

Package ‘onlineFDR’

September 17, 2019

Version 1.2.0

Date 2019-04-15

Title Online FDR control

Description This package allows users to control the false discovery rate for online hypothesis testing, where hypotheses arrive sequentially in a stream, as presented by Javanmard and Montanari (2015, 2018) and Ramdas et al. (2017). In this framework, a null hypothesis is rejected based only on the previous decisions, as the future p-values and the number of hypotheses to be tested are unknown.

License GPL-3

RoxygenNote 6.1.1

Encoding UTF-8

Imports stats

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

biocViews MultipleComparison, Software, StatisticalMethod

git_url <https://git.bioconductor.org/packages/onlineFDR>

git_branch RELEASE_3_9

git_last_commit 4693e46

git_last_commit_date 2019-05-02

Date/Publication 2019-09-16

Author David Robertson [aut, cre],

Adel Javanmard [aut],

Andrea Montanari [aut],

Natasha Karp [aut]

Maintainer David Robertson <david.robertson@mrc-bsu.cam.ac.uk>

R topics documented:

onlineFDR-package	2
bonfInfinite	3
LOND	4
LORD	6
LORDdep	7

Index	10
--------------	-----------

onlineFDR-package

onlineFDR: A package for online FDR control

Description

The onlineFDR package provides methods to control the false discovery rate (FDR) for online hypothesis testing, where hypotheses arrive sequentially in a stream. A null hypothesis is rejected based only on the previous decisions, as the future p-values and the number of hypotheses to be tested are unknown.

Details

Package: onlineFDR
Type: Package
Version: 1.1.3
Date: 2019-04-15
License: GPL-3

Javanmard and Montanari (2015, 2018) proposed two methods for online FDR control. The first is LORD, which stands for (significance) Levels based On Recent Discovery and is implemented by the function `LORD`. This function also includes the extension to the LORD procedure, called LORD++, proposed by Ramdas et al. (2017). `LORDdep` provides a modified version of LORD that is valid for dependent p-values.

The second method is LOND, which stands for (significance) Levels based On Number of Discoveries and is implemented by the function `LOND`. By specifying `dep = TRUE`, this function runs a modified version of LOND which is valid for dependent p-values.

As an alternative to these methods, a Bonferroni-like test is implemented by the function `bonfInfinite`. This procedure is also valid for dependent p-values.

Further details on all these procedures can be found in Javanmard and Montanari (2015, 2018) and Ramdas et al. (2017).

Author(s)

David Robertson (<david.robertson@mrc-bsu.cam.ac.uk>), Adel Javanmard, Andrea Montanari and Natasha Karp.

References

Javanmard, A. and Montanari, A. (2015) On Online Control of False Discovery Rate. *arXiv preprint*, <https://arxiv.org/abs/1502.06197>

Javanmard, A. and Montanari, A. (2018) Online Rules for Control of False Discovery Rate and False Discovery Exceedance. *Annals of Statistics*, 46(2):526-554.

Ramdas, A. et al. (2017). Online control of the false discovery rate with decaying memory. *Advances in Neural Information Processing Systems 30*, 5650-5659.

 bonfInfinite

Online FDR control based on a Bonferroni-like test

Description

Implements online FDR control using a Bonferroni-like test.

Usage

```
bonfInfinite(d, alpha = 0.05, alphai, random = TRUE,
  date.format = "%Y-%m-%d")
```

Arguments

d	Dataframe with three columns: an identifier ('id'), date ('date') and p-value ('pval'). If no column of dates is provided, then the p-values are treated as being ordered sequentially with no batches.
alpha	Overall significance level of the FDR procedure, the default is 0.05.
alphai	Optional vector of α_i , where hypothesis i is rejected if the i -th p-value is less than or equal to α_i . A default is provided as proposed by Javanmard and Montanari (2018), equation 31.
random	Logical. If TRUE (the default), then the order of the p-values in each batch (i.e. those that have exactly the same date) is randomised.
date.format	Optional string giving the format that is used for dates.

Details

The function takes as its input a dataframe with three columns: an identifier ('id'), date ('date') and p-value ('pval'). The case where p-values arrive in batches corresponds to multiple instances of the same date. If no column of dates is provided, then the p-values are treated as being ordered sequentially with no batches.

The procedure controls FDR for a potentially infinite stream of p-values by using a Bonferroni-like test. Given an overall significance level α , we choose a (potentially infinite) sequence of non-negative numbers α_i such that they sum to α . Hypothesis i is rejected if the i -th p-value is less than or equal to α_i .

