Random-Effects Models for Longitudinal Data

Presented by Rahul Kanekar and Ran Xie

Stanford University

January 28, 2024

Outline

Nan M. Laird

"She has worked in several areas of applications, including the quantification of adverse events in hospitals, childhood obesity, and genetic studies in Alzheimer's disease, bipolar disorder, asthma and lung disease. She has served on numerous committees for the NIH and NSF, including a National Academy of Sciences Committee on Cabin Air Quality which recommended the current ban on smoking in airplanes. "

"Her work on the EM Algorithm, with Art Dempster and Don Rubin, is among the top 100 most cited of all published articles in science. "

Maximum likelihood from incomplete data via the EM algorithm
AP Dempster, NM Laird, DB Rubin - Journal of the royal statistical society: series B …, 1977
Cited by 71905 Related articles All 66 versions

James H. Ware

"Ware joined the faculty in 1979 after receiving his PhD in statistics from Stanford University and spending eight years as a mathematical statistician at the National Heart, Lung, and Blood Institute."

"Ware had a long-standing interest in studies of pulmonary and cardiovascular disease. From 1980 to 1995, he was a co-investigator in the landmark Six Cities Study of Air Pollution and Health, which had a profound effect on Clean Air Act of 1970 regulations in the U.S. and on efforts to limit air pollution around the world."

– Magazine of the Harvard T.H. Chan School of Public Health

Outline

A motivating example

Harvard Six Cities Study

A cohort of 8000 people across 6 cities in the US was studied between 1980-95. The goal was to study the effect of air pollution on mortality rates.

This is a longitudinal study where multiple measurements are taken for each subject.

What does the data look like?

Number of subjects $\to m$.

Number of measurements for subject $i \to n_i$.

Measurements for subject $i \to y_i \in \mathcal{R}^{n_i}$.

Covariates for subject $i \to X_i$.

Initial approaches to modelling

Linear Model

$$y_i \sim X_i\beta + e_i, \quad e_i \sim \Sigma.$$

Issues with the Linear Model:

- Number of measurements per subject is often different.
- Experimental conditions are different.
- Accomodating generalized covariance structures makes parameter size too large.

Sources of variation

There are certain variables who's effect is the same for all subjects and some who's effect changes for each subject. We need to incorporate both in the model.

Initial approaches to modelling

Growth Curves

$$y_i = X_i b_i^* + e_i, \quad b_i^* \overset{i.i.d}{\sim} N(\beta, D), \quad e_i \overset{i.i.d}{\sim} N(0, R_i).$$

This can be rewritten as

$$y_i = X_i \beta + X_i b_i + e_i, \quad b_i \overset{i.i.d}{\sim} N(0, D), \quad e_i \overset{i.i.d}{\sim} N(0, R_i).$$

β is the population effect and b_i's are the subject-wise individual effects.

Remark: This forces the population and individual variation to be dependent on the same variables.

Two Stage Random-Effects Model

The Model

$$y_i = X_i\alpha + Z_i b_i + e_i \in \mathcal{R}^{n_i}, \quad b_i \sim N(0, D), \quad e_i \sim N(0, R_i).$$

This model gives us enough flexibility while still allowing computation.

- α is the vector of population effects and b_i's are individual effects.
- R_i's can be assumed to be diagonal or even $\sigma^2 I$.
- Columns size of Z_i can be made small which keeps D small.
- Parameter size remains in control.

Outline

Max. Likelihood vs Restricted Max. Likelihood (REML)

Goal: In two-stage mixed effects model

$$y_i = X_i\alpha + Z_i b_i + e_i \in \mathcal{R}^{n_i}, \quad b_i \sim N(0, D), \quad e_i \sim N(0, R_i),$$

construct estimators for regression parameters α and covariance matrices D, R_i, $i = 1, ..., m$ (if unknown).

Approach 1: Maximum likelihood with respect to marginal distribution of $y^T = (y_1^T, ..., y_m^T)$

$$\mathcal{L}(\alpha, D, R_i).$$

The estimator $\hat{\alpha}$ can be expressed in terms of $\hat{\alpha}(\hat{D}, \{\hat{R}_i\})$.

