

Package ‘ETASbootstrap’

June 9, 2024

Type Package

Title Bootstrap Confidence Interval Estimation for 'ETAS' Model
Parameters

Description The 2-D spatial and temporal Epidemic Type Aftershock Sequence ('ETAS') Model is widely used to 'decluster' earthquake data catalogs. Usually, the calculation of standard errors of the 'ETAS' model parameter estimates is based on the Hessian matrix derived from the log-likelihood function of the fitted model. However, when an 'ETAS' model is fitted to a local data set over a time period that is limited or short, the standard errors based on the Hessian matrix may be inaccurate. It follows that the asymptotic confidence intervals for parameters may not always be reliable. As an alternative, this package allows for the construction of bootstrap confidence intervals based on empirical quantiles for the parameters of the 2-D spatial and temporal 'ETAS' model. This version improves on Version 0.1.0 of the package by enabling the study space window (renamed 'study region') to be polygonal rather than merely rectangular. A Japan earthquake data catalog is used in a second example to illustrate this new feature.

Version 0.2.0

Author Renjie Peng [aut, cre],
Pierre Dutilleul [aut] (<<https://orcid.org/0000-0002-2381-3421>>),
Christian Genest [aut] (<<https://orcid.org/0000-0002-1764-0202>>)

Maintainer Renjie Peng <renjie.peng@mail.mcgill.ca>

Depends R (>= 3.3.0)

Imports ETAS (>= 0.5.1), MASS (>= 7.3.58.2), spatstat.geom (>= 3.2.8),
stats (>= 4.2.2), utils (>= 4.2.2)

License MIT + file LICENSE

LazyData true

Encoding UTF-8

RoxygenNote 7.3.1

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Repository CRAN

Date/Publication 2024-06-09 21:50:02 UTC

Contents

ETAS_Boots	2
JPN_earthquakes	7
simulate_aftershocks	8
simulate_background_earthquakes	9
VCI_earthquakes	11
VCI_earthquakes_plus	11
VCI_magnitude_sample	12
VCI_simulated_background_earthquakes	12
Index	13

ETAS_Boots	<i>Compute bootstrap confidence intervals</i>
------------	---

Description

A number (1000 by default) of earthquake data catalogs are simulated by bootstrap and recorded. A 2-D spatial and temporal ETAS model is fitted to each bootstrap-simulated earthquake data catalog, and the corresponding parameter estimates are recorded, which provides an empirical distribution for each estimate. For a given confidence level $1 - \alpha$ (0.95 by default), bootstrap confidence intervals are built from the empirical $\alpha/2$ (0.025) and $1 - \alpha/2$ (0.975) quantiles of the distributions of estimates for the parameters $(A, c, \alpha, p, D, q, \gamma)$ of the ETAS model.

Usage

```
ETAS_Boots(
  earthquake_data,
  longitude_boundaries = NULL,
  latitude_boundaries = NULL,
  study_region = NULL,
  time_begin = NULL,
  study_start = NULL,
  study_end = NULL,
  magnitude_threshold = NULL,
  time_zone = "GMT",
  round_off = FALSE,
  parameters_0 = NULL,
  number_simulations = 1000,
  confidence_level = 0.95,
  output_datasets = FALSE,
  output_estimates = FALSE
)
```

Arguments

`earthquake_data`

An object of class "data.frame" containing the following 5 columns:

- `date`: Occurrence date of earthquakes in the format "yyyy-mm-dd"
- `time`: Occurrence time of earthquakes in the format "hh:mm:ss"
- `longitude`: Longitude of the epicenter of earthquakes in decimal degrees
- `latitude`: Latitude of the epicenter of earthquakes in decimal degrees
- `magnitude`: Magnitude of earthquakes (Any type of magnitude is accepted as far as it is used consistently and thoroughly.)

See `VCI_earthquakes` for an example with a rectangular study region; for a more general, polygonal study region, see `JPN_earthquakes`. Note that West longitude and South latitude values should be negative, whereas East longitude and North latitude values are positive.

