fixed parameters, and therefore, \tilde{p}_{ij} is a function of the random effects u_i , thus

$$\tilde{p}_{ij} = \frac{\exp(x'_{ij}\tilde{\beta} + u_i)}{1 + \exp(x'_{ij}\tilde{\beta} + u_i)} \quad (5.3)$$

Using a classical theorem of statistical inference about the distribution of a transformation of a random variable [8], the respective density function of \tilde{p} (omitting the subscript ij for simplicity) is given by [13]:

$$f_{\tilde{p}}(\tilde{p}) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} \left\{ \text{logit}(\tilde{p}) - x'_{ij}\tilde{\beta} \right\}^2 \right] \frac{1}{\tilde{p}(1-\tilde{p})} \quad (5.4)$$

where $\frac{1}{\tilde{p}(1-\tilde{p})}$, $\tilde{p} \neq 0, 1$, is the term related with the transformation gradient in Equation 5.3. This density is important because it permits to determine the effect of variability between subjects over the conditional mean of the response variable across time. In addition, this density function shows an interesting form to interpret the marginal mean of response variable induced by the random effects. The formulation of the model in Equation 5.1 is based on the conditional mean, but in many times, it is necessary to analyze the marginal mean of Y_{ij} , that is, $E(Y_{ij}) = p_{ij}$. It is important to notice that \tilde{p}_{ij} and p_{ij} are not the same. \tilde{p}_{ij} represents the subject-specific or conditional mean, and p_{ij} is the marginal mean induced by the random effects.

Based on Equation 5.4 we can find, $F_{\tilde{p}}(\tilde{p})$, the cumulative distribution of \tilde{p} ,

$$F_{\tilde{p}}(t) = \int_0^t \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} \left\{ \text{logit}(\tilde{p}) - x'_{ij}\tilde{\beta} \right\}^2 \right] \frac{1}{\tilde{p}(1-\tilde{p})} dt \quad (5.5)$$

for some $0 < t < 1$. This integral must be solved using numerical methods.

5.1.1 Density Function and Cumulative Distribution Function of \tilde{p}

To illustrate the advantages of knowing the density function and cumulative distribution of \tilde{p} in a logistic model with random intercept, we will analyze some density functions, and cumulative distributions, across time by considering the following simple model for a longitudinal study with t repeated measures ($t = 1, 2, \dots, 15$):

$$\text{logit}(\tilde{p}_{ij}) = \tilde{\beta}_0 + \tilde{\beta}_1 \cdot t_{ij} + u_i \quad (5.6)$$

where $u_i \sim N(0, \sigma^2)$. σ takes different values ($\sigma = 0.15, 0.6$ and 1.5), $\tilde{\beta}_0 = -3$, and $\tilde{\beta}_1 = 0.5$, both arbitrary values. The σ values correspond to 5%, 20% and 50% of

the $|\beta_0|$ value. Also, these values imply that the odds ratio between two subjects with the same covariates values, differing by one unit in \mathbf{u} , are equal to 1.2, 1.8, and 4.5, respectively.

Figures 5–1, 5–2 and 5–3 show the density functions of \tilde{p} for the hypothetical model presented in Equation 5.6 for $t \leq 10$ ¹. The main advantage of knowing this density is that we can observe where the conditional probabilities are concentrated across time, and also, we can determine how large is the variability between the subject-specific probabilities. In practice this result is important because it permits to know how the subjects are responding as a function of the time and others covariates. Clearly, when σ takes a small value, the density function of \tilde{p} is concentrate around a specific value that corresponds to the marginal mean induced by the random intercept. In the literature this mean is known as the *induced marginal mean* [32, 35]. In contrast, when σ is large, the density of \tilde{p} is skewed towards the left or right, with certain concentration around zero or one, depending on the status response. Therefore, the variability between subjects determines the shape of the conditional probabilities distribution. Similarly, Figures 5–4, 5–5 and 5–6 show the cumulative distribution of \tilde{p} . Clearly, the cumulative distribution has a significative change around certain value (induced marginal mean) when σ is small. However, as σ increases, the cumulative distribution raises more slowly to 1. These curves are important in order to define the percentiles curves in these models.

5.1.2 Mean and Variance of \tilde{p}

An interesting question with respect to the previous distributions is related with the parameters of the conditional probabilities distribution: the mean and variance of \tilde{p} .

¹ These plots were generated using the `coplot` function from the `graphics` R library.

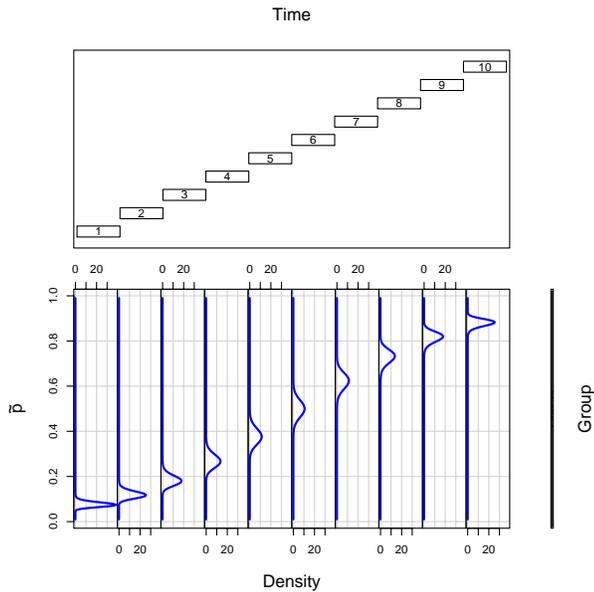


Figure 5–1: Density of \tilde{p} across time for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 0.15$

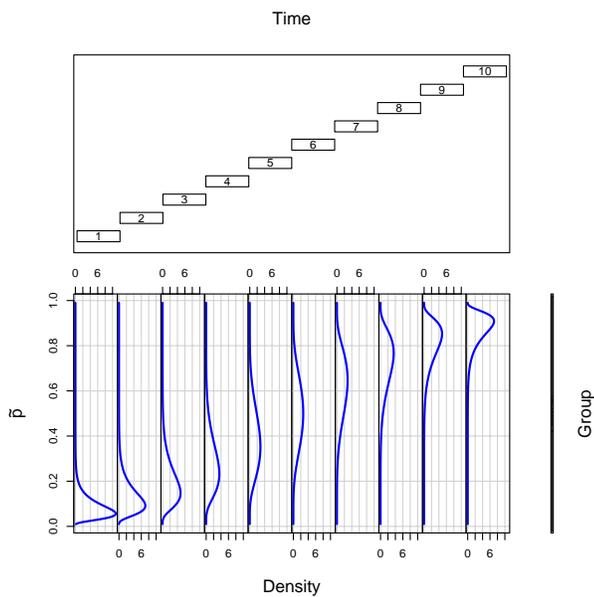


Figure 5–2: Density of \tilde{p} across time for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 0.6$

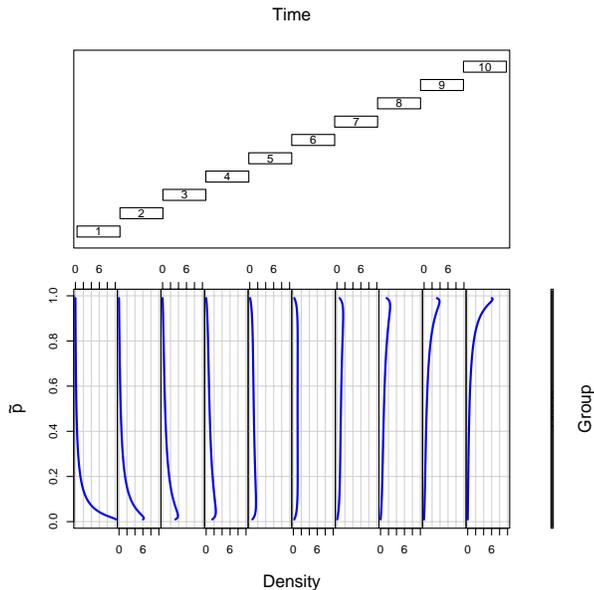


Figure 5–3: Density of \tilde{p} across time for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 1.5$

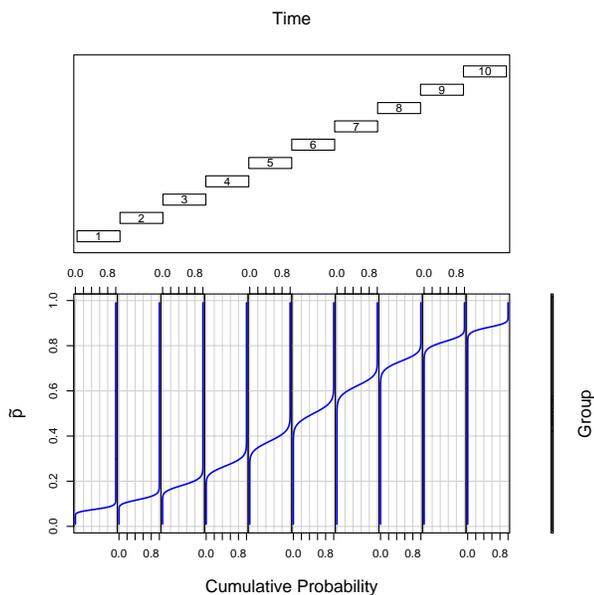


Figure 5–4: Cumulative Distribution of \tilde{p} for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 0.15$

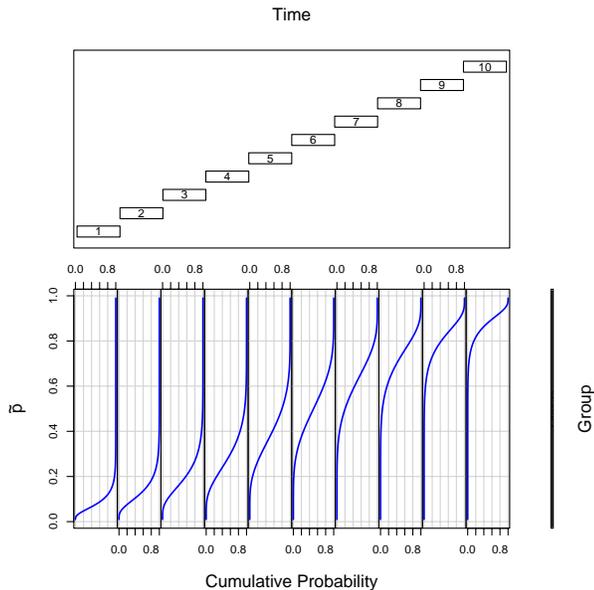


Figure 5–5: Cumulative Distribution of \tilde{p} for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 0.6$

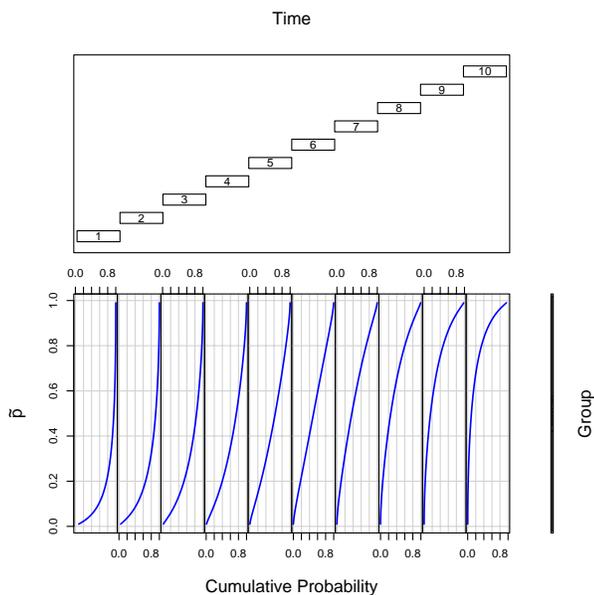


Figure 5–6: Cumulative Distribution of \tilde{p} for Logistic Model with Random Intercept (Eq. 5.1) and $\sigma = 1.5$

Based on the basic concepts of inference, the mean and variance of \tilde{p} are defined by the following equations, respectively:

$$E(\tilde{p}) = \int_0^1 \tilde{p} \cdot f_{\tilde{p}}(\tilde{p}) d\tilde{p} \quad (5.7)$$

and,

$$\text{Var}(\tilde{p}) = \int_0^1 \tilde{p}^2 \cdot f_{\tilde{p}}(\tilde{p}) d\tilde{p} - \left[\int_0^1 \tilde{p} \cdot f_{\tilde{p}}(\tilde{p}) d\tilde{p} \right]^2 \quad (5.8)$$

Since Y_{ij} is binary, it has a marginal Bernoulli distribution with mean $E(Y_{ij}) = p$ given by Equation 5.7. Its variance is therefore, $\text{Var}(Y_{ij}) = E(Y_{ij})[1 - E(Y_{ij})] = p(1 - p)$. $E(\tilde{p})$ represents the induced marginal mean, and $\text{Var}(\tilde{p})$ indicates the variability between subject-specific probabilities. $\text{Var}(\tilde{p})$ has only a theoretical interest because in practice this statistic can be inappropriate to measure the heterogeneity between the subject-specific probabilities, especially when the random effects have a large variance.

In the previous discussion, $E(\tilde{p})$ and $\text{Var}(\tilde{p})$ are the first and second population moments of the distribution of \tilde{p} only if the β values are known. Nevertheless, and in practice the β values are unknown, and must be estimated. Therefore, we define $E(\hat{\tilde{p}}) = \hat{p}$ as the estimated first moment when these are calculated by using $\hat{\beta}$. The estimated marginal mean induced by the random effect is \hat{p} .

Following with the hypothetical model presented in Section 5.1, the mean and variance of \tilde{p} for each time t , and each σ value are given by:

$$E(\tilde{p}) = \int_0^1 \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{1}{2\sigma^2} \left\{ \text{logit}(\tilde{p}) - (-3 + 0.5 \cdot t) \right\}^2 \right] \frac{1}{(1 - \tilde{p})} d\tilde{p}$$

and,

$$\text{Var}(\tilde{p}) = \int_0^1 \tilde{p} \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp[-B] \frac{1}{(1 - \tilde{p})} d\tilde{p} - \left[\int_0^1 \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp[-B] \frac{1}{(1 - \tilde{p})} d\tilde{p} \right]^2$$

where $B = \frac{1}{2\sigma^2} \left\{ \text{logit}(\tilde{p}) - (-3 + 0.5 \cdot t) \right\}^2$

We calculated the induced marginal mean and the variance of \tilde{p} for each σ value using the `integrate` function from `stats R` library as shown in Figures 5–7 and 5–8, respectively². The variance of \tilde{p} is almost zero when $\sigma = 0.1$, and it is larger for $\sigma = 0.5$ and $\sigma = 2$ around $t = 6$. This result is reasonable in several practical situations because the variability between subjects is larger in the central times of the study. In contrast, in the initial and final times the variability is smaller because the subjects present similar responses. Also, it can occur that the variance will be larger in the baseline and it will decrease across time, or conversely. However, the $\text{Var}(\tilde{p})$ does not provide a good quantification of the heterogeneity between \tilde{p} because the distribution of \tilde{p} is skewed [25], as shown in Figure 5–2. Hence, we will not focus on this dispersion measure in the future applications.

