Package `publipha`

January 15, 2020

**Title**  Bayesian Meta-Analysis with Publications Bias and P-Hacking

**Version**  0.1.1

**Description**  Tools for Bayesian estimation of meta-analysis models that account for publications bias or p-hacking. For publication bias, this package implements a variant of the p-value based selection model of Hedges (1992) <doi:10.1214/ss/1177011364> with discrete selection probabilities. It also implements the mixture of truncated normals model for p-hacking described in Moss and De Bin (2019) <arXiv:1911.12445>.

**License**  GPL-3

**Depends**  methods, R (>= 3.5.0), Rcpp (>= 0.12.19)

**Imports**  rstan (>= 2.18.1), rstantools (>= 1.5.1), loo, truncnorm

**LinkingTo**  BH (>= 1.72.0-2), Rcpp (>= 0.12.19), RcppEigen (>= 0.3.3.4.0), rstan (>= 2.18.1), StanHeaders (>= 2.18.0)

**Encoding**  UTF-8

**LazyData**  true

**NeedsCompilation**  yes

**SystemRequirements**  GNU make

**RoxygenNote**  7.0.2

**Suggests**  testthat (>= 2.1.0), covr, knitr, rmarkdown, metafor, qtl, spelling

**Collate**  'data-anderson2010.R' 'data-baskerville2012.R'
  'densities-helpers.R' 'densities-mpsnorm.R'
  'densities-phpnorm.R' 'densities-psnorm.R' 'densities-snorn.R'
  'ma.R' 'generics.R' 'publipha-package.R' 'stanmodels.R'
  'tools.R' 'utility.R' 'zzz.R'

**Language**  en-US

**Author**  Jonas Moss [aut, cre] (<https://orcid.org/0000-0002-6876-6964>),
  Trustees of Columbia University [cph]

**Maintainer**  Jonas Moss <jonas.gjertsen@gmail.com>

**Repository**  CRAN

**Date/Publication**  2020-01-15 00:20:07 UTC
R topics documented:

publipha-package .................................................. 2
dat.anderson2010 .................................................. 2
dat.baskerville2012 .................................................. 3
dat.cuddy2018 ....................................................... 4
dat.dang2018 ......................................................... 4
dat.motyl2017 ......................................................... 5
ExtractParameters .................................................... 6
loo,mafit-method ..................................................... 7
ma ................................................................. 8
mafit-class ........................................................ 10
mpsnorm ............................................................ 10
phnorm ............................................................. 11
psnorm ............................................................. 13
snorm ............................................................. 14

Index

publipha-package The 'publipha' package.

Description

Meta-analysis that corrects for publication selection bias and p-hacking.

References


dat.anderson2010 Studies on Effect of Violent Video Games on Negative Outcomes

Description

Results from 477 studies on the effect of violent video games on negative outcomes.

Usage

dat.anderson2010

Format

The data frame contains the following columns:
Studies on Practice Facilitation

Description

Results from 23 studies on the effect of practice facilitation in a primary care setting.

Usage
dat.baskerville2012

Format

The tibble contains the following columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>author</td>
<td>character</td>
<td>first author of study</td>
</tr>
<tr>
<td>year</td>
<td>numeric</td>
<td>publication year</td>
</tr>
<tr>
<td>outcome</td>
<td>character</td>
<td>one of seven outcomes</td>
</tr>
<tr>
<td>best</td>
<td>boolean</td>
<td>if TRUE, the was a best practice study</td>
</tr>
<tr>
<td>experimental</td>
<td>boolean</td>
<td>if TRUE, the study was experimental</td>
</tr>
<tr>
<td>adult</td>
<td>boolean</td>
<td>if TRUE, the study subjects were adults</td>
</tr>
<tr>
<td>country</td>
<td>character</td>
<td>country of study</td>
</tr>
<tr>
<td>ni</td>
<td>numeric</td>
<td>sample size</td>
</tr>
<tr>
<td>yi</td>
<td>numeric</td>
<td>observed mean difference in outcome (facilitated vs non-facilitated)</td>
</tr>
<tr>
<td>vi</td>
<td>numeric</td>
<td>corresponding sampling variance</td>
</tr>
</tbody>
</table>

Source

dat.cuddy2018  Studies on the Effect of Power Posing

Description

Results from 27 studies related to power posing.

Usage
dat.cuddy2018

Format

The data frame contains the following columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>author</td>
<td>character</td>
<td>first author</td>
</tr>
<tr>
<td>year</td>
<td>numeric</td>
<td>publication year</td>
</tr>
<tr>
<td>power</td>
<td>boolean</td>
<td>if TRUE, the outcome was feeling of power</td>
</tr>
<tr>
<td>ease</td>
<td>boolean</td>
<td>if TRUE, the outcome was an EASE variable</td>
</tr>
<tr>
<td>yi</td>
<td>numeric</td>
<td>standardized mean difference</td>
</tr>
<tr>
<td>vi</td>
<td>numeric</td>
<td>corresponding sampling variance</td>
</tr>
</tbody>
</table>

Details

The data points are taken from the p-curve analysis of Cuddy et al. (2018), restricted to 2 cell designs with mean difference as the outcome variable.