`longitude_boundaries`

A numerical vector of length 2 (`long_min`, `long_max`) with the longitude boundaries of a rectangular space window, for which the earthquake catalog data are contained in **`earthquake_data`**. If NULL (at the beginning of the execution of the program), `long_min` and `long_max` will be set (by the program) to the minimum and maximum values of the longitudes of earthquakes in **`earthquake_data`**. Together with **`latitude_boundaries`**, **`longitude_boundaries`** defines a region aimed to take edge effects into account in the analyses. This region includes the study region and approximately 20% more space around the study region.

`latitude_boundaries`

A numerical vector of length 2 (`lat_min`, `lat_max`) with the latitude boundaries of a rectangular space window, for which the earthquake catalog data are contained in **`earthquake_data`**. If NULL, `lat_min` and `lat_max` will be set to the minimum and maximum values of the latitudes of earthquakes in **`earthquake_data`**.

`study_region`

A list with two components (`lat`, `long`) of equal length specifying the coordinates of the vertices of a polygonal study region. The vertices must be written in anti-clockwise order. If NULL, `study_region` will be filled with boundaries defining a rectangular space window 20% narrower than the space window built from the `longitude_boundaries` and `latitude_boundaries`, while keeping the same center.

`time_begin`

A character string, in the date-time format (yyyy-mm-dd hh:mm:ss), which identifies the start of the time span in **`earthquake_data`**. If NULL, **`time_begin`** will be set to the date-time of the first event in **`earthquake_data`**.

`study_start`

A character string, in the date-time format, which specifies the start of the study period. If NULL, **`study_start`** will be set to the date-time corresponding to that of **`time_begin`** plus 20% of the length of the time span in **`earthquake_data`**.

`study_end`

A character string, in the date-time format, which specifies the end of the study period. If NULL, it will be set to the date-time of the last event in **`earthquake_data`**. Note: **`study_end`** coincides with the end of the time span in **`earthquake_data`**.

`magnitude_threshold`

A decimal number which specifies the threshold to be used for the magnitudes of earthquakes. Only earthquakes with a magnitude greater than or equal to

	magnitude_threshold will be considered, while the model is being fitted. If NULL, the minimum magnitude calculated from the events in earthquake_data will be used for magnitude_threshold .
time_zone	A character string specifying the time zone in which the occurrence times of earthquakes were recorded. The default "GMT" is the UTC (Universal Time Coordinates).
round_off	A logical flag indicating whether or not to account for round-off error in coordinates of epicenters.
parameters_0	A decimal vector of size 8 ($\nu, A, c, \alpha, p, D, q, \gamma$) to be used as an initial solution for the iterative maximum likelihood estimation of the ETAS model parameters. In particular, the values of parameters ν, A, c, α, D , and γ are positive, and those of p and q are strictly greater than 1. If NULL, the values recommended by Ogata (1998) will be used.
number_simulations	A positive integer which stands for the number of requested bootstrap simulations. The default value is 1000.
confidence_level	A decimal number in (0, 1) which specifies the confidence level associated with the bootstrap confidence intervals that are built for the ETAS model parameters, and saved as outputs. It is set to 0.95 by default.
output_datasets	A logical flag indicating whether or not the bootstrap-simulated earthquake data catalogs must be written in CSV files. The default setting is FALSE.
output_estimates	A logical flag indicating whether or not the maximum likelihood estimates of parameters from each bootstrap-simulated earthquake data catalog must be written in a CSV file. The default setting is FALSE.