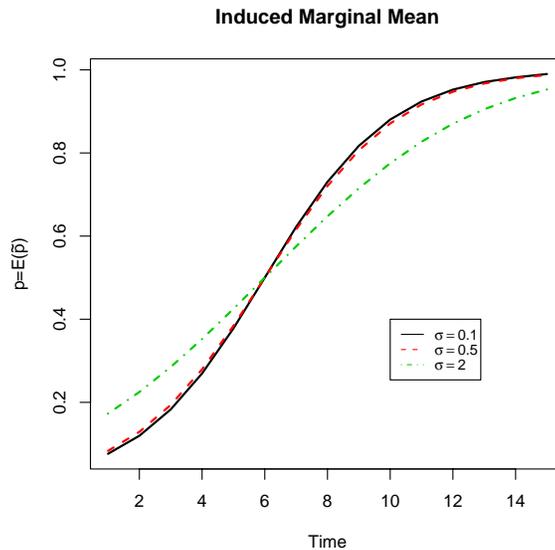


Figure 5–7: Induced Marginal Mean for the model defined in Equation (5.6)

In the previous results, the induced marginal mean and variance were calculated using the definition and the density function, $f_{\tilde{p}}(\tilde{p})$. Nevertheless, these estimates

² The approximation errors in these integrals are around 10^{-5} .

can be calculated directly using some results presented in Chapter 4. Note that the Equation 5.7 can be rewritten in terms of u_i as:

$$p = E_{\mathbf{u}}[E(Y_{ij} | \mathbf{u}_i)].$$

Hence,

$$p = \int_{-\infty}^{\infty} \tilde{p} \cdot f_{\mathbf{u}_i}(u_i) d\mathbf{u}_i \quad (5.9)$$

Rewriting Equation 5.9 in terms of the model presented in Equation 5.6, we obtain that

$$p = \int_{-\infty}^{\infty} \frac{\exp(-3 + 0.5 \cdot t + u_i)}{1 + \exp(-3 + 0.5 \cdot t + u_i)} \cdot f_{\mathbf{u}_i}(u_i) d\mathbf{u}_i \quad (5.10)$$

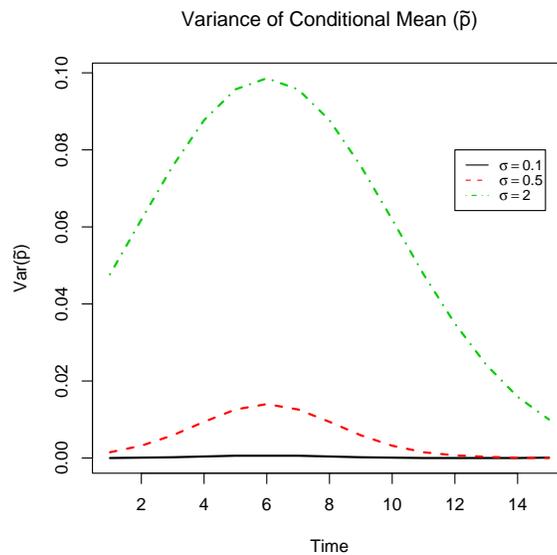


Figure 5–8: Variance of \tilde{p} for the model defined in Equation (5.6)

The equation 5.10 does not have a closed form and it is necessary to use numerical integration or another method to approximate it. To calculate this expression we use the `phermite` and `gaussh.int` R functions which can be found in *Appendix A*. These results are the same than the ones obtained with the `integrate` R function. These integrals also can be calculated in SAS PROC NL MIXED.

5.2 An Interpretation of Percentile Curves

The density function presented in Equation 5.4 permits to determine the influence of random effects on the conditional probabilities distribution. Given a random effect value, u_i , and certain values covariates, they produce a point estimate of the probability for the i – th subject, which corresponds to one observation of the distribution of \tilde{p} , as shown in Section 5.1.1. In addition, for each combination of the covariates, we can calculate the marginal mean by integrating out the random effects in the model. All these results motivated us to study the subject-specific curves across time, and interpret them. We will call these subject-specific curves *Subject-specific Curves Distribution*.

Suppose that u_q determines the q - th percentile of the random intercept distribution. Our objective is to prove that, under certain assumptions in the model, this u_q is equivalent to the q – th percentile of the *subject-specific curves distribution* induced by the random effects. Our proposal is the following:

Proposition 1 (Percentiles of the subject-specific Curves Distribution). *Let us consider the following logistic model with random intercept:*

$$\text{logit} [E(Y_{ij}) \mid \mathbf{u}_i] = x'_{ij}\tilde{\beta} + u_i \quad (5.11)$$

where $u_i \sim N(0, \sigma^2)$. Then the q – th percentile of \mathbf{u} is equivalent to the q – th percentile of the *subject-specific curves distribution*. The subject-specific curve C_q generated with u_q is called the q – th **Percentile Curve**. Particulary, for $u_{50} = 0$, the median of the random effect, the curve is called the **Median Curve** or **Typical Curve** and it is denoted by C_{med} . The curve generated by integrating out the random effects is called the **Marginal Mean Curve** and it is denoted by C_{mm} .

The proof of this *Proposition* is based on the properties of the link functions and the next classical theorem [8].

Theorem 1 (Distribution of a Transformation of Random Variables). *Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $D_X = \{x : f_X(x) > 0\}$ and $D_Y = \{y : f_Y(y) > 0\}$ then:*

- a. *If g is an increasing function on D_X , $F_Y(y) = F_X(g^{-1}(y))$ for $y \in D_Y$.*
- b. *If g is an decreasing function on D_X , $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in D_Y$.*

Proof. To prove the previous [Proposition 1](#), let u_q be the q -th percentile of the normal distribution $N(0, \sigma^2)$, with cumulative distribution $\Phi(\cdot)$. According to previous results,

$$\tilde{p}_{ij} = \text{logit}^{-1}(x'_{ij}\tilde{\beta} + u_i) = \frac{\exp(x'_{ij}\tilde{\beta} + u_i)}{1 + \exp(x'_{ij}\tilde{\beta} + u_i)}$$

Consider $X = u_i$ and $Y = \text{logit}^{-1}(u_i)$ in [Theorem 1](#). It is sufficient to prove that the function logit^{-1} is increasing with respect to u_i to prove the proposition:

$$\frac{\partial \tilde{p}_{ij}}{\partial u_i} = \frac{\exp(x'_{ij}\tilde{\beta} + u_i)}{[1 + \exp(x'_{ij}\tilde{\beta} + u_i)]^2} > 0$$

Therefore, $\text{logit}^{-1}(u_i)$ is an increasing function of u_i . [Theorem 1](#) permits to derive:

$$\begin{aligned} F_{\tilde{p}_{ij}(\tilde{p}_{ij})} &= P(\tilde{p}_{ij} \leq c) \\ &= P[\tilde{p}_{ij} \leq ([\text{logit}^{-1}(c) - x'_{ij}\tilde{\beta}]) \\ &= \Phi[\text{logit}^{-1}(c) - x'_{ij}\tilde{\beta}] \end{aligned}$$

which represents the q -th percentile curve, C_q . □

Hence, we prove that the q -th percentile of the random effects is equivalent to the q -th percentile curve of the subject-specific curves distribution. In other words, [Proposition 1](#) indicates that the logit transformation preserves the percentiles of the random effects. [Figure 5–10](#) illustrates this result. The previous proposition is valid for any link function, because these functions must be increasing [\[2\]](#).

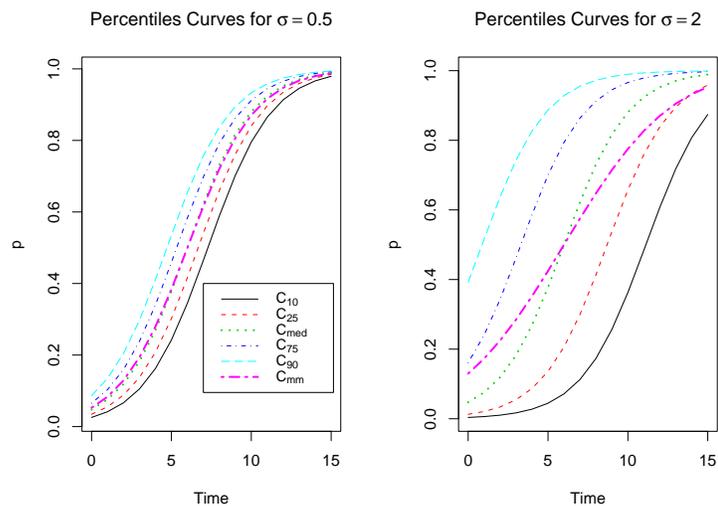


Figure 5–9: Percentiles Curves for the Model presented in Equation 5.6 with $\sigma = 0.5$ and $\sigma = 2$

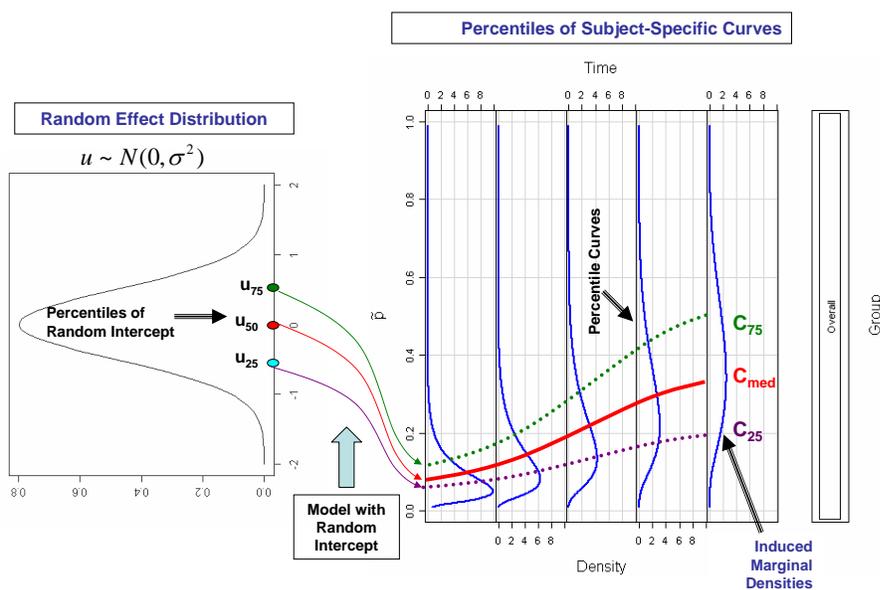


Figure 5–10: Percentiles of the Random Intercept Distribution, Densities of \tilde{p} , and Percentiles of Subject-Specified Curves for a Model with Random Intercept

To illustrate the previous proposition, consider the model defined in Equation 5.6 with $\sigma = 0.2$, and $\sigma = 0.5$. The subject-specific curves are determined by:

$$\tilde{p} = \frac{\exp(-3 + 0.5 \cdot t + u_q)}{1 + \exp(-3 + 0.5 \cdot t + u_q)}$$

where the u_q are the q -th percentiles of the normal distribution $N(0, \sigma^2)$. Figure 5-9 shows the percentiles curves for two hypothetic scenarios considering $q = 10, 25, 50, 75$ and 90 , that is, $u_q \approx (-1.2815, -0.6744, 0, 0.6745, 1.2815)$. Two important results can be extracted from these plots. First, larger values of σ disperse the curves more. This is consistent with the results about the density and cumulative distribution functions presented in Section 5.1.1. Second, when $\sigma = 0.5$ the curve with $u_{50} = 0$, C_{med} , and the marginal mean curve, C_{mm} , are similar. However, this is not true for $\sigma = 2$. In this situation the median, or typical curve, is different from the marginal mean curve. These plots could be useful to the researchers because they provide the mean and median subject-specific curves, and any other percentile of interest in the study. Under the percentiles curves interpretation, it is possible to determine what subjects have an abnormal response to the treatment, or what subjects are outside certain ranges of response probabilities. Thus, we can expand the percentiles curves concept used for continuous responses (i.e. anthropometry) to binary responses.

5.3 Application to Real Datasets

In this section we apply the percentiles curves concept presented previously to two datasets. We will describe the datasets used and we will show the respective results for each application. The estimates of the models were obtained by using SAS PROC NL MIXED [24].

5.3.1 Toenail Dataset

This data was obtained from a randomized, double-blind, parallel group, multicenter study for the comparison of two oral treatments (A and B , coded as 0 and

1, respectively) for toenail dermatophyte onychomycosis (TDO)³. The aim of the study was to compare the efficacy and safety of 12 weeks of continuous therapy with treatment *A*, or with treatment *B*. Subjects were followed during 3 months of treatment, and followed further up to a total of 12 months. Measurements were taken at baseline, every month during treatment, and every three months afterwards, resulting in a maximum of 7 measurements by subject. The severity of the infection was coded as 0 (not severe), or 1 (severe). The sample size is $N = 294$ patients. The variables in the dataset are: *id*, *response (0/1)*, *treat (0/1)*, *time (continuous) of each visit (1 thru 7)*. Table 5–1 and Figure 5–11 show the proportion of patients with severe Toenail infection for each treatment and visit.

Table 5–1: Proportion of Infected Patients, *Toenail Dataset*

<i>Visit</i>	<i>Time in Days,(min – max)</i>	<i>Treatment A</i>	<i>Treatment B</i>
1 (Baseline)	(0.0-0.0)	0.3699	0.3716
2	(0.7-1.6)	0.3475	0.3265
3	(1.5-3.5)	0.3188	0.2759
4	(2.5-4.8)	0.2197	0.2071
5	(4.0-8.8)	0.1077	0.0602
6	(5.3-12.5)	0.0855	0.0630
7	(10.5-18.5)	0.1053	0.0458

Model Equation: In order to analyze these data, we consider a model without time/treatment interaction:

$$\text{logit} [P(Y_{ij} | \mathbf{u}_i)] = \tilde{\beta}_0 + \tilde{\beta}_1 \cdot \text{time} + \tilde{\beta}_2 \cdot \text{treat} + u_i \quad (5.12)$$

where u_i is the random intercept distributed normally with zero mean and variance σ^2 , and $Y_{ij} | \mathbf{u}_i \sim \text{Bernoulli}(\tilde{p}_{ij})$. Note that we assume that the random intercept variance is constant throughout the environments.

