Guideline for codes

Restricted mean survival time to estimate an intervention effect in a cluster randomized trial

Floriane Le Vilain–Abraham, Elsa Tavernier, Etienne Dantan,
Solène Desmée, Agnès Caille

Contents

0   Introduction

The code is divided into two folders: one for the scenarios under proportional-hazards (PH) assumption and one under non-proportional-hazards (NPH) assumption. For both cases, there are 2 steps:

1. Generation of the datasets and statistical analysis

2. Estimation of the performance measures

The following sections describe how to used the R code for each step.

1   Generation of the datasets and statistical analysis

The `main_DA_PH.R` (for the PH assumption) and `main_DA_NPH.R` (for the NPH assumption) files are R codes to generate 1000 simulated datasets of a prefined scenario and analyse each dataset with the 5 methods (KM_{indep}, PV_{indep}, KM_{clust}, PV_{ECM} and PV_{ICM}). For the permutation test, see the subsection 1.

The necessary packages are: `doParalell` (line 14) and `doRNG` (line 15). The necessary R functions, saved in a R file with the same name, are loaded in the main script (`main_DA_PH.R` or `main_DA_NPH.R`):

- `sim_PH` or `sim_NPH` (line 20): simulates a dataset and estimates the difference in RMST, its variance and the 95% confidence interval with the 5 methods

- `generate_data_PH` or `generate_data_NPH` (line 21): generates one dataset

- `generate_cluster_PH` or `generate_cluster_NPH` (line 22): generates time-to-event data for one cluster

- `RMST_pseudo` (line 23): analyses one dataset with the pseudo-value regression-based methods (accounting and not accounting for clustering)

- `RMST_KM` (line 24): analyses one dataset with the Kaplan-Meier-based methods (accounting and not accounting for clustering)

- `variance_bootstrap` (line 25): resamples one dataset and estimates the difference in RMST with the Kaplan-Meier-based method (function to estimate the variance of the difference in RMST using bootstrap method)

The `RMST_pseudo`, `RMST_KM` and `variance_bootstrap` functions are identical for PH and NPH assumptions.

Before running, you have to specify in the main file (`main_DA_PH.R` or `main_DA_NPH.R`):

- The directory where you saved all the R files with the necessary R functions detailed previously (line 19).

- The parameters of the scenario (line 30 in the data frame `table_parameter`).

- The horizon time (line 41). Here, it is set at 365 days. All time are expressed in days.

- The directory where the 1000 simulated datasets and the estimations of the difference in RMST, the variance and 95% confidence interval will be saved (line 47). The 1000 datasets will be saved in a same automatically created folder. The estimations of the difference in RMST, variance and 95% confidence interval for the 5 methods and for the 1000 simulated datasets will be saved in one txt file. Table 1 and 2 described the content of one simulated dataset and of the txt file with all the estimations, respectively.

- The number of cores for the parallelisation (line 76).

Colnames	Variables
`time`	X_{lk}, the observed time of the individual l from the cluster k
`arm`	arm of the clusters k (= 1 for the intervention group, 0 for control group)
`cluster`	cluster's identifiant k
`id_patient`	patient's identifiant l
`status`	δ_{lk}, the event indicator of the individual l from the cluster k (= 1 for death, 0 for censor)

Table 1: Content of one simulated dataset

Colnames	Variables
	Parameters used to simulate the dataset
`d`	dataset number
`K`	total number of clusters
`m`	mean cluster size
`HR`	true HR
`t_delay`	change point (only for NPH assumption)
`tau`	Kendall's tau
`censoring`	censoring rate
	Analysis
`clustering`	if the method accounts (=`"yes"`) or does not account (=`"no"`) for clustering
`delta.rmst`	estimation of the difference in RSMT
`var`	variance estimation
`ci.low`	lower bound of the 95% confidence interval
`ci.up`	upper bound of the 95% confidence interval
`method`	type of method used for the analysis
`t_star`	horizon time of the analysis

Table 2: Content of txt file with the estimations of the difference in RMST, variance and 95% confidence intervals for the 5 methods and for the 1000 simulated datasets

Permutation test for the pseudo-values regression-based method: Statistical analysis

The permutation tests for the pseudo-values regression-based method with a independent and exchangeable working correlation matrix are performed separately. All the needed functions are available in the *Permutation test* forder (*/Proportional-hazards/1 - Generation of the datasets and statistical analysis/*).

The `main_permutation.R` file is the R code to analyse each simulated dataset with the permutation test. The necessary packages are: `doParalell` (line 14) and `doRNG` (line 15).