Source

https://osf.io/jx3av/

References


dat.dang2018  Meta-analysis on Ego Depletion

Description

Results from 150 studies of ego depletion, the claim that self-control is a limited resource which is tapped whenever self-control is exerted.

Usage
data(dat.dang2018)
Format

The tibble contains the following columns:

- **author** character: the last name of the first author and the first letter of the last name of the second author;
- **year** numeric: publication year;
- **in_carter** character: was the study in the meta-analysis of Carter et al. (2015);
- **study** character: the number given to the study in the original paper (0 = only one study was reported in the original paper);
- **dv** boolean: the dependent variable;
- **iv** boolean: the independent variable;
- **n1i** numeric: the number of participants in the depletion condition;
- **n2i** numeric: the number of participants in the control condition;
- **yi** numeric: the adjusted standardized mean difference;
- **vi** numeric: the variance.

Source

https://link.springer.com/article/10.1007%2Fs00426-017-0862-x#SupplementaryMaterial

References


Effect Sizes from 875 Studies in Psychology.

Description

Effect sizes from 875 studies in psychology. Adopted from Motyl et al. (2017).

Usage

data(dat.motyl2017)

Format

The tibble contains the following columns:

- **author** character: first author of study;
- **year** numeric: publication year;
- **study** numeric: the number given to the study in the original paper (0 = only one study was reported in the original paper);
- **journal** character: journal where the study was published;
- **concealed** character: design of the study; "Between", "Within", or "Mixed";
- **experimental** numeric: TRUE for an experimental study.
Extract Parameters

<table>
<thead>
<tr>
<th>ni</th>
<th>numeric</th>
<th>sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>yi</td>
<td>numeric</td>
<td>observed mean difference in outcome</td>
</tr>
<tr>
<td>vi</td>
<td>numeric</td>
<td>corresponding sampling variance</td>
</tr>
</tbody>
</table>

Source

https://osf.io/he8mu/

References

Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Mueller, A. B., ... & Yantis, C. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse?. Journal of personality and social psychology, 113(1), 34.

**ExtractParameters**  
*Extract Parameters from an mafit Object*

**Description**

Extract samples from a model of class `mafit` and apply a function `fun` to them.

**Usage**

- `extract_theta0(object, fun = mean)`
- `extract_theta(object, fun = mean, i)`
- `extract_taul(object, fun = mean)`
- `extract_eta(object, fun = mean, i)`
- `extract_isq(object, fun = mean)`

**Arguments**

- `object` : an object of class `mafit`.
- `fun` : the function to be applied to the fitted parameters.
- `i` : an optional index specifying which parameter to apply `fun` to. Only for `extract_eta` and `extract_theta`.

**Details**

Support parameters for extraction are: The meta-analytic mean `theta0`, the individual means `theta`, the heterogeneity parameter `tau`, the selection bias parameter `eta`, and the I squared `isq`. See Higgins and Thompson (2002) for details about I squared.

All `extract_` functions are wrappers around `rstan::extract`. 
### Description

Computes PSIS-LOO CV, approximate leave-one-out cross-validation using Pareto smoothed importance sampling, see `loo`.

### Usage

```r
## S4 method for signature 'mafit'
loo(x, ...)
```

### Arguments

- `x`: an object of class `mafit`.
- `...`: passed to `loo`. Only

### Details

... affect the function through two parameters, `marginal` and `lower_bound`. When `marginal` is `TRUE`, the PSIS-LOO CV is based on the marginal likelihood, i.e. with the dependence on `theta` integrated out. `marginal` defaults to `TRUE`. `lower_bound` species the lower bound where log-likelihoods are dropped; this is only used in the `p`-hacking model and defaults to `-6`.

### Value

A `loo` object.
Examples

```r
phma_model <- phma(yi, vi, data = metafor::dat.begg1989)
psma_model <- psma(yi, vi, data = metafor::dat.begg1989)
loo(phma_model)
loo(psma_model)
```

Meta-analysis Correcting for Publication Bias or p-hacking

Description

Bayesian random effects meta-analysis. Correct for publication bias, correct for p-hacking, or run an ordinary meta-analysis without any correction.

