8. Non-normal longitudinal data

Sonja Greven

Summer Term 2017

Introduction

So far we have always assumed the following normal distribution:

$$\mathbf{Y}_i \sim \mathcal{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i).$$

For non-normal, e.g.

- binary
- count
- non-normal continuous

responses, the models we discussed so far are not suitable. After some examples (8.1) we will thus discuss different extensions of generalized linear models (8.2) to the longitudinal setting (8.3).

Overview Chapter 8 - Non-normal longitudinal data

8.1 Data examples

8.2 Generalized linear models (GLMs)

8.3 Extending GLMs to longitudinal data

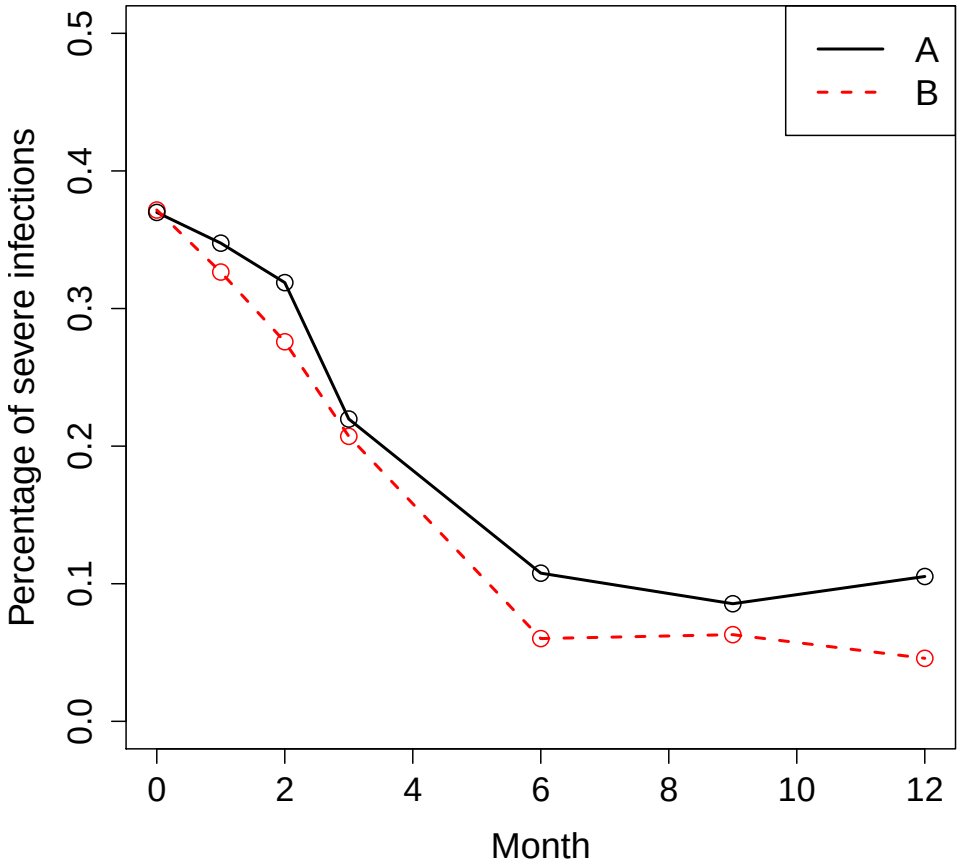
Binary example: The toenail data

- Randomized, double-blind, multi-center study for the comparison of two oral treatments (A and B) for the infection toenail dermatophyte onychomycosis (TDO) (Molenberghs & Verbeke, 2005)
- 2×189 patients
- 12 weeks of treatment, 48 weeks total.
- Measurements once per month during treatment, every 3 months afterwards $\rightarrow$ 7 measurements per patient (only 76% have all observations).
- Binary response Y: infection severe yes/no (1/0).

