ERP mass univariate analyses with LIMO EGG

June 14

2013

Description of files created and main function called during a standard mass univariate analysis of ERP

File descriptions and overview of 'standard' hierarchical modeling using a mass univariate approach in LIMO EEG

Cyril Pernet
cyril.pernet@ed.ac.uk

The LIMO toolbox is a set of Matlab functions primarily designed for hierarchical linear modeling, i.e. analyzes rely on statistical testing using the general linear model The 1st level analysis (per subject) relies on Ordinary Least Squares and 2nd level (group) analysis relies both on robust estimators like trimmed means to deal with outliers and on resampling techniques to ensure the validity of statistics even under non normality and heteroscedasticity of the parameter estimates, whilst keeping the type I error rate at the nominal level.

Contents

1st level analysis (for each subject)

The master file from which data are imported, analyzes 'called' and results displayed is *limo_eeg.m*. Following the EEG_lab convention, data are handled in matrices with dimensions [electrodes x time frames x trials]. During the analysis, many new files are created containing various statistical information (see below) and they almost all have for dimensions: [electrodes x frames x statistical information]. As data get imported, the design matrix created, parameters estimated, etc .. a LIMO.mat file is created. This file contains all relevant information about the data and analyzes.

Data import = limo_eeg(2) → *limo_import.m*

GUI which allows testing the data and setting up information about the analysis.

By convention, we distinguish between two types of variables: the categorical variable and the continuous variables. The *categorical variables* are defined by a vector or matrix identifying groups of trials. Each column is considered as a factor whilst each value in each column is considered as a specific condition. The *continuous variables* are defined by a vector or matrix such as each column contains values characterizing each trials. The software handles from 1 to (N trials - Number of categorical + continuous - 1) predictors. Above this limit, there are more columns in X (design matrix) than rows in Y (trials) and no solution can be found.

```
LIMO.Level                      = 1

LIMO.data                       = information about the data
LIMO.data.data_dir              = directory where to read them
LIMO.data.data                  = file name
LIMO.data.chanlocs              = import channel location information
LIMO.data.start                 = when to start the analysis
LIMO.data.end                   = when to stop the analysis
LIMO.data.sampling_rate         = sampliong rate of the data
LIMO.data.Cat                   = Categorical variable(s)
LIMO.data.Cont                  = Continuous variable(s)
LIMO.data.neighbouring_matrix   = matrix describing which electrodes
                                  are neighbourghs if bootstrap

LIMO.design                     = information about the design
LIMO.design.fullfactorial       = 0/1 specify if interaction should
                                  be included
LIMO.design.zscore              = 0/1 zscoring of continuous regressors
LIMO.design.method              = 'OLS','WLS' or 'IRLS' by default we
                                  use an ordinary least square approach
                                  but weighted least squares (one
                                  weight per trial - still in
                                  validationor iterative reweighted
                                  least squares (different weights per
                                  time frames) can be used
                                  - to be validated
LIMO.design.type_of_analysis    = 'Mass-univariate'
```

```
LIMO.design.bootstrap          = 0/1 indicates if bootstrap should be
                                 performed or not (by default 0 for
                                 group studies)
LIMO.design.tfce               = 0/1 indicates to compute TFCE or not
```

Create design matrix = limo_eeg(3) → *limo_design_matrix.m*

Create the design matrix using the information stored in LIMO.mat. At this stage, we read the data and reorganize the trial order to be grouped by categorical variable. This does not affect the analyses but allows visualizing the design matrix more efficiently. New files are created on the disk at this stage: Yr (the data), Yhat (the modeled data), Re (residuals), R2 (model fit), Beta (parameters).

```
LIMO.design.X                = 2 dimensional matrix that describes the
                               experiments' events
LIMO.design.nb_conditions    = vector that returns the number of
                               conditions per factor e.g. [2 2 2]
LIMO.design.nb_interactions = vector that returns the number of
                               conditions perinteraction e.g. [4 4 4]
LIMO.design.nb_continuous    = scalar that returns the number of
                               continuous variables e.g. [3]
LIMO.design.name             = name of the design
LIMO.design.status           = 'to do'
```

Compute parameter estimates = limo_eeg(4)

