

# Package ‘RecordTest’

October 5, 2020

**Type** Package

**Title** Inference Tools in Time Series Based on Record Statistics

**Version** 1.0.1

**Date** 2020-10-05

**Author** Jorge Castillo-Mateo [aut, cre, cph]  
(<<https://orcid.org/0000-0003-3859-0248>>),  
Ana C. Cebrián [aut],  
Jesús Asín [dct]

**Maintainer** Jorge Castillo-Mateo <[jorgecastillomateo@gmail.com](mailto:jorgecastillomateo@gmail.com)>

**Depends** R (>= 3.5.0)

**Imports** ggplot2, parallel, stats

**Suggests** ggpubr, knitr, rmarkdown

**Description** Statistical tools based on probabilistic properties of the occurrence of records in a sequence (Xt) of independent and identically distributed random variables. In particular, tools to pre-process a time series as well as distribution-free tests for trend in location and plots to visualize the behavior of the records.

**License** GPL-3

**URL** <https://github.com/JorgeCastilloMateo/RecordTest>

**BugReports** <https://github.com/JorgeCastilloMateo/RecordTest/issues>

**VignetteBuilder** knitr

**Encoding** UTF-8

**LazyData** true

**NeedsCompilation** no

**RoxygenNote** 7.1.1

**Repository** CRAN

**Date/Publication** 2020-10-05 13:20:14 UTC

## R topics documented:

RecordTest-package . . . . .	2
foster.plot . . . . .	3
foster.test . . . . .	5
I.record . . . . .	7
L.plot . . . . .	8
L.record . . . . .	9
L_global.test . . . . .	10
L_lm.test . . . . .	12
L_lr.test . . . . .	14
M.record . . . . .	16
N.plot . . . . .	17
N.record . . . . .	19
N_normal.test . . . . .	20
P_chisq.test . . . . .	22
P_exactPB.test . . . . .	23
P_regression.plot . . . . .	25
P_regression.test . . . . .	26
records . . . . .	28
series_double . . . . .	29
series_rev . . . . .	30
series_split . . . . .	31
series_uncor . . . . .	32
series_untie . . . . .	33
TX_Zaragoza . . . . .	34
ZaragozaSeries . . . . .	35
<b>Index</b>	<b>36</b>

---

RecordTest-package	<i>RecordTest: A package for testing the classical record model</i>
--------------------	---

---

## Description

RecordTest provides an extensive number of graphical tools and tests for randomness to detect trend in location in time series based on the classical record model. It also provides some preprocessing tools often needed to adequately prepare a dataset to apply those tests.

## Details

This package provides two families of tests to study the hypothesis of the classical record model, that is that the records from a series of values observed at regular time units (a vector) come from an i.i.d. series of continuous random variables. A particular problem, common for example in climate problems, where these tools can be useful, is to detect the existence of a trend in a series of variables. If we have sequences of uncorrelated variables with no seasonal component, the hypothesis of i.i.d. variables is equivalent to test the hypothesis of no trend.

A sample of  $M$  vectors uncorrelated between them is needed to apply the implemented tests. Then, the input of the functions to perform the statistical tools is a matrix  $\mathbf{XM}_T$  where each column corresponds to a vector formed by the values of a series  $X_t$ , from  $t = 1, \dots, T$ , so that each row of the matrix correspond to a time  $t$ .

In many real problems, such as those related to environmental phenomena, the series of variables to analyze show a seasonal behaviour, and only one realization is available. In order to be able to apply the suggested tools to detect the existence of a trend, the seasonal component has to be removed and a sample of  $M$  uncorrelated series has to be obtained, with a high enough value  $M$ . Those problems can be solved, preprocessing the data adequately. A wide set of tools to carry out a preliminary analysis and to preprocess daily data with a yearly seasonal pattern are implemented: [series\\_double](#), [series\\_rev](#), [series\\_split](#), [series\\_uncor](#) and [series\\_untie](#).

There is also available a set of functions to characterize the occurrence time, the value and other statistics related to the occurrence of records: [I.record](#), [L.record](#), [M.record](#), [N.record](#), [Nmean.record](#) and [records](#).

All the tests are based on the occurrence of records, but two families are distinguished: the first is based on the number of records up to time  $t$ ,  $N_t$  ([foster.test](#), [N\\_normal.test](#)). The second family is based on the record indicator random variables,  $I_t$  and record times,  $L_i$  ([L\\_lm.test](#), [L\\_lr.test](#), [L\\_global.test](#), [P\\_chisq.test](#), [P\\_exactPB.test](#), [P\\_regression.test](#)). All of them are distribution-free tests in time series for trend in location based on the null hypothesis that the record indicators are independent and the probabilities of record at time  $t$  are  $p_t = 1/t$ .

Finally, there is another set of functions aiming to plot different features related to records: [foster.plot](#), [L.plot](#), [N.plot](#) and [P\\_regression.plot](#).

All the tests and plots can be applied to both upper and lower records, positive or negative trends in location, and forward and backward series; using the corresponding value in the argument `record`, `trend`, or function [series\\_rev](#), respectively.

### Author(s)

Jorge Castillo-Mateo, Ana C. Cebrián

---

foster.plot

---

*Plot based on Foster-Stuart and Diersen-Trenkler statistics*


---

### Description

This function constructs a ggplot object to display two-sided confidence intervals based on Foster-Stuart and Diersen-Trenkler statistics for randomness.

### Usage

```
foster.plot(
  XM_T,
  statistic = c("D", "d", "TM"),
  weights = function(t) 1,
  interval = c("ribbon", "errorbar"),
  conf = 0.95,
```

```
    colour = "salmon"
  )
```

### Arguments

XM_T	A numeric vector, matrix (or data frame).
statistic	A character string indicating the statistic to be calculated, i.e., "D" (for $\overline{DM}$ or $\overline{DM}^\omega$ ), "d" (for $dM$ or $dM^\omega$ ) and "TM" (for $\overline{TM}$ or $\overline{TM}^\omega$ ).
weights	A function indicating the weight given to the different records according to their position in the series, e.g., if <code>function(t) t-1</code> then $\omega_t = t - 1$ .
interval	A character string indicating the type of display of the confidence intervals, "ribbon" (grey area) or "errorbar" (vertical lines).
conf	Numeric value in (0, 1). Confidence level of the two-sided confidence intervals.
colour	Colour used to plot the expected values and the CI.

### Details

See [foster.test](#).

### Value

A ggplot graph object.

### Author(s)

Jorge Castillo-Mateo

### References

- Diersen J, Trenkler G (1996). Records Tests for Trend in Location. *Statistics*, **28**(1), 1-12. doi:[10.1080/02331889708802543](#)
- Diersen J, Trenkler G (2001). *Weighted record tests for splitted series of observations*. En J Kunert, G Trenkler (Eds.), *Mathematical Statistics with Applications in Biometry. Festschrift in Honour of Prof. Dr. Siegfried Schach* (pp. 163-178). Lohmar: Josef Eul Verlag.
- Foster FG, Stuart A (1954). Distribution-Free Tests in Time-Series Based on the Breaking of Records. *Journal of the Royal Statistical Society. Series B (Methodological)*, **16**(1), 1-22.