Details

Ogata (1998) proposed the 2-D spatial and temporal ETAS model, which is now widely used to decluster earthquake catalogs and, to a lesser extent, make short-term forecasts. In the 2-D spatial and temporal ETAS model, the behavior of the point process for which $\{(t_i, x_i, y_i, m_i), i = 1, \dots, n\}$ is a partial realization is characterized by the conditional intensity function

$$\lambda_{\beta, \theta}(t, x, y, m | H_t) = s_{\beta}(m) \lambda_{\theta}(t, x, y | H_t),$$

where β and $\theta = (\nu, A, \alpha, c, p, q, D, \gamma)$ are the model parameters and s_{β} is the probability density function (pdf) associated with the distribution of earthquake magnitudes. It is assumed that the distribution of the magnitude of earthquakes is independent of the joint distribution of the occurrence time of earthquakes and the 2-D spatial location of their epicenters. It can be expressed, for arbitrary $\beta \in (0, \infty)$, as

$$s_{\beta}(m) = \beta \exp\{-\beta(m - m_0)\},$$

where m and m_0 represent the magnitude of the earthquake and the magnitude threshold, respectively. Moreover, $\lambda_{\theta}(t, x, y | H_t)$ represents the rate of observation of earthquakes in time and

space, given the information on earthquakes prior to time t . This rate is expressed as the sum of two terms, namely

$$\lambda_\theta(t, x, y | H_t) = \mu(x, y) + \sum_{i:t_i < t} k(m_i)g(t - t_i)f(x - x_i, y - y_i | m_i)$$

with

$$\mu(x, y) = \nu u(x, y),$$

where $\nu \in (0, \infty)$. The term $\mu(x, y)$ is usually called "background seismicity rate" and represents the rate at which earthquakes independently occur around longitude x and latitude y . The i th term of the summation in λ_θ , namely

$$k(m_i)g(t - t_i)f(x - x_i, y - y_i | m_i),$$

represents the effect of the i th earthquake before time t on the occurrence rate of earthquakes that would occur at time t , with an epicenter around (x, y) . Thus,

$$\sum_{i:t_i < t} k(m_i)g(t - t_i)f(x - x_i, y - y_i | m_i)$$

describes the total effect of all the earthquakes that occurred prior to time t , on the rate at which earthquakes would occur with an epicenter around (x, y) at time t . The expressions for k , g , and f are discussed individually as follows. First,

$$k(m) = Ae^{\alpha(m-m_0)}, \quad m \geq m_0,$$

can be interpreted as the expected number of earthquakes triggered by a previous earthquake with magnitude m , where $A \in (0, \infty)$ and $\alpha \in (0, \infty)$. Second, for all $t \in (t_i, \infty)$,

$$g(t - t_i) = \frac{p-1}{c} \left(1 + \frac{t - t_i}{c} \right)^{-p}$$

is the pdf for the occurrence time of an earthquake triggered by the i th earthquake in the catalog, which occurred at time t_i , where $c \in (0, \infty)$ and $p \in (1, \infty)$. Third,

$$f(x - x_i, y - y_i | m_i) = \frac{q-1}{\pi D e^{\gamma(m_i-m_0)}} \left\{ 1 + \frac{(x - x_i)^2 + (y - y_i)^2}{D e^{\gamma(m_i-m_0)}} \right\}^{-q}$$

is the pdf for the occurrence location (epicenter) of an earthquake triggered by the i th earthquake in the catalog, which occurred with magnitude m_i and an epicenter at (x_i, y_i) , where $D \in (0, \infty)$, $\gamma \in (0, \infty)$, and $q \in (1, \infty)$. For more details, see the articles of Zhuang et al. (2002, 2004).

Value

A list of three components:

- MLE: A numerical vector recording the maximum likelihood estimates of the ETAS model parameters $(A, c, \alpha, p, D, q, \gamma)$.
- ASE: A numerical vector recording the corresponding asymptotic standard errors.
- BootstrapCI: A matrix recording the corresponding bootstrap confidence intervals for the **confidence_level** entered as input and all the other input arguments of the [ETAS_Boots](#) function starting with **earthquake_data**.

When `output_datasets=TRUE`, the simulated earthquake data catalogs are written in "Boot_N.csv", where "N" denotes the number of bootstrap simulation runs.

When `output_estimates=TRUE`, the maximum likelihood estimates of parameters from each simulated earthquake data catalog are written in "estimates.csv".

