On likelihood methods for binary longitudinal data

Adelchi Azzalini

including work with Helena Gonçalves, Univ. Algarve, Portugal

51ª Reunião Anual da Região Brasileira
da Sociedade Internacional de Biometria (RBRAS)
24–26 May 2006

Binary longitudinal data

- A set of n individuals observed along time
- Response variable Y is binary (values: 0 and 1, say)
- Notation: response at time t from subject i is

$$y_{it} = \begin{cases} 0 \\ 1 \end{cases}$$

e.g. ith individual *profile* is $y_i = (1, 1, 0, 1, 0, 1)$

- Covariates, X_{it}, also recorded
- In general, want to relate X's and Y

Binary longitudinal data

- A set of n individuals observed along time
- Response variable Y is binary (values: 0 and 1, say)
- Notation: response at time t from subject i is

$$y_{it} = \begin{cases} 0 \\ 1 \end{cases}$$

  e.g. ith individual *profile* is $y_i = (1, 1, 0, 1, 0, 1)$
- Covariates, X_{it}, also recorded
- In general, want to relate X's and Y

Binary longitudinal data

- A set of n individuals observed along time
- Response variable Y is binary (values: 0 and 1, say)
- Notation: response at time t from subject i is

$$y_{it} = \begin{cases} 0 \\ 1 \end{cases}$$

e.g. ith individual *profile* is $y_i = (1, 1, 0, 1, 0, 1)$

- Covariates, X_{it}, also recorded
- In general, want to relate X's and Y

Binary longitudinal data

- A set of n individuals observed along time
- Response variable Y is binary (values: 0 and 1, say)
- Notation: response at time t from subject i is

$$y_{it} = \begin{cases} 0 \\ 1 \end{cases}$$

e.g. ith individual *profile* is $y_i = (1, 1, 0, 1, 0, 1)$

- Covariates, X_{it}, also recorded
- In general, want to relate X's and Y

Binary longitudinal data

- A set of n individuals observed along time
- Response variable Y is binary (values: 0 and 1, say)
- Notation: response at time t from subject i is

$$y_{it} = \begin{cases} 0 \\ 1 \end{cases}$$

e.g. ith individual *profile* is $y_i = (1, 1, 0, 1, 0, 1)$

- Covariates, X_{it}, also recorded
- In general, want to relate X's and Y

Binary longitudinal data

- A set of n individuals observed along time
- Response variable Y is binary (values: 0 and 1, say)
- Notation: response at time t from subject i is

$$y_{it} = \begin{cases} 0 \\ 1 \end{cases}$$

e.g. ith individual *profile* is $y_i = (1, 1, 0, 1, 0, 1)$

- Covariates, X_{it}, also recorded
- In general, want to relate X's and Y

What is the purpose of this talk

- Partly a review and discussion
- Partly presentation of specific results
  joint work with Helena Gonçalves (U. Algarve, Portugal)

What is the purpose of this talk

- Partly a review and discussion
- Partly presentation of specific results
  joint work with Helena Gonçalves (U. Algarve, Portugal)

What is the purpose of this talk

- Partly a review and discussion
- Partly presentation of specific results
  joint work with Helena Gonçalves (U. Algarve, Portugal)

Outline

Approaches: transition models

Transition models

- model transitions of an individual

$$\mathbb{P}\{Y_{it} = 1 | \texttt{past profile, covariates}\}$$

e. g.

$$\text{logit}\left(\mathbb{P}\left\{Y_{it} = 1 | y_{i,t-1}, x_{it}\right\}\right) = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x_{it}$$

- simple in formulation
- writing log-likelihood is immediate
  using direct Markov chain connection
- OK if we want to model transitions
- but often we want to model marginal probability

Approaches: transition models

Transition models

- model transitions of an individual

$$\mathbb{P}\{Y_{it} = 1|\texttt{past profile, covariates}\}$$

e. g.
$$\text{logit}\left(\mathbb{P}\left\{Y_{it} = 1|y_{i,t-1}, x_{it}\right\}\right) = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x_{it}$$