³ Source: Dermatological Data for Statistical Research, Novartis, Belgium, <http://www.blackwellpublishers.co.uk/rss/>

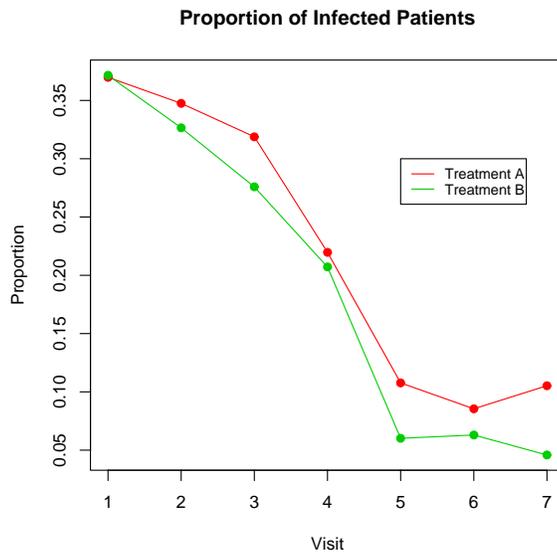


Figure 5–11: Proportion of Infected Patients for each treatment, *Toenail Dataset*

Results:

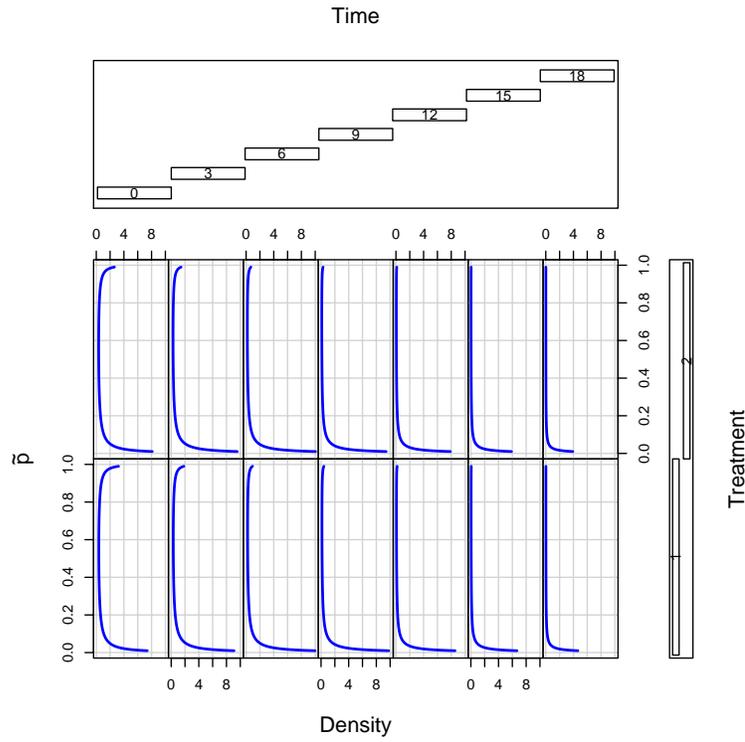
Table 5–2 shows the principal information on the fitted model for the *Toenail* dataset, and the parameter estimates are presented in Table 5–3. The log-likelihood for this model is equal to $\log L = -627.5$. The maximum repeated measurements by subject are 7, and 30 quadrature points were used to approximate the likelihood.

Table 5–2: Dimensions and Fit Statistics for model presented in Equation 5.12, *Toenail Dataset*

Dimensions	
<i>Aspect</i>	<i>Dimension</i>
Total Observations	1908
Subjects	294
Max Obs Per Subject	7
Parameters	4
Quadrature Points	30
Fit Statistics	
<i>Statistic</i>	<i>Value</i>
-2 Log Likelihood	1255.0
BIC	1277.7

Table 5–3: Estimates for model presented in Equation 5.12, *Toenail Dataset*

Parameter	Estimate	Standard Error	t-value	DF	P-value
$\tilde{\beta}_0$	-1.4470	0.4251	293	-3.40	0.0008
$\tilde{\beta}_1$	-0.4516	0.03669	293	-12.31	<.0001
$\tilde{\beta}_2$	-0.5225	0.5575	293	-0.94	0.3494
σ	4.0089	0.3819	293	10.50	< .0001

Figure 5–12: Density Function of \tilde{p} Induced by the Random Intercept, *Toenail Dataset*Table 5–4: Induced Marginal Mean, *Toenail Dataset*

<i>Time (Days)</i>	<i>Treat A</i>	<i>Treat B</i>
0	0.3707	0.3266
3	0.2614	0.2242
6	0.1717	0.1432
9	0.1047	0.0848
12	0.0591	0.0465
15	0.0309	0.0236
18	0.0149	0.0110

Instead of speaking about the estimates properties (i.e. significance), we are interested in plotting the estimated density function and cumulative distribution of

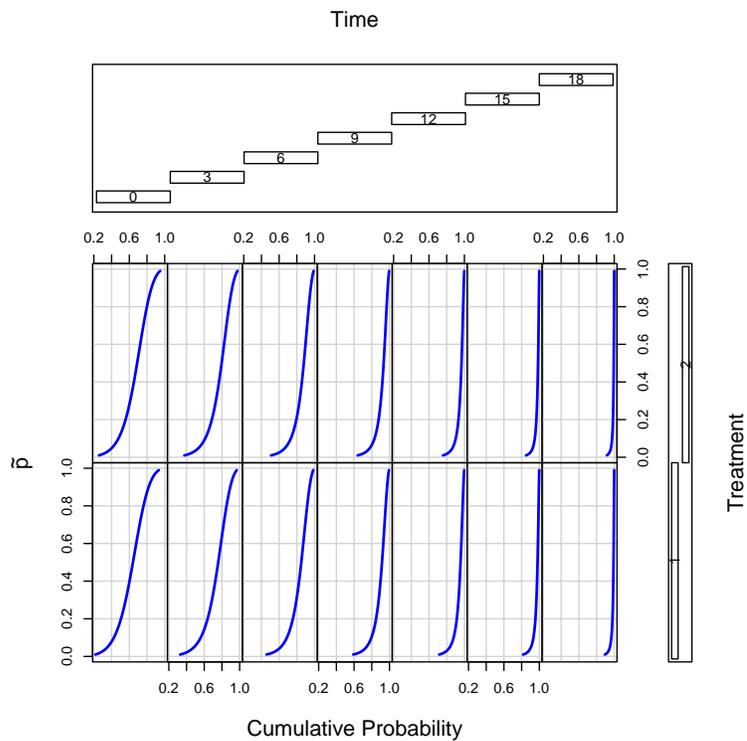


Figure 5–13: Cumulative Distribution of \hat{p} Induced by the Random Intercept, *Toenail Dataset*

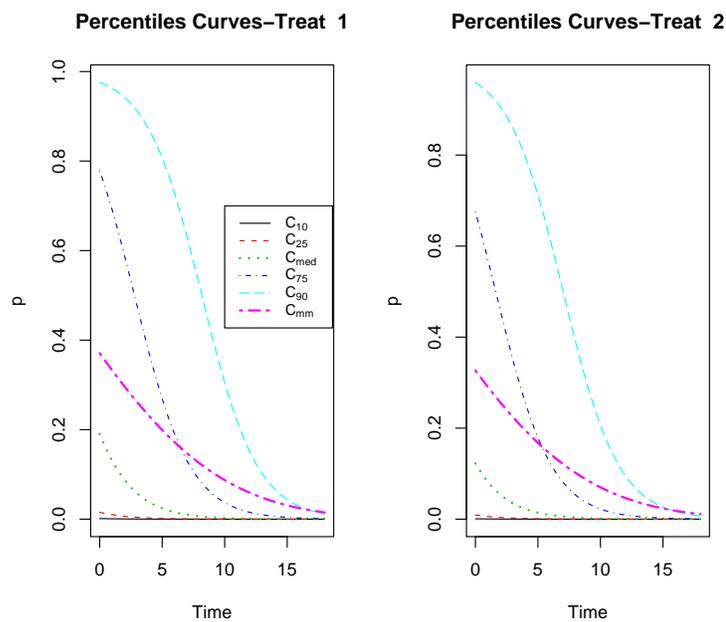


Figure 5–14: Percentiles Curves, *Toenail Dataset*

the conditional probabilities of Toenail infection in several times, as shown in Figures 5–12 and 5–13. In this case, the estimated standard error of random effect is equal to $\hat{\sigma} = 4.0089$. The density function of \tilde{p} at the beginning of the study is concentrated around $\tilde{p} = 0$ and $\tilde{p} = 1$, where the density is larger. This shape indicates a bimodal distribution. In contrast, it is more concentrated around $\tilde{p} = 0$ when time is near 18, which indicates a unimodal distribution. This behavior is stronger in treatment 2 because the conditional probabilities tend faster towards zero. In the initial times the cumulative function raises slowly to 1, whereas in final times it begins almost at 1 for both treatments.

Table 5–4 shows the induced marginal mean \hat{p} for seven specific time-points. It indicates the probability of infection for each treatment and time. These values have been calculated averaging the conditional probabilities over the random effects for each time and treatment.

The σ value for Toenail dataset induces disperse percentiles curves because of its magnitude. Also, it is possible that specific percentiles curves be different between treatments. For example, the C_{75} could be different between both treatments. Note that the median curve, or typical curve, differs notably from the marginal mean curve. In addition, all percentiles curves, except C_{90} , are similar between treatments. The C_{90} percentile curve in treatment 2 tends faster towards zero than the same curve in treatment 1.

Another important aspect about the percentiles curves consist in the possibility of calculating a $(1 - \alpha)100\%$ confidence interval for the infection subject-specific probabilities, \tilde{p} , which are given by the percentiles curves:

$$[C_{(\alpha/2)100\%}, C_{(1-\alpha/2)100\%}]$$

For example, a 95% confidence interval on \tilde{p} for $t = 8$ is given by $[0, 0.5192]$ and $[0, 0.3904]$ for treatment A and B, respectively. In addition, we can calculate the specific time in which there is a proportion of infected patients, for example, 50%.

5.3.2 Garlic Dataset

This data was obtained from a study about the epidemiology of White Rot in Garlic from five different environments⁴. Each environment is defined by the combination of two regions, *Jesús María* and *Cruz del Eje* in the *Córdoba* province (Argentina), and its years of study⁵. In *Cruz del Eje* region, the experimental field was divided in ten blocks. In each block six sampling points (*sites*) were randomly selected. Measurements of incidence over a total of fifty plants in each point (number of dead plants/total number of plants) were taken every t days, where t varies according to the environment. In *Jesús María* region the sampling design was similar, but five sampling points were selected instead of six. The objective is to characterize the progress curves of the infection. The variables in the dataset are: *year*, *region*, *site*, *block*, *incidence*, and *time (days)*. Figure 5–15 shows the incidences for each environment and block/site. The percent of dead plants varies notably between block and site combination and environments.

To analyze this data, we propose two models for each environment. In order to apply the proposed percentiles curves, each model is a logistic regression with random intercept. The differences between these models are based on the assumptions about the random intercept.

- **Model 1: Random Intercept with Constant Variance**

First, we define each combination of block and site as the analysis unit because

⁴ Source: Martha Conless, Universidad Nacional de Córdoba, Argentina.

⁵ The years of study were 2001 – 2003 for Cruz del Eje and 2001 – 2002 for Jesús María. They are denoted by *CE1*, *CE2*, *CE3*, *JM1*, and *JM2*, respectively.

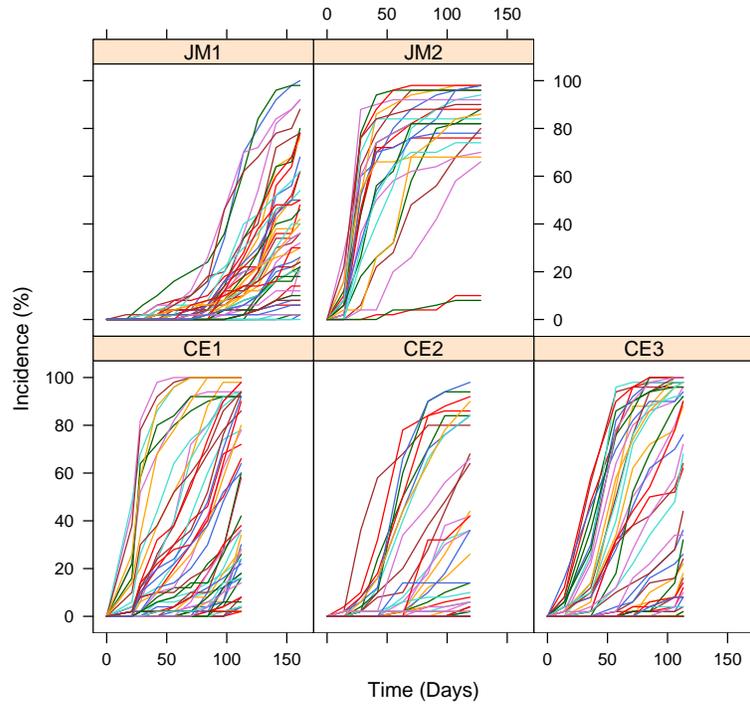


Figure 5–15: Infection Progress Curves (Incidence %) for the five environments in the white rot experiment, *Garlic Dataset*

the repeated measurements are taken over these. Suppose that Y_{ij} is the j – th measurement of incidence in the i – th block/site combination . Given the random intercept, u_i , the response in the k – th environment, $k = 1, \dots, 5$, has a Binomial distribution: $Y_{ijk} | \mathbf{u}_i \sim \text{Binomial}(50, \tilde{p}_{ijk})$. The equation model for each environment is given by:

$$\text{logit} [P(Y_{ijk} | \mathbf{u}_{ik})] = \tilde{\beta}_{0k} + \tilde{\beta}_{1k} \cdot \text{time} + u_{ik} \quad (5.13)$$

where $\mathbf{u}_{ik} \sim N(0, \sigma^2)$. The $time$ variable was transformed to $time/10$.

- **Model 2: Variance of Random Intercept varying across Environment**

In this case, the equation model for each environment is given by:

$$\text{logit} [P(Y_{ij} | \mathbf{u}_{ik})] = \tilde{\beta}_{0k} + \tilde{\beta}_{1k} \cdot \text{time} + u_{ik} \quad (5.14)$$

where $\mathbf{u}_{ik} \sim N(0, \sigma_k^2)$.

Results for Model 1: In order to plot the percentiles curves, we fitted *Model 1*.

Table 5–5 shows the main information on the fitted *Model 1*. The log-likelihood for this model is equal to $\log L(M1) = 4937.75$. The maximum repeated measurements of incidence in each block/site combination is 13, and 30 quadrature points were used to approximate the likelihood.

