

bild: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
Locust

References

# bild: a package for Binary Longitudinal Data

M. H. Gonçalves<sup>1</sup>   M. S. Cabral<sup>2</sup>   A. Azzalini<sup>3</sup>

<sup>1</sup>University of Algarve, Portugal

<sup>2</sup>University of Lisbon, Portugal

<sup>3</sup>Università di Padova, Italy

useR!2010, July 21-23

bild: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
Locust

References

- 1 Introduction
- 2 Parametric models for binary data
  - Binary Markov Chains
  - Random effects
- 3 bild function
- 4 Example
  - Locust
- 5 References

# What is bild?

bild: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
Locust

References

- an R package for Binary Longitudinal Data.
- parametric models with exact likelihood MLE.
- allows serial dependence and random effects.
- graphical analysis.
- allows missing data (with some restrictions).
- R code interfaced to some Fortran routines.
- S4 methods.

# Marginal Model

bild: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
Locust

References

Let  $y_{it}$  ( $t = 1, \dots, T_i$ )  $\in \{0, 1\}$  be the response value at time  $t$  from subject  $i$  ( $i = 1, \dots, n$ ). The parameter of interest is the marginal probability of success, that is related to the covariates via a logistic regression model,

$$\text{logit } P(Y_{it} = 1) = \text{logit}(\theta_{it}) = x_{it}^{\top} \beta \quad (1)$$

- $Y_{it}$  its generating random variable whose mean values is  $\theta_{it} = P(Y_{it} = 1)$ .
- $x_{it}$  a set of  $p$  covariates associated to each observation time and each subject.
- $\beta$  is a  $p$  vector of unknown parameters.

# Structure Dependence

bild: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
Locust

References

Taking into account that successive observations from the same individual cannot be assumed to be independent, the model considered for serial dependence is of Markovian type of first order ( $\psi_1$ ) or of second order ( $\psi_2$ ).

$$\begin{aligned}OR(Y_{t-1}, Y_{t-2}) &= \psi_1 = OR(Y_{t-1}, Y_t) \\OR(Y_{t-2}, Y_t | Y_{t-1} = 0) &= \psi_2 = OR(Y_{t-2}, Y_t | Y_{t-1} = 1)\end{aligned}$$

# Random intercept model

bild: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
Locust

References

This package allows individual random effects by adding the component  $b_i \sim N(0, \sigma^2)$  in (1) leading to the logistic model with random intercept

$$\text{logit } P(Y_{it} = 1) = x_{it}^T \beta + b_i, \quad (2)$$

the  $b_i$ 's are assumed to be independent from each other.

- $\omega = \log \sigma^2$  for numerical convenience and to improve accuracy of the asymptotic approximation to the distribution of MLEs.
- Integrals are computed using adaptive Gaussian quadrature.

# function bild()

bild: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

**bild function**

Example  
Locust

References

```
bild(formula=formula(data), data, time, id, subSET, aggregate=
  FALSE, start=NULL, trace=FALSE, dependence="ind", method=
  "BFGS", control=bildControl(), integrate=bildIntegrate())
```

bild() takes the following arguments:

- **formula** a description of the model to be fitted of the form response predictors
- **data** a data frame containing the variables in the formula. NA values are allowed.
- **time** a string that matches the name of the time variable in data.
- **id** a string that matches the name of the id variable in data.
- **subSET** an optional expression indicating the subset of the rows of data that should be used in the fit.

# function `build()`

`build`: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

`build` function

Example  
Locust

References

- **aggregate** a string that permits the user identify the factor to be used in the `plot`-methods.
- **start** a vector of initial values for the nuisance parameters of the likelihood.
- **trace** logical flag: if TRUE, details of the nonlinear optimization are printed.
- **dependence** expression stating which dependence structure should be used in the fit.
- **method** the method to be used in the optimization process for function `optim()`.
- **control** a list of algorithmic constants for the optimizer `optim`.
- **integrate** a list of algorithmic constants for the computation of a definite integral using a FORTRAN-77 subroutine.



# selecting the **dependence** structure

bild: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

**bild function**

Example  
Locust

References

bild() allows the following choices for the serial dependence structure:

- independence (**dependence**="ind")
- first order Markov Chain (**dependence**="MC1")
- second order Markov Chain (**dependence**="MC2")
- first order Markov Chain with random intercept (**dependence**="MC1R")
- second order Markov Chain with random intercept (**dependence**="MC2R")

# controlling `build()` fitting

`build`: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

**`build` function**

Example

Locust

References

`buildControl()` and `buildIntegrate()` are two auxiliary functions as user interface for `build()` fitting having the following arguments:

- `buildControl(maxit=100, abstol=1e-006, reltol=1e-006)`
- `buildIntegrate(li=-4, ls=4, epsabs=.Machine$double.eps^.25, epsrel=.Machine$double.eps^.25, limit=100, key=6, lig=-4, lsg=4)`

# Locust data

bild: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example

**Locust**

References

Observations on locomotory behaviour of 24 locusts observed at 161 time points: moving vs non-moving.

```
> str(locust)
'data.frame': 3864 obs. of 5 variables:
 $ id : int 1 1 1 1 1 1 1 1 1 1 ...
 $ move: num 0 0 0 0 0 0 0 0 0 0 ...
 $ sex : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
 $ time: num 0.00833 0.01667 0.025 0.03333 0.04167 ...
 $ feed: Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
```