- simple in formulation
- writing log-likelihood is immediate
using direct Markov chain connection
- OK if we want to model transitions
- but often we want to model marginal probability

Approaches: transition models

Transition models

- model transitions of an individual

$$\mathbb{P}\{Y_{it} = 1 | \texttt{past profile, covariates}\}$$

e. g.

$$\text{logit}\left(\mathbb{P}\{Y_{it} = 1 | y_{i,t-1}, x_{it}\}\right) = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x_{it}$$

- simple in formulation
- writing log-likelihood is immediate
  using direct Markov chain connection
- OK if we want to model transitions
- but often we want to model marginal probability

Approaches: transition models

Transition models

- model transitions of an individual

$$\mathbb{P}\{Y_{it} = 1 | \texttt{past profile, covariates}\}$$

e. g.

$$\operatorname{logit}\left(\mathbb{P}\left\{Y_{it} = 1 | y_{i,t-1}, x_{it}\right\}\right) = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x_{it}$$

- simple in formulation
- writing log-likelihood is immediate
  using direct Markov chain connection
- OK if we want to model transitions
- but often we want to model marginal probability

Approaches: transition models

Transition models

- model transitions of an individual

$$\mathbb{P}\{Y_{it} = 1|\texttt{past profile, covariates}\}$$

  e. g.
  $\text{logit}\left(\mathbb{P}\left\{Y_{it} = 1|y_{i,t-1}, x_{it}\right\}\right) = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x_{it}$
- simple in formulation
- writing log-likelihood is immediate
  using direct Markov chain connection
- OK if we want to model transitions
- but often we want to model marginal probability

Outline

Approaches: marginal models

Marginal models

- in many cases, interest is mostly in

$$\mathbb{P}\{Y_{it} = 1 | \texttt{covariates}\}$$

  allowing for dependence within a given profile y_i
- dependence structure is 'nuisance component'
- difficult to formulate fully-specified stochastic models with prescribed properties
- alternative route: do not attempt full stochastic specification

Approaches: marginal models

Marginal models

- in many cases, interest is mostly in

$$\mathbb{P}\{Y_{it} = 1 | \texttt{covariates}\}$$

  allowing for dependence within a given profile y_i
- dependence structure is 'nuisance component'
- difficult to formulate fully-specified stochastic models with prescribed properties
- alternative route: do not attempt full stochastic specification

Approaches: marginal models

Marginal models

- in many cases, interest is mostly in

$$\mathbb{P}\{Y_{it} = 1 | \texttt{covariates}\}$$

  allowing for dependence within a given profile y_i
- dependence structure is 'nuisance component'
- difficult to formulate fully-specified stochastic models with prescribed properties
- alternative route: do not attempt full stochastic specification

Approaches: marginal models via GEE

Generalized estimating equations (GEE)

- no full stochastic specification
- model quantity of interest: $\mathbb{P}\{Y_{it} = 1|\texttt{covariates}\}$
- requires specification of a 'working correlation structure', to accomodate correlation structure, compute std.errors 'adjusted' for presence of dependence
- Ok if we are only interested in the population behaviour
- cannot be used to tackle questions on individual profiles eg. $\mathbb{P}\{y_{i4} = 1|\text{past} = (1, 0, 1), x_{it}\} = ?$

Approaches: marginal models via GEE

Generalized estimating equations (GEE)

- no full stochastic specification
- model quantity of interest: $\mathbb{P}\{Y_{it} = 1 | \texttt{covariates}\}$
- requires specification of a ‘working correlation structure’, to accomodate correlation structure, compute std.errors ‘adjusted’ for presence of dependence
- Ok if we are only interested in the population behaviour
- cannot be used to tackle questions on individual profiles eg. $\mathbb{P}\{y_{i4} = 1 | \text{past} = (1, 0, 1), x_{it}\} = ?$