### See Also

[foster.test](#), [N.plot](#), [N\\_normal.test](#)

### Examples

```
foster.plot(ZaragozaSeries)
foster.plot(ZaragozaSeries, interval = 'error', conf = 0.9, colour = 1)
foster.plot(ZaragozaSeries, statistic = 'd', weights = function(t) t-1)
```

foster.test

*Foster-Stuart, Diersen-Trenkler or t Tests for Randomness***Description**

Performs Foster-Stuart, Diersen-Trenkler or t record tests for trend in location based on their normal or t asymptotic distribution, respectively. The null hypothesis of randomness is tested against the alternative hypothesis.

**Usage**

```
foster.test(
  XM_T,
  statistic = c("D", "d", "TM"),
  distribution = c("normal", "t"),
  weights = function(t) 1,
  trend = c("positive", "negative")
)
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
statistic	A character string indicating the type of statistic to be calculated, i.e., "D" (for $DM$ , $DM^\omega$ or $DM_S^\omega$ ), "d" (for $dM$ , $dM^\omega$ or $dM_S^\omega$ ) and "TM" (for $TM$ , $TM^\omega$ or $TM_S^\omega$ ).
distribution	A character string indicating the asymptotic distribution of the statistic, Normal distribution "normal" or Student's t-distribution "t".
weights	A function indicating the weight given to the different records according to their position in the series, e.g., if function(t) t-1 then $\omega_t = t - 1$ .
trend	A character string indicating the type of alternative hypothesis, positive trend in location "positive" or negative trend in location "negative".

**Details**

In this function, the tests are implemented as given by Foster and Stuart (1954), Diersen and Trenkler (1996, 2001) and some modifications standardizing the previous statistics. Let  $(I_{t,m})$  be the record indicator random variables of the  $m$ th series. The number of records up to time  $T$  of the  $m$ th series,  $N_{T,m} = \sum_{t=2}^T I_{t,m}$ , or the weighted number of records up to time  $T$  of the  $m$ th series,  $N_{T,m}^\omega = \sum_{t=2}^T \omega_t I_{t,m}$ ; for the upper and lower records and for the forward and backward series of the  $(X_{t,1}), \dots, (X_{t,M})$ ,  $t = 1, \dots, T$ , series are used for the statistics:

If statistic == "d", distribution = "normal",

$$dM^\omega = \sum_{m=1}^M \sum_{t=2}^T \omega_t (I_{t,m} - I_{t,m}^L).$$

If statistic == "D", distribution = "normal",

$$DM^{\omega} = \sum_{m=1}^M \sum_{t=2}^T \left( I_{t,m} - I_{t,m}^L - I_{t,m}^{rev} + I_{t,m}^{L,rev} \right).$$

If statistic == "TM", distribution = "normal",

$$TM^{\omega} = \sum_{m=1}^M \sum_{t=1}^T \left( I_{t,m} + I_{t,m}^{L,rev} \right).$$

While their means are very simple to calculate, their variances become unwieldy expressions and are given by Diersen and Trenkler (2001).

The p-value is calculated with the Normal asymptotic distribution in the usual way with a one-sided alternative based on the trend value.

If distribution = "t", for the above statistics it is calculated a new one, e.g. for  $DM^{\omega}$ ,

$$DM_S^{\omega} = \frac{DM^{\omega} - E(DM^{\omega})}{\sqrt{\widehat{Var}(DM^{\omega})}},$$

where  $E(DM^{\omega})$  is the expectation under the null, and  $\widehat{Var}(DM^{\omega})$  is an estimation of the variance obtained from the sample, which is why  $M > 1$  is required. The statistic above is asymptotically distributed and it has been proved that  $DM_S^{\omega}$  is highly robust against serial correlation.

## Value

A "htest" object with elements:

statistic	Value of the test statistic.
parameter	Only if distribution = 't', degrees of freedom of $t$ statistic equal to $M - 1$ .
p.value	P-value.
method	A character string indicating the type of test performed.
data.name	A character string giving the name of the data.

## Author(s)

Jorge Castillo-Mateo

## References

- Diersen J, Trenkler G (1996). Records Tests for Trend in Location. *Statistics*, **28**(1), 1-12. doi: [10.1080/02331889708802543](https://doi.org/10.1080/02331889708802543)
- Diersen J, Trenkler G (2001). *Weighted record tests for splitted series of observations*. En J Kunert, G Trenkler (Eds.), *Mathematical Statistics with Applications in Biometry*. Festschrift in Honour of Prof. Dr. Siegfried Schach (pp. 163-178). Lohmar: Josef Eul Verlag.
- Foster FG, Stuart A (1954). Distribution-Free Tests in Time-Series Based on the Breaking of Records. *Journal of the Royal Statistical Society. Series B (Methodological)*, **16**(1), 1-22.

**See Also**[foster.plot](#)**Examples**

```
foster.test(ZaragozaSeries)
foster.test(ZaragozaSeries, statistic = 'd')
foster.test(ZaragozaSeries, weights = function(t) t-1)
foster.test(ZaragozaSeries, distribution = 't')
foster.test(ZaragozaSeries, distribution = 't', weights = function(t) t-1)
```

I.record

*Record Indicator random variables***Description**

This function calculates for each value in the vector a binary variable which takes the value 1 if the corresponding value in the vector is a record and 0 otherwise. If the argument XM\_T is a matrix, each column is treated as a different vector.

**Usage**

```
I.record(XM_T, record = c("upper", "lower"))

## Default S3 method:
I.record(XM_T, record = c("upper", "lower"))

## S3 method for class 'numeric'
I.record(XM_T, record = c("upper", "lower"))

## S3 method for class 'matrix'
I.record(XM_T, record = c("upper", "lower"))
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
record	A character string indicating the type of record to be calculated, "upper" or "lower".

**Details**

Let  $\{X_1, \dots, X_T\}$  be a sequence of random variables of size  $T$ . An observation  $X_t$  will be called an upper record value if its value exceeds that of all previous observations. An analogous definition deals with lower record values. Here  $X_1$  is referred to as the reference value or the trivial record. Then, the sequence of record indicator random variables  $\{I_1, \dots, I_T\}$  is given by

$$I_t = \begin{cases} 1 & \text{if } X_t \text{ is a record,} \\ 0 & \text{if } X_t \text{ is not a record.} \end{cases}$$

The method `I.record` calculates the sample sequence above if the argument `XM_T` is a numeric vector. If the argument `XM_T` is a matrix (or data frame) with  $M$  columns, the method `I.record` calculates the sample sequence above for each column of the object as if all columns were different sequences.

Summarily:

$$\text{I.record : XM\_T} = \begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,M} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,M} \\ \vdots & \vdots & & \vdots \\ X_{T,1} & X_{T,2} & \cdots & X_{T,M} \end{pmatrix} \longrightarrow \begin{pmatrix} I_{1,1} & I_{1,2} & \cdots & I_{1,M} \\ I_{2,1} & I_{2,2} & \cdots & I_{2,M} \\ \vdots & \vdots & & \vdots \\ I_{T,1} & I_{T,2} & \cdots & I_{T,M} \end{pmatrix}.$$

Indicators of record occurrence can be calculated for both upper and lower records.

### Value

A binary matrix, indicating the record occurrence.

### Author(s)

Jorge Castillo-Mateo

### References

Arnold BC, Balakrishnan N, Nagaraja HN (1998). *Records*. New York: Wiley.

