

Package: EpiInvert (via r-universe)

August 31, 2024

Version 0.3.2

Type Package

Title Variational Techniques in Epidemiology

Maintainer Luis Alvarez <lalvarez@ulpgc.es>

Description Using variational techniques we address some epidemiological problems as the incidence curve decomposition by inverting the renewal equation as described in Alvarez et al. (2021) <[doi:10.1073/pnas.2105112118](https://doi.org/10.1073/pnas.2105112118)> and Alvarez et al. (2022) <[doi:10.3390/biology11040540](https://doi.org/10.3390/biology11040540)> or the estimation of the functional relationship between epidemiological indicators. We also propose a learning method for the short time forecast of the trend incidence curve as described in Morel et al. (2022) <[doi:10.1101/2022.11.05.22281904](https://doi.org/10.1101/2022.11.05.22281904)>.

URL <https://github.com/lalvarezmat/EpiInvert>

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports Rcpp (>= 1.0.8.3)

LinkingTo Rcpp

Suggests knitr, ggplot2, grid, dplyr, testthat (>= 3.0.0), rmarkdown, ggpubr

Config/testthat/edition 3

Depends R (>= 2.10)

LazyLoad yes

RoxygenNote 7.2.2

Repository <https://lalvarezmat.r-universe.dev>

RemoteUrl <https://github.com/lalvarezmat/epiinvert>

RemoteRef HEAD

RemoteSha 7b6afdda0b55e54c77f601239340552d64d9b132

Contents

apply_delay	2
EpiIndicators	3
EpiIndicators_params	4
EpiIndicators_plot	5
EpiInvert	5
EpiInvertForecast	9
EpiInvertForecast_plot	11
EpiInvert_plot	12
festives	13
incidence	13
incidence_weekly_aggregated	14
joint_indicators_by_date	15
owid	15
restored_incidence_database	16
select_params	17
si_distr_data	17
Index	18

apply_delay	apply_delay
-------------	-------------

Description

apply a delay vector, $s(t)$, to an indicator, $g(t)$

Usage

```
apply_delay(g, s)
```

Arguments

g	indicator.
s	delay to apply.

Value

A numeric vector with the result of apply the vector delay s to g .

EpiIndicators	EpiIndicators
---------------	---------------

Description

EpiIndicators estimates the ratio, $r(t)$, and shift(delay), $s(t)$, between 2 epidemiological indicators $f(t)$ and $g(t)$ following the relation $r(t)*f(t) = g(t+s(t))$

Usage

```
EpiIndicators(df, config = EpiIndicators_params())
```

Arguments

- | | |
|--------|--|
| df | a dataframe with 3 columns: the first column corresponds to the date of each indicator value, the second column is the value of the first indicator $f(t)$ and the third column is the value of the second indicator $g(t)$. A zero value is expected in the case that the real value of an indicator is not available. Indicators must be smooth functions. So, for instance, the raw registered number of cases or deaths are not adequate to run the function. These particular indicators should be smoothed before executing <code>EpiIndicators()</code> , for instance you can use the restored indicator values obtained by <code>EpiInvert()</code> |
| config | a list of the following optional parameters obtained using the function <code>EpiIndicators_params()</code> : $s_min = -10$, <ul style="list-style-type: none"> • s_min: min value allowed for the shift $s(t)$ (default value -10) • s_max: max value allowed for the shift $s(t)$ (default value 25) • wr: energy regularization parameter for the ratio $r(t)$ (default value 1000) • ws: energy regularization parameter for the shift $s(t)$ (default value 10) • s_init: manually fixed initial value (at time $t=0$) for $s(t)$ (default value $-1e6$) by default s_init is not fixed and it is automatically estimated • s_end: manually fixed final value (at the last time) for $s(t)$ (default value $-1e6$) by default s_end is not fixed and it is automatically estimated • r_init: manually fixed initial value (at time $t=0$) for $r(t)$ (default value $-1e6$) by default r_init is not fixed and it is automatically estimated • r_end: manually fixed final value (at the last time) for $r(t)$ (default value $-1e6$) by default r_end is not fixed and it is automatically estimated |

Details

EpiIndicators estimates the ratio, $r(t)$, and shift(delay), $s(t)$ between 2 epidemiological indicators $f(t)$ and $g(t)$ following the relation $r(t)*f(t) = g(t+s(t))$ a variational method is proposed to add regularity constraints to the estimates of $r(t)$ and $s(t)$.

