



Mixed Effects Machine Learning

Presented by

Htun Teza

Master of Science Program in Data Science for Health Care

Department of Clinical Epidemiology and Biostatistics

Faculty of Medicine Ramathibodi Hospital



Correlated observations

1. Studies with clustered data, such as students in classrooms
 2. Longitudinal or repeated-measures studies, in which subjects are measured repeatedly over time or under different conditions.
- It results in loss of independence between observations.
 - It violates i.i.d data assumption.

Linear mixed model

$$y_{ij} = X_{ij}b + Z_{ij}u$$

- where
- y is the target variable
- X is the fixed effects feature
- b is the fixed effects coefficient
- Zu is the random variable describing additive noise
- i is the cluster.
- j is the observation of i^{th} cluster.

Linear mixed model

- Considering that we assume random effects are different between each cluster,

$$y = Xb + Zu$$

- Xb is the population-average value, and it accounts for within-cluster variation.
- Zu is the subject-specific value, and it accounts for between-cluster variation.

Non-linear mixed model

$$y_{ij} = f(X_{ij}) + Z_{ij}u$$

- where
- y is the target variable
- f is the non-linear function
- X is the fixed effects feature
- Zu is the random variable describing additive noise
- i is the cluster
- j is the observation of i^{th} cluster



Machine Learning models

- Classical machine learning classification and regression algorithms do not generate high quality models on correlated data.
- This study develops an extension of traditional machine learning methods to longitudinal/clustered data.

Article

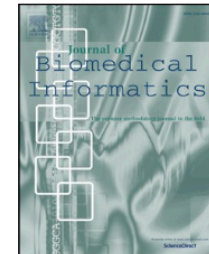
Journal of Biomedical Informatics 89 (2019) 56–67



Contents lists available at [ScienceDirect](#)

Journal of Biomedical Informatics

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



Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin A1c

Che Ngufor^{a,*}, Holly Van Houten^a, Brian S. Caffo^b, Nilay D. Shah^a, Rozalina G. McCoy^a