3

The necessary R functions, saved in a R file with the same name, are loaded in the main file (`main_permutation.R`):

- `sim_permutation` (line 20): to estimate the p-values for the pseudo-values regression-based methods

- `permutation` (line 21): to analyse of one dataset with the permutation test

- `permutation_test` (line 22): to estimate the test statistic for one permuted dataset

Before running, you have to specify in the `main_permutation.R` file:

- The directory where you saved all the R files with the necessary R functions detailed previously (line 19).

- The parameters of the scenario (line 27 in the data frame `table_parameter`).

- The horizon time (line 38). Here, it is set at 365 days. All time are expressed in days.

- The directory where the estimations of the p-values will be saved (line 45). The estimations of the p-values for the 1000 simulated datasets will be saved in one txt file. Table 3 describes this txt file.

- The number of cores for the parallelisation (line 69).

You also have to specify in the `sim_permutation.R` file, the directory where the simuled datasets have been saved (line 24).

Colnames	Variables
Parameters used to simulate the dataset	
d	dataset number
K	total number of clusters
m	mean cluster size
HR	true HR
tau	Kendall's tau
censoring	censoring rate
Analysis	
matrix	type of working correlation matrix used for the analysis ("ind" = independent, "exch" = exchangeable)
var	estimated p-value
t_star	horizon time of the analysis

Table 3: Content of txt file with the estimations of p-value for pseudo-values regression for the 1000 simulated datasets

2   Estimation of the performance measures

The `main_PM_PH.R` (for the PH assumption) and `main_PM_NPH.R` (for the NPH assumption) files are R codes to estimate the performance measures for the 5 methods (KM_{indep}, PV_{indep}, KM_{clust}, PV_{ECM} and PV_{ICM}). For the permutation test, see the subsection 2.

No package is necessary. The necessary R functions, saved in a R file with the same name, are loaded in the main script (`main_PM_PH.R` or `main_PM_NPH.R`):

- `pm_estimation_PH` or `pm_estimation_NPH` (line 15): estimates the performance measures for the 5 methods

- `performance_measures` (line 16): estimates the performance measure for one method

- `true_rmst_difference_PH` or `true_rmst_difference_NPH` (line 17): computes the true difference in RSMT

- `survival_function_PH` or `surival_function_NPH` (line 18): true survival function

Before running, you have to specify:

- The directory where you saved all the R files with the necessary R functions detailed previously (line 14).

- The directory and the name of the txt file where the estimations of the difference in RMST, variance and 95% confidence interval for the 5 methods and for the 1000 datasets have been saved in the step 1 (line 23).

When you run the main R script, the performance measures are summurized in a data frame. The content of this data frame is detailed in Table 4.

Colnames	Variables
method	method
tstar	horizon time for the analysis
K	total number of clusters
m	mean cluster size
HR	true HR
tau	Kendall's tau
censoring	censoring rate
RB	relative bias
RE	relative error
coverage	coverage rate
rejection.rate	type I error rate (for the null hypothesis, i.e. absence of intervention effect)
D	number of simulation iterations that converge

Table 4: Content of the data frame obtained when running the main scripts `main_PM_PH.R` (for the PH assumption) or `main_PM_NPH.R` (for the NPH assumption)

Permutation test for the pseudo-values regression-based method: Estimation of the type I error rate

The performance of the permutation test for the pseudo-values regression is computed separately. All the needed functions are available in the *Permutation test* forder (*/Proportional-hazards/2 - Estimation of the performance measures/*).

The `main_PM_permutation.R` file is R codes to estimate the type I error rate for the pseudo-values regression-based method. No package is necessary. The following necessary R function, saved in a R file with the same name, is loaded in the main script:

- `pm_estimation_permuation` (line 15): estimation of the type I error rate with the permutation test for the pseudo-values regression based-methods

Before running, you have to specify in the `main_PM_permutation.R` file:

- The directory where you saved all the R files with the necessary R function detailed previously (line 14).

- The directory and the name of the txt file where the estimations of the p-values for the the permutation test for the pseudo-values regression-based methods for the 1000 datasets have been saved in the step 1 (line 21).

When you run the main R script, the type I error rates are summurized in a data frame. The content of this data frame is detailed in Table 5

Colnames	Variables
tstar	horizon time for the analysis
K	total number of clusters
m	mean cluster size
HR	true HR
tau	Kendall's tau
censoring	censoring rate
matrix	type of working correlation matrix used for the analysis
tie	type I error rate

Table 5: Content of the data frame obtained when running the main script main_PM_permutation.R