If LIMO.design.type_of_analysis == 'Mass-univariate' → limo_glm1.m

Analyzes are carried out such as the effect of experimental variables on the EEG time are tested for each electrode and time frames independently (in practice however the full time course is analyzed simultaneously electrode by electrode). Files previously created are updated. Depending on the design, files of the statistical results are also saved on the disk: Condition_effect_X (refers to a factor effect), Interaction_effect_X (refers to an interaction between factors), Covariate_effect_X (refers to the effect of a continuous regressor).

```
LIMO.design.weights = matrix of trial weights

LIMO.model                  = information about the statistics
LIMO.model.conditions_df    = df [effect, error]
LIMO.model.interactions_df = df [effect, error]
LIMO.model.continuous_df    = df [effect, error]
LIMO.design.status          = 'done'
```

if LIMO.design.bootstrap == 1 → limo_glm1_boot.m

If the option of computing a bootstrap at this stage was selected, the above analyses are repeated under H0. Having the distribution of values under H0 allows computing, for a given threshold, the distribution of maximum values or maximum cluster mass under the null hypothesis, and therefore threshold the map of observed results based on this distribution. A new folder (H0) is created, the files saved onto the disk: boot_table (the resampling table), H0_Betas, H0_R2, H0_Condition_effect_X, H0_Interaction_effect_X, H0_Covariate_effect_X.

if LIMO.design.tfce == 1 → limo_tfce.m

The Threshold Free Cluster Enhancement technique allows to threshold the data cellule per cellule using the TFCE score. A TFCE score is the integral of T or F values weighted at multiple thresholds by the cluster extend. For each statistical effect (R2, Condition, etc) a TFCE score is computed, and similarly a TFCE score is obtained for each bootstrap computed under H0. TFCE scores can then be thresholded using the maximum of the distribution of TFCE scores under H0. A new folder (TFCE) is created and maps of TFCE scores saved onto the disk: tfce_R2, tfce_Condition_effect_X, tfce_Interaction_effect_X, tfce_Covariate_effect_X. Similarly, in the H0 folder, TFCE scores under H0 are saved onto the disk: tfce_H0_R2, tfce_H0_Condition_effect_X, tfce_H0_Interaction_effect_X, tfce_H0_Covariate_effect_X.

Visualize all results = limo_eeg(5)

Simply iteratively calls limo_display_results.m to plot F values at all electrodes and time frames for each effect, thresholding using whatever method is available (by preference order TFCE, spatial-temporal clustering, none).

Further analyses

One important feature of multiple regression / glm analyses is the ability to know the contribution of each regressor to the data. Via the result GUI (*limo_results.m*) one can make a contrast (*limo_contrast.m*) that will reveal the unique variance of a regressor given the presence of (or controlling for, regressing out) the other regressors. Of course, contrasts can also be used to compute effects of variables or differences between regressors. The result GUI also allows to compute semi-partial coefficients (*limo_semipartial_coef.m*) with represents the total variance of a regressor relative to the data.

Semi Partial correlation coefficients → limo_semi_partial_coef.m

Run the analysis automatically for each factor and covariate, and save files on the disk.

Contrasts → limo_eeg(6) // limo_contrast_manager.m

An automatic way to perform contrasts is to (1) create a .mat file of contrasts and (2) call limo_eeg(6). Alternatively, we can use the *limo_contrast_manager.m*

Output Files from 1st level analysis

In the subject analysis folder, the data, parameters and results are saved

Yr.mat: the EEG data from the .set reorganized to fit X, that is grouped by conditions if Cat ~=0; dimension [electrodes x time frames x trials]
Yhat.mat: the predicted data; dimension [electrodes x time frames x trials]
Beta.mat: the beta values (parameter estimates); dimension [electrodes x time frames x number of parameters in the model (columns of X)]
Res.mat: the residuals (non modelled) data; dimension [electrodes x time frames x trials]
R2.mat: the model fit statistic, i.e. percentage of variance explained; dimension [electrodes x time frames x R^2/F/p values]
Condition_effect_X: refers to a factor effect in categorical designs; dimension [electrodes x time frames x F/p values).
Interaction_effect_X: refers to an interaction between factors, dimension [electrodes x time frames x F/p values).
Covariate_effect_X: refers to the effect of a continuous regressor, dimension [electrodes x time frames x F/p values).
semi_partial_coef_X.mat: refers to the semi-partial coefficient of a factor (defined by LIMO.design.nb_conditions) or a covariate (defined in LIMO.design.nb_continuous), dimension [electrodes x frames x R2/F/p values].
con_X.mat: refers to a t contrast, dimension [electrodes x frames x cB/standard error/df/t/p values].
ess_X.mat: refers to a F contrast, dimension [electrodes x frames x cB/standard error/df/F/p values].

--- adding bootstrap

In the H0 subfolder, the data under H0 are saved

boot_table: the resampling table used, dimension [number of trials x number of bootstraps]
H0_Betas: dimension [electrodes x time frames x number of parameters in the model (columns of X) x number of bootstraps]
H0_R2: dimension: [electrodes x time frames x R^2/F/p values x number of bootstraps]
H0_Condition_effect_X: dimension [electrodes x time frames x F/p values x number of bootstraps]
H0_Interaction_effect_X: dimension [electrodes x time frames x F/p values x number of bootstraps]
H0_Covariate_effect_X: dimension [electrodes x time frames x F/p values x number of bootstraps]
H0_semi_partial_coef_X.mat: dimension [electrodes x frames x R2/F/p values x number of bootstraps].
H0_con_X.mat: dimension [electrodes x frames x cB/t/p values x number of bootstraps].
H0_ess_X.mat: dimension [electrodes x frames x cB/F/p values x number of bootstraps].