^a Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States

^b Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States



Fixed effects estimation

- By replacing the non-linear function with machine learning model,

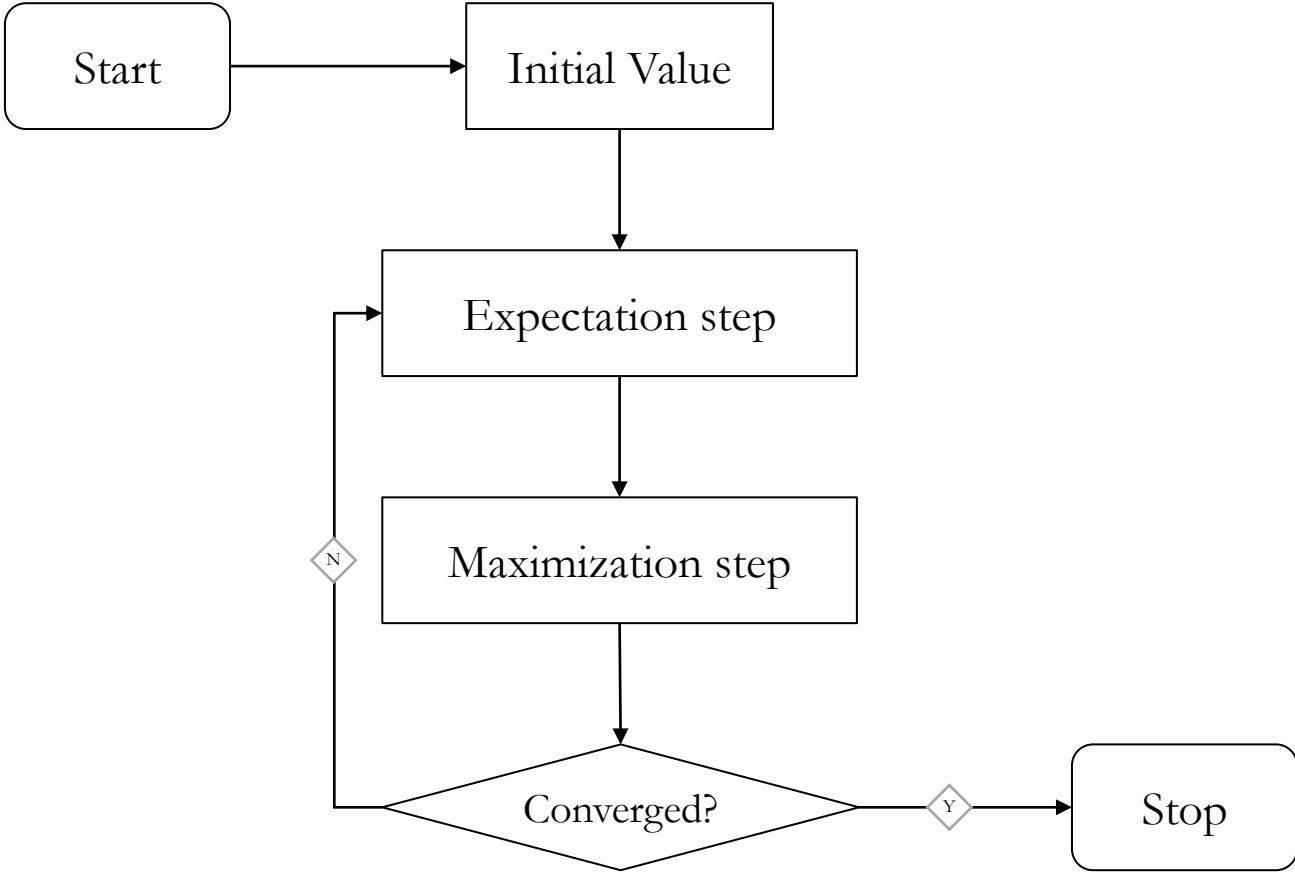
$$y = f(x) + Zu$$

- Machine learning models can be adapted to work with correlated observations as well.
- While the literature applies ensemble models such as random forest and gradient boost machine, it should work similarly with most regression models.

Expectation-Maximization

- an iterative algorithm
 1. Given a set of incomplete data, consider a set of starting parameters.
 2. Expectation step (E — step): Using the observed available data of the dataset, estimate (guess) the values of the missing data.
 3. Maximization step (M — step): Complete data generated after the expectation (E) step is used in order to update the parameters.
 4. Repeat step 2 and step 3 until convergence

Expectation-Maximization



Regression

Model Framework

Training

- Initial random effects value

Loop

- Fixed effects ; Machine learning regressor
- Random effects ; Linear mixed effects model