Table 5–5: Dimensions and Fit Statistics (*Model 1*), *Garlic Dataset*

Dimensions	
<i>Aspect</i>	<i>Dimension</i>
Observations Used	2375
Total Observations	2375
Subjects	255
Max Obs Per Subject	13
Parameters	11
Quadrature Points	30
Fit Statistics	
<i>Statistic</i>	<i>Value</i>
-2 Log Likelihood	9785.5
BIC	9846.5

Table 5–6 shows the estimates for fixed effects and the standard deviation of the random intercept. The estimates of the intercept and the slope, $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively, differ across environments. The estimated standard deviation of the random effect is equal to $\hat{\sigma} = 3.0999$, which is the same in all the environments. Figures 5–16 shows the estimated density functions for the five environments. Similar to the incidence plots, the estimated density functions and cumulative distribution between environments differ notably. For example, the similarity between the densities of Environment 1 and Environment 3 implies that the conditional probabilities across time are similar. In contrast, Environment 5 is different from the rest.

How can we compare these densities with the incidences in each environment? A direct form is to compare the densities with the histogram of incidences for each environment and specific times, such as shown in Figure 5–18. Note that the shape of the histograms are similar than the shape of the densities

Table 5–6: Estimates for *Model 1, Garlic Dataset*

Parameter	Estimate	Standard Error	t-value	DF	P-value
Environment 1: Cruz del Eje 2001					
$\tilde{\beta}_0$	-6.0418	0.4095	254	-14.76	< .0001
$\tilde{\beta}_1$	0.5534	0.009426	254	58.71	< .0001
Environment 2: Cruz del Eje 2002					
$\tilde{\beta}_0$	-9.0290	0.4556	254	-19.82	< .0001
$\tilde{\beta}_1$	0.5051	0.01216	254	41.53	< .0001
Environment 3: Cruz del Eje 2003					
$\tilde{\beta}_0$	-8.0871	0.4292	254	-18.84	< .0001
$\tilde{\beta}_1$	0.6497	0.01181	254	55.00	< .0001
Environment 4: Jesus María 2001					
$\tilde{\beta}_0$	-7.3829	0.4513	254	-16.36	< .0001
$\tilde{\beta}_1$	0.4256	0.007210	254	59.03	< .0001
Environment 5: Jesus María 2002					
$\tilde{\beta}_0$	2.3270	0.6222	254	-3.74	0.0002
$\tilde{\beta}_1$	0.4049	0.008171	254	49.55	< .0001
Standard Deviation of Random Intercept-Overall					
σ	3.0999	0.1636	254	19.07	< .0001

around the same times, which is an indicator of adequate fit. This similarity offers an interesting tool to analyze the properties of the incidence distribution.

Similarly, the cumulative distributions permit us to know how the infection probability evolves. For example, in Environment 5 the cumulative distribution raises slower to 1 as t increases. This indicates that for larger t , small incidences have a small probability, whereas a conditional probability of infection near 1 has a larger probability. This behavior means that the probability of a large incidence when t is large is more probable in the Environment 5 than in the rest, as can be observed in the histograms.

Finally, Figure 5–19 shows the percentiles curves for all environments. Graphically, the induced marginal incidence in Environment 1 and Environment 3 are similar. Likewise, the marginal incidence in Environment 2 and Environment 4 are similar. Environment 5 presents a marginal incidence different to the rest. Nevertheless, the fact that the marginal incidence is similar between certain environments, does not imply that the percentiles curves also are similar. The plots clearly show

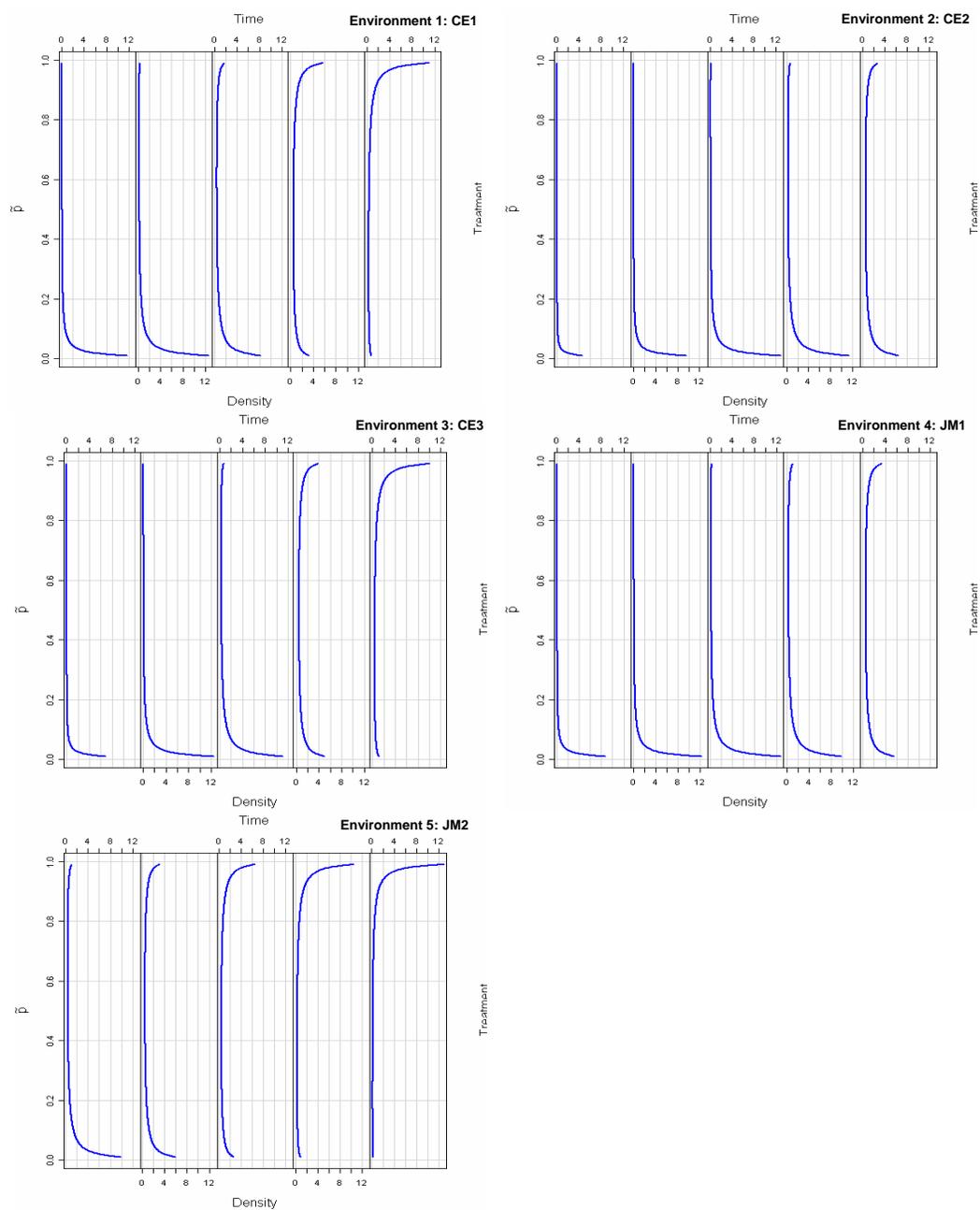


Figure 5–16: Density Functions of \tilde{p} for $t = 0, 40, 80, 120$ and 160 days (*Model 1*), *Garlic Dataset* [Each panel inside the plot represents a time-point]

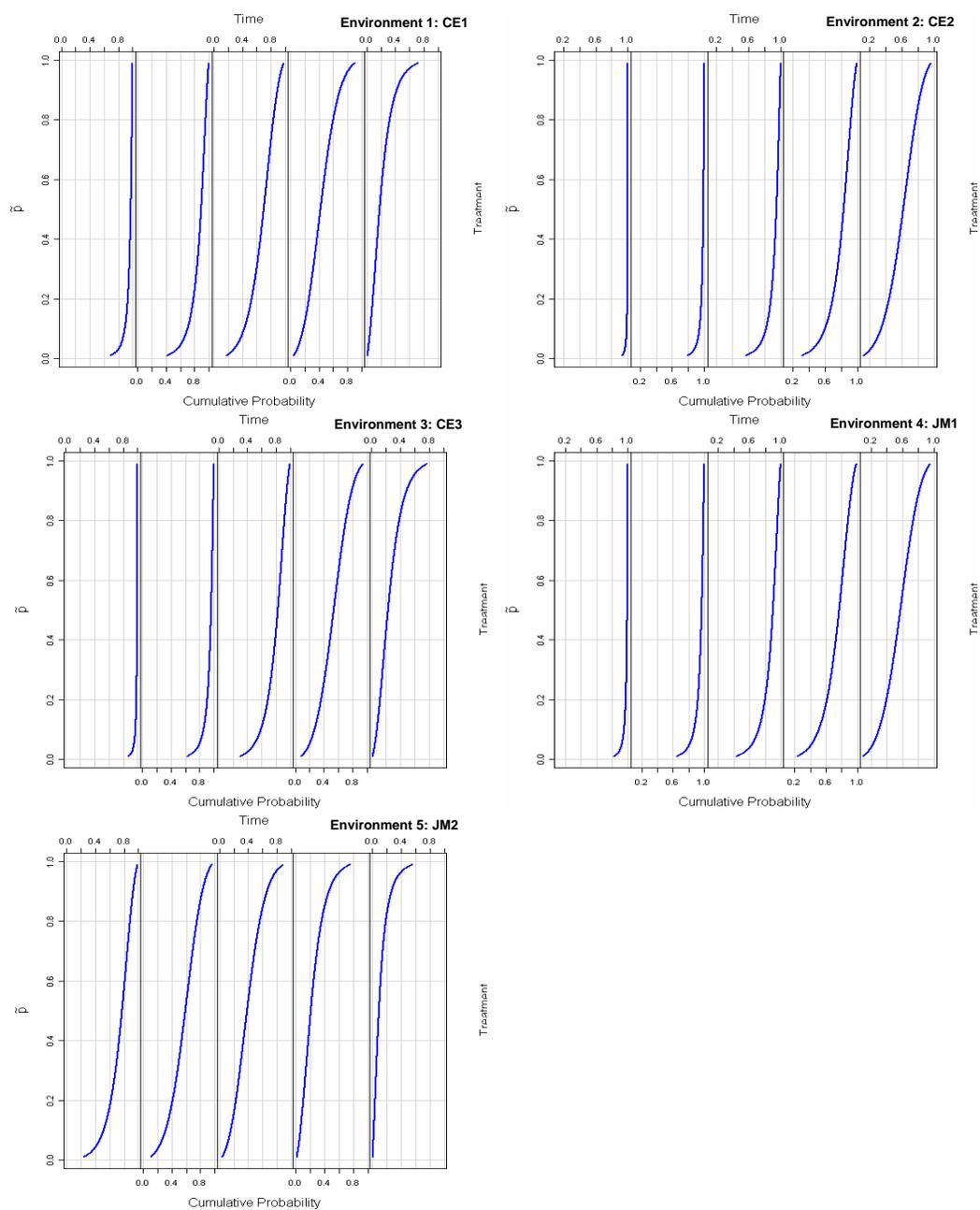


Figure 5–17: Cumulative Distribution of \tilde{p} for $t = 0, 40, 80, 120$ and 160 days (*Model 1*), *Garlic Dataset* [*Each panel inside the plot represents a time-point*]

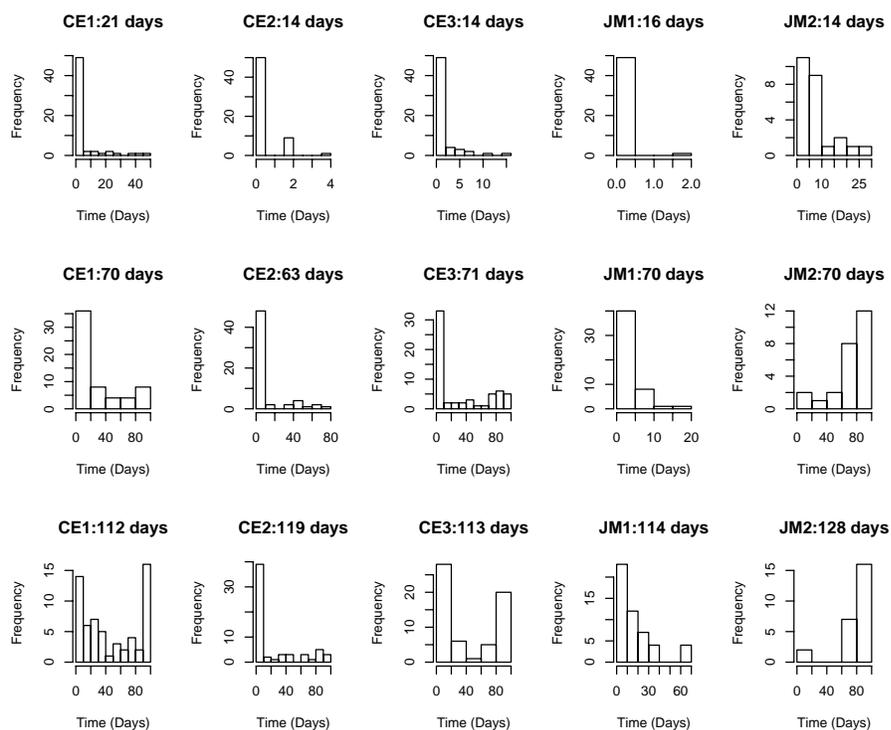


Figure 5–18: Histogram of Incidence in each Environment and Specific Time, *Garlic Dataset*

that the percentiles curves differ notably between environments, even though the marginal incidence among them is similar.