Value

A dataframe with the following columns :

- date: the date of the indicator values.
- f: the first indicator $f(t)$.
- g: the second indicator $g(t)$.
- r: the estimated ratio $r(t)$
- s: the estimated shift (delay) $s(t)$
- f.r: the result of $r(t)*f(t)$
- g.s: the result of $g(t)+s(t)$

Author(s)

Luis Alvarez <lalvarez@ulpgc.es>

Examples

```
## load data of epidemiological indicators obtained from the World in data
## organization
data("owid")

## Filter the data to get France epidemiological indicators
library(dplyr)
sel <- filter(owid,iso_code=="FRA")

## Generate a dataframe with the dates and the cases and deaths restored
## using EpiInvert()
df<-data.frame(sel$date,sel$new_cases_restored_EpiInvert,sel$new_deaths_restored_EpiInvert)

## Run EpiIndicators
res <- EpiIndicators(df)

## Plot the results
EpiIndicators_plot(res)
```

EpiIndicators_params EpiIndicators_params *function to select EpiIndicators parameters*

Description

EpiIndicators_params function to select EpiIndicators parameters

Usage

```
EpiIndicators_params(x = "")
```

Arguments

`x` a list with parameters of EpiIndicators() function

Value

the EpiIndicators() parameters

See Also

[EpiIndicators](#)

EpiIndicators_plot	EpiIndicators_plot
--------------------	--------------------

Description

EpiIndicators_plot() plots the results obtained by EpiIndicators()

Usage

```
EpiIndicators_plot(df)
```

Arguments

`df` : a dataframe obtained as the outcome of executing EpiIndicators()

Value

a combination of 3 plots: (A) the indicator $f(t)$ (in green in the main y-axis) and the indicator $g(t)$ (in red in the secondary axis). (B) $r(t)*f(t)$ (in green in the main y-axis) and $g(t+s(t))$ (in red in the secondary axis) and (C) $r(t)$ (in green in the main y-axis) and $s(t)$ (in red in the secondary axis)

EpiInvert	<i>EpiInvert estimates the reproduction number R_t and a restored incidence curve from the original daily incidence curve and the serial interval distribution. EpiInvert also corrects the festive and weekly biases present in the registered daily incidence.</i>
-----------	---

Description

EpiInvert estimates the reproduction number R_t and a restored incidence curve from the original daily incidence curve and the serial interval distribution. EpiInvert also corrects the festive and weekly biases present in the registered daily incidence.

Usage

```
EpiInvert(
  incid,
  last_incidence_date,
  festive_days = rep("1000-01-01", 2),
  config = EpiInvert::select_params()
)
```

Arguments

incid	The original daily incidence curve (a numeric vector).
last_incidence_date	The date of the last value of the incidence curve in the format YYYY-MM-DD. EpiInvert does not allow missing values. On days when a country does not report data, a zero must be registered as the value associated with the incidence of that day.
festive_days	The festive or anomalous dates in the format YYYY-MM-DD (a character vector). In these dates we "a priori" expect that the incidence value is perturbed.
config	An object of class <code>estimate_R_config</code> , as returned by function <code>select_params</code> . The element of config are : <ul style="list-style-type: none"> • <code>si_distr</code>: a numeric vector with the distribution of the serial interval (the default value is an empty vector). If this vector is empty, the serial interval is estimated using a parametric shifted log-normal. • <code>shift_si_distr</code>: shift of the above user provided serial interval. This shift can be negative, which means that secondary cases can show symptoms before the primary cases (the default value is 0). • <code>max_time_interval</code>: Maximum number of days used by EpiInvert (the default value is 150, which means that EpiInvert uses the last 150 days. The computational cost strongly depends on this value). • <code>mean_si</code>: mean of the parametric shifted log-normal to approximate the serial interval (in the case the above <code>si_distr</code> vector is empty), (the default value is 12.267893). • <code>sd_si</code>: standard deviation of the parametric shifted log-normal to approximate the serial interval (in the case the above <code>si_distr</code> vector is empty), (the default value is 5.667547). • <code>shift_si</code>: shift of the parametric shifted log-normal to approximate the serial interval (in the case the above <code>si_distr</code> vector is empty), (the default value is -5). • <code>Rt_regularization_weight</code>: regularization weight for R_t in the variational model used by EpiInvert (the default value is 5). • <code>seasonality_regularization_weight</code>: regularization weight for the weekly bias correction factors in the variational model used by EpiInvert (the default value is 5). • <code>incidence_weekly_aggregated</code>: a boolean value which determines if we use weekly aggregated incidence. In such case every week a single data is stored with the accumulated incidence in the last 7 days (the default value is FALSE).