- Updating random effects value

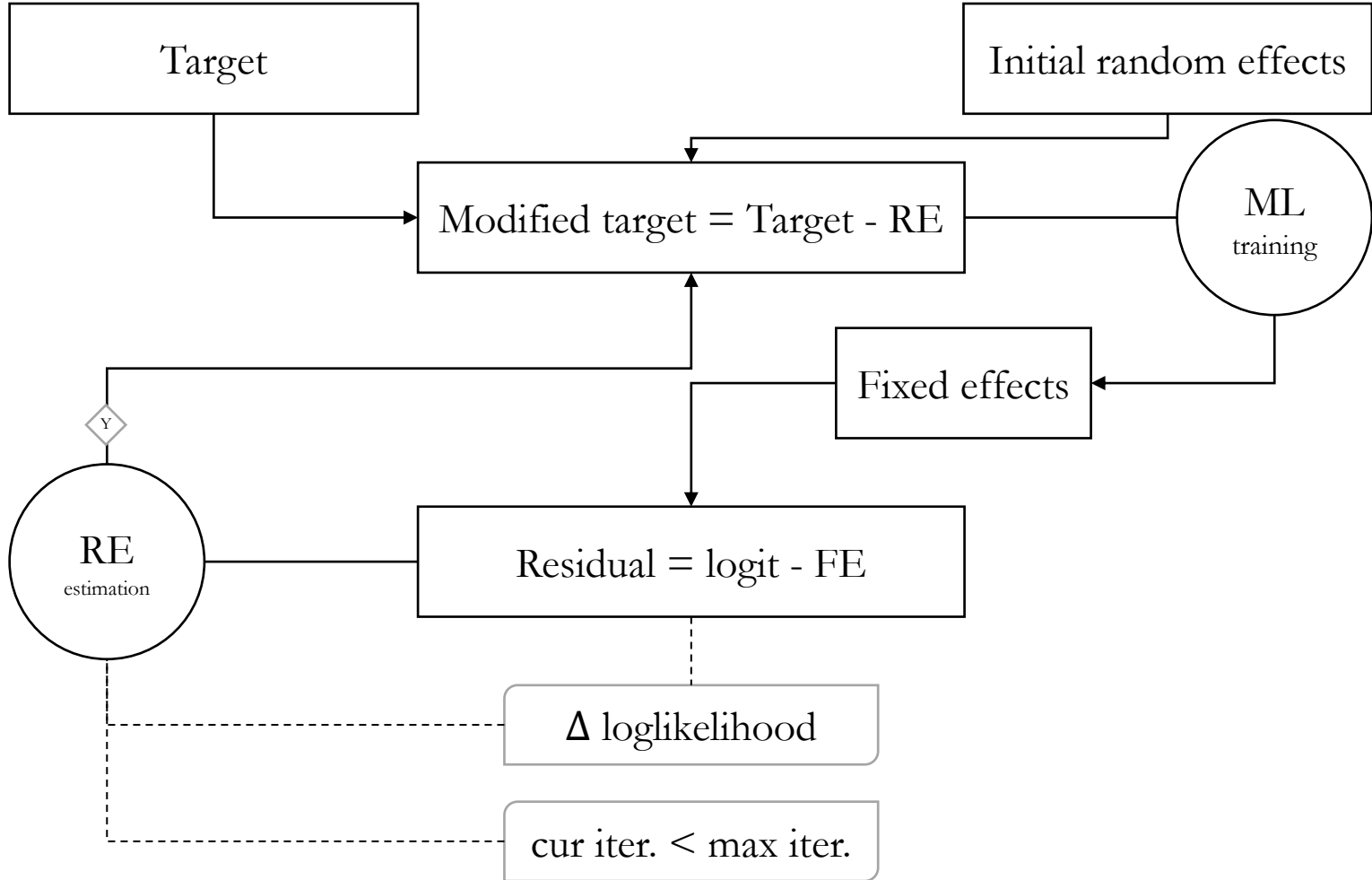
Final model

Fixed effects ;
Machine learning regressor

Random effects ;

- known cluster – LMM
- new cluster – 0

Loop



Stopping criteria

1. `max_iteration`
 2. tolerance with respect to absolute change of log likelihood
- Default values for the hyperparameters are not specified in the original literature.
 - Within the source code,
 - `max_iter = 100 / 1000`,
 - `tolerance = 0.01 / 0.001`
 - Literatures for tolerance with respect to log likelihood is used as convergence criterion for conventional statistical models as well.

Convergence criterion

1. Numerical convergence criterion;

- the iterations proceed until the relative change in the estimated parameter vector is negligible.
- What negligible is depends on: the numerical precision of the computer, the numerical stability of the algorithm and the preferred accuracy of the parameter estimates.

2. Statistical convergence criterion;

- the iterations are proceeded until the change of the estimated parameter vector is negligible with respect to the accuracy of the estimates.

- Moddemeijer, R. (2006). On the convergence of the iterative solution of the likelihood equations.

For conventional models

- When the change in log likelihood is monitored, both STATA and R (lme4) applies function
 - $|| \text{Log likelihood}_{\{k+1\}} - \text{Log likelihood}_{\{k\}} ||$
 - where $k = 0, 1, 2, \dots$, maximum iteration.
- In STATA, tolerance for change is $1e-6$ and maximum iteration is 300.
- In R(lme4), tolerance for change is $1e-6$ and maximum iteration is 50.

STATA - <https://www.stata.com/manuals/rmaximize.pdf>

R(lme4) - <http://search.r-project.org/R/library/lme4/html/convergence.html> (ftol_abs) (lmeControl)

Binary Classification

Model Framework

Training

Outer Loop

- Initial values (target + random effects)

Inner Loop

- Fixed effects ; Machine learning regressor
- Random effects ; Linear mixed effects model

- Updating values

Final model

Fixed effects ;

Machine learning regressor

Random effects ;