### See Also

[L.record](#), [M.record](#), [N.record](#), [Nmean.record](#), [P.record](#), [records](#)

### Examples

```
X <- c(1, 5, 3, 6, 6, 9, 2, 11, 17, 8)
I.record(X)

I.record(ZaragozaSeries)
# record argument can be shortened
I.record(ZaragozaSeries, record = 'l')
```

---

L.plot

---

*Plot of the record times*


---

### Description

This function constructs a ggplot object to display the upper and lower record times for both forward and backward sequences.



**Usage**

```
L.plot(XM_T, colour_point = "skyblue3", colour_line = "grey95")
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
colour_point	Colour to plot points.
colour_line	Colour to plot lines.

**Details**

The function can be applied to plot the record times in a vector (if argument XM\_T is a vector) or to plot and compare the record times in a set of vectors (if argument XM\_T is a matrix). In the latter case, the approach to obtain the record times is applied to each column of the matrix.

A matrix of four panels is displayed for upper and lower records, and for the forward and backward ([series\\_rev](#)) vectors.

**Value**

A ggplot object.

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[L.record](#), [L\\_lr.test](#)

**Examples**

```
Y <- c(1, 5, 3, 6, 6, 9, 2, 11, 17, 8)
L.plot(Y)

L.plot(ZaragozaSeries, colour_point = 1)
```

---

L.record

*Record Times*


---

**Description**

This function calculates the times (positions in the vector) where records occur.

**Usage**

```
L.record(XM_T, record = c("upper", "lower"))
```

**Arguments**

<code>XM_T</code>	A numeric vector, matrix (or data frame).
<code>record</code>	A character string indicating the type of record to be calculated, "upper" or "lower".

**Details**

If `XM_T` is a matrix, the approach to obtain record times is applied to each column of the matrix.

Record times can be calculated for both upper and lower records.

**Value**

If `XM_T` is a vector, the function returns a column matrix containing the record times. If `XM_T` is a matrix, the function returns a list where each element is a vector indicating the record times of the corresponding `XM_T` column.

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[I.record](#), [M.record](#), [N.record](#), [P.record](#), [Nmean.record](#), [records](#)

**Examples**

```
Y1 <- c( 1,  5,  3,  6,  6,  9,  2)
Y2 <- c(10,  5,  3,  6,  6,  9,  2)
Y3 <- c( 5,  7,  3,  6, 19,  2, 20)
Y  <- cbind(Y1, Y2, Y3)

L.record(Y1)
L.record(Y)
```

---

`L_global.test`

*LM and LR tests with Monte Carlo global statistic*

---

**Description**

This function performs a more powerful generalization of the Lagrange multiplier and likelihood ratio tests on records by means of the sum of the statistics of upper and lower records and forward and backward series to study the hypothesis of the classical record model.

**Usage**

```
L_global.test(
  XM_T,
  test = c("LM", "LR"),
  statistic = c("G1", "G2"),
  trend = c("positive", "negative"),
  B = 1000,
  rdist = stats::runif,
  parallel = FALSE,
  numCores = 2,
  seed = NULL
)
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
test	A character string indicating the type of test to perform. "LM" for <a href="#">L_lm.test</a> or "LR" for <a href="#">L_lr.test</a> .
statistic	A character string indicating the type of generalization to perform. "G1" or "G2" (see Details).
trend	A character string indicating the type of alternative hypothesis, positive trend in location "positive" or negative trend in location "negative".
B	An integer specifying the number of replicates used in the Monte Carlo approach.
rdist	function that simulates continuous random variables, e.g., <a href="#">runif</a> (fastest in stats package), <a href="#">rnorm</a> or <a href="#">rexp</a> .
parallel	If TRUE, then the Monte Carlo algorithm is done in parallel. This can give a significant speedup on multicore machines.
numCores	Allows the user to specify the amount of parallel processes to be used if parallel = TRUE. If NULL, then the number of logical cores is automatically detected and all available cores are used.
seed	A vector of the same length as the value of samples. Allows the user to specify the seed of each Monte Carlo simulation. If NULL, then no seed is specified.

**Details**

Here, the Monte Carlo statistics with general alternative  $\mathcal{LM}$  in [L\\_lm.test](#) or  $\mathcal{LR}$  in [L\\_lr.test](#) if test = 'LM' or test = 'LR' respectively, are joined for upper and lower records and forward and backward series as

$$\mathcal{LM}^{G1} = \mathcal{LM} - \mathcal{LM}^L - \mathcal{LM}^{rev} + \mathcal{LM}^{L,rev}$$

or

$$\mathcal{LM}^{G2} = \mathcal{LM} + \mathcal{LM}^{L,rev}$$

, where the superscripts  $L$  and  $rev$  means lower records and reversed (or backwards) series, respectively. Equivalently for  $\mathcal{LR}$ .

The distribution of this generalized statistics is unknown, but it can be estimated with a Monte Carlo approach.

**Value**

A list of class "htest" with the following elements:

statistic	Value of the statistic.
parameter	Number of replicates used in the Monte Carlo test.
p.value	P-value.
method	A character string indicating the type of test.
data.name	A character string giving the name of the data.

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[L\\_lm.test](#), [L\\_lr.test](#)

**Examples**

```
L_global.test(ZaragozaSeries, test = 'LM', B = 200)

L_global.test(ZaragozaSeries, test = 'LR', B = 200)
```

---

L\_lm.test

---

*Lagrange multiplier tests on record indicators*


---

**Description**

This function performs an asymptotic Lagrange multiplier test or a Monte Carlo approach based on the record indicators  $I_t$  to study the hypothesis of the classical record model.

**Usage**

```
L_lm.test(
  XM_T,
  record = c("upper", "lower"),
  alternative = c("!= ", "= "),
  trend = c("positive", "negative"),
  B = 1000,
  rdist = stats::runif,
  parallel = FALSE,
  numCores = 2,
  seed = NULL
)
```

## Arguments

XM_T	A numeric vector, matrix (or data frame).
record	A character string indicating the type of record, "upper" or "lower".
alternative	A character indicating if the alternative hypothesis assume all series with the same distribution "=" (asymptotic chi-squared) or different distribution "!=" (Monte Carlo simulated p-value).
trend	A character string indicating the type of alternative hypothesis, positive trend in location "positive" or negative trend in location "negative".
B	An integer specifying the number of replicates used in the Monte Carlo approach. Only used if alternative = '!='.
rdist	A function that simulates continuous random variables, e.g., <code>runif</code> (fastest in stats package), <code>rnorm</code> or <code>rexp</code> . Only used if alternative = '!='.
parallel	If TRUE, then the Monte Carlo algorithm is done in parallel. This can give a significant speedup on multicore machines. Only used if alternative = '!='.
numCores	Allows the user to specify the amount of parallel processes to be used if parallel = TRUE and alternative = '!='. If NULL, then the number of logical cores is automatically detected and all available cores are used.
seed	A vector of the same length as the value of samples. Allows the user to specify the seed of each Monte Carlo simulation. If NULL, then no seed is specified. Only used if alternative = '!='.