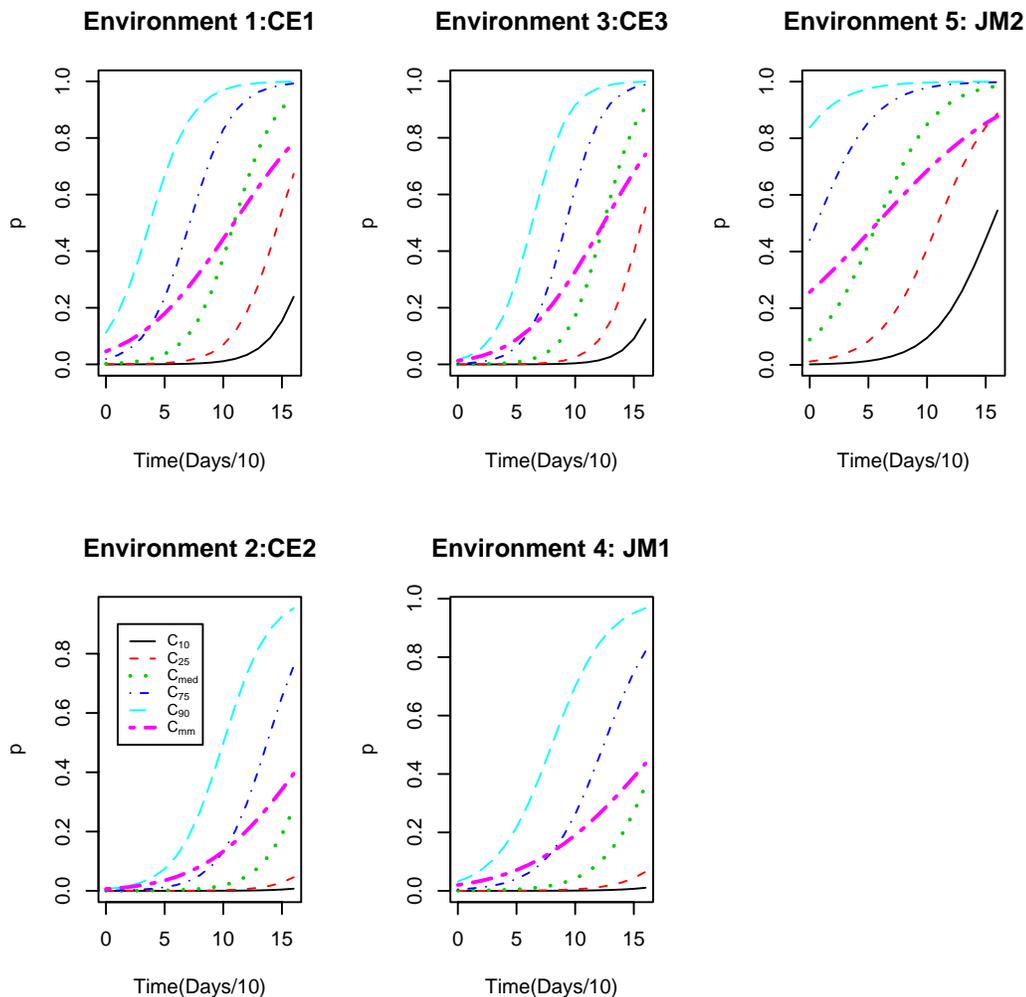


Figure 5–19: Percentile Curves (*Model 1*), *Garlic Dataset*

The marginal infection rates across time based on *Model 2* are shown in Table 5–7. These values represent the marginal mean⁶ incidences, \hat{p} , and permit to compare the incidences of garlic white rot among environments at a specific time. For

⁶ The marginal mean of Y_{ij} can be easily calculated because $Y_{ij} \sim \text{Binomial}(50, \hat{p}_{ij})$. Then, $E(Y_{ij}) = 50 \cdot \hat{p}_{ij}$.

example, Environments *CE1*, *CE3*, and *JM2* present most of the plants that are infected for $t = 180$ days. In contrast, Environments *CE2* and *JM1* have almost 50% of infected plants. Also, there are noticeable differences in the incidences for $t = 90$ days. Environment *JM2* raises up to 64% of infected plants, meanwhile the other environments have 38% or less.

Table 5–7: Induced Marginal Mean Incidences (*Model 1*), *Garlic Dataset*

<i>Time (Days)</i>	<i>CE1</i>	<i>CE2</i>	<i>CE3</i>	<i>JM1</i>	<i>JM2</i>
0	0.0461	0.0064	0.0125	0.0202	0.2570
30	0.1101	0.0185	0.0436	0.0444	0.3774
60	0.2227	0.0472	0.1205	0.0885	0.5115
90	0.3829	0.1049	0.2649	0.1598	0.6442
120	0.5668	0.2027	0.4675	0.2616	0.7612
150	0.7369	0.3417	0.6792	0.3896	0.8530
180	0.8638	0.5070	0.8439	0.5311	0.9172

Results for Model 2:

Table 5–8 shows the main information about the fitted *Model 2*. The log-likelihood for this model is equal to $\log L(M2) = -4860.95$. The likelihood ratio between the two models is equal to $-2[\log L(M2) - \log L(M1)] = 153.6$, which is significant if it is compared with χ_4^2 (p -value < 0.0001). This indicates that this model has a better fit. This result is confirmed by the differences that exist between the estimates of the standard deviation of the random intercept, as shown in Table 5–9. For example, Environments *CE2* and *CE3* present a larger random standard deviation in comparison with Environment *JM2*. A simple standard deviation ratio among Environments *CE1* and *CE3* with Environment *JM2* is found to be 2.8 and 3.02 times, respectively. The larger the standard deviation the larger the variability among the individual progress curves of infection.

Figure 5–20 shows the densities of \tilde{p} based on *Model 2*. These densities are different compared to the ones obtained in *Model 1*. The most relevant differences can be observed in Environments *JM1* and *JME2*. In Environment *JM1* the density

Table 5–8: Dimensions and Fit Statistics (*Model 2*), *Garlic Dataset*

Dimensions	
<i>Aspect</i>	<i>Dimension</i>
Observations Used	2375
Total Observations	2375
Subjects	255
Max Obs Per Subject	13
Parameters	15
Quadrature Points	30
Fit Statistics	
<i>Statistic</i>	<i>Value</i>
-2 Log Likelihood	9721.9
BIC	9805.0

in $t = 160$ is bimodal in *Model 1*, meanwhile it is skewed to the right in *Model 2*. Likewise, the densities in Environment *JM2* show differences especially at $t = 0, 40$ and 80 days. For example, the density at $t = 80$ in *Model 2* is skewed to the left, meanwhile the one in *Model 1* is most concentrated around 0 and 1 (bimodal). The cumulative distributions also present the same differences, as shown in Figure 5–21. Note that Environments *JM1* and *JM2* have the smallest σ values, and it is smaller to the common one calculated in *Model 1* ($\hat{\sigma} = 3.0999$). These differences between the densities and cumulative distributions will permit us to analyze the differences in the percentiles curves among both Models.

Clearly the percentiles curves are different in both models. First, the marginal incidences induced by the random effects are different, as shown in Table 5–10. Second, the dispersion of the percentiles curves is smaller in Environments *JM1* and *JM2* for the *Model 2*, as indicated by the $\hat{\sigma}$ values on Table 5–9. In general, these percentiles curves explain better the progress infection curves and the variability among them in all the environments for the garlic white rot experiment.

Another important result in the percentiles curves is the time-point where the marginal mean and median incidence curves are equal. In this time-point, the skewness of the density function is switched to left or right depending on the positions of

both curves, as shown in Figure 5–22 for Environment *JM2*. In this case, the curves are equal in $t \approx 57$ days. Before $t = 57$ days, the density function is skewed to the right, and after this point it is skewed to the left. This time-point can be calculated using the following formula:

$$t = \frac{-\hat{\beta}_0}{\hat{\beta}_1}$$

This result can be used to characterize the density function of \tilde{p} . In practice, this time-point corresponds to the number of days where there is 50% of infected plants. Clearly, this value varies among environments.

Table 5–9: Estimates for *Model 2, Garlic Dataset*

Parameter	Estimate	Standard Error	t-value	DF	P-value
Environment 1: Cruz del Eje 2001					
$\tilde{\beta}_0$	-6.0218	0.3581	254	-16.82	< .0001
$\tilde{\beta}_1$	0.5524	0.009428	254	58.59	< .0001
σ_1	2.6909	0.2524	254	10.66	< .0001
Environment 2: Cruz del Eje 2002					
$\tilde{\beta}_0$	-9.5725	0.6620	254	-14.46	< .0001
$\tilde{\beta}_1$	0.5070	0.01222	254	41.48	< .0001
σ_2	4.3361	0.5829	254	7.44	< .0001
Environment 3: Cruz del Eje 2003					
$\tilde{\beta}_0$	-8.5011	0.6455	254	-13.17	< .0001
$\tilde{\beta}_1$	0.6535	0.01191	254	54.85	< .0001
σ_3	4.6502	0.5477	254	8.49	< .0001
Environment 4: Jesus María 2001					
$\tilde{\beta}_0$	-7.2929	0.2580	254	-28.27	< .0001
$\tilde{\beta}_1$	0.4244	0.007193	254	59.01	< .0001
σ_4	1.6710	0.1813	254	9.22	< .0001
Environment 5: Jesus María 2002					
$\tilde{\beta}_0$	-2.3056	0.3114	254	-7.40	0.0002
$\tilde{\beta}_1$	0.4033	0.008153	254	49.47	< .0001
σ_5	1.5348	0.2222	254	6.91	< .0001

5.4 Heterogeneity in a Logistic Model with Random Intercept and Slope

In the previous sections we talked about the percentiles curves in a logistic models with random intercept, and how this concept can be directly studied because the conditional probabilities are a increasing univariate function of the random intercept and therefore, these conserve the percentiles of the random effects distribution.

In this section, we will expand the percentiles curves concept to a logistic model

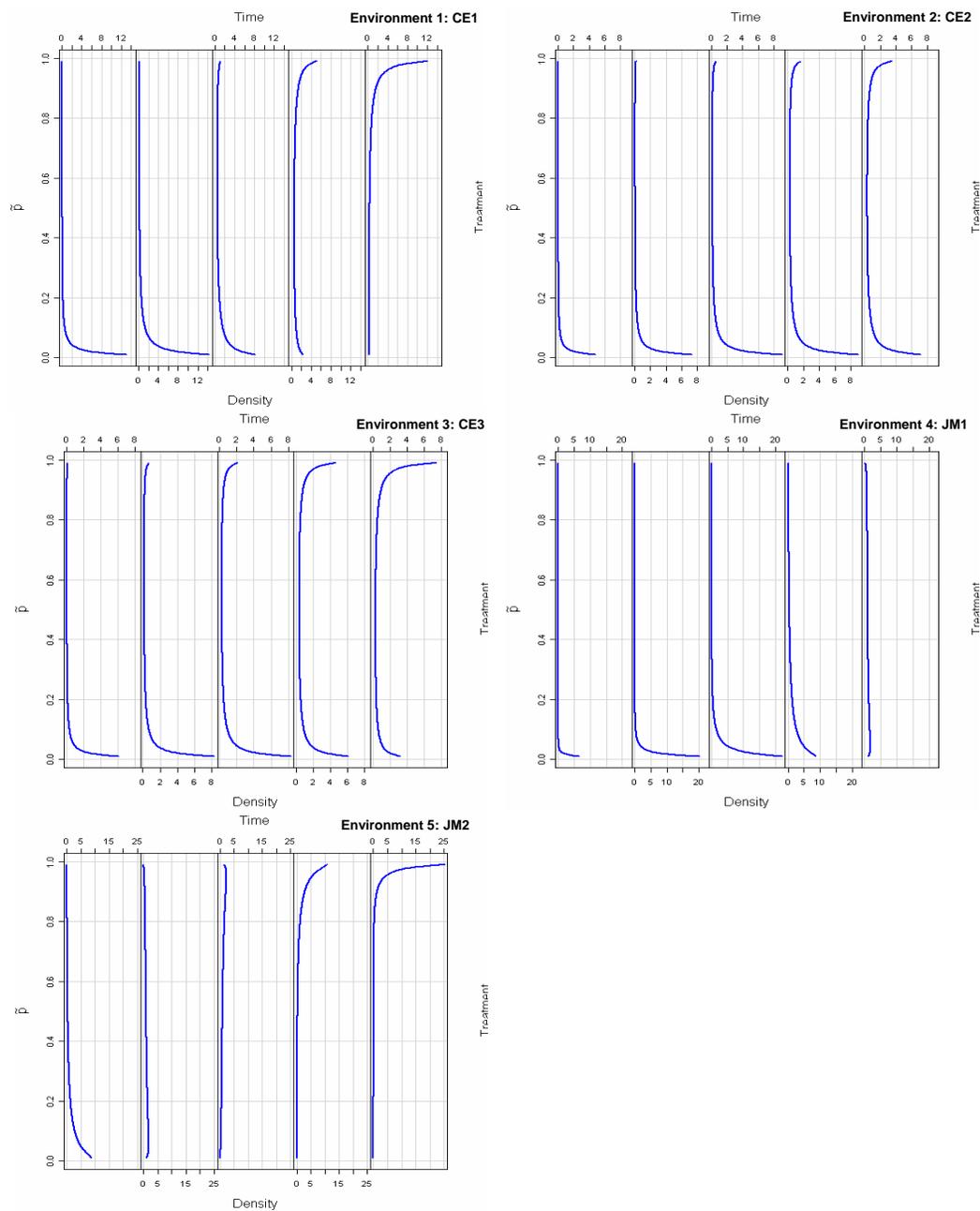


Figure 5–20: Density Functions of \tilde{p} for $t = 0, 40, 80, 120$ and 160 days (*Model 2*), *Garlic Dataset*. [Each panel inside the plot represents a time-point]

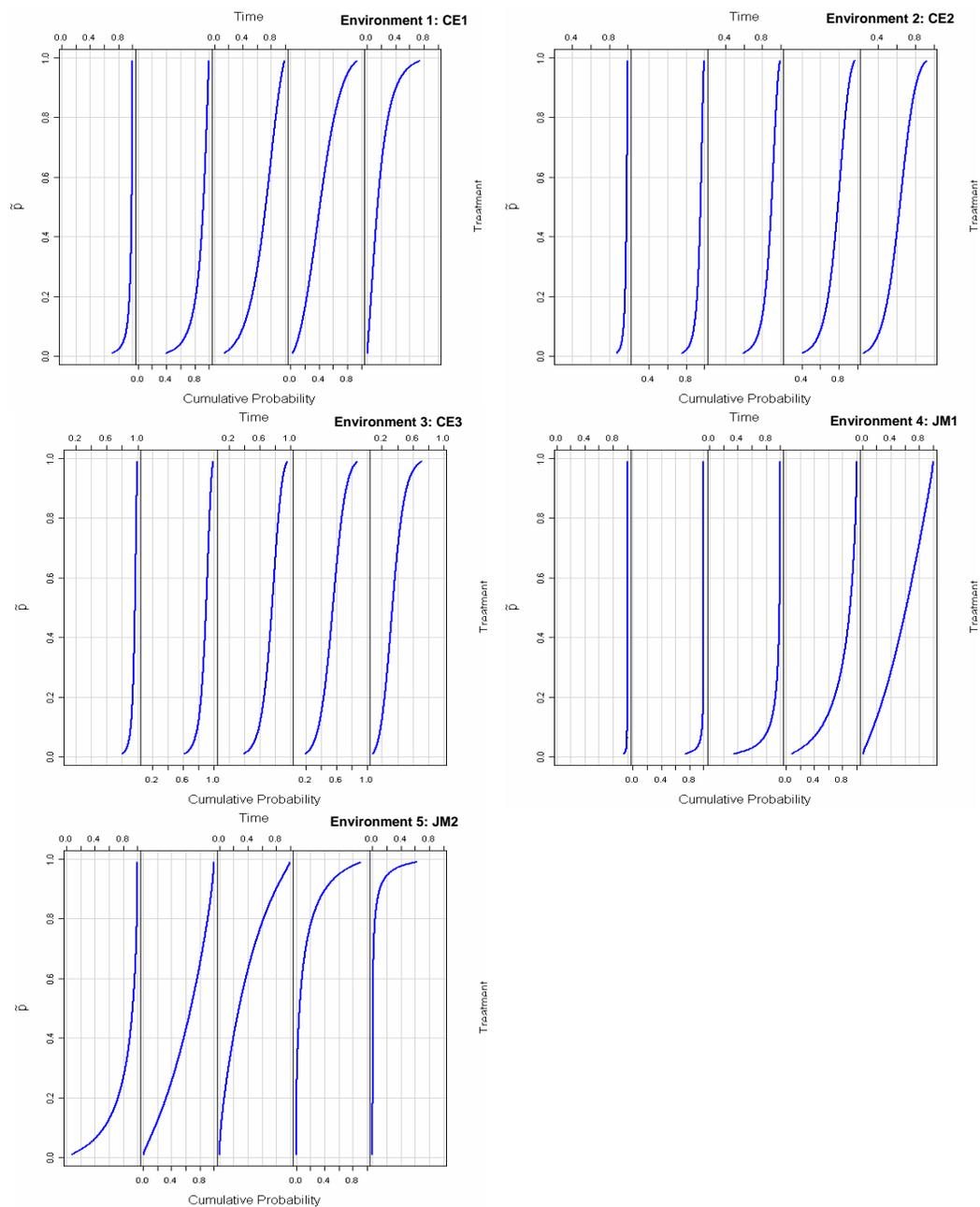


Figure 5–21: Cumulative Distribution of \tilde{p} for $t = 0, 40, 80, 120$ and 160 days (*Model 1*), *Garlic Dataset*. [Each panel inside the plot represents a time-point]

with random intercept and slope. We will apply the proposed method to the *Garlic dataset*.