- known cluster – LMM
- new cluster – 0

Initial logit value

```
# Get initial values
```

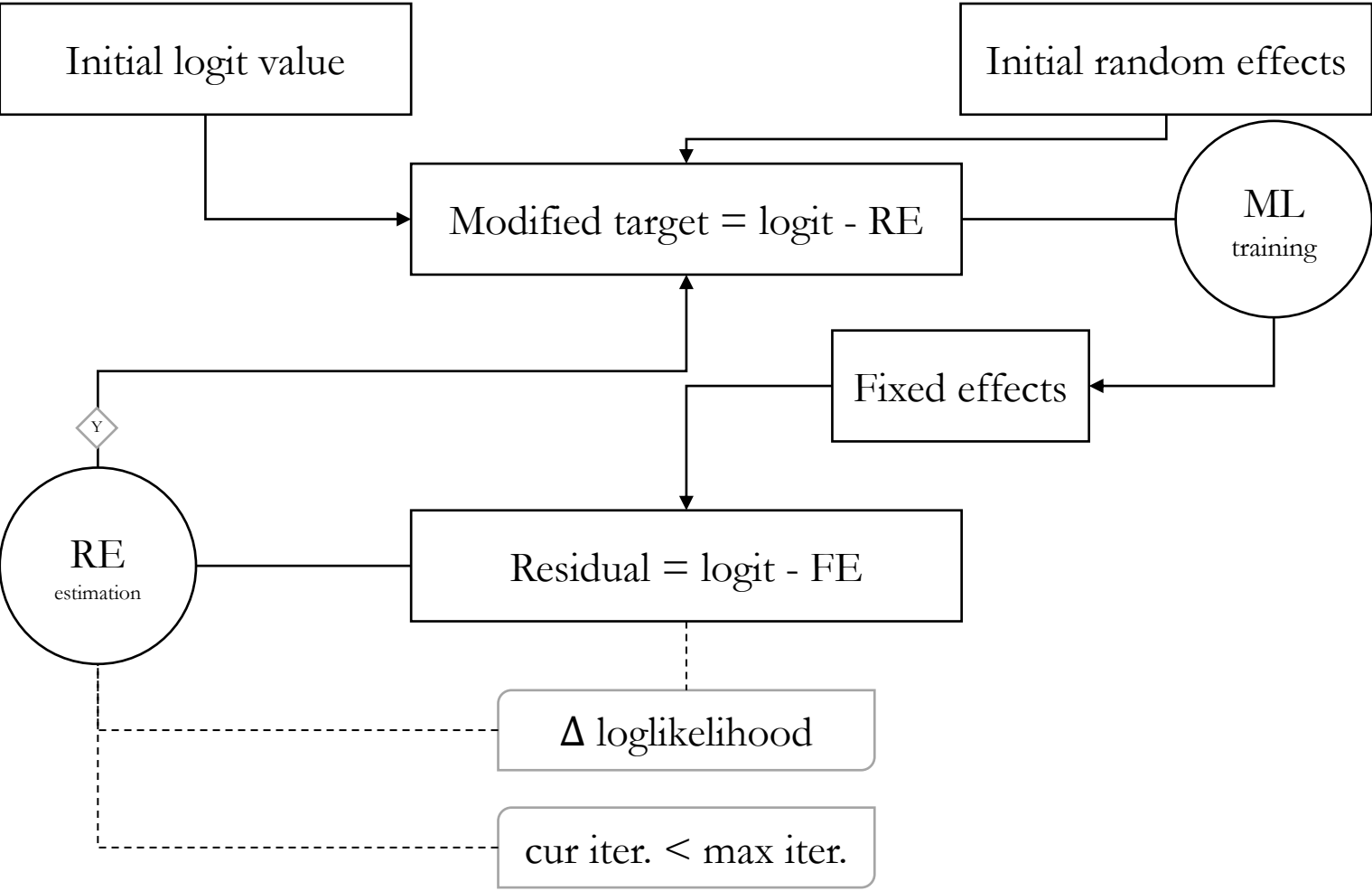
```
mu = rep(mean(Y), length(Y))
```

```
eta = log(mu/(1-mu))
```