## Details

The null hypothesis of this likelihood ratio tests is that in all the series,  $m = 1, \dots, M$ , the probability of record at time  $t$  is  $1/t$ , and the alternative depends on the alternative argument. The probability at time  $t$  is any value, but equal in the  $M$  series if alternative = '=' or different in the  $M$  series if alternative = '!='. The alternative hypothesis is more specific in the first case than in the second one.

If alternative = '=', under the null, the likelihood ratio statistic has an asymptotic  $\chi^2$  distribution with  $T - 1$  degrees of freedom. It has been seen that for the approximation to be adequate  $M$  must be between 4 and 5 times greater than  $T$ .

If alternative = '!', the asymptotic behaviour is not fulfilled, but a Monte Carlo approach can be applied.

## Value

A list of class "htest" with the following elements:

statistic	Value of the statistic.
parameter	Degrees of freedom of the approximate $\chi^2$ distribution, or number of replicates used in the Monte Carlo test.
p.value	P-value.
method	A character string indicating the type of test.
data.name	A character string giving the name of the data.
alternative	A character string indicating one of both alternative hypothesis.

**Author(s)**

Jorge Castillo-Mateo

**See Also**[L\\_lr.test](#), [L\\_global.test](#)**Examples**

```
L_lm.test(ZaragozaSeries, B = 200)
```

```
L_lm.test(series_rev(ZaragozaSeries), alternative = '=', trend = 'negative')
```

---

`L_lr.test`*Likelihood ratio tests on record times*

---

**Description**

This function performs an asymptotic likelihood ratio test or a Monte Carlo approach based on the record times  $L_i$  to study the hypothesis of the classical record model.

**Usage**

```
L_lr.test(
  XM_T,
  record = c("upper", "lower"),
  alternative = c("!= ", "= "),
  trend = c("positive", "negative"),
  B = 1000,
  rdist = stats::runif,
  parallel = FALSE,
  numCores = 2,
  seed = NULL
)
```

**Arguments**

<code>XM_T</code>	A numeric vector, matrix (or data frame).
<code>record</code>	A character string indicating the type of record, "upper" or "lower".
<code>alternative</code>	A character indicating if the alternative hypothesis assume all series with the same distribution "=" (asymptotic chi-squared) or different distribution "!=" (Monte Carlo simulated p-value).
<code>trend</code>	A character string indicating the type of alternative hypothesis, positive trend in location "positive" or negative trend in location "negative".
<code>B</code>	An integer specifying the number of replicates used in the Monte Carlo approach. Only used if <code>alternative = '!='</code> .

<code>rdist</code>	A function that simulates continuous random variables, e.g., <code>runif</code> (fastest in stats package), <code>rnorm</code> or <code>rexp</code> . Only used if <code>alternative = '!='</code> .
<code>parallel</code>	If TRUE, then the Monte Carlo algorithm is done in parallel. This can give a significant speedup on multicore machines. Only used if <code>alternative = '!='</code> .
<code>numCores</code>	Allows the user to specify the amount of parallel processes to be used if <code>parallel = TRUE</code> and <code>alternative = '!='</code> . If NULL, then the number of logical cores is automatically detected and all available cores are used.
<code>seed</code>	A vector of the same length as the value of <code>samples</code> . Allows the user to specify the seed of each Monte Carlo simulation. If NULL, then no seed is specified. Only used if <code>alternative = '!='</code> .

### Details

The null hypothesis of this likelihood ratio tests is that in all the series,  $m = 1, \dots, M$ , the probability of record at time  $t$  is  $1/t$ , and the alternative depends on the `alternative` argument. The probability at time  $t$  is any value, but equal in the  $M$  series if `alternative = '='` or different in the  $M$  series if `alternative = '!='`. The alternative hypothesis is more specific in the first case than in the second one.

If `alternative = '='`, under the null, the likelihood ratio statistic has an asymptotic  $\chi^2$  distribution with  $T - 1$  degrees of freedom. It has been seen that for the approximation to be adequate  $M$  must be between 4 and 5 times greater than  $T$ .

If `alternative = '!='`, the asymptotic behaviour is not fulfilled, but a Monte Carlo approach can be applied.

### Value

A list of class "htest" with the following elements:

<code>statistic</code>	Value of the statistic.
<code>parameter</code>	Degrees of freedom of the approximate $\chi^2$ distribution, or number of replicates used in the Monte Carlo test.
<code>p.value</code>	P-value.
<code>method</code>	A character string indicating the type of test.
<code>data.name</code>	A character string giving the name of the data.
<code>alternative</code>	A character string indicating one of both alternative hypothesis.

### Author(s)

Jorge Castillo-Mateo

### See Also

[L\\_lm.test](#), [L\\_global.test](#)

### Examples

```
L_lr.test(ZaragozaSeries, B = 200)

L_lr.test(series_rev(ZaragozaSeries), alternative = '=', trend = 'negative')
```

---

M.record	<i>Number of Records and Estimated Record Probability</i>
----------	---

---

### Description

M.record and P.record calculates the number of records and mean number of records at each time  $t$  in a set of  $M$  vectors (columns of the matrix), respectively. In particular, P.record is the estimated record probability at each time  $t$ .

### Usage

```
M.record(XM_T, record = c("upper", "lower"))

P.record(XM_T, record = c("upper", "lower"))
```

### Arguments

XM_T	A numeric vector, matrix (or data frame).
record	A character string indicating the type of record to be calculated, "upper" or "lower".

### Details

Given a matrix formed by  $M$  vectors (columns), measured at  $T$  times (rows), M.record calculates the number of records in the  $M$  vectors at each observed time  $t$ ,  $M_t$ .

The function P.record is equivalent, but calculates the proportion of records at each time  $t$ , that is the ratio:

$$\hat{p}_t = \frac{M_t}{M} = \frac{I_{t,1} + \dots + I_{t,M}}{M},$$

this proportion is an estimation of the probability of record at that time.

Following the notation in [I.record](#), summarily:

$$XM\_T = \begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,M} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,M} \\ \vdots & \vdots & & \vdots \\ X_{T,1} & X_{T,2} & \cdots & X_{T,M} \end{pmatrix} \begin{matrix} \xrightarrow{\text{M.record}} \\ \xrightarrow{\text{P.record}} \end{matrix} \begin{matrix} (M_1, M_2, \dots, M_T) \\ (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_T) \end{matrix}$$

Summaries for both upper and lower records can be calculated.

### Value

A vector with the number (or proportion in the case of P.record) of records at each time  $t$  (row).



**Author(s)**

Jorge Castillo-Mateo

**See Also**[I.record](#), [L.record](#), [N.record](#), [Nmean.record](#), [records](#)**Examples**

```

Y1 <- c( 1,  5,  3,  6,  6,  9,  2)
Y2 <- c(10,  5,  3,  6,  6,  9,  2)
Y3 <- c( 5,  7,  3,  6, 19,  2, 20)
Y  <- cbind(Y1, Y2, Y3)

M.record(Y)
P.record(Y)

M.record(ZaragozaSeries)
P.record(ZaragozaSeries, record = 'l')
```

N.plot

*Plot of the mean number of records up to time  $t$* **Description**

This function constructs a ggplot object to compare the sample means of the number of records in a vector up to time  $t$ ,  $\bar{N}_t$ , and the expected values  $E(N_t)$  under the classical record model.