Binary example: The toenail data

	Treatment A			Treatment B		
	$\#Y = 1$	N	%	$\#Y = 1$	N	%
Baseline	54	146	37.0%	55	148	37.2%
Month 1	49	141	34.7%	48	147	32.6%
Month 2	44	138	31.9%	40	145	27.6%
Month 3	29	132	22.0%	29	140	20.7%
Month 6	14	130	10.8%	8	133	6.0%
Month 9	10	117	8.5%	8	127	6.3%
Month 12	14	133	10.5%	6	131	4.6%

Binary example: The toenail data



Count example: Epileptic seizures

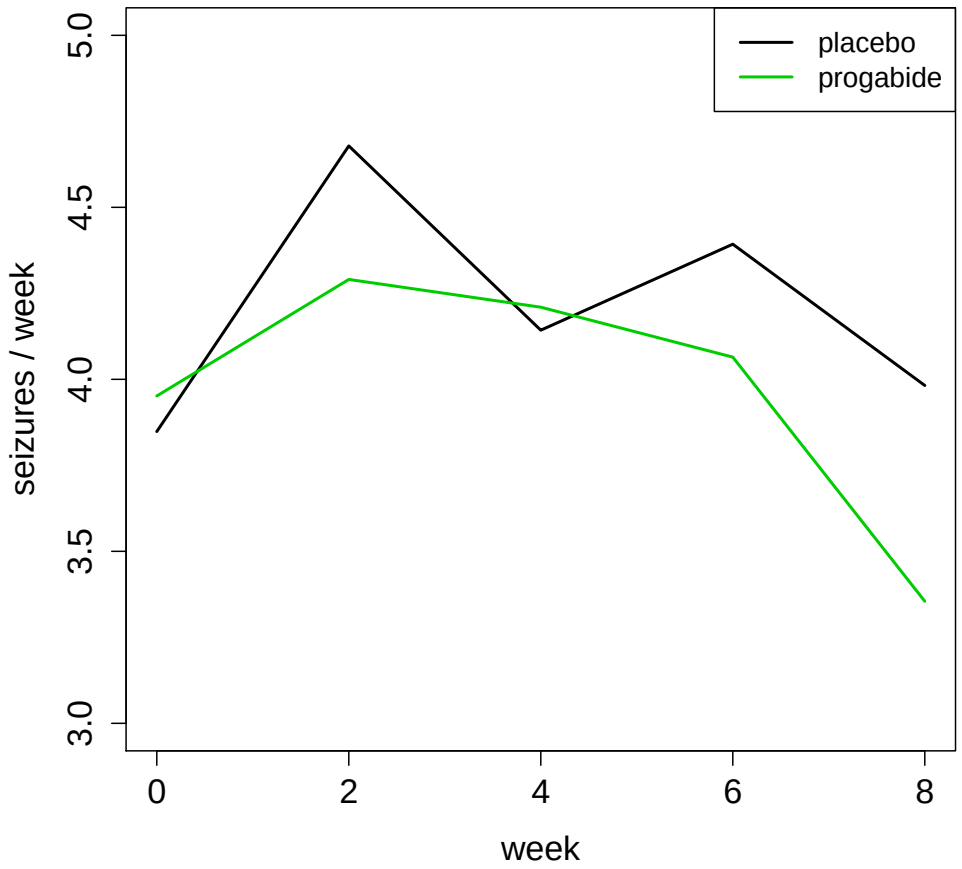
- Randomized, double-blind, multi-center study on the effectiveness of a new anti-epileptic drug (Diggle et al, 2002)
- Two groups: progabide / placebo (+ standard treatment)
- $N = 59$ patients
- Baseline Y_{i1}: Number of seizures during baseline period of eight weeks
- $Y_{i2}, Y_{i3}, Y_{i4}, Y_{i5}$: Number of seizures in the two weeks preceeding the planned visits at weeks 2, 4, 6, 8.

Count example: Epileptic seizures

Average number of seizures (and standard deviation) per week:

Treatment	Baseline	Week 2	Week 4	Week 6	Week 8
Progabide	3.95	4.29	4.21	4.06	3.35
	(3.5)	(9.1)	(5.9)	(6.9)	(5.6)
Placebo	3.85	4.68	4.14	4.39	3.98
	(3.3)	(5.1)	(4.1)	(7.3)	(3.8)

Count example: Epileptic seizures



Further examples

Count data

- number of side effects after treatment
- number of episodes in multiple sclerosis patients

Ordinal data

- 5 category assessment scale from 'very good' to 'very bad' for satisfaction with pain relief

Binary data

- Presence (yes/no) of respiratory infection in children
- Pain relief (yes/no) after analgesic treatment
- Correct solution (yes/no) to a task or test question
- . . .