In order to define the percentiles curves in these type of models, consider the following logistic model for each environment in the Garlic dataset:

$$\text{logit} [P(Y_{ijk} | \mathbf{u}_{i0k}, \mathbf{u}_{i1k})] = \text{logit} (\tilde{p}_{ijk}) = \tilde{b}_{0k} + \tilde{b}_{1k} \text{time} \quad (5.15)$$

where $\tilde{b}_{0k} = \tilde{\beta}_{0ik} + u_{i0k}$ and $\tilde{b}_{1k} = \tilde{\beta}_{1ik} + u_{i1k}$ are the random intercept and slope for each block/site in the respective environment. The random effects, \mathbf{u}_{ik} , is a random bivariate vector distributed normally with zero mean vector and covariance matrix D_k . In general, the matrix D contains in its principal diagonal the variance of the random intercept and slope, and the other component is the covariance between the random intercept and slope:

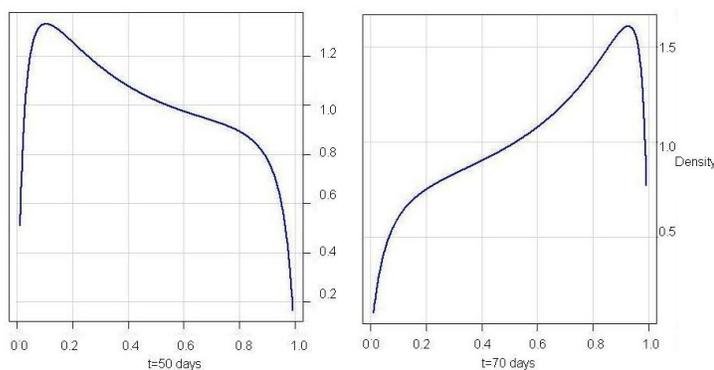


Figure 5–22: Density Function of \tilde{p} for $t = 50$ and $t = 70$ days, before and after the time-point where marginal mean curve is equal to median curve (*Model 2*), Environment *JM2*, *Garlic Dataset*

Table 5–10: Induced Marginal Mean Incidences (*Model 2*), *Garlic Dataset*

<i>Time (Days)</i>	<i>CE1</i>	<i>CE2</i>	<i>CE3</i>	<i>JM1</i>	<i>JM2</i>
0	0.0319	0.0210	0.0442	0.0027	0.1590
30	0.0881	0.0433	0.0948	0.0089	0.3160
60	0.1998	0.0821	0.1790	0.0278	0.5200
90	0.3721	0.1428	0.2995	0.0767	0.7183
120	0.5751	0.2284	0.4474	0.1799	0.8632
150	0.7596	0.3373	0.6031	0.3488	0.9441
180	0.8881	0.4621	0.7437	0.5578	0.9801

$$D = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

Using the correlation coefficient, ρ , between the intercept and slope, D can also be written as:

$$D = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

where σ_1 and σ_2 are the corresponding standard deviations of the random effects.

Similarly to Section 5.3.2, we consider two models with random intercept and slope:

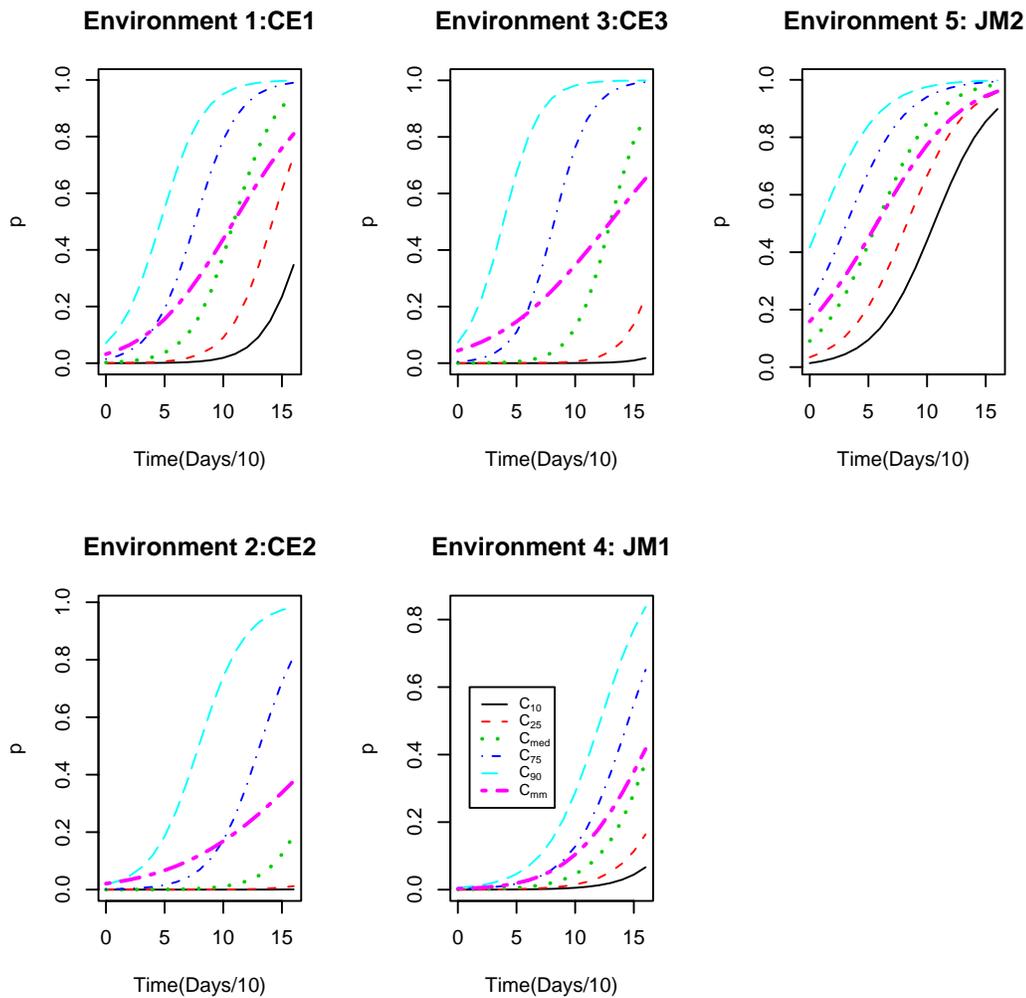


Figure 5–23: Percentile Curves (*Model 2*), *Garlic Dataset*

- *Model 3*: The variance of the random intercept vary across environment, but the variance of the random slope is constant.
- *Model 4*: The variance of the random intercept and the slope vary across environments.

The fit statistics for both models are shown in Table 5–11. Two important conclusions can be extracted from these results. First, a model with random intercept and slope has a better fit than a model with only random intercept. Second, the model with the variance of the random intercept and slope varying across environment is better compared with the model that assumes constant variance for the random slope. The *BIC* criterion is smaller in *Model 4*, and the likelihood ratio test is favorable to this model. The likelihood ratio is equal to $-2[\log L(M4) - \log L(M3)] = 26.3$. The p-value associated to this value is less to 0.0001 compared with a χ_4^2 . Therefore, we will use this model to analyze the density function of \tilde{p} and percentiles curves in the case of two random effects.

Table 5–11: Fit Statistics for *Model 3* and *Model 4*, *Garlic Dataset*

Fit Statistics		
<i>Statistic</i>	<i>Model 3</i>	<i>Model 4</i>
-2 Log Likelihood	8856.1	8829.8
BIC	8950.3	8946.2

5.4.1 Density Function of \tilde{p} and Percentiles Curves

In a model with random intercept, the density and cumulative function of \tilde{p} were easily deduced using the classical theorem of the transformation of random variables. However, in the case of two random effects, intercept and slope, we have the following situation:

$$\tilde{p}_{ij} = P(Y_{ij} \mid \mathbf{u}_{i0}, \mathbf{u}_{i1}) = \varphi(\mathbf{u}_{i0}, \mathbf{u}_{i1}) \quad (5.16)$$

\tilde{p} is a bivariate function of the random effects, as shown in Figure 5–24. Therefore, it is complicated to find the density function of \tilde{p} for a given time t with the

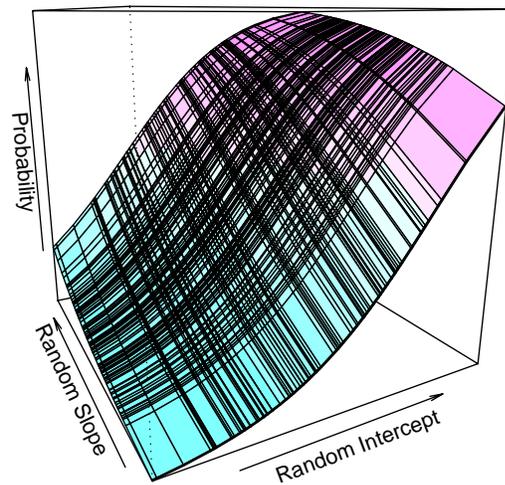


Figure 5–24: Conditional Probability in Function of Random Intercept and Slope

model presented in Equation 5.15 because of the transformation between the two random effects and the probability. We propose an empirical approximation to the density function and percentiles curves using the following methodology.

Steps to Calculate the Density Function of \tilde{p} and Percentiles Curves

- Generate a number, sufficiently large, of samples of the a bivariate normal distribution with $(\tilde{\beta}_0, \tilde{\beta}_1)$ mean vector, and covariance matrix D using the Monte Carlo Method. We used the `rmvnorm` function from the `mvtnorm` R library. We generated $N=100,000$ samples of the bivariate normal distribution with the covariance matrix \hat{D} for each environment. In situations with a large number of random effects, a Cholesky decomposition can be useful in order to simplify the computational task

[26]. We used the density function from stats R library to estimate the density function.

- Calculate the conditional probabilities in each time t using the inverse logit transformation:

$$\tilde{p}_{ij} = \frac{\exp(\tilde{b}_0 + \tilde{b}_1 t)}{1 + \exp(\tilde{b}_0 + \tilde{b}_1 t)} \quad (5.17)$$

- Estimate the density function of \tilde{p} using a kernel density estimator (i.e. Gaussian kernel) [38]. Then, plot the density function for each time.
- Calculate the percentiles of the empirical distribution of \tilde{p} . For each time it is necessary to sort the \tilde{p} values calculated previously, and select the respective percentiles. To find the $k - th$ percentile, take the correspondent $\frac{Nk}{100}$ \tilde{p} value. We developed this procedure because in \Re^2 (intercept and slope), to define an order is not as trivial as in \Re (only intercept). This methodology can be expanded for more than two random effects.
- Finally, estimate the marginal mean induced by the random effects in a specific time averaging by the subject-specific probabilities estimated with the N random effects \mathbf{u}_i generated in the first step, as was proposed by Molenbergs and Verbeke [34]:

$$\hat{p}_{ij} = \frac{1}{N} \sum_{i=1}^N \frac{\exp(\tilde{b}_0 + \tilde{b}_1 t)}{1 + \exp(\tilde{b}_0 + \tilde{b}_1 t)} \quad (5.18)$$

Note that the previous integral is a Monte Carlo integral for the bivariate case [16].

In general, this is a Monte Carlo integration with Gaussian weighted function.

Using the previous methodology we can obtain an approximation to the density of \tilde{p} and the marginal mean \tilde{p} , such as in Section 5.1.1, where they were derived theoretically. The β 's and variance components estimates with its respective standard errors for *Model 4* are shown in Table 5-12 ⁷.

⁷ All these parameters are significant ($\alpha = 0.01$).

Table 5–12: Estimates and Standard Errors for *Model 4*, *Garlic Dataset*

<i>Environment</i>	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
<i>CE1</i>	-5.6389 (0.3280)	0.5462 (0.0351)	2.2579 (0.2793)	0.2403 (0.0303)
<i>CE2</i>	-7.7526 (0.6847)	0.3399 (0.0556)	2.9130 (0.6330)	0.2015 (0.0396)
<i>CE3</i>	-7.1399 (0.5983)	0.5508 (0.0521)	3.0347 (0.6033)	0.2464 (0.0432)
<i>JM1</i>	-6.8402 (0.2184)	0.3932 (0.0167)	1.2281 (0.1960)	0.0971 (0.0137)
<i>JM2</i>	-2.2993 (0.2325)	0.4423 (0.0391)	1.1219 (0.1894)	0.1886 (0.0309)
Correlation between intercepts and slopes: $\hat{\rho} = 0.2603$				

In this case, the covariance matrices for the random effects in each environment is given by:

$$\hat{D}_{CE1} = \begin{bmatrix} 5.0981 & 0.1412 \\ 0.1412 & 0.0578 \end{bmatrix} \quad \hat{D}_{CE2} = \begin{bmatrix} 8.4856 & 0.1528 \\ 0.1528 & 0.0406 \end{bmatrix} \quad \hat{D}_{CE3} = \begin{bmatrix} 9.2094 & 0.1946 \\ 0.1946 & 0.0607 \end{bmatrix}$$

$$\hat{D}_{JM1} = \begin{bmatrix} 1.5082 & 0.0311 \\ 0.0311 & 0.0094 \end{bmatrix} \quad \hat{D}_{JM2} = \begin{bmatrix} 1.2586 & 0.0551 \\ 0.0551 & 0.0356 \end{bmatrix}$$

The infection progress rates in the environments of the *Jesús María* province has a smaller variability in comparison with the ones in *Cruz del Eje* province. Likewise, it occurs with the random intercept.