**Usage**

```

N.plot(
  XM_T,
  weights = function(t) 1,
  record = c("both", "upper", "lower"),
  interval = c("ribbon", "errorbar"),
  conf = 0.95,
  colour = "salmon"
)
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
weights	A function indicating the weight given to the different records according to their position in the series, e.g., if function(t) $t-1$ then $\omega_t = t - 1$ .
record	A character string indicating the type of record, "upper", "lower" or "both".

interval	A character string indicating the type of display of the confidence intervals, "ribbon" (grey area) or "errorbar" (vertical lines).
conf	Confidence level of the confidence intervals.
colour	Colour used to plot the expected values of $N_t$ , and the CI.

## Details

First, this function calculates the sample means of the number of records in a vector up to time  $t$ . These sample means  $\bar{N}_t$  are calculated from the sample of  $M$  values obtained from  $M$  vectors, the columns of matrix  $\mathbf{XM}_T$ . Then, these values are plotted and compared with the expected values  $E(N_t)$  and their confidence intervals (CI), under the hypothesis of the classical record model.

The CI of  $E(N_t)$  uses the fact that, under the classical record model, the statistic  $\bar{N}_t$  is asymptotically Normal.

## Value

A ggplot object.

## Author(s)

Jorge Castillo-Mateo

## See Also

[N\\_normal.test](#), [foster.test](#)

## Examples

```
N.plot(ZaragozaSeries, weights = function(t) t-1, interval = 'errorbar')

Zplot <- N.plot(ZaragozaSeries, record = 'lower')
Zplot

library(ggplot2)
# change the colour of the points
colour_point <- 'black'
Zplot + scale_colour_manual(name = '', values = c('Lower record' = colour_point))

# remove legend
Zplot + theme(legend.position = 'none')
```

---

N.record	<i>Number and Mean Number of Records up to time <math>t</math></i>
----------	--

---

### Description

N.record and Nmean.record calculates the number of records up to time  $t$  in a vector (or every matrix column) of length  $T$  and the mean number of records up to time  $t$  in  $M$  vectors (or matrix columns), respectively.

### Usage

```
N.record(XM_T, record = c("upper", "lower"))
```

```
Nmean.record(XM_T, record = c("upper", "lower"))
```

### Arguments

XM_T	A numeric vector, matrix (or data frame).
record	A character string indicating the type of record to be calculated, "upper" or "lower".

### Details

The record counting process  $\{N_1, \dots, N_T\}$  is defined by the number of records up to time  $t$ , and can be expressed in terms of the record indicator random variables [I.record](#) by

$$N_t = I_1 + I_2 + \dots + I_t.$$

If XM\_T is a matrix with  $M > 1$  columns, each column is treated as a vector and Nmean.record calculates for each  $t$ ,

$$\bar{N}_t = \frac{N_{t,1} + \dots + N_{t,M}}{M}.$$

Summarily:

$$\text{N.record : XM\_T} = \begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,M} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,M} \\ \vdots & \vdots & & \vdots \\ X_{T,1} & X_{T,2} & \cdots & X_{T,M} \end{pmatrix} \longrightarrow \begin{pmatrix} N_{1,1} & N_{1,2} & \cdots & N_{1,M} \\ N_{2,1} & N_{2,2} & \cdots & N_{2,M} \\ \vdots & \vdots & & \vdots \\ N_{T,1} & N_{T,2} & \cdots & N_{T,M} \end{pmatrix}$$

and

$$\text{Nmean.record : XM\_T} \longrightarrow (\bar{N}_1, \bar{N}_2, \dots, \bar{N}_T).$$

Number and mean number of records for both upper and lower records can be calculated.

### Value

N.record returns a numeric matrix with the number of records up to each time (row)  $t$  for a vector or each column in XM\_T. Nmean.record returns a numeric vector with the mean number of records up to each time (row)  $t$ .

**Note**

If `XM_T` is a vector both functions return the same values, `N.record` as a matrix and `Nmean.record` as a vector.

**Author(s)**

Jorge Castillo-Mateo

**References**

Arnold BC, Balakrishnan N, Nagaraja HN (1998). *Records*. New York: Wiley.

**See Also**

[I.record](#), [L.record](#), [M.record](#), [P.record](#), [records](#)

**Examples**

```
Y1 <- c( 1,  5,  3,  6,  6,  9,  2)
Y2 <- c(10,  5,  3,  6,  6,  9,  2)
Y3 <- c( 5,  7,  3,  6, 19,  2, 20)
Y  <- cbind(Y1, Y2, Y3)

N.record(Y)
Nmean.record(Y)

N.record(ZaragozaSeries)
Nmean.record(ZaragozaSeries, record = 'l')
```

---

`N_normal.test`

*Normal test on the number of records*

---

**Description**

This function performs a test based on  $N_T$ , the number of records in the observation period, to study the hypothesis of the classical record model.

**Usage**

```
N_normal.test(
  XM_T,
  weights = function(t) 1,
  record = c("upper", "lower"),
  distribution = c("normal", "t"),
  trend = c("positive", "negative")
)
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
weights	A function indicating the weight given to the different records according to their position in the series, e.g., if function(t) t-1 then $\omega_t = t - 1$ .
record	A character string indicating the type of record to be calculated, "upper" or "lower".
distribution	A character string indicating the asymptotic distribution of the statistic, Normal distribution "normal" or Student's t-distribution "t".
trend	A character string indicating the type of alternative hypothesis, positive trend in location "positive" or negative trend in location "negative".

**Details**

In this test the null hypothesis is that the expected value of  $N_T$ , the number of records in the observation period  $t = 1, \dots, T$  is  $\mu_T = \sum_{i=1}^T 1/i$  and the variance  $\sigma_T^2 = \sum_{i=1}^T (1/i - 1/i^2)$ ; these are the values obtained under the classical record model (see Arnold et al. (1998)). The test statistic is based on  $\bar{N}_T$ , the mean of the number of records up to time t, calculated from a sample of  $M$  vectors (columns in XM\_T). The distribution of the statistic under the null is asymptotically Normal.

If the sequences of variables in the vectors are not i.i.d., but they have a monotonous positive trend, an unilateral alternative hypothesis must be stated, which in the case of upper records is  $\mu_T > \sum_{i=1}^T 1/i$  and  $\sigma_T^2 > \sum_{i=1}^T (1/i - 1/i^2)$ , and in the case of lower records is  $\mu_T < \sum_{i=1}^T 1/i$  and  $\sigma_T^2 < \sum_{i=1}^T (1/i - 1/i^2)$ . The opposite happens if the trend is negative.

The statistic is calculated as

$$Z = \frac{\bar{N}_T - E(\bar{N}_T)}{\sqrt{Var(\bar{N}_T)}},$$

which follows a standard Normal distribution.

If distribution = "t", an estimation of the variance  $\widehat{Var}(DM^\omega)$  is computed from the sample, which is why  $M > 1$  is required in this case, and the statistic is asymptotically t distributed.

**Value**

A "htest" object with elements:

statistic	Value of the test statistic.
parameter	Only if distribution = 't', degrees of freedom of $t$ statistic equal to $M - 1$ .
p.value	P-value.
method	A character string indicating the type of test performed.
data.name	A character string giving the name of the data.

**Author(s)**

Jorge Castillo-Mateo

## References

Arnold BC, Balakrishnan N, Nagaraja HN (1998). *Records*. New York: Wiley.