Overview Chapter 8 - Non-normal longitudinal data

8.1 Data examples

8.2 Generalized linear models (GLMs)

8.3 Extending GLMs to longitudinal data

Recap: The exponential family

A distribution belongs to the **exponential family** if its density is of the form

$$f(y|\theta,\phi) = \exp\{[y\theta - \psi(\theta)]/\phi + c(y,\phi)\},$$

where θ (natural parameter) and ϕ (scale parameter) are unknown and $\psi(.)$ and $c(.,.)$ are known functions. The first two moments are given by

$$\begin{array}{lcccc} \mathsf{E}(Y) & = & \mu & = & \psi'(\theta) \\ \mathsf{Var}(Y) & = & \sigma^2 & = & \phi\psi''(\theta). \end{array}$$

Importantly, μ and σ^2 are connected via

$$\sigma^2 = \phi\psi''(\psi'^{-1}(\mu)) =: \phi v(\mu),$$

where v then corresponds to the variance function.

Recap: Generalized linear models (GLMs)

- $Y_1, \ldots, Y_N$ independent outcome variables
- $\mathbf{x}_1, \ldots, \mathbf{x}_N$ corresponding vectors of p covariates

Assumptions:

- Each Y_i has a density $f(y_i|\theta_i, \phi)$ from the exponential family with obervation-specific θ_i.
- $\mu_i = E(Y_i)$ is of the form

$$\mu_i = h(\eta_i) = h(\mathbf{x}_i^T \boldsymbol{\beta}),$$

where $h(.)$ is a known function and $\boldsymbol{\beta}$ is a p-vector of regression coefficients. The function $g = h^{-1}$ is called **link function**. If the natural (or canonical) link function $h(.) = \psi'(.)$ is used, the natural parameter is assumed to satisfy the linear relationship $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta}$.

Special cases

distribution	ϕ	variance function	canonical link function
Normal	σ^2	$v(\mu) = 1$	$\mu = \eta$ (identity)
Bernoulli	1	$v(\mu) = \mu(1-\mu)$	$\log\left(\frac{\mu}{1-\mu}\right) = \eta$ (logit)
Poisson	1	$v(\mu) = \mu$	$\log(\mu) = \eta$ (log)

ML estimation for GLMs

- For independent observations, the log-likelihood is given by

$$l(\boldsymbol{\beta}, \phi) = \frac{1}{\phi} \sum_{i=1}^{N} [y_i \theta_i - \psi(\theta_i)] + \sum_{i=1}^{N} c(y_i, \phi),$$

  where θ_i depends on $\boldsymbol{\beta}$.

- Taking the derivative w.r.t. $\boldsymbol{\beta}$ and equating it to zero ($\frac{\partial l}{\partial \boldsymbol{\beta}} = 0$) yields the ML estimator for $\boldsymbol{\beta}$ as solution to the score equations

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{\partial \theta_i}{\partial \boldsymbol{\beta}} [y_i - \psi'(\theta_i)] = \mathbf{0}.$$

- Standard algorithm: Newton-Raphson (iterative).

ML estimation for GLMs

As $\mu_i = \psi'(\theta_i)$ and $v_i := v(\mu_i) = \psi''(\theta_i)$, we have

$$\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \psi''(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\beta}} = v_i \frac{\partial \theta_i}{\partial \boldsymbol{\beta}}$$

and thus

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} v_i^{-1}(y_i - \mu_i) = \mathbf{0}.$$

We will need this in Chapter 10 for the generalized estimating equations.

For the canonical link function, we have the special case $\partial\theta_i/\partial\boldsymbol{\beta} = \mathbf{x}_i^T$, i.e.

$$\sum_{i=1}^{N} \mathbf{x}_i^T(y_i - \mu_i) = \mathbf{0}.$$

Overview Chapter 8 - Non-normal longitudinal data

8.1 Data examples

8.2 Generalized linear models (GLMs)

8.3 Extending GLMs to longitudinal data

Overview non-normal longitudinal data

For the extenstion of generalized linear models (GLMs) to longitudinal data there are essentially three approaches:

1. **Marginal models** (Chapter 10): Model marginal correlation and/or account for it with robust standard errors (GEE).

2. **Mixed models** (Chapter 9): Observations are correlated, because they are from the same subject and share the same underlying processes.

3. **Transition/Markov models** (at the end if there is time): Observations are correlated, because the past influences the presence.
   (Typical here: Past $=$ last q observations $\rightarrow$ Markov property.)

Difference to the normal response case

- For the linear mixed model we had

$$\mathsf{E}(\mathbf{Y}_i) = \mathsf{E}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathsf{E}(\mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta}.$$

- Marginal view:

$$\mathbf{Y}_i \sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i)$$

- $\boldsymbol{\beta}$ can be interpreted as a population parameter.