Approaches: marginal models via GEE

Generalized estimating equations (GEE)

- no full stochastic specification
- model quantity of interest: $\mathbb{P}\{Y_{it} = 1 | \texttt{covariates}\}$
- requires specification of a ‘working correlation structure’, to accomodate correlation structure, compute std.errors ‘adjusted’ for presence of dependence
- Ok if we are only interested in the population behaviour
- cannot be used to tackle questions on individual profiles eg. $\mathbb{P}\{y_{i4} = 1 | \text{past} = (1, 0, 1), x_{it}\} = ?$

Approaches: marginal models via GEE

Generalized estimating equations (GEE)

- no full stochastic specification
- model quantity of interest: $\mathbb{P}\{Y_{it} = 1 | \texttt{covariates}\}$
- requires specification of a 'working correlation structure', to accomodate correlation structure, compute std.errors 'adjusted' for presence of dependence
- Ok if we are only interested in the population behaviour
- cannot be used to tackle questions on individual profiles eg. $\mathbb{P}\{y_{i4} = 1 | \text{past} = (1, 0, 1), x_{it}\} = ?$

Approaches: marginal models via GEE

Generalized estimating equations (GEE)

- no full stochastic specification
- model quantity of interest: $\mathbb{P}\{Y_{it} = 1 | \texttt{covariates}\}$
- requires specification of a 'working correlation structure', to accomodate correlation structure, compute std.errors 'adjusted' for presence of dependence
- Ok if we are only interested in the population behaviour
- cannot be used to tackle questions on individual profiles eg. $\mathbb{P}\{y_{i4} = 1 | \text{past} = (1, 0, 1), x_{it}\} = ?$

Outline

Approaches: marginal likelihood-based models

Want (full) stochastic formulation for marginal modelling

- purpose:
  (a) use standard likelihood-based inferences
  (b) model population as well as individual behaviour
- aim at stochastic model for profile Y_i such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it}$$

  is represented by

$$\text{logit}\, \theta_{it} = x^\top \beta\,,$$

  or possibly other link function in place of logit
- from full stochastic specification, modelling of (serial) dependence must be allowed
- special interest for individual random effects

Approaches: marginal likelihood-based models

Want (full) stochastic formulation for marginal modelling

- purpose:
  (a) use standard likelihood-based inferences
  (b) model population as well as individual behaviour
- aim at stochastic model for profile Y_i such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it}$$

  is represented by

$$\text{logit}\, \theta_{it} = x^\top \beta\,,$$

  or possibly other link function in place of logit
- from full stochastic specification, modelling of (serial) dependence must be allowed
- special interest for individual random effects

Approaches: marginal likelihood-based models

Want (full) stochastic formulation for marginal modelling

- purpose:
  (a) use standard likelihood-based inferences
  (b) model population as well as individual behaviour
- aim at stochastic model for profile Y_i such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it}$$

  is represented by

$$\operatorname{logit} \theta_{it} = x^\top \beta\,,$$

  or possibly other link function in place of logit
- from full stochastic specification, modelling of (serial) dependence must be allowed
- special interest for individual random effects

Approaches: marginal likelihood-based models

Want (full) stochastic formulation for marginal modelling

- purpose:
  (a) use standard likelihood-based inferences
  (b) model population as well as individual behaviour
- aim at stochastic model for profile Y_i such that

  $$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it}$$

  is represented by

  $$\operatorname{logit} \theta_{it} = x^\top \beta \,,$$

  or possibly other link function in place of logit
- from full stochastic specification, modelling of (serial) dependence must be allowed
- special interest for individual random effects

Approaches: marginal likelihood-based models

Want (full) stochastic formulation for marginal modelling

- purpose:
  (a) use standard likelihood-based inferences
  (b) model population as well as individual behaviour
- aim at stochastic model for profile Y_i such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it}$$

  is represented by

$$\operatorname{logit} \theta_{it} = x^\top \beta \,,$$

  or possibly other link function in place of logit
- from full stochastic specification, modelling of (serial) dependence must be allowed
- special interest for individual random effects