# function `build()` to fit the model

`build`: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

`build` function

Example

**Locust**

References

the function `build()` is called to fit the model

$$\text{logit}(\theta_{it}^b) = \beta_0 + b_i + \beta_1 \text{time} + \beta_2 \text{time}^2 + \beta_3 \text{feed} + \beta_4 \text{time} \times \text{feed} + \beta_5 \text{time}^2 \times \text{feed}$$

using a dependence structure MC2R

```
> Integ <- buildIntegrate(li=-2.5, ls=2.5, lig=-2.5, lsg=2.5)

> locust2r <- build(move ~ (time + I(time^2))*feed, data=locust,
+   aggregate=feed, dependence="MC2R", integrate=Integ)
```

# summary

bild: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
Locust

References

```
> summary(locust2r)
Call: bild(formula=move ~ (time + I(time^2))*feed, data=locust,
  aggregate=feed, dependence="MC2R", integrate=Integ)
```

```
Number of profiles in the dataset: 24
Number of profiles used in the fit: 24
Log likelihood: -1495.802
AIC: 3009.603
```

Coefficients:

	Label	Value	Std. Error	t value	p-value
(Intercept)	1	-1.8145720	0.3680344	-4.930	0.000001
time	2	5.3239922	0.7594622	7.010	0.000000
I(time^2)	3	-2.9392803	0.5315684	-5.529	0.000000
feed1	4	-2.3050095	0.6616485	-3.484	0.000494
time:feed1	5	-6.0737691	1.5647952	-3.882	0.000104
I(time^2):feed1	6	5.0664871	1.0435033	4.855	0.000001
log.psi1	7	0.9811151	0.1125461	8.717	0.000000
log.psi2	8	0.5583951	0.1030314	5.420	0.000000

```
Random effect (omega):      Value      Std. Error
                             -0.002794812  0.350137839
```

```
Message: 0
```

# probability plot

```
> plot(locust2r, which = 5, ylab = "probability of locomoting")
```

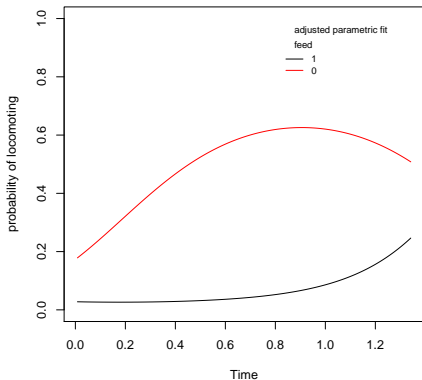


Figure 1: probability of locomoting of locust data for MC2R

bild: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example

**Locust**

References

# residual plots

bild: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example

**Locust**

References

The residual analysis can be summarized using plot-methods setting `which=1` for Residuals vs Fitted, `which=2` for Residuals vs Time, `which=3` for ACF residuals and `which=4` for PACF residuals

```
> par(mfrow = c(2, 2))  
> plot(locust2r, which = 1)  
> plot(locust2r, which = 2)  
> plot(locust2r, which = 3)  
> plot(locust2r, which = 4)  
> par(mfrow = c(1, 1))
```

# residual plots

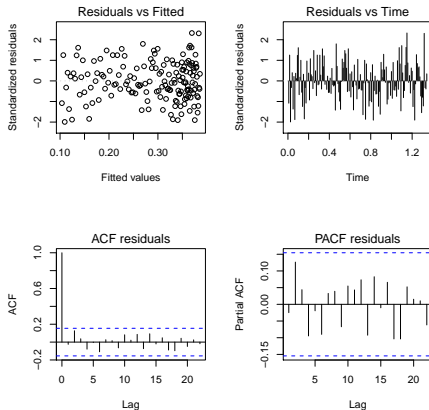


Figure 2: residual plots of locust data for MC2R

bild: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example

**Locust**

References



# individual mean profiles

bild: a  
package for  
Binary  
Longitudinal  
Data

Goncalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
**Locust**

References

The individual mean profile to all subjects (by default) or to a subset (`subSET=`) is obtained choosing `which=6` in `plot-methods`. The identification of the subjects is also allowed by setting `ident=TRUE`. These options are only available for random intercept models.

```
> par(mfrow = c(1, 2))
> plot(locust2r, which = 6, ylab = "probability of locomoting",
      main = "Feed & Unfeed groups")
> plot(locust2r, which = 6, ident = TRUE, subSET = feed == "0",
      ylab = "probability of locomoting", main = "Unfeed group")
> par(mfrow = c(1, 1))
```

# individual mean profiles

bild: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example

**Locust**

References

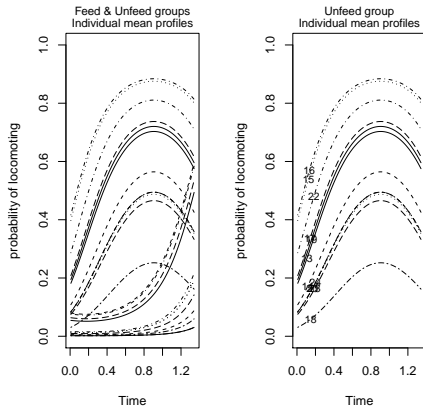


Figure 3: individual mean profiles of locust data for MC2R

# References

bild: a  
package for  
Binary  
Longitudinal  
Data

Gonçalves et  
al.

Outline

Introduction

Parametric  
models for  
binary data

Binary  
Markov  
Chains  
Random  
effects

bild function

Example  
Locust

References



Azzalini, A. (1994). Logistic Regression for Autocorrelated Data with Application to Repeated Measures. *Biometrika*, 81, 767-775. Amendment: vol. 84 (1997), 989.



Gonçalves, M. H. and Azzalini, A. (2008). Using Markov chains for marginal modelling of binary longitudinal data in an exact likelihood approach. *Metron*, LXVI, n.2, 157-181.



Gonçalves, M. H., Cabral, M. S., and Azzalini, A. (2010). bild: a package for Binary Longitudinal Data, Version 1.0. URL: <http://CRAN.R-project.org/package=bild>