References

Dutilleul, P., Genest, C., Peng, R., 2024. Bootstrapping for parameter uncertainty in the space-time epidemic-type aftershock sequence model. *Geophysical Journal International* 236, 1601–1608.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics* 50(2), 379–402.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association* 97(458), 369–380.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth* 109(B05301).

Examples

```
set.seed(23)
ETAS_Boots(earthquake_data = VCI_earthquakes,
           longitude_boundaries = c(-131, -126.25),
           latitude_boundaries = c(48, 50),
           study_region = list(long = c(-130.5, -130.5, -126.75, -126.75),
                              lat = c(49.75, 48.25, 48.25, 49.75)),
           time_begin = "2000/01/01 00:00:00",
           study_start = "2008/04/27 00:00:00",
           study_end = "2018/04/27 00:00:00",
           magnitude_threshold = 4,
           time_zone = "GMT",
           parameters_0 = c(0.65, 0.24, 0.0068, 0.97, 1.22, 0.0033, 2.48, 0.17),
           number_simulations = 4,
           confidence_level = 0.95,
           output_datasets = FALSE,
           output_estimates = FALSE)
```

```
ETAS_Boots(
  earthquake_data=JPN_earthquakes,
  longitude_boundaries = c(128, 145),
  latitude_boundaries = c(27, 45),
  study_region = list(long=c(130,135,145,140),
                    lat=c(33,30,40,43)),
  time_begin = "1926-01-08",
  study_start = "1953-05-26",
  study_end = "1990-01-08",
  magnitude_threshold = 5.5,
  time_zone = "GMT",
  round_off = FALSE,
  parameters_0 = c(0.524813924, 0.09, 0.045215442, 1.970176559,
                 1.249620329, 0.002110203, 1.910492169,1.763149113 ),
```

```
number_simulations = 2,  
confidence_level = 0.95,  
output_datasets = FALSE,  
output_estimates = FALSE  
)
```

JPN_earthquakes

New example data for the function [ETAS_Boots 0.2.0](#)

Description

This earthquake data catalog is an excerpt from the dataset "japan.quakes" in the ETAS package (Jalilian, 2019), excluding the last column "depth". Its space-time window covers 128°E to 145°E in longitude and 27°N to 45°N in latitude and the period from 1926-01-08 00:00:00 to 2007-12-29 23:59:59 (UTC). Note: The hypocenter depth of the earthquakes of interest is less than 100 km.

Usage

```
JPN_earthquakes
```

Format

An object of class `data.frame` with 5 columns:

- `date`: Occurrence date of earthquakes in the format "yyyy-mm-dd"
- `time`: Occurrence time of earthquakes in the format "hh-mm-ss"
- `longitude`: Longitude of the epicenter of earthquakes in decimal degrees
- `latitude`: Latitude of the epicenter of earthquakes in decimal degrees
- `magnitude`: Magnitude of earthquakes

Source

<https://CRAN.R-project.org/package=ETAS>

References

Jalilian, A. (2019). ETAS: An R package for fitting the space-time ETAS model to earthquake data. *Journal of Statistical Software, Code Snippets*, 88(1), 1–39. doi:10.18637/jss.v088.c01.

simulate_aftershocks *Simulate a catalog of aftershocks*

Description

When a catalog of background earthquakes is given, this function can be applied to simulate aftershocks under the intensity function $\sum_{i:t_i < t} \hat{k}(m_i) \hat{g}(t - t_i) \hat{f}(x - x_i, y - y_i | m_i)$, which is determined by the target parameter values given by the user.