--- adding tfce

In the TFCE subfolder, the tfce score maps are saved

tfce_R2: dimension: dimension [electrodes x time frames]
tfce_Condition_effect_X: dimension [electrodes x time frames]

tfce_Interaction_effect_X: dimension [electrodes x time frames]
tfce_Covariate_effect_X: dimension [electrodes x time frames]
tfce_semi_partial_coef_X.mat: dimension [electrodes x frames].
tfce_con_X.mat: dimension [electrodes x frames].
tfce_ess_X.mat: dimension [electrodes x frames].

In the H0 subfolder, the tfce score maps under H0 are saved

tfce_H0_R2: dimension [electrodes x time frames x number of bootstraps]
tfce_H0_Condition_effect_X: dimension [electrodes x time frames x number of bootstraps]
tfce_H0_Interaction_effect_X: dimension [electrodes x time frames x number of bootstraps]
tfce_H0_Covariate_effect_X: dimension [electrodes x time frames x number of bootstraps]
tfce_H0_semi_partial_coef_X.mat: dimension [electrodes x frames x number of bootstraps].
tfce_H0_con_X.mat: dimension [electrodes x frames x number of bootstraps].
tfce_H0_ess_X.mat: dimension [electrodes x frames x number of bootstraps].

2nd level analysis (group)

To test for effects across subjects, the beta parameters or con values computed for each subjects are combined. This is similar in spirit to using mean values (ERPs) obtained for each subjects. However, for each subject the beta values reflect the adjusted mean of each effect such as it is weighted by the within subject variance. At the 2nd level, the analysis take into account the between subjects variance and therefore, under the assumption that errors are i.i.d. at the 1st level, these analyses are real random effects (within and between variance is accounted for). 2nd level analyses uses robust estimators (trimmed means) and bootstrap thus dealing with non normality and heteroscedasticity.

LIMO_random_effect.m is the GUI allowing selecting the appropriate statistics (t-tests, regression, ANOVA). The data selection is performed by *limo_random_select.m* and tests performed via *limo_random_robust.m*. In this last function one calls the various tests to be performed on the actual data.

One sample t-test (*limo_trimci.m*)

Computes a one-sample t-test using 20% trimmed mean and winsorized variance. Once the data are selected the LIMO.mat contains the following information is created

```
LIMO.dir: where data are stored
LIMO.Level = 2;
LIMO.data.chanlocs: channel locations for the cap
LIMO.data.neighbouring_matrix: neighbourhood matrix
LIMO.data.data: names of the files to read
LIMO.data.data_dir: directories of the files to read
LIMO.data.sampling_rate: taken across all subjects
LIMO.data.trim1: starting frame (the latest across subjects)
LIMO.data.start: starting time
LIMO.data.trim2: ending frame (the earliest across subjects)
LIMO.data.end: ending time
```

Output Files:
Yr: the data used for the analysis
one_sample_ttest_parameter_X.mat: returns the parameter values and associated statistics (dimensions electrodes x time frames x 5). The last dimension codes for trimmed mean values, standard error, degrees of freedom, t and p.
H0_one_sample_ttest_parameter_X.mat: this file constrains the T and p values obtained under H0 for each bootstrap (dimensions electrodes x time frames x 2 x nboot)
tfce_one_sample_ttest_parameter_X.mat: tfce scores of the t-test (dimensions electrodes x time frames)
H0_tfce_one_sample_ttest_parameter_X.mat: tfce scores of the t-test under H0 (dimensions electrodes x time frames x nboot)

Two samples t-test (*limo_yuen_ttest.m*):

Computes a two-samples t-test based on 20% trimmed mean and winsorized variances across samples.

Once the data are selected the LIMO.mat contains the following information is created

```
LIMO.dir: where is the LIMO.mat and Yr
LIMO.data.chanlocs: chanel locations from the expected electrode file
LIMO.data.neighbouring_matrix: binary matrix of neighbourhood
LIMO.data.data: 2 sets of cells e.g. {{1x10 cell}  {1x8 cell}} with
the full name of the Betas or con files
LIMO.data.data_dir: cells with directories of the Betas or con files
LIMO.data.sampling_rate: should be the same across subjects
LIMO.data.trim1: 1st data point to analyze
LIMO.data.start: 1st data point to analyze in sec
LIMO.data.trim2: last data point to analyze
LIMO.data.end: last data point to analyze in sec
LIMO.design.bootstrap: nb of bootstrap to perform (0 if none)
LIMO.design.tfce: 0 or 1
LIMO.design.name: 'one 'two samples t-test all electrodes'
LIMO.design.electrode: [] (or 1value or set of values for optimized
electrode analysis)
LIMO.design.X: []
LIMO.Level = 2;
```

