MLwiN: software for fitting multilevel models

History

- Software developed by the Centre for Multilevel Modelling at the Institute of Education.
- Forerunners over the last 15 years or so include ML2, ML3 and MLn.
- Main Programmer: Jon Rasbash.
- MLwiN is first version to include Windows interface and MCMC estimation procedures.
- Currently around 3,000 users (mainly academic) worldwide.
- Software consists of user-friendly Visual Basic Windows interface on top of fast estimation engines written in C++.
- MCMC theory and programming by William Browne (with advice on earlier versions from David Draper).
- Version 1.2 with extensive new MCMC features released in April 2002

Estimation Methods available:

- IGLS and RIGLS for Normal response models.
- MQL and PQL for discrete response models.
- Gibbs sampling for Normal response models.
- Hybrid MH-Gibbs sampling for discrete response models.
- Parametric and non-parametric bootstrapping.

Web Sites:
- http://tramss.data-archive.ac.uk – Free educational version of MLwiN available from here
- http://multilevel.ioe.ac.uk/team/bill.html – Research papers available from here
Equations Window Interface

Model specification and modification is done via an easy to use equation based Windows interface. An example of this equations window with a Normal model set up and run using MCMC is shown below:

\[
\begin{align*}
\text{normexam}_{ij} & \sim N(X_{i0}, \Omega) \\
\text{normexam}_{ij} & = \beta_{0ij} \text{cons} + 0.563(0.012)\text{standlrt}_{ij} \\
\beta_{0ij} & = 0.005(0.042) + u_{0ij} + e_{0ij} \\
\left[u_{0j}\right] & \sim N(0, \Omega_{u}) : \Omega_{u} = \begin{bmatrix} 0.097(0.021) \end{bmatrix} \\
\left[e_{0j}\right] & \sim N(0, \Omega_{e}) : \Omega_{e} = \begin{bmatrix} 0.566(0.013) \end{bmatrix}
\end{align*}
\]

Prior specifications:
\[
\begin{align*}
p(\beta_0) & \propto 1 \\
p(\beta_1) & \propto 1 \\
p(1/\sigma_{u0}^2) & \sim \text{Gamma}(0.001, 0.001) \\
p(1/\sigma_{e0}^2) & \sim \text{Gamma}(0.001, 0.001)
\end{align*}
\]

Deviance (MCMC) = 9209.146 (4059 of 4059 cases in use)

This window shows MCMC estimates for a variance components model fitted to an educational example of 4059 pupils in 65 schools. The response is the (normalised) total exam scores for the pupils at age 16. The main predictor is a (standardised) London Reading Test score at age 11.
Gibbs Sampling in a simple 2 level multilevel model

A simple two level Normal model can be written as

\[ y_{ij} = X_{ij} \beta + Z_{ij} u_j + e_{ij} \]

where \( u_j \sim N(0, \Omega_u) \) and \( e_{ij} \sim N(0, \sigma_e^2) \)

To use Gibbs sampling on this problem we need to simulate from the conditional distributions of the groups of parameters in turn:

Step 1: \( p(\beta \mid y, u, \Omega_u, \sigma_e^2) \sim MVN(\beta^*, D^*) \)

Step 2: \( p(u_j \mid y, \beta, \Omega_u, \sigma_e^2) \sim MVN(u^*_j, D_j^*) \)

Step 3: \( p(\Omega_u \mid y, \beta, u, \sigma_e^2) \sim InvWishart(\nu^*, S^*) \)

Step 4: \( p(\sigma_e^2 \mid y, \beta, u, \Omega_u) \sim InvGamma(a^*, b^*) \)

The parameters of these 4 distributions vary at each iteration.

These 4 steps are repeated until the Markov chains converge to the joint distribution of the unknown parameters. The first \( B \) values of a chain are thrown away as we need to allow the chain to reach the joint distribution from its starting values. These \( B \) values are known as the ‘burnin’ period. In MLwiN we always have good starting values as we can use the estimates from the IGLS and RIGLS estimation methods as starting values.
Adaptive hybrid Metropolis Hastings (MH) - Gibbs sampling

In MLwiN we offer this alternative hybrid estimation procedure. It consists of Gibbs sampling for the variance parameters and MH Sampling for other parameters. It has the advantage that it can be used for non-Normal response models.

The MH steps involve for each parameter, \( \theta \) drawing from a Normal proposal distribution \( p(\theta) \sim N(\theta^{(t)}, \sigma^2_\theta) \) and accepting / rejecting the new value as in the MH algorithm. The difficulty is in choosing the value \( \sigma^2_\theta \). One approach is to use a scaled version of the IGLS/RIGLS estimated parameter standard errors. An alternative is the following adapting procedure, which aims to get a acceptance rate of R\% with a tolerance of T\%.

**Adaptive Procedure**

Start with a value of the proposal variance \( \sigma^2_{\text{old}} \), then repeat the following:

Run the sampler for 100 iterations and calculate the acceptance rate, N\%.