The densities of \tilde{p} for all environments using *Model 4* are shown in Figure 5–25 but in $t = 40, 80$ and 120 days. They are similar to the ones for *Model 2* for the equivalent times. However, among them there are differences that can be observed in the marginal mean incidences in Table 5–13. These values differ from the ones presented in Tables 5–7 and 5–10. For example, the incidence in Environment *CE1* is smaller in comparison with the other models.

The percentiles curves are also similar to the ones presented in the previous models. Note that the marginal mean incidence curve and the median curve are equal in $\tilde{p} = 0.5$, which is an evidence that the proposed method to find the curves with two random effects is adequate. The consistency of the \tilde{p} estimates also indicates the adequacy of the methodology.

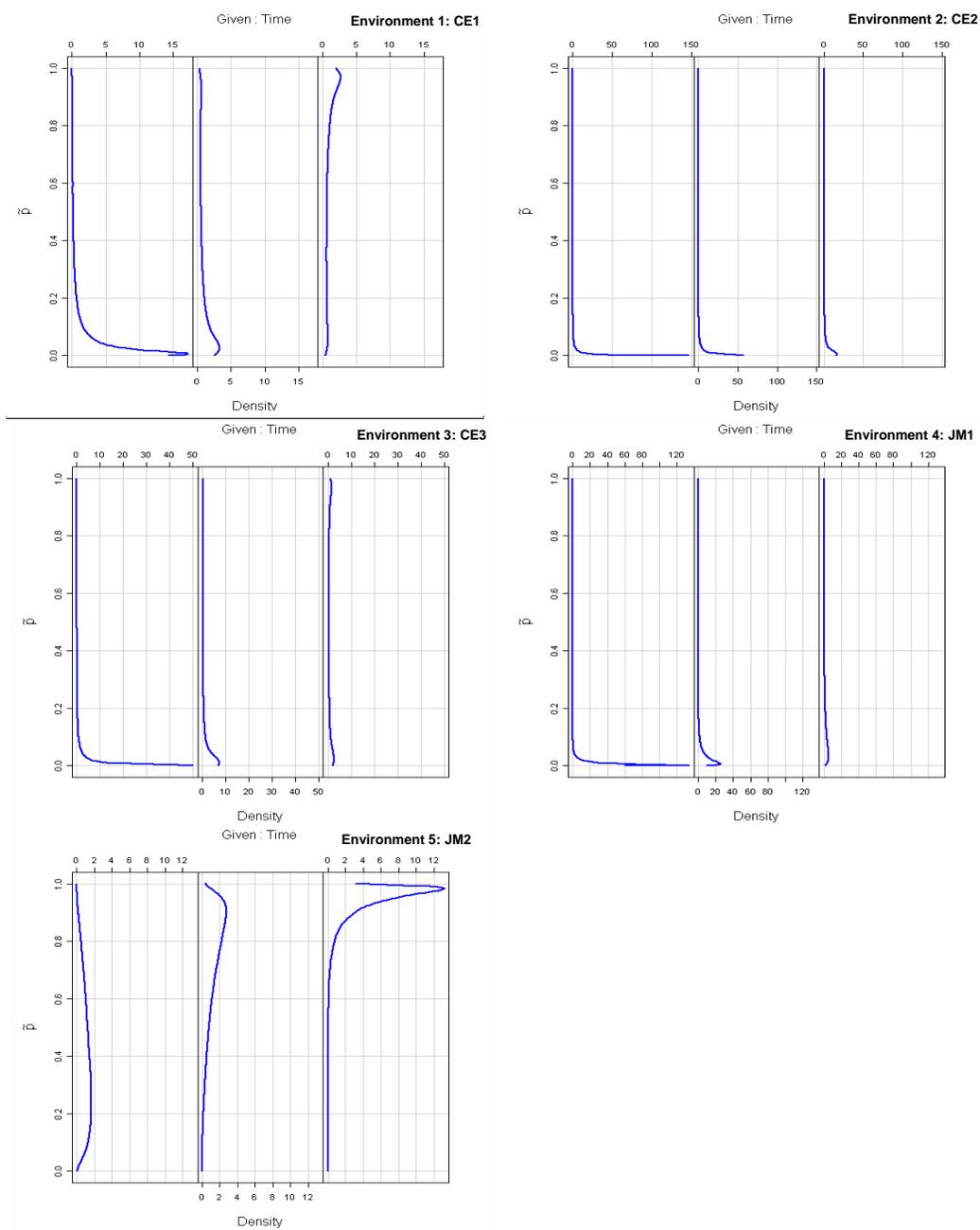
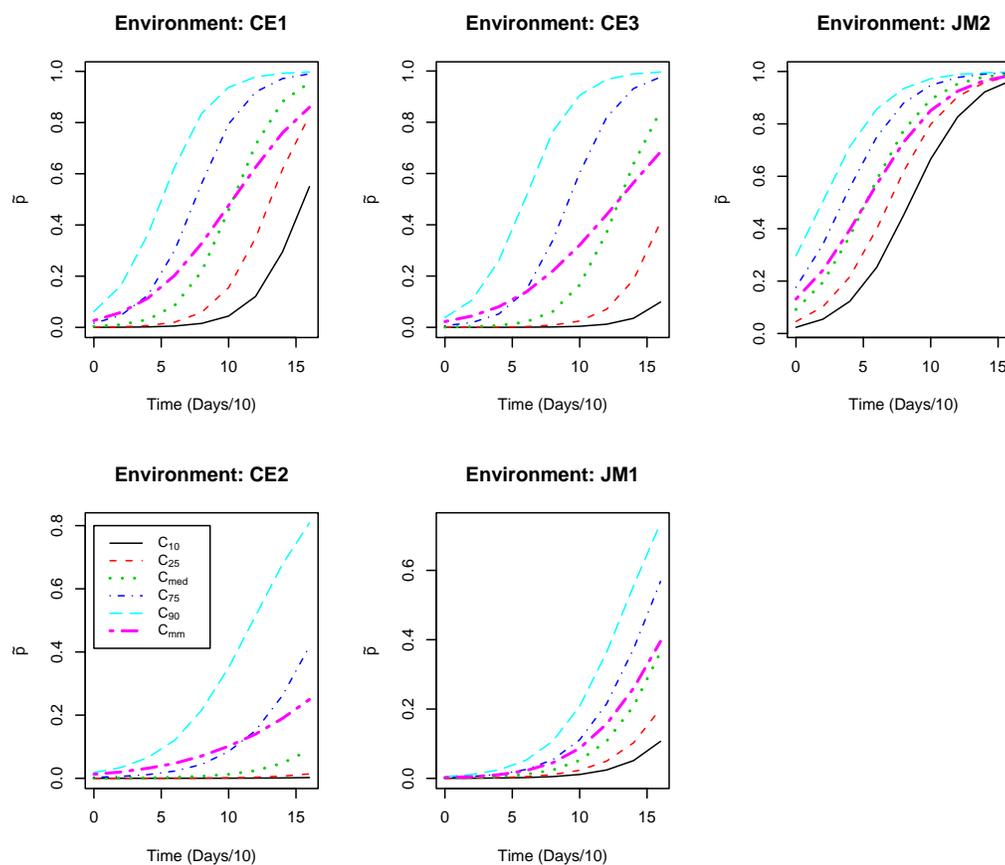


Figure 5–25: Density Functions of \tilde{p} for $t = 40, 80$ and 120 days (*Model 4*), *Garlic Dataset*. [Each panel inside the plot represents a time-point]

Table 5–13: Induced Marginal Mean Incidences (*Model 4*), *Garlic Dataset*

<i>Time (Days)</i>	<i>CE1</i>	<i>CE2</i>	<i>CE3</i>	<i>JM1</i>	<i>JM2</i>
0	0.0260	0.0125	0.0222	0.0022	0.1311
30	0.0818	0.0251	0.0600	0.0071	0.3149
60	0.2038	0.0482	0.1369	0.0220	0.5695
90	0.3987	0.0844	0.2671	0.0630	0.7971
120	0.6244	0.1413	0.4386	0.1572	0.9250
150	0.8137	0.2177	0.6254	0.3254	0.9771
180	0.9276	0.3152	0.7840	0.5440	0.9935

Figure 5–26: Percentile Curves (*Model 4*), *Garlic Dataset*

CHAPTER 6

CONCLUSIONS

Longitudinal data analysis requires special statistical techniques because of the dependence between repeated measurements. A model to analyze longitudinal data can explain the dependence using the following three sources: *random effects*, *serial correlation* and *measurement error*. Depending on the sources included in the model, it is possible to talk about *Covariance Pattern Models* (CPM) and *Mixed Models*. In the first case, the CPM models include correlation matrices to explain the dependence between observations. In the second case, the mixed models include random effects to induce the dependence between observations. The type of inference and the interpretation of the estimates are important points to compare these models. The *CPM* models are based on a marginal inference and the estimates of the fixed effects in the model have a population-average interpretation. Precisely, these models are called *marginal models*. In these models, we can talk about the *marginal mean*. Meanwhile, the mixed models, also called conditional models, are based on a subject-specific inference and the estimates have a subject-specific interpretation. In these models, we can talk about the *conditional mean* because the mean of the response depends on the values of the random effects. However, using mixed models it is possible to obtain the marginal mean induced by the random effects. The ability of the mixed models to produce subject-specific and population-average interpretations is an advantage of these models. In contrast, the marginal models are criticized by many authors because they only yield estimates with a

population-average interpretation. In addition, the relationship between marginal and subject-specific models is not easily to derived in some situations.

For normal responses, the linear relationship between the response and the mean and the properties of the multivariate normal distribution, permit to derive the theory related to the linear mixed models. For example, the relationship between the marginal and subject-specific models can be easily derived using these properties. Likewise, the marginal mean and the conditional mean with the random effects equal to zero (typical mean) are the same. This result is different in linear models with non-normal response. In categorical longitudinal data, such as binary and count data, the non-linearity in the mean-response relationship does not permit to derive analytically the marginal mean induced by a subject-specific model. Also, the marginal mean and the typical mean are different. Generalized linear mixed models (GLMM) can be used to analyze this type of data. For these models the estimation methods are more complicated and more computationally intensive than the ones in linear normal response. Numerical integration and optimization algorithms are necessary to compute the estimates in GLMMs. For binary longitudinal data, a logistic model with a correlation structure (marginal models) can be used. However, the adequacy of the correlations matrices in binary data have been questioned by some authors. Therefore, a logistic model with random intercept is a better alternative to analyze binary longitudinal data. This model includes a random intercept to induce the dependence between the observations. This is a subject-specific model and belongs to the GLMM family.

Percentile Curves is an original concept in the longitudinal data analysis using a logistic model with random effects. This formulation permits us to interpret the subject-specific probability curves as percentile curves across time. This concept have been used by the researcher in continuous longitudinal data analysis, specially in the medicine area. We expand the percentiles curves to binary longitudinal data.

This concept was applied to a logistic model with random intercept, where the density and cumulative distribution of the subject-specific probabilities across time can be derived using statistical inference.

For a generalized linear model with random intercept, we conclude that the percentiles random effects are conserved by the link function, and therefore, are equivalent to the same percentiles in the conditional probabilities. This is because the link functions are increasing with respect to the random effect. These results permit us to study the properties of the subject-specific probabilities, such as the shape distribution and the marginal mean induced by the random effects. The marginal mean has a population-average interpretation, and it is important in order to study the population-average effects. The *Toenail* data set was analyzed using this model, and the results obtained permit us to evidence our proposal as an important tool in studying binary longitudinal data. The percentiles curves for the *Toenail* data set were calculated. Using these curves, it is possible to analyze the mean and median response probability of infection for the two treatments, as well as another percentile of interest in the study. In addition, it is possible to establish differences between treatments for a specific percentile curve.

Also, we expanded the *Percentiles Curves* to a logistic model with random intercept and slope, and we proposed a methodology to calculate the density distribution and the *Percentiles Curves* by using the Monte Carlo method and Gaussian kernel. We applied this methodology to white rot in garlic data set, and we can establish that the results are adequacy, offer a interesting tool for the researcher. Using the fitter model, a logistic model with random intercept and slope, the percentiles curves obtained permit us to detect important differences in the infection progress curves within and between environments.

The percentile curves concept to analyze binary longitudinal data analysis could be applied to other link functions, such as probit and complement log-log. It is possible apply the percentile curves for other types of outcome, such as count (i.e. Poisson). In any outcome type, it is possible analyze the curves for more than two random effects using the Monte Carlo methodology.

Percentile Curves concept is a good tool in order to interpret results in a subject-specific model when categorical longitudinal data are analyzed.