Approach 2: Restricted maximum likelihood with respect to marginal distribution of $U^T y$ s.t. $E[U^T y] = 0$,

$$\mathcal{L}(D, \{R_i\}).$$

Plug in $\hat{D}$, $\{\hat{R}_i\}$ to obtain empirical Bayes estimator $\hat{\alpha}(\hat{D}, \{\hat{R}_i\})$.

Restricted Maximum Likelihood (REML)

- REML produces less biased (sometimes unbiased) estimators compared to classical maximum likelihood.
- Note that $E[U_i^T y_i] = 0 \Leftrightarrow U_i^T X_i \alpha = 0$. Therefore,

$$U_i^T y_i = U_i^T Z_i b_i + e_i$$

  is **unrelated** to α.
  Assume the design matrix $X \in R^{n \times p}$ has full column rank, then by design U has rank $n - p$.

Restricted Maximum Likelihood (REML)

Toy example: Consider the model for $i = 1, ..., n$,

$$y_i = \mu + e_i, \quad e_i \sim N(0, \sigma^2).$$

From classical maximum likelihood,

$$\hat{\mu}_{ML} = \bar{x}, \quad \hat{\sigma}^2_{\mathsf{ML}} = \frac{1}{n} \sum_i (x_i - \bar{x})^2.$$

Let $C \in \mathcal{R}^{n \times (n-1)}$ s.t. $C^T 1_n = 0$. Then $C^T y \sim N(0, \sigma^2 I_{n-1})$. From REML, we have

$$\hat{\sigma}^2_{\mathsf{REML}} = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2.$$

EM Algorithm for ML Estimates

In the mixed effects model,

$$y_i = X_i\alpha + Z_i b_i + e_i \in \mathcal{R}^{n_i}, \quad b_i \sim N(0, D), \quad e_i \sim N(0, R_i),$$

Let θ be the parameters that determine covariance matrices D, R_i. If θ is known, we can easily obtain the MLE $\hat{\alpha}(\theta)$.
If we "augment" the data to (y_i, b_i, e_i) (in which only y_i is actually observed), we can easily compute the MLE if θ as a function of sufficient statistics $t_1 = \sum e_i^T e_i$ and $t_2 = \sum b_i^T b_i$, this yields the M-step:

$$\text{M-step: } \hat{\theta} = M(t).$$

With estimates $\hat{\theta}$, $\hat{\alpha}(\hat{\theta})$, we can estimate the sufficient statistics via an E-step:

$$\text{E-step: } \hat{t} = \mathrm{E}[t | y, \hat{\theta}, \hat{\alpha}(\hat{\theta})].$$

EM Algorithm for REML estimates

Bayesian formulation of REML: Assume $\alpha \sim N(0, \Gamma)$, then the marginal likelihood of y as $\Gamma^{-1} \to 0$ is the same as the REML likelihood and hence the estimate $\hat{\theta}$ is the REML estimate(Harville 1976 [2]).

The M-Step remains the same.

The E-Step in the EM algorithm now is

$$\hat{t} = \mathrm{E}[t|y, \hat{\theta}, \Gamma^{-1} = 0].$$

$\hat{t}$ does not depend on $\hat{\alpha}$ which is also what happens in REML.

At convergence, we should be close to the REML estimate of θ.

Advantage: In ML and REML, $\hat{b}_i = \mathrm{E}(b_i \mid y, \hat{\theta}, \Gamma^{-1} = 0)$ can serve as an estimate of individual effect. This can be used for follow-up studies.

EM Algorithm for ML/REML Estimates

- A unified approach to fitting the model.
- An iterative procedure to compute the maximum likelihood estimator (MLE).
- "Augments" the observable data to a set of comlete data, and treat the problem as an incomplete-data problem to facilitate computation.
- Easy to implement and friendly to applied statisticians, in contrast to previous works at the time which were very difficult to implement.
- At the time, sensitive to initial estimates and converges slowly.