Value

d.out	A dataframe with the original dataframe d (which will be reordered if there are batches and random = TRUE), the adjusted significance thresholds alphai and the indicator function of discoveries R, where $R[i] = 1$ corresponds to hypothesis i being rejected (otherwise $R[i] = 0$).
-------	---

References

Javanmard, A. and Montanari, A. (2018) Online Rules for Control of False Discovery Rate and False Discovery Exceedance. *Annals of Statistics*, 46(2):526-554.

Examples

```

sample.df <- data.frame(
  id = c('A15432', 'B90969', 'C18705', 'B49731', 'E99902',
        'C38292', 'A30619', 'D46627', 'E29198', 'A41418',
        'D51456', 'C88669', 'E03673', 'A63155', 'B66033'),
  date = as.Date(c(rep("2014-12-01",3),
                   rep("2015-09-21",5),
                   rep("2016-05-19",2),
                   "2016-11-12",
                   rep("2017-03-27",4))),
  pval = c(2.90e-17, 0.06743, 0.01514, 0.08174, 0.00171,
           3.60e-05, 0.79149, 0.27201, 0.28295, 7.59e-08,
           0.69274, 0.30443, 0.00136, 0.72342, 0.54757))

set.seed(1); bonfInfinite(sample.df)
bonfInfinite(sample.df, random=FALSE)
set.seed(1); bonfInfinite(sample.df, alpha=0.1)

```

LOND

*Online FDR control based on number of discoveries***Description**

Implements the LOND algorithm for online FDR control, where LOND stands for (significance) Levels based On Number of Discoveries, as presented by Javanmard and Montanari (2015).

Usage

```

LOND(d, alpha = 0.05, beta, dep = FALSE, random = TRUE,
     date.format = "%Y-%m-%d")

```

Arguments

<code>d</code>	Dataframe with three columns: an identifier ('id'), date ('date') and p-value ('pval'). If no column of dates is provided, then the p-values are treated as being ordered sequentially with no batches.
<code>alpha</code>	Overall significance level of the FDR procedure, the default is 0.05.
<code>beta</code>	Optional vector of β_i . A default is provided as proposed by Javanmard and Montanari (2018), equation 31.
<code>dep</code>	Logical. If TRUE, runs the modified LOND algorithm which guarantees FDR control for <i>dependent</i> p-values. Defaults to FALSE.
<code>random</code>	Logical. If TRUE (the default), then the order of the p-values in each batch (i.e. those that have exactly the same date) is randomised.
<code>date.format</code>	Optional string giving the format that is used for dates.

Details

The function takes as its input a dataframe with three columns: an identifier ('id'), date ('date') and p-value ('pval'). The case where p-values arrive in batches corresponds to multiple instances of the same date. If no column of dates is provided, then the p-values are treated as being ordered sequentially with no batches.

The LOND algorithm controls FDR for independent p-values. Given an overall significance level α , we choose a sequence of non-negative numbers β_i such that they sum to α . The values of the adjusted significance thresholds α_i are chosen as follows:

$$\alpha_i = (D(i - 1) + 1)\beta_i$$

where $D(n)$ denotes the number of discoveries in the first n hypotheses.

For dependent p-values, LOND controls FDR if it is modified with $\beta_i/H(i)$ in place of β_i , where $H(j)$ is the j -th harmonic number.

Further details of the LOND algorithm can be found in Javanmard and Montanari (2015).

Value

d.out A dataframe with the original dataframe d (which will be reordered if there are batches and random = TRUE), the LOND-adjusted significance thresholds α_i and the indicator function of discoveries R. Hypothesis i is rejected if the i -th p-value is less than or equal to α_i , in which case $R[i] = 1$ (otherwise $R[i] = 0$).

References

Javanmard, A. and Montanari, A. (2015) On Online Control of False Discovery Rate. *arXiv preprint*, <https://arxiv.org/abs/1502.06197>

Javanmard, A. and Montanari, A. (2018) Online Rules for Control of False Discovery Rate and False Discovery Exceedance. *Annals of Statistics*, 46(2):526-554.