'Mixed parameter' formulation

'Mixed parameter' formulation (Fitzmaurice & Laird, 1993)

- based on mixed mean and canonical association parameters in exponential families
- orthogonal regression parameters, β, and association parameters (α)
- various desirable features:
  - robustness to misspecification of time dependence
  - $\mathrm{var}\left\{\hat{\beta}\right\}$ not influenced by knowledge of α, at least asymptotically
- some drawbacks:
  - association parameters are *conditional* log-odds ratios
  - distribution is not "reproducible", as profile length varies

‘Mixed parameter’ formulation

‘Mixed parameter’ formulation (Fitzmaurice & Laird, 1993)

- based on mixed mean and canonical association parameters in exponential families
- orthogonal regression parameters, β, and association parameters (α)
- various desirable features:
  - robustness to misspecification of time dependence
  - $\text{var}\left\{\hat{\beta}\right\}$ not influenced by knowledge of α, at least asymptotically
- some drawbacks:
  - association parameters are *conditional* log-odds ratios
  - distribution is not “reproducible”, as profile length varies

‘Mixed parameter’ formulation

‘Mixed parameter’ formulation (Fitzmaurice & Laird, 1993)

- based on mixed mean and canonical association parameters in exponential families
- orthogonal regression parameters, β, and association parameters (α)
- various desirable features:
  - robustness to misspecification of time dependence
  - $\mathrm{var}\left\{\hat{\beta}\right\}$ not influenced by knowledge of α, at least asymptotically
- some drawbacks:
  - association parameters are *conditional* log-odds ratios
  - distribution is not “reproducible”, as profile length varies

‘Mixed parameter’ formulation

‘Mixed parameter’ formulation (Fitzmaurice & Laird, 1993)

- based on mixed mean and canonical association parameters in exponential families
- orthogonal regression parameters, β, and association parameters (α)
- various desirable features:
  - robustness to misspecification of time dependence
  - $\text{var}\left\{\hat{\beta}\right\}$ not influenced by knowledge of α, at least asymptotically
- some drawbacks:
  - association parameters are *conditional* log-odds ratios
  - distribution is not “reproducible”, as profile length varies

Markov chain re-parametrization

Markov chain re-parametrization (Azzalini, 1994)

- build 1st order Markov chain such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it} = \operatorname{logit}^{-1}(x^\top \beta)$$

  and

$$OR(Y_{it}, Y_{i,t-1}) = \psi$$

  by solving equation to get suitable transition matrix, which depends on $(y_{i,t-1}, \theta_{it}, \theta_{i,t-1}, \psi)$
- desirable features:
  - likelihood expression is simple
  - parameter interpretation is transparent
  - orthogonal parameters, β and ψ
  - time length T_i unrestricted, possibly non-constant
- drawback: only one parameter to regulate dependence

Markov chain re-parametrization

Markov chain re-parametrization (Azzalini, 1994)

- build 1st order Markov chain such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it} = \operatorname{logit}^{-1}(x^\top \beta)$$

  and

$$OR(Y_{it}, Y_{i,t-1}) = \psi$$

  by solving equation to get suitable transition matrix, which depends on $(y_{i,t-1}, \theta_{it}, \theta_{i,t-1}, \psi)$
- desirable features:
  - likelihood expression is simple
  - parameter interpretation is transparent
  - orthogonal parameters, β and ψ
  - time length T_i unrestricted, possibly non-constant
- drawback: only one parameter to regulate dependence

Markov chain re-parametrization

Markov chain re-parametrization (Azzalini, 1994)

- build 1st order Markov chain such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it} = \mathrm{logit}^{-1}(x^\top \beta)$$

  and

$$OR(Y_{it}, Y_{i,t-1}) = \psi$$

  by solving equation to get suitable transition matrix, which depends on $(y_{i,t-1}, \theta_{it}, \theta_{i,t-1}, \psi)$
- desirable features:
  - likelihood expression is simple
  - parameter interpretation is transparent
  - orthogonal parameters, β and ψ
  - time length T_i unrestricted, possibly non-constant
- drawback: only one parameter to regulate dependence