Output Files:
Y1r and Y2r: this is the data for each group (electrodes * time frames * subjects)
two_samples_ttest_parameter_X.mat: returns the parameter values and associated statistics (dimensions electrodes x time frames x 5). The last dimension codes for trimmed mean difference, df, standar error, t and p.
H0_two_samples_ttest_parameter_X.mat: this file constrains the trimmed mean differences, t and p values obtained under H0 for each bootstrap (dimensions electrodes x time frames x 3 x nboot)
tfce_two_sample_ttest_parameter_X.mat: tfce scores of the t-test (dimensions electrodes x time frames)
H0_tfce_two_sample_ttest_parameter_X.mat: tfce scores of the t-test under H0 (dimensions electrodes x time frames x nboot)

Paired t-test (*limo_yuend_ttest.m*):

Computes a paired-samples t-test using 20% trimmed mean and winsorized variance. Once the data are selected the LIMO.mat contains the following information is created

```
LIMO.dir: where is the LIMO.mat and Yr
LIMO.data.chanlocs: chanel locations from the expected electrode file
LIMO.data.neighbouring_matrix: binary matrix of neighbourhood
LIMO.data.data: cells with the full name of the Betas or con files
LIMO.data.data_dir: cells with directories of the Betas or con files
LIMO.data.sampling_rate: should be the same across subjects
LIMO.data.trim1: 1st data point to analyze
LIMO.data.start: 1st data point to analyze in sec
LIMO.data.trim2: last data point to analyze
```

```
LIMO.data.end: last data point to analyze in sec
LIMO.design.bootstrap: nb of bootstrap to perform (0 if none)
LIMO.design.tfce: 0 or 1
LIMO.design.name: 'paired t-test all electrodes'
LIMO.design.electrode: [] (or 1value or set of values for optimized
electrode analysis)
LIMO.design.X: []
LIMO.Level = 2;
```

Output Files:
Y1r and Y2r: this is the data for each parameter (electrodes * time frames * subjects)
paired_ttest_parameter_X.mat: returns the parameter values and associated statistics (dimensions electrodes x frames x 5). The last dimension codes for trimmed mean difference, df, standar error, t and p.
H0_paired_ttest_parameter_X.mat: this file constrains the trimmed mean differences, t and p values obtained under H0 for each bootstrap (dimensions electrodes x time frames x 3 x nboot)
tfce_paired_ttest_parameter_X.mat: tfce scores of the t-test (dimensions electrodes x time frames)
H0_tfce_paired_ttest_parameter_X.mat: tfce scores of the t-test under H0 (dimensions electrodes x time frames x nboot)

Regression, ANOVA and ANCOVA analyzes (*limo_glm1.m*):

This is the same information as for 1st level analysis.

Repeated measure ANOVA *(limo_rep_anova.m):*

Once the selection is done a LIMO.mat file is created with the following information

```
LIMO.dir: where is the LIMO.mat and Yr
LIMO.data.chanlocs: chanel locations from the expected electrode file
LIMO.data.neighbouring_matrix: binary matrix of neighbourhood
LIMO.data.data: cells with the full name of the Betas or con files
LIMO.data.data_dir: cells with directories of the Betas or con files
LIMO.data.sampling_rate: should be the same across subjects
LIMO.data.trim1: 1st data point to analyze
LIMO.data.start: 1st data point to analyze in sec
LIMO.data.trim2: last data point to analyze
LIMO.data.end: last data point to analyze in sec
LIMO.design.bootstrap: nb of bootstrap to perform (0 if none)
LIMO.design.tfce: 0 or 1
LIMO.design.name: 'Repeated measures ANOVA all electrodes'
LIMO.design.electrode: [] (or 1value or set of values for optimized
electrode analysis)
LIMO.design.X: []
LIMO.Level = 2;
```

On the drive, Yr is saved of dimensions electrodes * time frames * subjects * repeated measures. The order of repeated measures follows the 'usual' order of statistical software. For instance for a 3*2 design we have 6 conditions in the f1(1)f2(1) f1(1)f2(2) f1(2)f2(1) f1(2)f2(2) f1(3)f2(1) f1(3)f2(2).

Depending on the design are also saved
Rep_ANOVA_factor_X.mat: dimension electrodes * time frames * F/p values
Rep_ANOVA_gp_effect.mat: dimension electrodes * time frames * F/p values
Rep_ANOVA_Interaction_gp_Factor_X.mat: dimension electrodes * time frames * F/p values