## See Also

[N.plot](#)

## Examples

```
N_normal.test(ZaragozaSeries)
```

---

P_chisq.test	<i>Chi-Square test on record probabilities</i>
--------------	--

---

## Description

This function performs a chi-square test based on the record probabilities  $p_t$  to study the hypothesis of the classical record model.

## Usage

```
P_chisq.test(XM_T, record = c("upper", "lower"), correction = TRUE)
```

## Arguments

XM_T	A matrix.
record	A character string indicating the type of record to be calculated, "upper" or "lower".
correction	Logical flag. If TRUE, a generalization of the record indicator random variables is calculated by making all expected values greater than 5.

## Details

The null hypothesis of this chi-square test is that in all the vectors (columns of matrix XM\_T), the probability of record at time  $t$  is  $1/t$ , and the alternative that the probabilities are not equal to those values. First, the chi-square goodness of fit statistics to study the null hypothesis  $H_0 : p_t = 1/t$  are calculated for each time  $t = 2, \dots, T$ , where the observed value is the number of records at time  $t$  in the  $M$  vectors and the expected value under the null is  $M/t$ . The test statistic is the sum of the previous  $T - 1$  statistics and its distribution under the null is approximately  $\chi^2_{T-1}$ .

The chi-square approximation may not be valid with low  $M$ , since it requires expected values  $> 5$  or up to 20 % of the expected values is between 1 and 5. If this condition is not satisfied, a warning is displayed. In order to avoid this problem, a correction can be made.

**Value**

A "htest" object with elements:

statistic	Value of the chi-squared statistic.
df	Degrees of freedom of the approximate chi-squared.
p.value	P-value.
method	A character string indicating the type of test performed.
data.name	A character string giving the name of the data.

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[P\\_exactPB.test](#), [P\\_regression.test](#)

**Examples**

```
P_chisq.test(ZaragozaSeries)
P_chisq.test(ZaragozaSeries, correction = FALSE)
```

---

P_exactPB.test	<i>Exact Poisson binomial test on record probabilities</i>
----------------	--

---

**Description**

This function performs an exact test based on the record probabilities  $p_t$  to study the hypothesis of the classical record model.

**Usage**

```
P_exactPB.test(
  XM_T,
  record = c("upper", "lower"),
  method = c("mixed 1", "mixed 2", "dft", "butler")
)

N_exactPB.test(
  XM_T,
  record = c("upper", "lower"),
  method = c("mixed 1", "mixed 2", "dft", "butler")
)
```

**Arguments**

<code>XM_T</code>	A numeric vector, matrix (or data frame).
<code>record</code>	A character string indicating the type of record to be calculated, "upper" or "lower".
<code>method</code>	A character string that indicates the method by which the cdf of the Poisson-Binomial distribution is calculated and therefore the p-value. <code>mixed 1</code> is the preferred (and default) method, it is a more efficient combination of the other algorithms. <code>mixed 2</code> is a less efficient version of <code>mixed 1</code> . <code>dft</code> uses the discrete Fourier transform which algorithm is given in Hong (2013). <code>butler</code> use the algorithm given by Butler and Stephens (2016).

**Details**

The null hypothesis of this likelihood ratio test is that in all the vectors (columns of matrix `XM_T`), the probability of record at time  $t$  is  $1/t$ . The test statistic is the total number of records at times  $t = 2, \dots, T$ , in the  $M$  vectors. Under the null, this is the sum of  $(T - 1)M$  independent Bernoulli variables, with probabilities  $p_2, \dots, p_2, \dots, p_T, \dots, p_T$  with  $p_t = 1/t$ , so that its distribution is a Poisson-Binomial.

Only unilateral alternative hypothesis  $p_t > 1/t, t = 2, \dots, T$  or  $p_t < 1/t, t = 2, \dots, T$  are valid, since otherwise the statistic is not able to detect deviations from the null hypothesis.

`N_exactPB.test` is the same test, but applied to only one vector, instead of  $M$ . Note in this case this test considers the probability at time  $t = 1$  (by definition of  $N_t$ ), but the p-value is the same.

**Value**

A "htest" object with elements:

<code>statistic</code>	Value of the likelihood ratio statistic.
<code>parameter</code>	Length $T$ of the probability vector $(p_2, \dots, p_T)$ and number $M$ of Poisson-Binomial i.i.d. variables summed with the same probability vector in the statistic.
<code>p.value</code>	P-value.
<code>method</code>	A character string indicating the type of test performed.
<code>data.name</code>	A character string giving the name of the data.

**Author(s)**

Jorge Castillo-Mateo

**References**

- Butler, K. and Stephens, M.A. (2016). The Distribution of a Sum of Independent Binomial Random Variables. *Methodology and Computing in Applied Probability*, **19**(2), 557-571. doi:[10.1007/s11009-016-9533-4](https://doi.org/10.1007/s11009-016-9533-4)
- Hong, Y. (2013). On computing the distribution function for the Poisson binomial distribution. *Computational Statistics & Data Analysis*, **59**, 41-51. doi:[10.1016/j.csda.2012.10.006](https://doi.org/10.1016/j.csda.2012.10.006)



**See Also**

[P\\_chisq.test](#), [P\\_regression.test](#)

**Examples**

```
P_exactPB.test(ZaragozaSeries)
N_exactPB.test(ZaragozaSeries[, 23])
```

---

P_regression.plot	<i>Plot of record probabilities</i>
-------------------	-------------------------------------

---

**Description**

This function constructs a ggplot object to display different functions of the record probabilities at time  $t$ ,  $p_t$ .

**Usage**

```
P_regression.plot(
  XM_T,
  record = c("upper", "lower"),
  plot = 1,
  interval = c("ribbon", "errorbar"),
  conf = 0.95,
  weight = TRUE,
  colour_point = "black",
  colour_CI = "salmon",
  colour_lm = "royalblue4"
)
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
record	A character string indicating the type of record to be calculated, "upper" or "lower".
plot	One of the values 1, 2 or 3. It determines the type of plot to be displayed. See Details.
interval	A character string indicating the type of display of the confidence intervals, "ribbon" (grey area) or "errorbar" (vertical lines).
conf	Numeric value in (0,1). Confidence level of the confidence intervals.
weight	logical. If TRUE (the default) the regression line is estimated with weights. Only used if plot = 1.
colour_point	Colour used to plot the points.
colour_CI	Colour used to plot the expected values and the CI.
colour_lm	Colour used to plot the regression line. Only used if plot = 1.

## Details

Three different types of plots which aim to analyse the hypothesis of the classical record model using the record probabilities are implemented. Estimations of the record probabilities  $\hat{p}_t$  used in the plots are obtained as the proportion of records at time  $t$  in  $M$  vectors (columns of matrix  $\mathbf{XM\_T}$ ).

Type 2 is the plot of the estimated record probabilities  $p_t$  versus time. The expected probabilities under the classical record model,  $p_t = 1/t$ , are also plotted, together with binomial confidence intervals. Type 3 is the same plot but on a logarithmic scale, so that the expected value is  $-\log(t)$ .

Type 1 is the plot of the observed values  $t\hat{p}_t$  versus time. The expected values under the classical record model are 1 for any value  $t$ , so that a cloud of points around 1 and with no trend should be expected. The estimated values are plotted, together with binomial confidence intervals. In addition, a regression line is fitted to the cloud of points and plotted together with confidence intervals of the response. If the classical record model is true, the confidence band (in grey) should contain the horizontal line equal to 1. Plots of type 1 are easier to interpret than types 2 and 3.

## Value

A ggplot object.

## Author(s)

Jorge Castillo-Mateo

## See Also

[P\\_regression.test](#)

## Examples

```
P_regression.plot(ZaragozaSeries, plot = 1)
P_regression.plot(ZaragozaSeries, plot = 2)
P_regression.plot(ZaragozaSeries, plot = 3)
P_regression.plot(series_rev(ZaragozaSeries), record = 'lower')
```

---

P\_regression.test

*Regression test on record probabilities*

---

## Description

This function performs a test based on a regression on the record probabilities  $p_t$  to study the hypothesis of the classical record model.

## Usage

```
P_regression.test(XM_T, record = c("upper", "lower"))
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
record	A character string indicating the type of records to be calculated, "upper" or "lower".

