

Package ‘abcsmcrf’

October 16, 2025

Type Package

Version 1.0.0

Date 2024-07-10

Title Approximate Bayesian Computation sequential Monte Carlo via random forests

Description Parameter inference using random forests in the framework of Approximate Bayesian Computation sequential Monte Carlo.

License GPL (≥ 2)

Imports abcrf,
crayon,
data.table,
dplyr,
drf,
EasyABC,
ggplot2,
Hmisc,
MASS,
matrixStats,
parallel,
pbapply,
RColorBrewer

Suggests coala,
deSolve,
invgamma,
knitr,
LaplacesDemon,
Pareto,
patchwork,
reshape2,
rmarkdown,
sensitivity,
tidyverse,
tseries,
weights

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

R topics documented:

plot_compare_joint	2
plot_compare_marginal	3
plot_compare_qqplot	4
plot_smcrgf_joint	5
plot_smcrgf_marginal	6
smcrgf	8

Index	11
--------------	-----------

plot_compare_joint	<i>Plot and compare joint posterior distribution(s) from ABC-SMC-(D)RF result</i>
--------------------	---

Description

[plot_compare_joint](#) plots the joint posterior distribution(s) for the provided ABC-SMC-(D)RF result. It can also compare the joint posterior distributions for the provided ABC-SMC-(D)RF result with several ABC-SMC-(D)RF plots results.

Usage

```
plot_compare_joint(
  plots = NULL,
  abc_results,
  parameters_truth = NULL,
  parameters_labels = NULL,
  lims = NULL,
  nBins = 4
)
```

Arguments

plots	An existed ABC-SMC-(D)RF joint plots result. If provided, plot_compare_joint will plot the new ABC-SMC-(D)RF result and compare it with provided plots results. If plots = NULL, plot_compare_joint will make a new plot for the ABC-SMC-(D)RF result.
abc_results	An ABC-SMC-(D)RF result. Will be plotted out by the function.
parameters_truth	A dataframe containing true values of parameters from the ground-truth distributions. If provided, the function will plot the true values or distributions of parameters.
parameters_labels	A dataframe containing labels in the plots for corresponding parameters. If provided, parameter labels will be exhibited on the plots' axes.
lims	A dataframe containing the maximum and minimum bounds for parameters. If provided, x-axis and y-axis will be scaled by them.
nBins	Number of contour bins shown in the plot. Default is 5.

Value

A list of ggplot2 objects containing the joint plots results of posterior distributions. The user can use the function to compare ABC-SMC-(D)RF joint plots with the joint posterior distribution(s) of other ABC-SMC-(D)RF result(s).

plot_compare_marginal	<i>Plot and compare marginal posterior distribution(s) from ABC-SMC-(D)RF result</i>
-----------------------	--

Description

`plot_compare_marginal` plots the marginal posterior distribution(s) for the provided ABC-SMC-(D)RF result. It can also compare the marginal posterior distributions for the provided ABC-SMC-(D)RF result with several ABC-SMC-(D)RF plots results.

Usage

```
plot_compare_marginal(
  plots = NULL,
  abc_results,
  parameters_truth = NULL,
  parameters_labels = NULL,
  statistics_labels = NULL,
  xlimit = NULL,
  plot_statistics = FALSE,
  plot_truth_hist = TRUE,
  plot_hist = FALSE,
  plot_hist_point = FALSE,
  breaks = NULL,
  alpha_truth = 0.8,
  alpha = 0.3,
  plot_prior = FALSE
)
```

Arguments

<code>plots</code>	An existed ABC-SMC-(D)RF marginal plots result. If provided, <code>plot_compare_marginal</code> will plot the new ABC-SMC-(D)RF result and compare it with provided plots results. If <code>plots = NULL</code> , <code>plot_compare_marginal</code> will make a new plot for the ABC-SMC-(D)RF result.
<code>abc_results</code>	An ABC-SMC-(D)RF result. Will be plotted out by the function.
<code>parameters_truth</code>	A dataframe containing true values of parameters from the ground-truth distributions. If provided, the function will plot the true values or distributions of parameters.
<code>parameters_labels</code>	A dataframe containing labels in the plots for corresponding parameters. If provided, parameter labels will be exhibited on the plots' axes.