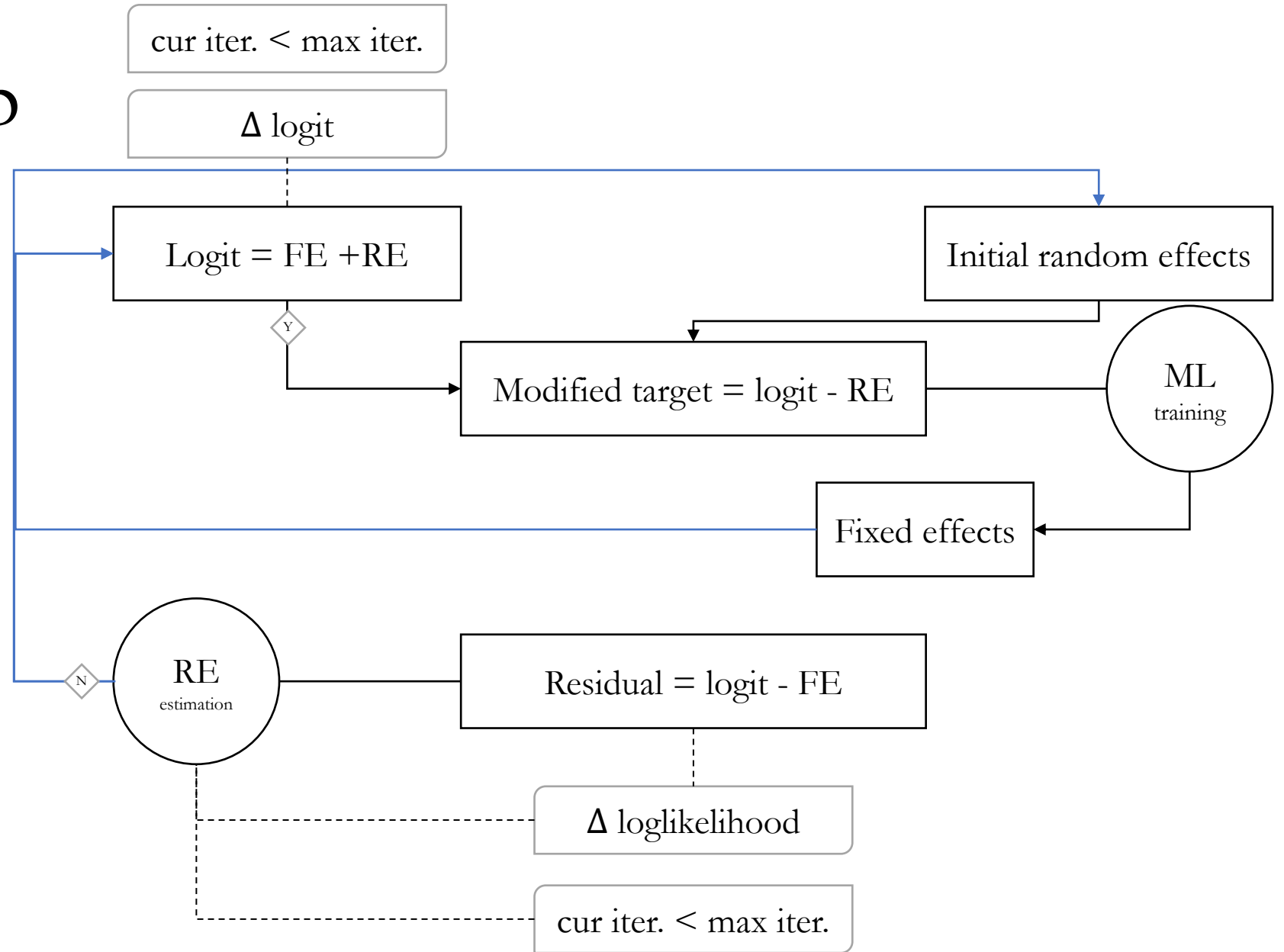
```
y = eta + (Y-mu)/(mu*(1-mu))
```

target class (Y)	probability (mu)	odds	log odds (eta)	initial logit value (y)
1	0.3	0.428	-0.848	2.485
0	0.3	0.428	-0.848	-2.277
1	0.3	0.428	-0.848	2.485
0	0.3	0.428	-0.848	-2.277
0	0.3	0.428	-0.848	-2.277
1	0.3	0.428	-0.848	2.485
0	0.3	0.428	-0.848	-2.277
0	0.3	0.428	-0.848	-2.277
0	0.3	0.428	-0.848	-2.277
0	0.3	0.428	-0.848	-2.277

Inner Loop



Outer Loop



Stopping criteria

- Inner Loop – max_iteration, tolerance with respect to absolute change of log likelihood
- Outer Loop – max_iteration, tolerance with respect to the maximum of the absolute change of logit value
- Default values for the hyperparameters are not specified in the original literature.
- Within the source code,
 - inner loop – max_iter = 200, tolerance = 0.3 / 0.001
 - outer loop – max_iter = 15, tolerance = 1 / 0.001
- Recommendations for tolerance with respect to logit value are not found.
- Recommendations for tolerance with respect to log likelihood is the same as regression tasks.



Result

- longitudinal/clustered supervised machine learning,
- as that of learning the two components of a non-linear mixed model
- separately through an iterative expectation maximization-like algorithm,
- in which the fixed-effect component is estimated using machine learning methods and
- the random-effect component is estimated using linear mixed model.



Discussion

- Significant improvement of Mixed effects machine learning over classical machine learning methods.
- As the number of repeated observations increased,
 - the performance of Mixed effects machine learning increased
 - the performance of classical methods deteriorated.



Discussion

- By including random effects within the model, Mixed effects machine learning is resistant to variabilities introduced by correlated data.
- Mixed effects machine learning can take advantage of dependencies between the observations to generate more robust and accurate models.
- However, improvement over generalized linear mixed model was not observed.



Thank You