**Details**

The null hypothesis of this regression test is that in all the vectors (columns in matrix XM\_T), the probability of record at time  $t$  is  $1/t$ , so that  $tp_t = 1$ . Then, hypothesis  $H_0 : p_t = 1/t, t = 2, \dots, T$  is equivalent to  $H_0 : \beta_0 = 1, \beta_1 = 0$  where  $\beta_0$  and  $\beta_1$  are the coefficients of the regression model  $tp_t = \beta_0 + \beta_1 t$ . The model has to be estimated by weighted least squares since the response is heteroskedastic.

The F statistic is used to compare the regression model under the null hypothesis and a linear regression model with no restriction (the alternative hypothesis is then that  $tp_t$  is a linear function of time). This alternative hypothesis may be reasonable in many real examples, but not always.

**Value**

A "htest" object with elements:

null.value	Value of $\beta_0$ and $\beta_1$ under the null hypothesis.
alternative	Character string indicating the type of alternative hypothesis (two-sided).
method	A character string indicating the type of test performed.
estimate	Value of $\hat{\beta}_0$ and $\hat{\beta}_1$ .
data.name	A character string giving the name of the data.
statistic	Value of the F statistic.
parameters	Degrees of freedom of the F statistic.
p.value	P-value.

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[P\\_exactPB.test](#), [P\\_chisq.test](#), [P\\_regression.plot](#)

**Examples**

```
P_regression.test(ZaragozaSeries)
```

---

records	<i>Record values and record times</i>
---------	---------------------------------------

---

### Description

This function identifies (and plot if argument `plot = TRUE`) the record values ( $R_i$ ), and the record times ( $L_i$ ) in a vector, for both upper and lower records.

### Usage

```
records(
  X_T,
  plot = TRUE,
  variables = NULL,
  colour = c("black", "skyblue2", "limegreen", "salmon"),
  alpha = c(1, 1, 0.1, 1)
)
```

### Arguments

<code>X_T</code>	A numeric vector.
<code>plot</code>	logical. If TRUE (the default) the records are plotted.
<code>variables</code>	Optional. A matrix, containing other variables related to <code>X_T</code> and measured at the same times. Only used if <code>plot = FALSE</code> .
<code>colour, alpha</code>	Character and numeric vectors of length 4, respectively. This arguments represent respectively the colour and transparency of the points: trivial record, upper records, observations and lower records, respectively.

### Value

If `plot = TRUE` a ggplot object, otherwise a list with two data frames where the first column are the record times, the second the record values and, if `variables` is not null, the third column are their values at the record times, respectively for upper and lower records.

### Author(s)

Jorge Castillo-Mateo

### See Also

[I.record](#), [series\\_double](#), [series\\_rev](#), [series\\_split](#), [series\\_uncor](#), [series\\_untie](#)

**Examples**

```
Y <- c(5, 7, 3, 6, 19, 2, 20)
records(Y, plot = FALSE)

# compute tables
records(TX_Zaragoza$TX, plot = FALSE, variables = TX_Zaragoza$DATE)
# compute plot
records(TX_Zaragoza$TX, alpha = c(1, 1, 0.01, 1))
```

---

series_double	<i>Transforms a <math>T \times M</math> in a <math>\lfloor T/k \rfloor \times k M</math> matrix</i>
---------------	---

---

**Description**

This function changes the format of a matrix, in the following way. First, the matrix is divided into  $k$  matrices  $\lfloor T/k \rfloor \times M$ , containing the rows whose remainder of the division of the row number by  $k$  is  $1, 2, \dots, k-1, 0$ , respectively; and secondly those matrices are cbinded.

**Usage**

```
series_double(XM_T, k = 2)
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
k	Integer $> 1$ , times to increase the number of columns.

**Details**

This function is used in the data preparation (or pre-processing) often required to apply the record inference tools in this package.

Most of the record inference tools require a high number of independent series  $M$  (number of columns) to be applied. If  $M$  is low and the time period of observation,  $T$ , is high enough, the following procedure can be applied in order to multiply by  $k$  the value  $M$ . The approach consists of considering that the observations at two (or more) consecutive times,  $t$  and  $t+1$  (or  $t+k-1$ ), are independent observations measured at the same time unit. That means that we are doubling (or multiplying by  $k$ ) the original time unit of the records, so that the length of the observation period will be  $\lfloor T/k \rfloor$ . This function rearranges the original data matrix into the new format.

If the number of rows of the original matrix is not divisible by  $k$ , the first  $nrow(XM\_T) \% k$  rows are deleted.

**Value**

A  $\lfloor T/k \rfloor \times k M$  matrix.

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[series\\_rev](#), [series\\_split](#), [series\\_uncor](#), [series\\_untie](#)

**Examples**

```
series_double(matrix(1:100, 10, 10))
```

```
series_double(ZaragozaSeries, k = 4)
```

---

series\_rev

*Reverse elements by columns*


---

**Description**

Result from apply rev function by columns to the matrix. This allows studying the series backwards and not just forward.

**Usage**

```
series_rev(XM_T)
```

**Arguments**

XM\_T            A numeric matrix (or data frame).

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[series\\_double](#), [series\\_split](#), [series\\_uncor](#), [series\\_untie](#)

**Examples**

```
series_rev(matrix(1:100, 10, 10))
```

```
series_rev(ZaragozaSeries)
```

series\_split

*Splitted series***Description**

The vector  $X$  of length  $T$  is broken into  $Mcols$  blocks, each part containing  $T/Mcols$  elements.

If the vector  $X$  represents consecutive daily values, e.g., and then  $Mcols = 365$  is preferred; this function rearranges  $X$  into a matrix format, where each column is the vector of values at the same day of the year. If it were monthly data it would be preferred, e.g.,  $Mcols = 12$ .

**Usage**