- For non-normal responses this is no longer the case:
  - In the marginal model inference is about the population ("population-average").
  - In mixed models the focus is on the subjects ("subject-specific").

Short overview marginal models

Similarly to 8.2:

$$\text{Var}(Y_{ij}) = \phi v(\mu_{ij})$$

where $\mu_{ij} = \text{E}(Y_{ij})$ and

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

Difference to the GLM: $Y_{i1}, \ldots, Y_{in_i}$ are (conditional on the covariates) associated.

Assumption: This association is a function of additional parameters $\boldsymbol{\alpha}$. For example, $\boldsymbol{\alpha}$ can include pairwise correlations or log-odds-ratios.

Example logistic model

Consider as an example a logistic model for the probability of having an infection (Y), given (no) vitamin A deficiency (x). Cross-sectional model:

$$\text{logit Pr}(Y_i = 1) = \beta_0 + \beta_1 x_i.$$

Marginal model: Models mean, variance and correlation

$$\begin{aligned} \text{logit Pr}(Y_{ij} = 1) = \text{logit } \mu_{ij} &= \beta_0 + \beta_1 x_{ij} \\ \text{Var}(Y_{ij}) &= \mu_{ij}(1 - \mu_{ij}) \\ \text{Corr}(Y_{ij}, Y_{ik}) &= \rho(\mu_{ij}, \mu_{ik}; \boldsymbol{\alpha}) \end{aligned}$$

Example logistic model

Mixed model: Each individual has its own propensity for an infection

$$\text{logit Pr}(Y_{ij} = 1|b_i) = (\beta_0^* + b_i) + \beta_1^* x_{ij},$$
$$b_i \overset{iid}{\sim} (0, d^2)$$

Transition model: The probability for an infection depends on whether there was an infection at the last visit

$$\text{logit Pr}(Y_{ij} = 1|Y_{ij-1}, \ldots, Y_{i1}) \quad = \quad \beta_0^{**} + \beta_1^{**} x_{ij} + \gamma Y_{ij-1}$$

Example logistic model

Now, the β parameters have quite different interpretations and will typically differ.

- β_1 is the log-odds ratio of infection between vitamin A deficient and replete children. It is a **population-averaged parameter** (marginal interpretation).

- β_1^* is the log-odds ratio of infection when a child is deficient relative to when **that same child** is not (conditional interpretation, conditional on individual propensity to infection). The resulting change in absolute risk depends on the baseline rate for that child.

Example logistic model

- β_1^{**} is the log-odds ratio of infection for vitamin A deficiency versus repletion **among the group of children free of infection** [the infected group] at the last visit (conditional interpretation, conditional on infection status at last visit).

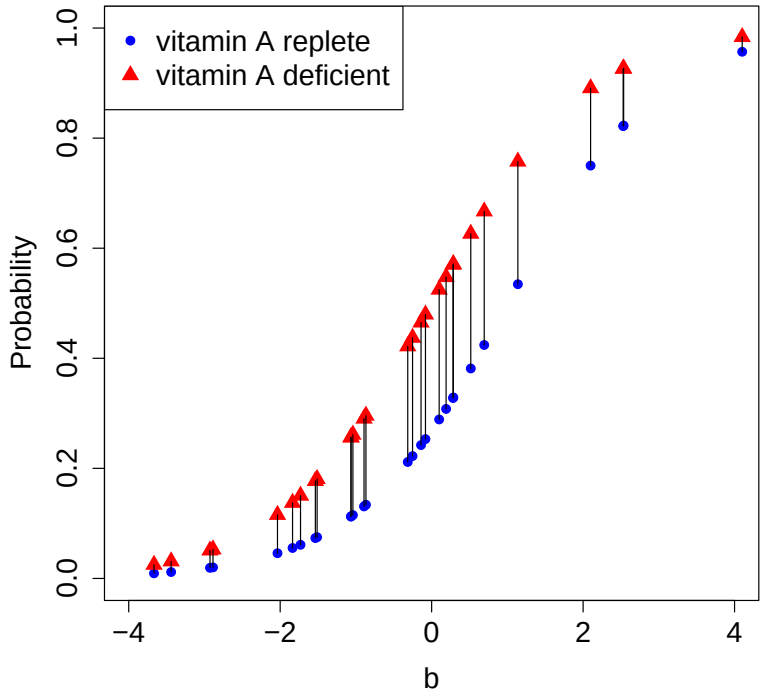
For the logistic regression: (Neuhaus et al, 1991; Zeger et al, 1988)

$|\beta_1| \leq |\beta_1^*|$ with equality iff $\beta_1^* = 0$, and an increase in discrepancy with Var(b_i).