If \( N < R \), \( \sigma_{\text{new}} = \sigma_{\text{old}} / (2 - (N/R)) \)
Else \( \sigma_{\text{new}} = \sigma_{\text{old}} * (2 - ((100-N)/(100-R))) \)

When three consecutive values of N lie within the interval (R-T\%,R+T\%) for this parameter it is marked. When all parameters are marked the adaptive method ends and the burn-in proper begins followed by the main run.
Trajectories Window

The trajectories window in MLwiN displays the chains of MCMC values generated over time. It is updated as the estimation proceeds allowing the user to view the progress of the estimation routine and identify problems in the MCMC mixing.

These chains are for the Normal response educational example shown earlier. The chain has here been run for 5000 iterations after a ‘burn-in’ of 500 iterations.
MCMC Convergence Diagnostics

To check MCMC convergence and to establish suitable chain run length of a particular parameter chain is available via the point and click interface.

The above window shows diagnostics and summary statistics for the level-2 variance parameter in the educational example.
Residuals and School Rankings

MLwiN allows the user to calculate the residuals after running a model using any estimation method. These residuals can then be used for checking model assumptions and for comparison between higher-level units.

Chains of the residuals can be stored and from these, and an MLwiN macro, confidence intervals for the institution ranks can be created.
Extensive Interactive Graphics facilities

MLwiN allows many types of graphs to be plotted and links highlighted observations across graphs as shown below:

Here the individual pupil scores in the top and bottom achieving schools are highlighted.
Informative Prior Distributions

MLwiN features a prior specification window for inputting informative prior distributions.

The diagnostics screen incorporates the prior distribution in the kernel density plot for fixed-effects parameters.
Multilevel logistic regression models

MLwiN can also fit both Binomial and Poisson response models using MCMC methods.

The above equations show a model for a problem from a study in demography. 1934 women in Bangladesh were selected from 60 districts and were asked if they currently used contraception. Here we see the effects of number of children and age of the woman on the likelihood of using contraceptive plus we differentiate between usage in urban and rural areas within a district. We see that the difference in usage in urban and rural areas also varies across the districts.
Heteroscedasticity/ Complex level 1 variation

Typically linear statistical modelling assumes a constant variance for the response variable and interest lies in modelling the mean function. The variance may however not be a constant but instead could be a function of predictor variables. This is known as heteroscedasticity.

Here we have fitted a 1-level linear regression model to the education dataset, where the variance is assumed to be a quadratic function of the pupils intake score (standlrt). We use a truncated Normal proposal MH algorithm described in Browne et al. (2002)
WinBUGS Output/Input Interface

WinBUGS (Spiegelhalter et al. 1998) uses the Adaptive Rejection (AR) sampler as the basis for much of its sampling. Rather than also code up our own AR sampler we instead offer options to convert the MLwiN model to BUGS code and to read in BUGS output files:

Example output

# WINBUGS code generated from MLwiN program

model
{
  # Level 1 definition
  for(i in 1:N) {
    normexam[i] ~ dnorm(mu[i],tau)
    mu[i]<- beta[1] * cons[i] +
    beta[2] * standlrt[i] +
    u2[school[i]] * cons[i]
  }
  # Higher level definitions
  for (j in 1:n2) {
    u2[j] ~ dnorm(0,tau.u2)
  }
  # Priors for fixed effects
  for (k in 1:2) { beta[k] ~ dnorm(0, 0.0001) }
  # Priors for random terms
  sigma2 ~ 1/tau
  tau ~ dgamma(0.001,0.001)
  tau.u2 ~ dgamma(0.001,0.001)
  sigma2.u2 ~ 1/tau.u2
}
MVN response multilevel models using MCMC estimation with missing data

Educational example featuring two responses, a written and a coursework score for 1905 students in 73 schools.

Here girls are generally doing better than boys at coursework but not at written work and there is significant variation in both scores at both the school and student level with positive correlation at each level. The 382 missing responses are given uniform priors and updated, at each iteration, as an additional Gibbs sampling step in the MCMC algorithm.
Cross-classified and multiple-membership models.

MLwiN is historically a multilevel/hierarchical modelling package that relies on nesting of the levels in the model for example pupils nested in classes in schools. Social science data often have more complex structures than a simple nested structure.

Cross-classified models

Here the classifications are not nested for example pupils are nested within schools and live in neighbourhoods but the neighbourhoods and schools may be crossed rather than nested.

Multiple-membership models

A pupil may move school during their study and both schools will have an effect on their study. Consequently a pupil can be a member of more than 1 school. Then each school has a (weighted) effect on the pupils progress.

Multiple-membership multiple classification (MMMC) models

Browne, Goldstein and Rasbash (2001) consider combining crossings and multiple-membership to form such a family.
Measurement Errors in Multilevel Models

Generally statistical modelling assumes that predictor variables are measured without any errors. This is of course generally a false assumption and we should account for the errors in our predictor variables. This can be done easily using MCMC estimation.

Above we can see the educational example with an observed predictor ‘obslrt’ which has been observed with error. The model is then accounting for the error through some additional Gibbs sampling steps.
Multilevel Factor Analysis Models

These models combine fitting factor variables to multiple responses with a multilevel data structure to produce factors at different levels of the data structure.

The above is a 1-level 1-factor model but we can extend both the number of levels and the number of factors at each level. Such models become very complicated and care has to be taken with the constraints used to ensure that the model is uniquely identified. The dataset consists of exam marks on 6 papers for 2,439 students who sat between 3 and 5 of the papers. The factor can be interpreted as an overall academic achievement measure.