```
series_split(X, Mcols = 365)
```

**Arguments**

$X$                       A numeric vector.  
 $Mcols$                   An integer number, giving the number of columns in the final matrix.

**Details**

This function is used in the data preparation (or pre-processing) often required to apply the record inference tools in this package.

This function transforms a vector into a matrix, applying the following procedure: the first row of the matrix is made up of the first  $Mcols$  elements of the vector, the second row by the  $Mcols$  following elements, and so on. The length of the vector must be a multiple of  $Mcols$  (see Note otherwise).

In the case of a vector of daily values,  $Mcols$  is usually 365, so that the first column corresponds to all the values observed at the 1st of January, the second to the 2nd of January, etc.

If  $X_{i,j}$  represents the value in day  $j$  of year  $i$ , then if

$$X = (X_{1,1}, X_{1,2}, \dots, X_{1,365}, X_{2,1}, X_{2,2}, \dots, X_{T,365}),$$

applying `splitSeries` to  $X$  returns the following matrix:

$$\begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,365} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,365} \\ \vdots & \vdots & & \vdots \\ X_{T,1} & X_{T,2} & \cdots & X_{T,365} \end{pmatrix}_{T \times 365}.$$

**Value**

A matrix with  $Mcols$  columns.

**Note**

[series\\_double](#) can be implemented for the same purpose as this function but without requiring that the length of  $X$  be divisible by  $Mcols$ . It removes the first elements of  $X$  until its length is divisible by  $Mcols$ .

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[series\\_double](#), [series\\_rev](#), [series\\_uncor](#), [series\\_untie](#)

**Examples**

```
series_split(1:100, Mcols = 10)
```

```
series_split(TX_Zaragoza$TX)
```

---

series\_uncor

---

*Extracts a subset of uncorrelated vectors*


---

**Description**

Given a set of  $M$  vectors, this function extracts a subset of them which are uncorrelated.

**Usage**

```
series_uncor(XM_T, m = 1, alpha = 0.05)
```

**Arguments**

XM_T	A numeric matrix (or data frame) where the uncorrelated vectors are extracted from.
m	Integer value giving the starting column.
alpha	Numeric value in $(0, 1)$ . It gives the significance level of the correlation test where alternative hypothesis is that the true correlation is not equal to 0.

**Details**

This function is used in the data preparation (or pre-processing) often required to apply the record inference tools in this package.

Given a set of  $M$  vectors, which are the columns of matrix  $XM\_T$ , this function extracts the biggest subset of uncorrelated vectors (columns), using the following procedure: starting from column  $m$ , the test [cor.test](#) is applied to study the correlation between columns  $m$  and  $m + 1, m + 2, \dots$  and so on up to find a column  $m + k$  which is not significantly correlated with column  $m$ . Then, the process is repeated starting at column  $m + k$ .



**Value**

A vector with the index of the uncorrelated columns in the matrix.

**Author(s)**

Jorge Castillo-Mateo

**See Also**

[series\\_double](#), [series\\_rev](#), [series\\_split](#), [series\\_untie](#)

**Examples**

```
ZM_T <- series_split(TX_Zaragoza$TX)

series_uncor(ZM_T)
```

---

series_untie	<i>Avoid record ties</i>
--------------	--------------------------

---

**Description**

Avoids record ties when observations have been rounded.

**Usage**

```
series_untie(XM_T, a = -0.5, b = 0.5, seed = NULL)
```

**Arguments**

XM_T	A numeric vector, matrix (or data frame).
a, b	Lower and upper limits of the uniform distribution.
seed	Allows the user to specify the seed of the random sampling. If NULL, then no seed is specified.

**Details**

This function is used in the data preparation (or pre-processing). Its use is recommended but not necessary to apply the record inference tools in this package.

The record theory on which the hypothesis tests are based assumes that the random variables are continuous, proving that the probability that two observations take the same value is zero. Most of the data collected is rounded, giving a certain probability to the tie between records, slightly skewing the results, in general obtaining a fewer amount of records if it has been rounded a lot.

This function avoids ties by adding a sample from a uniform variable in the interval (a,b) to each element of XM\_T. If rounding is in units it is recommended to take  $a = -0.5$ ,  $b = 0.5$ , while if it is in tenths it is recommended  $a = -0.05$ ,  $b = 0.05$  and so on.

If the function is used correctly (appropriate a,b values) the records in the original series will also be in the new series.

### Value

A matrix equal to XM\_T whose elements have been added a sample from a uniform variable.

### Author(s)

Jorge Castillo-Mateo

### See Also

[series\\_double](#), [series\\_rev](#) [series\\_split](#), [series\\_uncor](#)

### Examples

```
set.seed(23)
series_untie(matrix(round(rnorm(100), digits = 1), 10, 10), a = -0.05, b = 0.05)

series_untie(ZaragozaSeries, seed = 23)
```

---

TX\_Zaragoza

*Time series of daily maximum temperature at Zaragoza (Spain)*

---

### Description

A dataset containing the series of daily maximum temperature at Zaragoza (Spain), from 01/01/1953 to 31/12/2018. This series is obtained from the ECA series but it has been transformed, by removing days February 29th and filling the missing values. The variables are the following:

- STAID : Station identifier
- SQUID : Source identifier
- DATE : Date YYYYMMDD
- TX : Maximum temperature in 0.1 °C
- Q\_TX : quality code for TX (0='valid'; 1='suspect'; 9='missing')

### Usage

```
data(TX_Zaragoza)
```

### Format

A data frame with 28670 rows and 5 variables.

**Source**

EUROPEAN CLIMATE ASSESSMENT & DATASET (ECA&D)

**References**

Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. of Climatol.*, 22, 1441-1453.

---

ZaragozaSeries

*Splitted and uncorrelated time series* [TX\\_Zaragoza](#)

---

**Description**

The matrix resulting from the preprocessing of TX\_Zaragoza\$TX. The matrix is the result from applying consecutively: [series\\_split](#) and [series\\_uncor](#).

**Usage**

```
data(ZaragozaSeries)
```

**Format**

A matrix with 66 rows and 75 columns.

# Index

## \* datasets

TX\_Zaragoza, 34

ZaragozaSeries, 35

cor.test, 32

foster.plot, 3, 3, 7

foster.test, 3, 4, 5, 18

I.record, 3, 7, 10, 16, 17, 19, 20, 28

L.plot, 3, 8

L.record, 3, 8, 9, 9, 17, 20

L\_global.test, 3, 10, 14, 15

L\_lm.test, 3, 11, 12, 12, 15

L\_lr.test, 3, 9, 11, 12, 14, 14

M.record, 3, 8, 10, 16, 20

N.plot, 3, 4, 17, 22

N.record, 3, 8, 10, 17, 19

N\_exactPB.test, 24

N\_exactPB.test (P\_exactPB.test), 23

N\_normal.test, 3, 4, 18, 20

Nmean.record, 3, 8, 10, 17

Nmean.record (N.record), 19

P.record, 8, 10, 20

P.record (M.record), 16

P\_chisq.test, 3, 22, 25, 27

P\_exactPB.test, 3, 23, 23, 27

P\_regression.plot, 3, 25, 27

P\_regression.test, 3, 23, 25, 26, 26

records, 3, 8, 10, 17, 20, 28

RecordTest-package, 2

rexp, 11, 13, 15

rnorm, 11, 13, 15

runif, 11, 13, 15

series\_double, 3, 28, 29, 30, 32–34

series\_rev, 3, 9, 28, 30, 30, 32–34

series\_split, 3, 28, 30, 31, 33–35

series\_uncor, 3, 28, 30, 32, 32, 34, 35

series\_untie, 3, 28, 30, 32, 33, 33

TX\_Zaragoza, 34, 35

ZaragozaSeries, 35