APPENDICES

APPENDIX A

CODES OF R FUNCTIONS USED IN THIS THESIS

PHERMITE FUNCTION

```

function(n){
#-----#
# CALCULATE THE ABCISSAS FOR GAUSS-HERMITE QUADRATURE
# n: Number of points of quadrature (Q)
#-----#
pim4=1/(pi^(1/4))
eps=3e-14
maxit=10
x=rep(0,n)
w=rep(0,n)
m=(n+1)/2
for(i in 1:m){
  if(i==1){z=sqrt(2*n+1)-(1.85575)*((2*n+1)^(-1/6))}
  else if(i==2){z=z-(1.14*(n^(0.426))/z)}
  else if(i==3){z=1.86*z-0.86*x[1]}
  else if(i==4){z=1.91*z-0.91*x[2]}
  else{z=2.0*z-x[i-2]}
for(j in 1:maxit){
  p1=pim4
  p2=0.0
  for(k in 1:n){
    p3=p2
    p2=p1
    p1=z*sqrt(2.0/k)*p2-sqrt((k-1)/k)*p3
    pp=sqrt(2*n)*p2
  }
  z1=z
  z=z1-p1/pp
  if(abs(z-z1)<=eps)
    stop
  }
  x[i]=z
  x[n+1-i]=-z
  w[i]=2.0/(pp^2)
  w[n+1-i]=w[i]
}
list(x=x,w=w)
}

```

GAUSSH.INT FUNCTION

```
function(f, q=10, start, ...){
#-----#
# CALCULATE THE INTEGRAL USING GAUSS-HERMITE QUADRATURE
#-----#
library(MASS)
fn <- function(x, f, ...) {-log(f(x,...))}
if(missing(start)) start <- runif(1)
opt <- optim(start, fn=fn,
+ method="CG", control=list(maxit=500), hessian=T,f=f,...)
u<-phermite(q)
xq <- u$x sigma <- sqrt(c(1/opt$hes))
r2 <- sqrt(2)*sigma
g <- r2*exp(xq*xq)*f(opt$par+r2*xq, ...)
return(sum(u$w*g) ) }
```

APPENDIX B

CODES TO FIT A LOGISTIC MODEL WITH RANDOM INTERCEPT FOR TOENAIL DATASET

glmmML R function (glmmML library)

Code:

```
> library(glmmML)
> glmmML(response ~ treat + time + time * treat, data = infect,
+ family = binomial, cluster = id, n.points = 50)
```

Output:

	coef	se(coef)	z	Pr(> z)
(Intercept)	-1.6295	0.42682	-3.8177	0.000135
treat	-0.1465	0.57400	-0.2553	0.798000
time	-0.3909	0.04430	-8.8242	0.000000
treat:time	-0.1379	0.06798	-2.0284	0.042500

Standard deviation in mixing distribution: 4.020
Std. Error: 0.3853

Residual deviance: 1251 on 1903 degrees of freedom AIC: 1261

glmmPQL R function (MASS library)

Code:

```
> library(MASS)
> glmmPQL(response ~ treat + time + time * treat, data = infect,
+ family = binomial, random = ~1 | id, verbose = F)
```

Output:

Loading required package: nlme

Attaching package: 'nlme'

The following object(s) are masked from package:Matrix :

BIC fixef ranef VarCorr

Linear mixed-effects model fit by maximum likelihood

Data: infect

Log-likelihood: -5849.194

Fixed: response ~ treat + time + time * treat

```
(Intercept)      treat      time  treat:time
-0.74324709 -0.03480232 -0.29469100 -0.10017514
```

```
Random effects:
Formula: ~1 | id
      (Intercept) Residual
StdDev:   2.317071 0.9362732
```

```
Variance function:
Structure: fixed weights
Formula: ~invwt
Number of Observations: 1908
Number of Groups: 294
```

lmer R function (Matrix library)

Code:

```
> library(Matrix)
> lmer(response ~ treat + time + time * treat + (1 | id), family = binomial,
+       data = infect)
```

Output:

```
Generalized linear mixed model fit using PQL
Formula: response ~ treat + time + time * treat + (1 | id)
Data: infect
Family: binomial(logit link)
      AIC      BIC  logLik deviance
1305.676 1333.445 -647.838 1295.676
Random effects:
      Groups      Name      Variance  Std.Dev.
      id (Intercept)      6.24      2.498
# of obs: 1908, groups: id, 294

Estimated scale (compare to 1) 0.9291529

Fixed effects:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.745071  0.264086 -2.8213  0.004783 **
treat        -0.035512  0.374394 -0.0949  0.924433
time         -0.295909  0.033892 -8.7311 < 2.2e-16 ***
treat:time   -0.100492  0.053772 -1.8689  0.061640 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) treat  time
treat  -0.705
time   -0.372  0.262
treat:time 0.234 -0.370 -0.630
```

glmm R function (Repeated library)

Code:

```
> library(repeated)
> glmm(response~treat+time+time*treat, nest=id, family = binomial,
+ data = infect)
```

Output:

```
Coefficients:
(Intercept)      treat      time      sd  treat:time
-0.7992      -0.3070     -0.3675     3.2788     -0.1775
```

```
Degrees of Freedom: 1907 Total (i.e. Null); 1903 Residual
Null Deviance:      1980
Residual Deviance: 1268      AIC: 1278
Normal mixing variance: 10.75045
```

SAS PROC NL MIXED with Normal Random Effects

Code:

```
title "Logistic Model with Random Intercept - Normal distribution";
proc sort data=tesis.infection;
by id;
run;
proc nlmixed data=tesis.infection qpoints=30;
parms beta0=0 beta1=0 beta2=0 beta3=0 sigma=1;
eta = beta0 + beta1*treat + beta2*time + beta3*time*treat + u;
expeta = exp(eta);
p = expeta/(1+expeta);
model response binary(p);
random u normal(0,sigma*sigma) subject=id;
run;
```

Output:

Specifications	
Data Set	TESIS.INFECTION
Dependent Variable	RESPONSE
Distribution for Dependent Variable	Binary
Random Effects	b
Distribution for Random Effects	Normal
Subject Variable	ID
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions	
Observations Used	1908
Observations Not Used	0
Total Observations	1908
Subjects	294
Max Obs Per Subject	7
Parameters	5
Quadrature Points	50

Parameters					
beta0	beta1	beta2	beta3	sigma	NegLogLike
0	0	0	0	1	1025.69308

NOTE: GCONV convergence criterion satisfied.

Fit Statistics	
-2 Log Likelihood	1250.8
AIC (smaller is better)	1260.8
AICC (smaller is better)	1260.8
BIC (smaller is better)	1279.2

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
beta0	-1.6183	0.4343	293	-3.73	0.0002	0.05	-2.4731	-0.7636	-2.27E-6
beta1	-0.1608	0.5840	293	-0.28	0.7833	0.05	-1.3101	0.9885	5.153E-6
beta2	-0.3910	0.04438	293	-8.81	<.0001	0.05	-0.4783	-0.3037	0.00004
beta3	-0.1368	0.06801	293	-2.01	0.0452	0.05	-0.2706	-0.00293	0.00007
sigma	4.0066	0.3799	293	10.55	<.0001	0.05	3.2590	4.7543	-1.89E-6

SAS PROC NLMIXED with Bridge Random Effects

Code:

```

title "Logistic Model with Random Intercept - Bridge distribution";
proc sort data=tesis.infection;
by id;run;
proc nlmixed data=tesis.infection qpoints=50;
*parms beta0=-1.63 beta1=0.19 beta2=-0.40 beta3=-0.14 s1=4;
parms beta0=0 beta1=0 beta2=0 beta3=0 s1=1;
pi=2*arcsin(1);
uni=probnorm(u/s1);
phi=1/sqrt(1+3/pi/pi*s1*s1);
ul=1/phi*log(sin(pi*uni*phi)/sin(phi*pi*(1-uni)));
eta = beta0 + beta1*treat + beta2*time + beta3*time*treat + ul;
expeta = exp(eta);
p = expeta/(1+expeta);
model response binary(p);
random u normal(0,s1*s1) subject=id;
estimate 'phi' phi;
run;

```

Output:

Specifications	
Data Set	TESIS.INFECTION
Dependent Variable	RESPONSE
Distribution for Dependent Variable	Binary
Random Effects	b
Distribution for Random Effects	Normal
Subject Variable	ID
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions	
Observations Used	1908
Observations Not Used	0
Total Observations	1908
Subjects	294
Max Obs Per Subject	7
Parameters	5
Quadrature Points	50

Parameters					
beta0	beta1	beta2	beta3	s1	NegLogLike
0	0	0	0	1	1054.23871

Fit Statistics	
-2 Log Likelihood	1252.7
AIC (smaller is better)	1262.7
AICC (smaller is better)	1262.8
BIC (smaller is better)	1281.2

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
beta0	-1.4966	0.4106	293	-3.65	0.0003	0.05	-2.3047	-0.6886	-1.77E-6
beta1	-0.1255	0.5665	293	-0.22	0.8248	0.05	-1.2405	0.9894	-4.99E-7
beta2	-0.4010	0.04598	293	-8.72	<.0001	0.05	-0.4915	-0.3105	0.000013
beta3	-0.1410	0.07177	293	-1.96	0.0504	0.05	-0.2823	0.000235	0.000013
s1	4.3597	0.4176	293	10.44	<.0001	0.05	3.5379	5.1816	-8.98E-7

Additional Estimates									
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	
phi	0.3841	0.03136	293	12.25	<.0001	0.05	0.3224	0.4458	

SAS PROC GLIMMIX

Code:

```

title "Logistic Model with Random Intercept - Normal distribution";
proc glimmix data=tesis.infection;
class treat id;
model response (event='1') = treat time treat*time /dist=binary solution;
random intercept / subject=id;
run;

```

Output:

Model Information	
Data Set	TESIS.INFECTION
Response Variable	RESPONSE
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	ID
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information		
Class	Levels	Values
TREAT	2	0 1
ID	294	1 2 ... 382 383

Number of Observations Read	1908
Number of Observations Used	1908

Response Profile			
Ordered Value	RESPONSE	Total	Frequency
1	0		1500
2	1		408
The GLIMMIX procedure is modeling the probability that RESPONSE='1'.			

Dimensions	
G-side Cov. Parameters	1
Columns in X	6
Columns in Z per Subject	1
Subjects (Blocks in V)	294
Max Obs per Subject	7

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	1
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	11159.19
Generalized Chi-Square	1489.85
Gener. Chi-Square / DF	0.78

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	ID	4.7095	0.6024

Solutions for Fixed Effects						
Effect	TREAT	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-0.7464	0.2382	292	-3.13	0.0019
TREAT	0	0.02594	0.3360	1612	0.08	0.9385
TREAT	1	0
TIME		-0.3740	0.03960	1612	-9.45	<.0001
TIME*TREAT	0	0.09583	0.05105	1612	1.88	0.0607
TIME*TREAT	1	0

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TREAT	1	1612	0.01	0.9385
TIME	1	1612	163.24	<.0001
TIME*TREAT	1	1612	3.52	0.0607

REFERENCE LIST

- [1] N. Chaganty and H. Joe. Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Society*, 66(4):851–860, B 2004.
- [2] A. Agresti. *Categorical Data Analysis*. John Wiley & Sons, Inc., second edition, 2002.
- [3] N. Breslow. Whither pql? Paper Series 192, University of Washington, 2003.
- [4] N. Breslow and D. Clayton. Approximate inference in generalized linear models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.
- [5] G. Brostrom. *Generalized linear models with random intercepts*, 2003.
- [6] H. Brown and R. Prescott. *Applied Mixed Models in Medicine*. John Wiley & Sons, first edition, 1999.
- [7] R. Carroll. Variances are not always nuisance parameters. *Biometrics*, June 2003.
- [8] G. Casella and R. Berger. *Statistical Inference*. Thompson Learning, second edition, 2001.
- [9] E. Cruch and D. Spiegelman. The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(t) \exp(-t^2) dt$: Application to logistic-normal models. *Journal of the American Statistical Association*, 85(410):464–469, June 1990.
- [10] M. Davidian and D. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC, Boca Raton edition, 1995.
- [11] P. Davis and Rabinowitz P. *Methods of Numerical Integration*. Academic Press, second edition, 1984.
- [12] E. Demidenko. *Mixed Models. Theory and Applications*. John Wiley & Sons, first edition, 2004.

- [13] L. Duchateau and P. Janssen. Understanding heterogeneity in generalized mixed and frailty models. *The American Statistician*, 59(2):143–6, May 2005.
- [14] M. Evans and T. Swart. Methods for approximating integral in statistics with special emphasis on bayesian integration problems. *Statistical Science*, 10(3):254–272, 1995.
- [15] G. Fitzmaurice and N. Lard. A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, 80(1):141–51, 1993.
- [16] J. E. Gentle. *Random Number Generation and Monte Carlo Methods*. Springer, second edition, 2000.
- [17] H. Goldstein. Software for multilevel modelling, resources and further developments. Technical report, Published by Hodder Arnold, 2003.
- [18] M. Griswold and S. Zeger. On marginalized multilevel models and their computation. Paper series, Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, 2004.
- [19] P. Heagerty. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, 55:688–698, Sep. 1999.
- [20] P. Heagerty and B. Kurland. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88(4):973–985, 2001.
- [21] P. Heagerty and S. Zeger. Marginalized multilevel models and likelihood inference. *Statistical Science*, 15(1):1–19, Feb. 2000.
- [22] D. Hedeker. Longitudinal data analysis, 2005. Reading Material for Course of Longitudinal Data Analysis.
- [23] SAS Inc. *SAS online documentation: The NLMIXED PROCEDURE*, 2005.
- [24] SAS Institute. *SAS System for Windows v9.1.3*. Cary, NC, USA, 2002-2003.
- [25] E. Budtz-Jorgensen and L. Endahl K. Larsen, J. Holm. Interpreting parameters in the logistic model with random effects. *Biometrics*, 56:909–914, September 2000.

- [26] K. Lange. *Numerical Analysis for Statistician*. Springer, first edition, 1988.
- [27] Y. Lee and J. Nelder. Conditional and marginal models: Another view. *Statistical Science*, 19(2):219–238, May 2004.
- [28] E. Lessafre and B. Spiessens. On the effect of the number of quadrature points in logistic random-effects model: an example. *Applied Statistics*, 50(3):325–335, 2001.
- [29] K. Liang and S. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.
- [30] Q. Liu and D. Pierce. A note on gauss-hermite quadrature. *Biometrika*, 81(2):624–9, 1994.
- [31] C. Davis M. Stokes and G. Koch. *Categorical data Analysis using the SAS System*. SAS Institute, Cary, NC, USA, second edition, 2000.
- [32] C. McCulloch. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437):162–170, March 1997.
- [33] C. McCulloch and S. Searle. *Generalized, Linear and Mixed Models*. John Wiley & Sons, first edition, 2001.
- [34] G. Molenberghs and G. Verbeke. *Models for Discrete Longitudinal Data*. Springer, first edition, 2005.
- [35] K. Liang P. Diggle, P. Heagerty and S. Zeger. *Analysis of Longitudinal Data*. Oxford University Press, second edition, 2002.
- [36] J. Pinheiro and D. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4(1):12–35, 1995.
- [37] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.

- [38] D. Scott. *Multivariate Density Estimation. Theory, Practice, and Visualization*. John Wiley & Sons, first edition, 1992.
- [39] P. Torres. Numerical integration: Its use in glmms. Technical report, Advanced Topics of Computation Course, Department of Mathematics, University of Puerto Rico at Mayaguez, 2004.
- [40] W. Venables and B. Ripley. *Modern Applied Statistics with S-PLUS*. Springer, fourth edition, 2002.
- [41] W. Vetterling W. Press, S. Teukolsky and B. Flannery. *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press, first edition, 1992.
- [42] Z. Wang and T. Louis. Matching conditional and marginal shapes in binary random intercept models using bridge distribution function. *Biometrika*, 90(4):765–775, 2003.
- [43] R. Wolfinger and M. O’Connell. Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computational Simulation*, 48:233–43, 1993.
- [44] H. Yu and W. Cumberland. Logistic regression models for analysis of multistage survey data. Technical report, University of California.