- `NweeksToKeepIncidenceSum`: number of weeks to keep the value of the incidence accumulation. The default is 2.

Details

EpiInvert estimates the reproduction number R_t and a restored incidence curve by inverting the renewal equation :

$$i(t) = \sum_k i(t-k)R(t-k)\Phi(k)$$

using a variational formulation. The theoretical foundations of the method can be found in references [1] and [2].

Value

an object of class `estimate_EpiInvert`, given by a list with components:

- `i_original`: the original daily incidence curve. In the case of weekly aggregated incidence, we initialize the original curve assigning each day of the week $1/7$ of the weekly aggregated value.
- `i_festive`: the incidence after correction of the festive days bias.
- `i_bias_free`: the incidence after correction of the festive and weekly biases.
- `i_restored`: the restored incidence obtained using the renewal equation.
- `Rt`: the reproduction number R_t obtained by inverting the renewal equation.
- `Rt_CI95`: 95% confidence interval radius for the value of R_t taking into account the variation of R_t when more days are added to the estimation.
- `seasonality`: the weekly bias correction factors.
- `dates`: a vector of dates corresponding to the incidence curve.
- `festive`: boolean associated to each incidence value to check if it has been considered as a festive or anomalous day.
- `epsilon`: normalized error curve obtained as $(i_bias_free - i_restored) / i_restored^a$.
- `power_a`: the power, a , which appears in the above expression of the normalized error.
- `si_distr`: values of the distribution of the serial interval used in the EpiInvert estimation.
- `shift_si_distr`: shift of the above distribution of the serial interval `si_distr`.

Author(s)

Luis Alvarez <lalvarez@ulpgc.es>

References

- [1] Alvarez, L.; Colom, M.; Morel, J.D.; Morel, J.M. Computing the daily reproduction number of COVID-19 by inverting the renewal equation using a variational technique. Proc. Natl. Acad. Sci. USA, 2021.
- [2] Alvarez, Luis, Jean-David Morel, and Jean-Michel Morel. "Modeling COVID-19 Incidence by the Renewal Equation after Removal of Administrative Bias and Noise" Biology 11, no. 4: 540. 2022.
- [3] Ritchie, H. et al. Coronavirus Pandemic (COVID-19), OurWorldInData.org. Available online: <https://ourworldindata.org/coronavirus-source-data> (accessed on 5 May 2022).

