

Jonathan Chauvin, Géraldine Ayrat, Pauline Traynard

(1) Lixoft, Antony, France. [Contact: jonathan.chauvin@lixoft.com](mailto:jonathan.chauvin@lixoft.com)



Online documentation for Rsmplx package containing the COSSAC algorithm

Introduction

Covariate selection is a key element in the modeling process. As it involves a large number of runs, automatic covariate selection procedures have been developed. The most commonly used method is SCM (stepwise covariate modeling). The SCM procedure includes a forward selection, in which covariates are added one by one after evaluation of all possible additions, followed by a backward elimination. This method is effective but expensive in terms of number of runs. We propose an alternative method called COSSAC based on the individual parameter-covariate correlations observed in the current model. The method is available as an R script to be used with MonolixSuite2018.

COSSAC for covariate selection in R

The SCM and COSSAC covariate selection algorithms are implemented in the Rsmplx package available on the CRAN, to be used with Monolix projects.

```

covariateSearch(project, final.project, covToTest,
covToTransform, paramToUse, test_relations, settings)
- project: path to the base Monolix project
- final.project: name of the final project to save
- covToTest: list of covariates to test,
- covToTransform: list of covariates to logtransform and center
- paramToUse: list of parameters on which the covariates should be tried
- test_relations: dedicated parameter-covariate relationship to test
- method: method used for the covariate selection. It can be 'COSSAC' or 'SCM'.
- settings: some settings for the algorithms, the choice of criteria, direction of the algorithm, ...
    
```

SCM and COSSAC will be available through the Monolix GUI in the 2019 version.

Concept

COSSAC makes use of the correlation between the random effects and the covariates of the current model to choose which parameter-covariate relationship to try next. To avoid biased correlations that may appear (for instance in case of shrinkage), samples from the a posteriori conditional distribution are used instead of the EBES.

COSSAC algorithm

Initialization:

- Run the base model (population parameter estimation, conditional distribution sampling, and log-likelihood estimation).
- Calculate the p-values of all the parameter-covariate relationships using Pearson's correlation tests and ANOVA (automatically done in Monolix 2018)

Forward selection:

1. Add the covariate with the smallest correlation p-value (among the remaining parameter-covariate relationships) to the model
2. Run the model
3. Accept/reject the relationship based on the likelihood ratio test or BIC
4. Calculate all parameter-covariate correlation p-values. Go back to step 1 until no correlation p-values above a threshold remain

Backward selection:

5. Among the covariates present in the model, remove the covariate with the highest (less significant) correlation p-value
6. Run the model
7. Accept/reject the relationship removal based on the likelihood ratio test or BIC
8. Calculate all parameter-covariate relationship correlation p-values. Go back to step 5 until no correlation p-values below a threshold remain

Comparison of SCM and COSSAC

The SCM and COSSAC covariate selection algorithms are compared on three representative data sets. The same final model is obtained with both method, but the number of iterations required with COSSAC is much lower than with SCM.

Data set	Base model	Available covariates	Final covariate model	Number of iterations	
				SCM	COSSAC
Remifentanil: dense PK data	6 parameters: 3-cpt PK model with infusion	6 covariates: age, sex, height, weight, BSA (body surface area), LBM (lean body mass)	<i>Same for SCM and COSSAC:</i> - age and BSA on CI - age and LBM on V1 - age on V2 and Q3 - age and LBM on V3	298	15
Tobramycin: sparse PK data	4 parameters: 2-cpt model with bolus	4 covariates: age, sex, weight, creatinine clearance	<i>Same for SCM and COSSAC:</i> - weight and creatine clearance on k - weight on k12	63	17
NCCTG lung cancer: time-to-death data	2 parameters: Gompertz model	5 covariates: age, sex, ECOG, Karno score by patient, Karno score by physician	<i>Same for SCM and COSSAC:</i> - ECOG and sex on Te - age on parameter k	39	14