statistics_labels	A dataframe containing labels in the plots for corresponding statistics. If provided, statistics labels will exhibit on the plots' axes.
xlimit	A dataframe containing the maximum and minimum bounds for parameters. If provided, the x-axis will be scaled by them.
plot_statistics	A logic variable (plot_statistics = FALSE by default). If plot_statistics = TRUE, the marginal distributions in each iteration for corresponding statistics will also be output.
plot_hist	A logic variable (plot_hist = FALSE by default). If plot_hist = TRUE, marginal distributions will be plotted in histograms.
plot_hist_point	A logic variable (plot_hist_point = FALSE by default). If plot_hist_point = TRUE, marginal distributions will be plotted in histograms with points in the middle.
alpha	The numeric number to modify transparency. Default is 0.3.
plot_prior	A logic variable (plot_prior = FALSE by default) If plot_prior = TRUE, the prior distribution will be plotted out.

Value

A list of ggplot2 objects containing the marginal plots results of posterior distributions. The user can use the function to compare ABC-SMC-(D)RF marginal plots with the marginal posterior distribution(s) of other ABC-SMC-(D)RF result(s).

See Also

[smcrf](#)

plot_compare_qqplot	<i>Plot and compare marginal quantile-quantile plots from ABC-SMC-(D)RF result</i>
---------------------	--

Description

[plot_compare_qqplot](#) plots the marginal quantile-quantile plots for inferred parameters and parameters from ground-truth distributions.

Usage

```
plot_compare_qqplot(
  plots = NULL,
  abc_results,
  parameters_truth,
  parameters_labels = NULL,
  lims = NULL
)
```

Arguments

plots	An existed ABC-SMC-(D)RF quantile-quantile plots result. If provided, plot_compare_qqplot will plot the quantile-quantile plot for inferred parameters in the new ABC-SMC-(D)RF result and compare it with provided quantile-quantile plots result. If plots = NULL, plot_compare_qqplot will make a new quantile-quantile plot for inferred parameters in the ABC-SMC-(D)RF result and true parameters.
abc_results	An ABC-SMC-(D)RF result. The function will plot the quantile-quantile plot between the inferred parameters from the ABC-SMC-(D)RF result and true parameters.
parameters_truth	A dataframe containing true values of parameters from the ground-truth distributions.
parameters_labels	A dataframe containing labels in the plots for corresponding parameters. If provided, parameter labels will be exhibited on the plots' axes.
lims	A dataframe containing the maximum and minimum bounds for parameters. If provided, x-axis and y-axis will be scaled by them.

Value

A list of ggplot2 objects containing the quantile-quantile plots results. The user can use the function to compare ABC-SMC-(D)RF quantile-quantile plots with the quantile-quantile plots of other ABC-SMC-(D)RF result(s).

Examples

```
# parameters_truth is optional here!
parameters_truth <- data.frame(
  theta = rnorm(10000, mean = 2, 1)
)
# Dataframe containing the parameter labels
parameters_labels <- data.frame(
  parameter = c("theta"),
  label = c(deparse(expression(theta)))
)
# Dataframe containing the x-axis bounds
xlimit <- data.frame(
  parameter = c("theta"),
  min = c(1),
  max = c(20)
)
```

plot_smcrf_joint	<i>Plot joint distribution(s) of each iteration from ABC-SMC-(D)RF result</i>
------------------	---

Description

[plot_smcrf_joint](#) plots the joint distribution(s) of two parameters for each iteration from an Approximate Bayesian Computation sequential Monte Carlo via random forest result.

Usage

```
plot_smcrf_joint(
  smcrf_results,
  parameters_truth = NULL,
  parameters_labels = NULL,
  lims = NULL,
  nBins = 5
)
```

Arguments

<code>smcrf_results</code>	An ABC-SMC-(D)RF result containing the inference distributions of parameters from each iteration.
<code>parameters_truth</code>	A dataframe containing true values of parameters from the ground-truth distributions. If provided, the function will plot the true values or distributions of parameters.
<code>parameters_labels</code>	A dataframe containing labels in the plots for corresponding parameters. If provided, parameter labels will be exhibited on the plots' axes.
<code>lims</code>	A dataframe containing the maximum and minimum bounds for parameters. If provided, x-axis and y-axis will be scaled by them.
<code>nBins</code>	Number of contour bins shown in the plot. Default is 5.