Examples

```
## load data on COVID-19 daily incidence up to 2022-05-05 for France,
## and Germany (taken from the official government data) and for UK and
## the USA taken from reference [3]
data(incidence)

## EpiInvert execution for USA with no festive days specification
## using the incidence 70 days in the past
res <- EpiInvert(incidence$USA,
"2022-05-05",
"1000-01-01",
select_params(list(max_time_interval = 70))
)

## Plot of the results
EpiInvert_plot(res)

## load data of festive days for France, Germany, UK and the USA
data(festives)

## EpiInvert execution for France with festive days specification using
## 70 days in the past
res <- EpiInvert(incidence$FRA,"2022-05-05",festives$FRA,
select_params(list(max_time_interval = 70)))

## Plot of the incidence between "2022-04-01" and "2022-05-01"
EpiInvert_plot(res,"incid","2022-04-01","2022-05-01")

## load data of a serial interval
data("si_distr_data")

## EpiInvert execution for Germany using the uploaded serial interval shifted
## -2 days, using the incidence 70 days in the past
res <- EpiInvert(incidence$DEU,"2022-05-05",festives$DEU,
EpiInvert::select_params(list(si_distr = si_distr_data,
shift_si_distr=-2,max_time_interval = 70)))

## Plot of the serial interval used (including the shift)
EpiInvert_plot(res,"SI")
```



```

## EpiInvert execution for UK changing the default values of the parametric
## serial interval (using a shifted log-normal) using 70 days in the past
res <- EpiInvert(incidence$UK,"2022-05-05",festives$UK,
  EpiInvert::select_params(list(mean_si = 11,sd_si=6,shift_si=-1,max_time_interval = 70))
)

## Plot of the reproduction number Rt including an empiric 95% confidence
## interval of the variation of Rt. To calculate Rt on each day t, EpiInvert
## uses the past days (t'<=t) and the future days (t'>t) when available.
## Therefore, the EpiInvert estimate of Rt varies when there are more days
## available. This confidence interval reflects the expected variation of Rt
## as a function of the number of days after t available.
EpiInvert_plot(res,"R")

```

EpiInvertForecast	<i>EpiInvertForecast computes a 28-day forecast of the restored incidence curve including a 95 using the weekly seasonality, from the forecasted restored incidence curve we also estimate a 28-day forecast of the original incidence curve.</i>
-------------------	---

Description

EpiInvertForecast computes a 28-day forecast of the restored incidence curve including a 95 using the weekly seasonality, from the forecasted restored incidence curve we also estimate a 28-day forecast of the original incidence curve.

Usage

```

EpiInvertForecast(
  EpiInvert_result,
  restored_incidence_database,
  type = "median",
  trend_sentiment = 0,
  NumberForecastAdditionalDays = 0
)

```

Arguments

EpiInvert_result	output list of the EpiInvert execution including, in particular, the restored incidence curve and the seasonality.
restored_incidence_database	a database including 27,418 samples of different restored incidence curves computed by EpiInvert using real data. Each restored incidence curve includes the last 56 values of the sequence. That is this database can be viewed as a matrix of size 27,418 X 56
type	string with the forecast option. It can be "mean" or "median".

trend_sentiment

"a priori" knowledge about the future incidence evolution. == 0 means that you are neutral about the future trend > 0 means that you expect that the future trend is higher than the expected one using all database curves. the value represents the percentage of database curves removed before computing the forecast The curves removed are the ones with lowest growth in the last 28 days. < 0 means that you expect that the future trend is higher than the expected one using all database curves. the meaning of the value is similar to the previous case, but removing the curves with the highest growth in the last 28 days.

NumberForecastAdditionalDays

The number of forecast days is 28. With this parameter you can add extra forecast days using linear extrapolation.

Details

EpiInvertForecast estimates a forecast of the restored incidence curve using a weighted average of 27,418 restored incidence curves previously estimated by EpiInvert and stored in the database "restored_incidence_database". The weight, in the average computation, of each restored incidence curve of the database depends on the similarity between the current curve in the last 28 days and the first 28 days of the database curve. Each database curve contains 56 days. The first 28 days are used for comparison with the current curve and the last 28 days are used for forecasting.

Value

a list with components:

- **dates**: a vector of dates corresponding to the forecast days.
- **i_restored_forecast**: a numeric vector with the forecast of the restored incidence curve for the next 28 days.
- **i_original_forecast**: a numeric vector with the forecast of the original incidence curve for the next 28 days.
- **i_restored_forecast_CI50**: radius of an empiric confidence interval, with percentile 50, for the restored incidence forecast following the number of days passed since the current day.
- **i_restored_forecast_CI75**: radius of an empiric confidence interval, with percentile 75, for the restored incidence forecast following the number of days passed since the current day.
- **i_restored_forecast_CI90**: radius of an empiric confidence interval, with percentile 90, for the restored incidence forecast following the number of days passed since the current day.
- **i_restored_forecast_CI95**: radius of an empiric confidence interval, with percentile 95, for the restored incidence forecast following the number of days passed since the current day.