If $b_i \overset{iid}{\sim} N(0, d^2)$, then $\beta_1 \approx (c^2 d^2 + 1)^{-1/2} \beta_1^*$ with $c^2 \approx 0.346$.

(Analogously component-wise for a vector $\boldsymbol{\beta}$.)

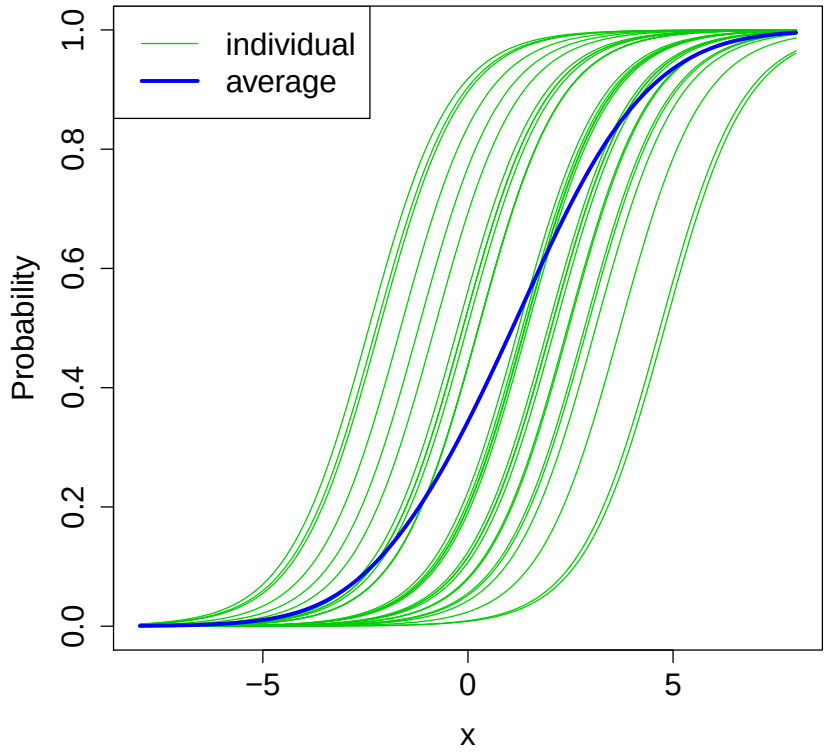
Marginal and mixed logistic model



Marginal probability: (Thus: marginal odds-ratio $\neq \exp(\beta_1^*)$)

$$\Pr(Y_{ij}=1) = \int \Pr(Y_{ij}=1|b_i)dF(b_i) = \int \frac{\exp(\beta_0^* + b_i + \beta_1^* x_{ij})}{1+\exp(\beta_0^* + b_i + \beta_1^* x_{ij})} f(b_i)db_i$$

Marginal and mixed logistic model



Steep rate of increase in individual curves attenuated in average curve (graph for continuous x).

The three approaches - some pros & cons

Marginal models (Chapter 10):

+ Separate modeling of mean and correlation. Correlation model does not change interpretation of β parameters.

± Are appropriate for inference about the population mean.

+ Only requires specification of the first two moments, not the entire likelihood (less assumptions).

− No likelihood-based inference (but instead: GEE).

+ Can easily accommodate unequally spaced time points and unbalanced data.

The three approaches - some pros & cons

Mixed models (Chapters 3-7, 9):

$\pm$ Can make inference about individuals rather than population averages.

$+$ Can easily accommodate unequally spaced time points, unbalanced data.

$+$ Flexible, can additonally model clustered data, smooth functions (penalized spline smoothing) etc.

$\pm$ Parsimonious modeling of covariance. But random effects imply specific correlation structure, less flexibility.

$+$ Allows likelihood-based inference.

$-$ Fitting of models often hard (especially for generalized case).

The three approaches - some pros & cons

Transition/Markov models:

– Modeling and inference easiest for equally spaced time points t_{ij}.

\+ Might be a very meaningful way to think of the underlying process in some cases (e.g. for categorical data when thinking of transitions between "states")

\+ Allows likelihood-based inference (typically conditional on first q observations).