Markov chain re-parametrization

Markov chain re-parametrization (Azzalini, 1994)

- build 1st order Markov chain such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it} = \text{logit}^{-1}(x^\top \beta)$$

and

$$OR(Y_{it}, Y_{i,t-1}) = \psi$$

by solving equation to get suitable transition matrix, which depends on $(y_{i,t-1}, \theta_{it}, \theta_{i,t-1}, \psi)$

- desirable features:
  - likelihood expression is simple
  - parameter interpretation is transparent
  - orthogonal parameters, β and ψ
  - time length T_i unrestricted, possibly non-constant
- drawback: only one parameter to regulate dependence

Markov chain re-parametrization

Markov chain re-parametrization (Azzalini, 1994)

- build 1st order Markov chain such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it} = \mathrm{logit}^{-1}(x^\top \beta)$$

  and

$$OR(Y_{it}, Y_{i,t-1}) = \psi$$

  by solving equation to get suitable transition matrix, which depends on $(y_{i,t-1}, \theta_{it}, \theta_{i,t-1}, \psi)$
- desirable features:
  - likelihood expression is simple
  - parameter interpretation is transparent
  - orthogonal parameters, β and ψ
  - time length T_i unrestricted, possibly non-constant
- drawback: only one parameter to regulate dependence

Markov chain re-parametrization

Markov chain re-parametrization (Azzalini, 1994)

- build 1st order Markov chain such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it} = \mathrm{logit}^{-1}(x^\top \beta)$$

  and

$$OR(Y_{it}, Y_{i,t-1}) = \psi$$

  by solving equation to get suitable transition matrix, which depends on $(y_{i,t-1}, \theta_{it}, \theta_{i,t-1}, \psi)$
- desirable features:
  - likelihood expression is simple
  - parameter interpretation is transparent
  - orthogonal parameters, β and ψ
  - time length T_i unrestricted, possibly non-constant
- drawback: only one parameter to regulate dependence

Markov chain re-parametrization

Markov chain re-parametrization (Azzalini, 1994)

- build 1st order Markov chain such that

$$\mathbb{P}\{Y_{it} = 1 | X_{it} = x\} = \theta_{it} = \text{logit}^{-1}(x^\top \beta)$$

  and

$$OR(Y_{it}, Y_{i,t-1}) = \psi$$

  by solving equation to get suitable transition matrix, which depends on $(y_{i,t-1}, \theta_{it}, \theta_{i,t-1}, \psi)$
- desirable features:
  - likelihood expression is simple
  - parameter interpretation is transparent
  - orthogonal parameters, β and ψ
  - time length T_i unrestricted, possibly non-constant
- drawback: only one parameter to regulate dependence

Random effects

- random effects, in a simple case:

$$\text{logit}(\mathbb{P}\{Y_{it}|x, b_i\}) = x^\top\beta + b_i$$

where $b_i \sim N(0, \sigma^2)$

- problems:
(a) computational, due to integration wrt dist'n of b_i
(b) interpretation of parameters, since

$$\mathbb{E}\left\{\frac{e^{\eta+b}}{1+e^{\eta+b}}\right\} \neq \frac{e^\eta}{1+e^\eta}$$

where $\eta = x^\top\beta$, hence meaning of β changes

Random effects

- random effects, in a simple case:

$$\operatorname{logit}(\mathbb{P}\{Y_{it}|x, b_i\}) = x^\top \beta + b_i$$

where $b_i \sim N(0, \sigma^2)$

- problems:

(a) computational, due to integration wrt dist'n of b_i

(b) interpretation of parameters, since

$$\mathbb{E}\left\{\frac{e^{\eta+b}}{1+e^{\eta+b}}\right\} \neq \frac{e^{\eta}}{1+e^{\eta}}$$

where $\eta = x^\top \beta$, hence meaning of β changes

Marginalized random effects

Alternative formulation to random effects
(Heagerty, 1999; Heagerty & Zeger, 2000)