Usage

```
simulate_aftershocks(
  parameters_target,
  background_catalog,
  time_begin_background = NULL,
  longitude_limit = NULL,
  latitude_limit = NULL,
  time_limit = NULL,
  magnitude_sample = NULL,
  magnitude_threshold = NULL,
  time_zone = "GMT"
)
```

Arguments

- parameters_target** A numerical vector of size 7, $(\hat{A}, \hat{c}, \hat{\alpha}, \hat{p}, \hat{D}, \hat{q}, \hat{\gamma})$, specifying the target values of parameters in the ETAS model.
- background_catalog** An object of class "data.frame" with 5 columns: recording date, time, longitude, latitude, and magnitude of the background events, in this order and in a format consistent with that of **earthquake_data** in the function [ETAS_Boots](#).
- time_begin_background** A character string, in the date-time format, that specifies the beginning of the time span in **background_catalog**. If NULL, it will be set by the program to the date-time of the first earthquake in **background_catalog**.
- longitude_limit** A vector of size 2 (xlim_min, xlim_max) specifying the longitude boundaries for the simulated aftershocks. If NULL, xlim_min and xlim_max will be set by the program to the minimum and maximum values of the longitude for the earthquakes in **background_catalog**, respectively. Only the simulated aftershocks with a longitude inside **longitude_limit** will be kept.
- latitude_limit** A vector of size 2 (ylim_min, ylim_max) specifying the latitude boundaries for the simulated aftershocks. If NULL, ylim_min and ylim_max will be set by the program to the minimum and maximum values of latitude for the earthquakes in **background_catalog**, respectively. Only the simulated aftershocks with a latitude inside **latitude_limit** will be kept.

time_limit	A vector of size 2 (tlim_min, tlim_max) specifying the time span for the simulated aftershocks. If NULL, tlim_min and tlim_max will be set by the program to the date-time of the first and last earthquakes (in chronological order) in background_catalog , respectively. Only the simulated aftershocks inside the specified time span will be kept.
magnitude_sample	A vector recording the sample from the distribution of earthquake magnitudes ($s_{\beta}(m)$). If NULL, the magnitudes of earthquakes in background_catalog will be used.
magnitude_threshold	A decimal value specifying the magnitude threshold to be applied. Only the simulated aftershocks with a magnitude of at least mag_threshold will be kept. If NULL, the minimum magnitude of the events in background_catalog will be used as magnitude_threshold .
time_zone	A character string specifying the time zone. The default setting "GMT" is the UTC (Universal Time Coordinated).

Value

aftershocks_simulated: An object of class "data.frame" with 5 columns: recording the date, time, longitude, latitude and magnitude of the simulated aftershocks, in this order and a consistent format.

References

Dutilleul, P., Genest, C., Peng, R., 2024. Bootstrapping for parameter uncertainty in the space-time epidemic-type aftershock sequence model. *Geophysical Journal International* 236, 1601–1608.

Examples

```
set.seed(1)
simulate_aftershocks(parameters_target = c(0.2424, 0.0068, 0.9771, 1.2200,
                                         0.0033, 2.4778, 0.1718),
                    background_catalog = VCI_simulated_background_earthquakes,
                    time_begin_background = "2000/01/01",
                    longitude_limit = c(-131, -126.25),
                    latitude_limit = c(48, 50),
                    time_limit = c("2000/01/01", "2018/04/27"),
                    magnitude_sample = VCI_magnitude_sample,
                    magnitude_threshold = 3.5,
                    time_zone="GMT")
```

simulate_background_earthquakes

Simulate a catalog of background earthquakes

Description

In fitting the ETAS model to the earthquake data catalog of interest (**earthquake_data**), the background intensity function μ is estimated. This function performs a simulation of background events based on the estimate $\hat{\mu}$. The time period for the simulated background catalog is consistent with that of **earthquake_data**.

Usage

```
simulate_background_earthquakes(earthquake_data_plus)
```

Arguments

earthquake_data_plus

An object of data.frame with 7 columns: date, time, longitude, latitude, magnitude, bandwidth, and probability, in this order and in a consistent format for the first 5 columns. The columns bandwidth and probability are two numeric vectors. The column bandwidth records the smoothness bandwidths used in variable kernel estimation and the column probability contains the probability for each earthquake in the catalog of interest (observed earthquakes) to be a background event; see the [etas](#) function in the ETAS package (Jalilian, 2019) and the articles of Zhuang et al. (2002, 2004).