Examples

```
sample.df <- data.frame(
  id = c('A15432', 'B90969', 'C18705', 'B49731', 'E99902',
        'C38292', 'A30619', 'D46627', 'E29198', 'A41418',
        'D51456', 'C88669', 'E03673', 'A63155', 'B66033'),
  date = as.Date(c(rep("2014-12-01", 3),
                  rep("2015-09-21", 5),
                  rep("2016-05-19", 2),
                  "2016-11-12",
                  rep("2017-03-27", 4))),
  pval = c(2.90e-17, 0.06743, 0.01514, 0.08174, 0.00171,
          3.60e-05, 0.79149, 0.27201, 0.28295, 7.59e-08,
          0.69274, 0.30443, 0.00136, 0.72342, 0.54757))

set.seed(1); LOND(sample.df)
LOND(sample.df, random=FALSE)
set.seed(1); LOND(sample.df, alpha=0.1)
```

LORD

*Online FDR control based on recent discovery***Description**

Implements the LORD procedure for online FDR control where LORD stands for (significance) Levels based On Recent Discovery, as presented by Javanmard and Montanari (2018) and Ramdas et al. (2017).

Usage

```
LORD(d, alpha = 0.05, gammai, version = 3, w0, b0, random = TRUE,
      date.format = "%Y-%m-%d")
```

Arguments

d	Dataframe with three columns: an identifier ('id'), date ('date') and p-value ('pval'). If no column of dates is provided, then the p-values are treated as being ordered sequentially with no batches.
alpha	Overall significance level of the FDR procedure, the default is 0.05.
gammai	Optional vector of γ_i . A default is provided as proposed by Javanmard and Montanari (2018), equation 31.
version	Either an integer from 1 to 3, or the character '++'. This specifies the version of LORD to use, and defaults to 3.
w0	Initial 'wealth' of the procedure, defaults to $\alpha/10$.
b0	The 'payout' for rejecting a hypothesis in LORD 1/2/3. Defaults to $\alpha - w_0$.
random	Logical. If TRUE (the default), then the order of the p-values in each batch (i.e. those that have exactly the same date) is randomised.
date.format	Optional string giving the format that is used for dates.

Details

The function takes as its input a dataframe with three columns: an identifier ('id'), date ('date') and p-value ('pval'). The case where p-values arrive in batches corresponds to multiple instances of the same date. If no column of dates is provided, then the p-values are treated as being ordered sequentially with no batches.

The LORD procedure controls FDR for independent p-values. Given an overall significance level α , we choose a sequence of non-negative numbers γ_i such that they sum to 1, and $\gamma_i \geq \gamma_j$ for $i \leq j$.

Javanmard and Montanari (2018) present three versions of LORD which differ in the way the adjusted significance thresholds α_i are calculated. The significance thresholds for LORD 1 are based on the time of the last discovery (i.e. hypothesis rejection), LORD 2 are based on all previous discovery times, and LORD 3 are based on the time of the last discovery as well as the 'wealth' accumulated at that time. LORD 2 was extended by Ramdas et al. (2017) to give a procedure called LORD++.

LORD depends on constants w_0 and b_0 , where $0 \leq w_0 \leq \alpha$ represents the initial 'wealth' of the procedure and $b_0 > 0$ represents the 'payout' for rejecting a hypothesis. We require $w_0 + b_0 \leq \alpha$ for FDR control to hold for LORD 1/2/3.

Note that FDR control also holds for the LORD procedure if only the p-values corresponding to true nulls are mutually independent, and independent from the non-null p-values.

Further details of the LORD procedures can be found in Javanmard and Montanari (2018) and Ramdas et al. (2017).

Value

d.out A dataframe with the original dataframe d (which will be reordered if there are batches and random = TRUE), the LORD-adjusted significance thresholds α_i and the indicator function of discoveries R. Hypothesis i is rejected if the i -th p-value is less than or equal to α_i , in which case $R[i] = 1$ (otherwise $R[i] = 0$).

References

Javanmard, A. and Montanari, A. (2018) Online Rules for Control of False Discovery Rate and False Discovery Exceedance. *Annals of Statistics*, 46(2):526-554.

Ramdas, A. et al. (2017). Online control of the false discovery rate with decaying memory. *Advances in Neural Information Processing Systems* 30, 5650-5659.

See Also

[LORDdep](#) uses a modified version of the LORD algorithm that is valid for *dependent* p-values.