Speeding up the EM Algorithm

J. R. Statist. Soc. B (1982),
44, *No.* 2, *pp.* 226–233

Finding the Observed Information Matrix when Using the *EM* Algorithm

By Thomas A. Louis

Harvard School of Public Health, Mass., USA

[Received July 1980. Revised January 1981]

Summary

A procedure is derived for extracting the observed information matrix when the *EM* algorithm is used to find maximum likelihood estimates in incomplete data problems. The technique requires computation of a complete-data gradient vector or second derivative matrix, but not those associated with the incomplete data likelihood. In addition, a method useful in speeding up the convergence of the *EM* algorithm is developed. Two examples are presented.

Speeding up the EM Algorithm

J. R. Statist. Soc. B (1982),
44, *No.* 2, *pp.* 226–233

Finding the Observed Information Matrix when Using the *EM* Algorithm

Biometrika (1993), **80**, 2, *pp.* 267–78
Printed in Great Britain

A procedure is
algorithm is use
The technique
derivative matri
addition, a met
developed. Two

Maximum likelihood estimation via the ECM algorithm: A general framework

BY XIAO-LI MENG
Department of Statistics, University of Chicago, Chicago, Illinois 60637, U.S.A.

AND DONALD B. RUBIN
Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, U.S.A.

SUMMARY

Two major reasons for the popularity of the EM algorithm are that its maximum step involves only complete-data maximum likelihood estimation, which is often computationally simple, and that its convergence is stable, with each iteration increasing the

Speeding up the EM Algorithm

J. R. Statist. Soc. B (1982),
44, No. 2, pp. 226–233

Finding the Observed Information Matrix when Using the *EM* Algorithm

Biometrika (1993), **80**, 2, *pp.* 267–78
Printed in Great Britain

Maximum likelihood estimation via the ECM algorithm: A general framework

Biometrika (1994), **81**, 4, *pp.* 633–48
Printed in Great Britain

7, U.S.A.

2138, U.S.A.

aximum step
ten computa-
ncreasing the

The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence

BY CHUANHAI LIU AND DONALD B. RUBIN

Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, U.S.A.

SUMMARY

A generalisation of the ECM algorithm (Meng & Rubin, 1993), which is itself an extension of the EM algorithm (Dempster, Laird & Rubin, 1977), can be obtained by replacing some CM-steps of ECM, which maximise the constrained expected complete-data loglikelihood function, with steps that maximise the correspondingly constrained actual likelihood

Software for Linear Mixed Models

Fitting linear mixed-effects models using lme4
D Bates, M Mächler, B Bolker, S Walker - arXiv preprint arXiv:1406.5823, 2014
Cited by 70973 Related articles All 93 versions

ASReml user guide release 1.0 *
AR Gilmour, BJ Gogel, BR Cullis, SJ Welham… - 2002
Cited by 3873 Related articles All 18 versions

[BOOK] HLM 6: Hierarchical linear and nonlinear modeling
SW Raudenbush - 2004 - books.google.com
HLM 6 greatly broadens the range of hierarchical models that can be estimated. It also offers greater convenience of use than previous versions. Here is a quick overview of key new …
☆ Save 99 Cite Cited by 7538 Related articles »

Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models
JD Singer - Journal of educational and behavioral statistics, 1998
Cited by 3826 Related articles All 16 versions

Outline

Extensions and Improvements

1. Extension from linear models to generalized linear models to allow applications to discrete/categorical data

   Longitudinal data analysis using generalized linear models
   KY Liang, SL Zeger - Biometrika, 1986
   Cited by 21254 Related articles All 21 versions

2. Allow for missing data (e.g. dropout probability related to individual history) and relax assumption of Gaussianity

   Estimation of regression coefficients when some regressors are not always observed
   JM Robins, A Rotnitzky, LP Zhao - Journal of the American statistical Association, 1994
   Cited by 3305 Related articles All 10 versions

   Analysis of semiparametric regression models for repeated outcomes in the presence of missing data
   JM Robins, A Rotnitzky, LP Zhao - Journal of the american statistical association, 1995
   Cited by 1919 Related articles All 9 versions