- similar logic of MC marginalisation is applied to random effects: find $\Delta = \Delta(x, \sigma)$ such that

$$\mathrm{logit}^{-1}(\eta) = \int_{-\infty}^{\infty} \mathrm{logit}^{-1}(\Delta + \sigma z)\, \phi(z)\, \mathrm{d}z$$

- requires repeated solution of integral equation

Outline

2nd order MC

- general idea: extend approach of Azzalini (1994) to 2nd order dependence
- specifically: formulate 2nd order MC such that

$$\mathbb{P}\{Y_t = 1 | X_t = x\} = \theta_t$$

(index i dropped) is given by

$$\operatorname{logit} \theta_t = x^\top \beta$$

allowing dependence on (Y_{t-2}, Y_{t-1})

- in $2 \times 2 \times 2$ probability table of (Y_{t-2}, Y_{t-1}, Y_t) the above condition sets 3 probabilities, hence 4 parameters left
- parsimoniuos choice: use two parameters for modelling dependence, and add two constraints

2nd order MC

- general idea: extend approach of Azzalini (1994) to 2nd order dependence
- specifically: formulate 2nd order MC such that

  $$\mathbb{P}\{Y_t = 1 | X_t = x\} = \theta_t$$

  (index i dropped) is given by

  $$\operatorname{logit} \theta_t = x^\top \beta$$

  allowing dependence on (Y_{t-2}, Y_{t-1})
- in $2 \times 2 \times 2$ probability table of (Y_{t-2}, Y_{t-1}, Y_t) the above condition sets 3 probabilities, hence 4 parameters left
- parsimoniuos choice: use two parameters for modelling dependence, and add two constraints

2nd order MC

- general idea: extend approach of Azzalini (1994) to 2nd order dependence
- specifically: formulate 2nd order MC such that

  $$\mathbb{P}\{Y_t = 1 | X_t = x\} = \theta_t$$

  (index i dropped) is given by

  $$\operatorname{logit} \theta_t = x^\top \beta$$

  allowing dependence on (Y_{t-2}, Y_{t-1})
- in $2 \times 2 \times 2$ probability table of (Y_{t-2}, Y_{t-1}, Y_t) the above condition sets 3 probabilities, hence 4 parameters left
- parsimoniuos choice: use two parameters for modelling dependence, and add two constraints

2nd order MC

- general idea: extend approach of Azzalini (1994) to 2nd order dependence
- specifically: formulate 2nd order MC such that

  $$\mathbb{P}\{Y_t = 1 | X_t = x\} = \theta_t$$

  (index i dropped) is given by

  $$\operatorname{logit} \theta_t = x^\top \beta$$

  allowing dependence on (Y_{t-2}, Y_{t-1})
- in $2 \times 2 \times 2$ probability table of (Y_{t-2}, Y_{t-1}, Y_t) the above condition sets 3 probabilities, hence 4 parameters left
- parsimoniuos choice: use two parameters for modelling dependence, and add two constraints

2nd order MC (cntd)

- Choice among possible parameters & constraints: impose

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

- analogy with Gaussian AR(2) models, referred to OR in place of partial correlations
- technical problem: solve elements of MC transition matrix for given β, ψ_1, ψ_2
- very lengthy algebra but explicit solutions available

2nd order MC (cntd)

- Choice among possible parameters & constraints: impose

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

- analogy with Gaussian AR(2) models,
  referred to OR in place of partial correlations
- technical problem:
  solve elements of MC transition matrix for given β, ψ_1, ψ_2
- very lengthy algebra but explicit solutions available

2nd order MC (cntd)

- Choice among possible parameters & constraints: impose

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

- analogy with Gaussian AR(2) models,
  referred to OR in place of partial correlations
- technical problem:
  solve elements of MC transition matrix for given β, ψ_1, ψ_2
- very lengthy algebra but explicit solutions available