Author(s)

Luis Alvarez <lalvarez@ulpgc.es>

References

- [1] Alvarez, L.; Colom, M.; Morel, J.D.; Morel, J.M. Computing the daily reproduction number of COVID-19 by inverting the renewal equation using a variational technique. Proc. Natl. Acad. Sci. USA, 2021.
- [2] Alvarez, Luis, Jean-David Morel, and Jean-Michel Morel. "Modeling COVID-19 Incidence by the Renewal Equation after Removal of Administrative Bias and Noise" Biology 11, no. 4: 540. 2022.
- [3] Ritchie, H. et al. Coronavirus Pandemic (COVID-19), OurWorldInData.org. Available online: <https://ourworldindata.org/coronavirus-source-data> (accessed on 5 May 2022).
- [4] Alvarez, Luis, Jean-David Morel, and Jean-Michel Morel. EpiInvertForecast Available online: <https://ctim.ulpgc.es/covid19/EpiInvertForecastPaper.html>

Examples

```
## load data on COVID-19 daily incidence up to 2022-05-05 for France,  
## and Germany (taken from the official government data) and for UK and  
## the USA taken from reference [3]  
data(incidence)  
  
## load of the database of restored incidence curves.  
data("restored_incidence_database")  
  
## EpiInvert execution for USA with no festive days specification  
## using the incidence 90 days in the past  
res <- EpiInvert(incidence$USA,  
"2022-05-05",  
"1000-01-01",  
select_params(list(max_time_interval = 90))  
)  
  
## EpiInvertForecast execution using the EpiInvert results obtained by USA  
forecast <- EpiInvertForecast(res,restored_incidence_database)
```

```
EpiInvertForecast_plot  
EpiInvertForecast_plot
```

Description

EpiInvertForecast_plot() plot the restored incidence forecast

Usage

```
EpiInvertForecast_plot(EpiInvert_results, Forecast)
```

Arguments

<code>EpiInvert_results</code>	the list returned by the <code>EpiInvert</code> execution
<code>Forecast</code>	the list returned by the <code>EpiInvertForecast</code> execution

Value

a plot with the last 28 days of the original and restored incidence curves and a 28-day forecast of the same curves. It also includes a shaded area with a 95 of the restored incidence forecast estimation

<code>EpiInvert_plot</code>	<code>EpiInvert_plot</code>
-----------------------------	-----------------------------

Description

`EpiInvert_plot()` plots the results obtained by `EpiInvert`

Usage

```
EpiInvert_plot(
  x,
  what = "all",
  date_start = "1000-01-01",
  date_end = "3000-01-01"
)
```

Arguments

<code>x</code>	an object of class <code>estimate_EpiInvert</code> .
<code>what</code>	one of the following drawing options: <ul style="list-style-type: none"> • <code>all</code>: a plot combining the main <code>EpiInvert</code> results. • <code>R</code>: a plot of the reproduction number <code>Rt</code> estimation. • <code>incid</code>: a plot combining the obtained incidence curves. • <code>SI</code>: the serial interval used in the <code>EpiInvert</code> estimation.
<code>date_start</code>	the start date to plot
<code>date_end</code>	the final date to plot

Value

a plot.

See Also

[EpiInvert](#)

festives	<i>A dataset containing festive days in France, Germany, UK and the USA</i>
----------	---

Description

A dataset containing festive days in France, Germany, UK and the USA

Usage

festives

Format

A list with 4 variables:

FRA festive day of France

DEU festive day of Germany

UK festive day of UK

USA festive day of USA

incidence	<i>A dataset containing daily incidence of COVID-19 for France, Germany, UK and the USA</i>
-----------	---

Description

A dataset containing daily incidence of COVID-19 for France, Germany, UK and the USA

Usage

incidence

Format

A data frame with 5 variables:

date date of the incidence.

FRA incidence of France.

DEU incidence of Germany.

UK incidence of UK.

USA incidence of USA.

An updated version of this dataset can be found at <https://www.ctim.es/covid19/incidence.csv>

Source

<https://github.com/owid/covid-19-data/tree/master/public/data>

<https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-chiffres-cles-et-evolu>

https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4/page/page_1/

incidence_weekly_aggregated

A dataset containing weekly aggregated incidence of COVID-19 for France, Germany, UK and the USA

Description

A dataset containing weekly aggregated incidence of COVID-19 for France, Germany, UK and the USA

Usage

incidence_weekly_aggregated

Format

A data frame with 5 variables:

date date of the weekly aggregated incidence.