Value

background_catalog: An object of data.frame with 5 columns: date, time, longitude, latitude, and magnitude of the simulated background earthquakes, in this order and a consistent format.

References

Dutilleul, P., Genest, C., Peng, R., 2024. Bootstrapping for parameter uncertainty in the space-time epidemic-type aftershock sequence model. *Geophysical Journal International* 236, 1601–1608.

Jalilian, A. (2019). ETAS: An R package for fitting the space-time ETAS model to earthquake data. *Journal of Statistical Software, Code Snippets*, 88(1), 1–39. doi:10.18637/jss.v088.c01.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association* 97(458), 369–380.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth* 109(B05301).

Examples

```
set.seed(1)
simulate_background_earthquakes(VCI_earthquakes_plus)
```

VCI_earthquakes *Original example data for the function* [ETAS_Boots 0.1.0](#)

Description

The source organization for this earthquake data catalog is the Canadian National Earthquake Database. Its space-time window covers 126.25°W to 131°W in longitude and 48°N to 50°N in latitude and the period from 2000-01-01 00:00:00 to 2019-12-31 23:59:59 (UTC). Note: The hypocenter depth of the earthquakes of interest ranges from about −5 km to 1000 km.

Usage

VCI_earthquakes

Format

An object of class `data.frame` with 5 columns:

- `date`: Occurrence date of earthquakes in the format "yyyy-mm-dd"
- `time`: Occurrence time of earthquakes in the format "hh-mm-ss"
- `longitude`: Longitude of the epicenter of earthquakes in decimal degrees
- `latitude`: Latitude of the epicenter of earthquakes in decimal degrees
- `magnitude`: Magnitude (Moment magnitude) of earthquakes

Source

<https://www.earthquakescanada.nrcan.gc.ca/stndon/NEDB-BNDS/bulletin-en.php>

VCI_earthquakes_plus *Example data for the function* [simulate_background_earthquakes](#)

Description

This catalog contains the earthquakes from [VCI_earthquakes](#), for which the magnitude is greater than or equal to 3.5 or the magnitude threshold that has been chosen.

Usage

VCI_earthquakes_plus

Format

An object of class `data.frame` with 7 columns, formatted like the 7 columns of `earthquake_data_plus`, the argument of the function [simulate_background_earthquakes](#). The magnitudes are moment magnitudes, as in [VCI_earthquakes](#).

VCI_magnitude_sample *First example data for the function [simulate_aftershocks](#)*

Description

A data set containing only the moment magnitudes from [VCI_earthquakes](#).

Usage

VCI_magnitude_sample

Format

A numerical vector

VCI_simulated_background_earthquakes
Second example data for the function [simulate_aftershocks](#)

Description

This earthquake data catalog is used as **background_catalog** in an example of application of the function [simulate_aftershocks](#) and was obtained by running the example for the function [simulate_background_earthquakes](#).

Usage

VCI_simulated_background_earthquakes

Format

An object of class `data.frame` with 5 columns, formatted like other data frames with the same structure described above, including [VCI_earthquakes](#).

Index

* datasets

JPN_earthquakes, [7](#)
VCI_earthquakes, [11](#)
VCI_earthquakes_plus, [11](#)
VCI_magnitude_sample, [12](#)
VCI_simulated_background_earthquakes,
[12](#)

etas, [10](#)

ETAS_Boots, [2](#), [5](#), [7](#), [8](#), [11](#)

JPN_earthquakes, [7](#)

simulate_aftershocks, [8](#), [12](#)

simulate_background_earthquakes, [9](#), [11](#),
[12](#)

VCI_earthquakes, [11](#), [11](#), [12](#)

VCI_earthquakes_plus, [11](#)

VCI_magnitude_sample, [12](#)

VCI_simulated_background_earthquakes,
[12](#)