Applications: Policy Making

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

ELSEVIER

ENERGY POLICY

Should we build more large dams? The actual costs of hydropower megaproject development ☆

CrossMark

Atif Ansar [a,b,*], Bent Flyvbjerg [b], Alexander Budzier [b], Daniel Lunn [c]

[a] *Blavatnik School of Government, University of Oxford, Oxford OX1 4JJ, UK*
[b] *Saïd Business School, University of Oxford, Oxford OX1 1HP, UK*
[c] *Department of Statistics, University of Oxford, Oxford OX1 3GT, UK*

Applications: Policy Making

ELSEVIER

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

ENERGY POLICY

Should we build more large da
megaproject development☆

Atif Ansar [a,b,*], Bent Flyvbjerg [b], Alexanc

[a] Blavatnik School of Government, University of Oxford, Oxford OX1
[b] Saïd Business School, University of Oxford, Oxford OX1 1HP, UK
[c] Department of Statistics, University of Oxford, Oxford OX1 3GT, UK

3.3. Multilevel regression analysis of cost and schedule performance

Means, standard deviations, and correlations of the variables used in the multilevel regressions are shown in Table 2.

We fitted multilevel regression models with projects nested by country as a second level to incorporate within-country correlation. The models were fitted using the “lme” procedure in the “nlme” package in R software. This function fits a linear mixed-effects model in the formulation described in Laird and Ware (1982) but allowing for nested random effects. The within-group errors are allowed to be correlated and/or have unequal variances. We found it necessary to transform variables to remove excessive skewness as noted in Table 2. Using stepwise variable selection, we are not only able to fit explanatory models for cost and overruns and estimated duration but also practically parsimonious models for predicting them.

Applications: Education

Developmental Psychology
2001, Vol. 37, No. 2, 231–242

0012-1649/01/$5.00 DOI: 10.1037//0012-1649.37.2.231

The Development of Cognitive and Academic Abilities: Growth Curves From an Early Childhood Educational Experiment

Frances A. Campbell, Elizabeth P. Pungello,
Shari Miller-Johnson, and Margaret Burchinal
University of North Carolina at Chapel Hill

Craig T. Ramey
University of Alabama at Birmingham

In the Abecedarian Project, a prospective randomized trial, the effects of early educational intervention on patterns of cognitive and academic development among poor, minority children were examined. Participants in the follow-up were 104 of the original 111 participants in the study (98% African American). Early treatment was full-time, high-quality, educational child care from infancy to age 5. Cognitive test scores collected between the ages of 3 and 21 years and academic test scores from 8 to 21

Applications: Education

Developmental Psychology
2001, Vol. 37, No. 2, 231–242

0012-1649/01/$5.00 DOI: 10.1037//0012-1649.37.2.231

The Development of Cognitive and Academic Abilities: Growth Curves From an Early Childhood Educational Experiment

Frances A
Shari Mille
Univer

Data Analyses

Hierarchical linear models (HLMs) were used to test the longitudinal hypotheses in this study, using the MIXED procedure in SAS (Singer, 1998). This analytic method estimates individual and group growth curves to describe patterns of change over time and factors associated with those patterns (see Byrk & Raudenbush, 1987, 1992; Laird & Ware, 1982). That is, this method permits examination of regression relationships among dependent and predictor variables and the ways in which these relationships vary over time. This approach accommodates randomly missing data points and inconsistent timing in data collection (i.e., unequal time intervals between data points) and allows for flexible specification of the within- and between-subjects variance.

References

Douglas W Dockery, C Arden Pope, Xiping Xu, John D Spengler, James H Ware, Martha E Fay, Benjamin G Ferris Jr, and Frank E Speizer.
An association between air pollution and mortality in six us cities.
New England journal of medicine, 329(24):1753–1759, 1993.

David A. Harville.
Maximum likelihood approaches to variance component estimation and to related problems.
Journal of the American Statistical Association, 72(358):320–338, 1977.

Nan M Laird and James H Ware.
Random-effects models for longitudinal data.
Biometrics, pages 963–974, 1982.