Examples

```
sample.df <- data.frame(
  id = c('A15432', 'B90969', 'C18705', 'B49731', 'E99902',
        'C38292', 'A30619', 'D46627', 'E29198', 'A41418',
        'D51456', 'C88669', 'E03673', 'A63155', 'B66033'),
  date = as.Date(c(rep("2014-12-01", 3),
                  rep("2015-09-21", 5),
                  rep("2016-05-19", 2),
                  "2016-11-12",
                  rep("2017-03-27", 4))),
  pval = c(2.90e-17, 0.06743, 0.01514, 0.08174, 0.00171,
           3.60e-05, 0.79149, 0.27201, 0.28295, 7.59e-08,
           0.69274, 0.30443, 0.00136, 0.72342, 0.54757))

LORD(sample.df, random=FALSE)
set.seed(1); LORD(sample.df, version='++')
set.seed(1); LORD(sample.df, alpha=0.1, w0=0.05)
```

LORDdep

Online FDR control based on recent discovery for dependent p-values

Description

Implements the LORD procedure for online FDR control for dependent p-values, where LORD stands for (significance) Levels based On Recent Discovery, as presented by Javanmard and Montanari (2018).

Usage

```
LORDdep(d, alpha = 0.05, xi, w0 = alpha/10, b0 = alpha - w0,
        random = TRUE, date.format = "%Y-%m-%d")
```

Arguments

d	Dataframe with three columns: an identifier ('id'), date ('date') and p-value ('pval'). If no column of dates is provided, then the p-values are treated as being ordered sequentially with no batches.
alpha	Overall significance level of the FDR procedure, the default is 0.05.
xi	Optional vector of ξ_i . A default is provided to satisfy the condition given in Javanmard and Montanari (2018), example 3.7.
w0	Initial 'wealth' of the procedure. Defaults to $\alpha/10$.
b0	The 'payout' for rejecting a hypothesis. Defaults to $\alpha - w_0$.
random	Logical. If TRUE (the default), then the order of the p-values in each batch (i.e. those that have exactly the same date) is randomised.
date.format	Optional string giving the format that is used for dates.

Details

The function takes as its input a dataframe with three columns: an identifier ('id'), date ('date') and p-value ('pval'). The case where p-values arrive in batches corresponds to multiple instances of the same date. If no column of dates is provided, then the p-values are treated as being ordered sequentially with no batches.

This modified LORD procedure controls FDR for dependent p-values. Given an overall significance level α , we choose a sequence of non-negative numbers ξ_i such that they satisfy a condition given in Javanmard and Montanari (2018), example 3.8.

The procedure depends on constants w_0 and b_0 , where $w_0 \geq 0$ represents the initial 'wealth' and $b_0 > 0$ represents the 'payout' for rejecting a hypothesis. We require $w_0 + b_0 \leq \alpha$ for FDR control to hold.

Further details of the modified LORD procedure can be found in Javanmard and Montanari (2018).

Value

d.out	A dataframe with the original dataframe d (which will be reordered if there are batches and random = TRUE), the LORD-adjusted significance thresholds α_i and the indicator function of discoveries R. Hypothesis i is rejected if the i -th p-value is less than or equal to α_i , in which case $R[i] = 1$ (otherwise $R[i] = 0$).
-------	---

References

Javanmard, A. and Montanari, A. (2018) Online Rules for Control of False Discovery Rate and False Discovery Exceedance. *Annals of Statistics*, 46(2):526-554.

Examples

```
sample.df <- data.frame(
  id = c('A15432', 'B90969', 'C18705', 'B49731', 'E99902',
        'C38292', 'A30619', 'D46627', 'E29198', 'A41418',
        'D51456', 'C88669', 'E03673', 'A63155', 'B66033'),
  date = as.Date(c(rep("2014-12-01", 3),
```

```
      rep("2015-09-21",5),
      rep("2016-05-19",2),
      "2016-11-12",
      rep("2017-03-27",4))),
pval = c(2.90e-17, 0.06743, 0.01514, 0.08174, 0.00171,
        3.60e-05, 0.79149, 0.27201, 0.28295, 7.59e-08,
        0.69274, 0.30443, 0.00136, 0.72342, 0.54757))

set.seed(1); LORDdep(sample.df)
LORDdep(sample.df, random=FALSE)
set.seed(1); LORDdep(sample.df, alpha=0.1, w0=0.05)
```

Index

[bonfInfinite](#), [2](#), [3](#)

[LOND](#), [2](#), [4](#)

[LORD](#), [2](#), [6](#)

[LORDdep](#), [2](#), [7](#), [7](#)

[onlineFDR-package](#), [2](#)