See Also

[smcrf](#)

Examples

```
lims <- data.frame(
  parameter = c("theta", "mu"),
  min = c(0, 0),
  max = c(10, 10)
)
parameters_truth <- data.frame(
  theta = rnorm(10000, mean = 2, sd = 1),
  mu = rnorm(10000, mean = 1, sd = 2)
)
```

<code>plot_smcrf_marginal</code>	<i>Plot distribution(s) in each iteration from ABC-SMC-(D)RF result</i>
----------------------------------	---

Description

[plot_smcrf_marginal](#) plot the marginal distribution(s) for each iteration from an Approximate Bayesian Computation sequential Monte Carlo via random forest result.

Usage

```
plot_smcrg_marginal(
  smcrg_results,
  parameters_truth = NULL,
  parameters_labels = NULL,
  statistics_labels = NULL,
  plot_statistics = FALSE,
  xlimit = NULL,
  alpha = 0.3,
  plot_hist = FALSE
)
```

Arguments

<code>smcrg_results</code>	An ABC-SMC-(D)RF result containing the inference distributions of parameters from each iteration.
<code>parameters_truth</code>	A dataframe containing true values of parameters from the ground-truth distributions. If provided, the function will plot the true values or distributions of parameters.
<code>parameters_labels</code>	A dataframe containing labels in the plots for corresponding parameters. If provided, parameter labels will be exhibited on the plots' axes.
<code>statistics_labels</code>	A dataframe containing labels in the plots for corresponding statistics. If provided, statistics labels will exhibit on the plots' axes.
<code>plot_statistics</code>	A logic variable (<code>plot_statistics = FALSE</code> by default). If <code>plot_statistics = TRUE</code> , the marginal distributions in each iteration for corresponding statistics will also be output.
<code>xlim</code>	A dataframe containing the maximum and minimum bounds for parameters. If provided, the x-axis will be scaled by them.
<code>alpha</code>	The numeric number to modify transparency. Default is 0.3.
<code>plot_hist</code>	A logic variable (<code>plot_hist = FALSE</code> by default). If <code>plot_hist = TRUE</code> , marginal distributions will be plotted in histograms.

See Also

[smcrg](#)

Examples

```
# Dataframe containing the true parameters
parameters_truth <- data.frame(
  theta = 2 # this will plot a vertical line
)
if you prefer to have a histogram of true posterior
parameters_truth <- data.frame(
  theta = rnorm(10000, mean = 2, 1)
)
# Dataframe containing the parameter labels
parameters_labels <- data.frame(
```

```

    parameter = c("theta"),
    label = c(deparse(expression(theta)))
  )
#   Dataframe containing the x-axis bounds
xlimit <- data.frame(
  parameter = c("theta"),
  min = c(1),
  max = c(20)
)

```

smcrf

Approximate Bayesian Computation sequential Monte Carlo via random forests

Description

`smcrf` uses random forests to find the posterior distribution(s) for one or more parameters in a model. It implements the sequential Monte Carlo framework, where each iteration uses either ABC-RF (functions `regAbcrf` and `predict` in R package `abcrf`) or ABC-DRF (functions `drf` and `predict` in R package `drf`) to update the posterior distribution(s).

Usage

```

smcrf(
  method = "smcrf-single-param",
  statistics_target = NULL,
  statistics_selection = NULL,
  smcrf_results = NULL,
  model,
  rprior,
  dprior,
  rperturb,
  dperturb,
  nParticles,
  final_sample = TRUE,
  model_redo_if_NA = FALSE,
  verbose = TRUE,
  parallel = FALSE,
  save_model = TRUE,
  save_rds = FALSE,
  filename_rds = "ABCSMCDRF.rds",
  ...
)

```

Arguments

<code>method</code>	Random forest method to implement in each iteration ("smcrf-single-param" by default). <code>method = "smcrf-single-param"</code> : implements ABC-RF for each parameter and results in their marginal posterior distributions. <code>method = "smcrf-multi-param"</code> : implements ABC-DRF for all parameters and results in the joint posterior distribution.
---------------------	--