2nd order MC (cntd)

- Choice among possible parameters & constraints: impose

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

- analogy with Gaussian AR(2) models, referred to OR in place of partial correlations
- technical problem: solve elements of MC transition matrix for given β, ψ_1, ψ_2
- very lengthy algebra but explicit solutions available

2nd order MC (cntd)

Further work & related work

- obtain derivatives of logL to improve optimisation (even more messy algebra)
- allow for missing data (further complications. . .)
- random effects: possible but desirable to incorporate with Heagerty's (1999) approach
- another 2nd order model (Heagerty & Zeger, 2000):

$$\text{logit}(p_{hj}) = \Delta + \gamma_1 j + \gamma_2 h$$

turns out to be formally equivalent, but
(a) parameter interpretation is simpler for above OR
(b) transition probabilities not given explicitly

2nd order MC (cntd)

Further work & related work

- obtain derivatives of logL to improve optimisation (even more messy algebra)
- allow for missing data (further complications. . .)
- random effects: possible but desirable to incorporate with Heagerty's (1999) approach
- another 2nd order model (Heagerty & Zeger, 2000):

$$\mathrm{logit}(p_{hj}) = \Delta + \gamma_1 j + \gamma_2 h$$

turns out to be formally equivalent, but
(a) parameter interpretation is simpler for above OR
(b) transition probabilities not given explicitly

2nd order MC (cntd)

Further work & related work

- obtain derivatives of logL to improve optimisation (even more messy algebra)
- allow for missing data (further complications. . .)
- random effects: possible but desirable to incorporate with Heagerty's (1999) approach
- another 2nd order model (Heagerty & Zeger, 2000):

$$\mathrm{logit}(p_{hj}) = \Delta + \gamma_1 j + \gamma_2 h$$

turns out to be formally equivalent, but
(a) parameter interpretation is simpler for above OR
(b) transition probabilities not given explicitly

2nd order MC (cntd)

Further work & related work

- obtain derivatives of logL to improve optimisation (even more messy algebra)
- allow for missing data (further complications. . .)
- random effects: possible but desirable to incorporate with Heagerty's (1999) approach
- another 2nd order model (Heagerty & Zeger, 2000):

$$\text{logit}(p_{hj}) = \Delta + \gamma_1 j + \gamma_2 h$$

turns out to be formally equivalent, but
(a) parameter interpretation is simpler for above OR
(b) transition probabilities not given explicitly

2nd order MC (cntd)

Further work & related work

- obtain derivatives of logL to improve optimisation (even more messy algebra)
- allow for missing data (further complications...)
- random effects: possible but desirable to incorporate with Heagerty's (1999) approach
- another 2nd order model (Heagerty & Zeger, 2000):

$$\text{logit}(p_{hj}) = \Delta + \gamma_1 j + \gamma_2 h$$

turns out to be formally equivalent, but
(a) parameter interpretation is simpler for above OR
(b) transition probabilities not given explicitly

References

AZZALINI, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika* **81**, 767–775. Amendment: vol. 84 (1997), 989

FITZMAURICE, G. & LAIRD, N. (1993). A likelihood-based method for analyzing longitudinal binary responses. *Biometrika* **80**, 141–151

FITZMAURICE, G., LAIRD, N., & ROTNITZKY, A. (1993). Regression models for discrete longitudinal responses. *Statistical Science* **8**, 284–309.

GONÇALVES, M. H. (2002). *Likelihood methods for discrete longitudinal data*. PhD thesis, University of Lisbon.

HEAGERTY, P. (1999). Marginally specified logistic-normal models for longitudinal data. *Biometrics* **55**, 668–698.

HEAGERTY, P. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics* **58**, 342–351.

HEAGERTY, P. & ZEGER, S. (2000). Marginalized multilevel models and likelihood inference. *Statistical Science* **15**, 1–26.