FRA weekly aggregated incidence of France.

DEU weekly aggregated incidence of Germany.

UK weekly aggregated incidence of UK.

USA weekly aggregated incidence of USA.

An updated version of this dataset can be found at <https://www.ctim.es/covid19/incidence.csv>

Source

<https://github.com/owid/covid-19-data/tree/master/public/data>

<https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-chiffres-cles-et-evolu>

https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4/page/page_1/

```
joint_indicators_by_date
      joint_indicators_by_date
```

Description

generates a dataframe joining the dates and values of 2 indicators

Usage

```
joint_indicators_by_date(date0, i0, date1, i1)
```

Arguments

date0	the dates of the first indicator.
i0	the values of the first indicator.
date1	the dates of the second indicator.
i1	the values of the second indicator.

Value

A dataframe with the following columns :

- date: all dates presented in any of the indicators.
- f: the values of the first indicator. We assign 0 in the case the data is not available for a given day.
- g: the values of the second indicator. We assign 0 in the case the data is not available for a given day

owid	<i>A dataset with COVID-19 indicators</i>
------	---

Description

A dataset containing COVID-19 epidemiological indicators for Canada, France, Germany, Italy, UK and the USA from Our World in data organization <https://github.com/owid/covid-19-data/tree/master/public/data> up to 2022-11-28. In the case a data value is not available for a given day we assign the value 0 to the indicator

Usage

```
owid
```

Format

A dataframe with 13 variables:

iso_code iso code of the country

location country name

date date of the indicator value

new_cases new confirmed cases

new_cases_smoothed new confirmed cases smoothed

new_cases_restored_EpiInvert new confirmed cases restored using EpiInvert

new_deaths new deaths attributed to COVID-19

new_deaths_smoothed new deaths smoothed

new_deaths_restored_EpiInvert new deaths restored using EpiInvert

icu_patients number of COVID-19 patients in intensive care units (ICUs) on a given day

hosp_patients number of COVID-19 patients in hospital on a given day

weekly_icu_admissions number of COVID-19 patients newly admitted to intensive care units (ICUs) in a given week (reporting date and the preceding 6 days)

weekly_hosp_admissions number of COVID-19 patients newly admitted to hospitals in a given week (reporting date and the preceding 6 days)

restored_incidence_database

A dataset of restored daily incidence trend curves

Description

A dataset including 27,418 samples of different restored incidence curves computed by EpiInvert using real data. Each restored incidence curve includes the last 56 values of the sequence.

Usage

restored_incidence_database

Format

A 27,418 X 56 numeric matrix

select_params	<i>select_params function to select EpiInvert parameters</i>
---------------	--

Description

select_params function to select EpiInvert parameters

Usage

```
select_params(x = "")
```

Arguments

x a list with elements of the class estimate_R_config

Value

an object of class estimate_R_config.

See Also

[EpiInvert](#)

si_distr_data	<i>A dataset containing the values of a serial interval</i>
---------------	---

Description

A dataset containing the values of a serial interval

Usage

```
si_distr_data
```

Format

A numeric vector with 1 variable representing the serial interval

Index

* datasets

- festives, [13](#)
- incidence, [13](#)
- incidence_weekly_aggregated, [14](#)
- owid, [15](#)
- restored_incidence_database, [16](#)
- si_distr_data, [17](#)

apply_delay, [2](#)

EpiIndicators, [3](#), [5](#)

EpiIndicators_params, [4](#)

EpiIndicators_plot, [5](#)

EpiInvert, [5](#), [12](#), [17](#)

EpiInvert_plot, [12](#)

EpiInvertForecast, [9](#)

EpiInvertForecast_plot, [11](#)

festives, [13](#)

incidence, [13](#)

incidence_weekly_aggregated, [14](#)

joint_indicators_by_date, [15](#)

owid, [15](#)

restored_incidence_database, [16](#)

select_params, [17](#)

si_distr_data, [17](#)