statistics_target	A dataframe containing statistics from data. Column names are the statistics IDs. <code>smcrf</code> only supports one row of statistics. If there are multiple observations, we recommend applying <code>smcrf</code> to each row individually.
statistics_selection	A dataframe indicating selection of statistics for fitting individual parameters (only works for method "smcrf-single-param"; NULL by default). Each column's name matches one statistic ID, and each row's name matches one parameter ID. The value is 1 if the statistic is used for the parameter, 0 otherwise.
smcrf_results	An existing ABC-SMC-RF result. If provided, <code>smcrf</code> will continue ABC-SMC-RF from the last iteration of the previous run.
model	Model for the statistics. The function must take two inputs: a dataframe <code>parameters</code> and logic variable <code>parallel</code> . The model must output a reference table, where each row contains parameters for each simulation and corresponding statistics. The column names of the reference table must match the parameter and statistics IDs.
rprior	Function to generate particles from the prior distribution. The function must take one input: <code>Nparameters</code> , the number of particles to generate. The output is a dataframe where column names match parameter IDs, and each row contains one parameter set.
dprior	Function to compute the prior density. The function must take two inputs: <code>parameters</code> and <code>parameter_id</code> . The dataframe <code>parameters</code> contains parameter sets in each row, with column names as parameter IDs. The <code>parameter_id</code> is either "all" or one of the parameter IDs. The output is a vector of prior probabilities corresponding to rows in <code>parameters</code> , either for the parameter indicated by <code>parameter_id</code> or jointly for all parameters (if <code>parameter_id</code> = "all").
rperturb	Function to generate perturbed particles. The function must take three inputs: <code>parameters_unperturbed</code> , <code>parameters_previous_sampled</code> , and <code>iteration</code> . The dataframe <code>parameters_unperturbed</code> contains unperturbed parameter sets in each row. The dataframe <code>parameters_previous_sampled</code> contains parameter sets sampled from the previous iteration in each row. In both dataframes, the column names match parameter IDs. The integer <code>iteration</code> indicates the index for current iteration. The output is a dataframe where column names match parameter IDs, and each row contains one perturbed parameter set corresponding to each row in <code>parameters_unperturbed</code> . A popular choice for perturbation is the normal distribution centered at the unperturbed parameters, with standard deviation equal to twice the empirical standard deviation of the parameters sampled from the previous iteration, truncated to within the prior distribution.
dperturb	Function to compute the perturbation density. The function must take five inputs: <code>parameters</code> , <code>parameters_previous</code> , <code>parameters_previous_sampled</code> , <code>iteration</code> , and <code>parameter_id</code> . The dataframe <code>parameters</code> contains parameter sets in each row. The dataframe <code>parameters_previous</code> contains one parameter set from the previous iteration. The dataframe <code>parameters_previous_sampled</code> contains parameter sets sampled from the previous iteration in each row. In all three dataframes, the column names match parameter IDs. The integer <code>iteration</code> indicates the index for current iteration. The <code>parameter_id</code> is either "all" or one of the parameter IDs. The output is a vector of perturbation probabilities from <code>parameters_previous</code> to rows in <code>parameters</code> , either for the parameter indicated by <code>parameter_id</code> or jointly for all parameters (if <code>parameter_id</code> = "all").

nParticles	A vector of particle counts. Each entry indicates the number of simulations (e.g. particles) in the corresponding iteration.
final_sample	A logic variable (TRUE by default). If final_sample = TRUE, parameters will be sampled from the final iteration and saved in iteration length(nParticles)+1.
model_redo_if_NA	A logic variable (FALSE by default). If model_redo_if_NA = TRUE, the particles where model returns NA will be simulated again.
verbose	A logic variable (TRUE by default). If verbose = TRUE, progress messages will be printed during the fitting process.
parallel	A logic variable (FALSE by default). If parallel = TRUE, the ABC-RF functions will be computed in parallel.
save_model	A logic variable (FALSE by default). If save_model = TRUE, the random forest will be saved in smcrf's output.
save_rds	A logic variable (FALSE by default). If save_rds = TRUE, the ABC-SMC-RF results will be saved in an rds file.
filename_rds	A string ("ABCSCDRF.rds" by default). If save_rds = TRUE, the output from ABC-SMC-(D)RF will be saved in a file with this name.
...	Additional arguments to be passed to abcrf or drf.

Value

An object smcrf_results containing the results of the inference. If the posterior distributions have not converged to a satisfactory level, the user may continue with smcrf(smcrf_results = smcrf_results, ...), in which case ABC-SMC-(D)RF will continue iterating from the last run in smcrf_results.

Index

`plot_compare_joint`, [2](#), [2](#)
`plot_compare_marginal`, [3](#), [3](#)
`plot_compare_qqplot`, [4](#), [4](#), [5](#)
`plot_smcrf_joint`, [5](#), [5](#)
`plot_smcrf_marginal`, [6](#), [6](#)

`smcrf`, [4](#), [6–8](#), [8](#), [9](#), [10](#)