Usage

```r
ma(
  yi,
  vi,
  bias = c("publication selection", "p-hacking", "none"),
  data,
  alpha = c(0, 0.025, 0.05, 1),
  prior = NULL,
  ...
)

psma(yi, vi, data, alpha = c(0, 0.025, 0.05, 1), prior = NULL, ...)

phma(yi, vi, data, alpha = c(0, 0.025, 0.05, 1), prior = NULL, ...)

cma(yi, vi, data, prior = NULL, ...)

allma(yi, vi, data, alpha = c(0, 0.025, 0.05, 1), prior = NULL, ...)
```

Arguments

- **yi**: Numeric vector of length codek with observed effect size estimates.
- **vi**: Numeric vector of length codek with sampling variances.
- **bias**: String; If "publication bias", corrects for publication bias. If "p-hacking", corrects for p-hacking.
- **data**: Optional list or data frame containing yi and vi.
- **alpha**: Numeric vector; Specifies the cutoffs for significance. Should include 0 and 1. Defaults to (0, 0.025, 0.05, 1).
- **prior**: Optional list of prior parameters. See the details.
- **...**: Passed to `rstan::sampling`. 
Details

ma does a Bayesian meta-analysis with the type of correction used specified by bias. psma is a wrapper for ma with bias = "publication selection", phma is a wrapper with bias = "p-hacking", while cma has bias = "none". The function allma runs all bias options and returns a list.

The bias options are:

1. publication selection: The model of publication bias described in Hedges (1992).
2. p-hacking: The model for p-hacking described in Moss & De Bin (2019).
3. none: Classical random effects meta-analysis with no correction for selection bias.

The effect size distribution is normal with mean theta0 and standard deviation tau. The prior for theta0 is normal with parameters theta0_mean (default: 0), theta0_sd (default: 1). The prior for tau is half normal with parameters tau_mean (default: 1), tau_sd (default: 1). eta is the vector of K normalized publication probabilities (publication bias model) or K p-hacking probabilities (p-hacking model). The prior of eta is Dirichlet with parameter eta0, which defaults to rep(1, K) for the publication bias model and the p-hacking model. eta0 is the prior for the Dirichlet distribution over the non-normalized etas in the publication bias model, and they are forced to be decreasing. To change the prior parameters, pass them to prior in a list.

Value

An S4 object of class mafit when ma, psma, phma or cma is run. A list of mafit objects when allma is run.

References


Examples

phma_model <- phma(yi, vi, data = metafor::dat.begg1989)

prior <- list(
  eta0 = c(3, 2, 1),
  theta0_mean = 0.5,
  theta0_sd = 10,
  tau_mean = 1,
  tau_sd = 1
)

psma_model <- psma(yi, vi, data = metafor::dat.begg1989, prior = prior)

cma_model <- psma(yi, vi, data = metafor::dat.begg1989, prior = prior)

model <- all(yi, vi, data = metafor::dat.begg1989, prior = prior)
Class mafit: Fitted Meta-analysis Model

Slots

bias  The kind of bias modelled. Can be one of publication_selection, p-hacking or none.
alpha  Ordered numeric vector of cutoffs including 0 and 1.
yi  Numeric vector of estimated effect sizes.
vi  Numeric vector of study-specific variances.
parameters  The list of prior parameters used in the fitting.

Marginal Publication Selection Meta-analysis Model

Description

Density, distribution, and random variate generation for the marginalized distribution of the publication selection meta-analysis model

Usage

dmpsnorm(x, theta0, tau, sigma, alpha = c(0, 0.025, 0.05, 1), eta, log = FALSE)

pmpsnorm(q, theta0, tau, sigma, alpha = c(0, 0.025, 0.05, 1), eta, lower.tail = TRUE, log.p = FALSE)

rmepsnorm(n, theta0, tau, sigma, alpha = c(0, 0.025, 0.05, 1), eta)
**Arguments**

- `x, q` vector of quantiles.
- `theta0` vector of means.
- `tau` vector of heterogeneity parameters.
- `sigma` vector of study standard deviations.
- `alpha` vector of thresholds for publication bias.
- `eta` vector of publication probabilities, normalized to sum to 1.
- `log, log.p` logical; If TRUE, probabilities are given as log(p).
- `lower.tail` logical; If TRUE (default), the probabilities are \( P[X \leq x] \) otherwise, \( P[X \geq x] \).
- `n` number of observations. If `length(n) > 1`, the length is taken to be the number required.

**Details**

These functions assume a normal underlying effect size distribution and one-sided selection on the effects. For the fixed effects publication bias model see `psnorm`.

**Value**

dmpsnpnorm gives the density, pmpsnpnorm gives the distribution function, and rmpsnpnorm generates random deviates.

**References**


Moss, Jonas and De Bin, Riccardo. "Modelling publication bias and p-hacking" Forthcoming (2019)

**Examples**

```r
rmpsnpnorn(100, theta0 = 0, tau = 0.1, sigma = 0.1, eta = c(1, 0.5, 0.1))
```

---

**Description**

Density, distribution, and random variate generation for the p-hacking meta-analysis model.
Usage

dphnorm(x, theta, sigma, alpha = c(0, 0.025, 0.05, 1), eta, log = FALSE)

rphnorm(n, theta, sigma, alpha = c(0, 0.025, 0.05, 1), eta)

pphnorm(q, theta, sigma, alpha = c(0, 0.025, 0.05, 1), eta, lower.tail = TRUE, log.p = FALSE)

Arguments

x, q vector of quantiles.
theta vector of means.
sigma vector of study standard deviations.
alpha vector of thresholds for p-hacking.
eta vector of p-hacking probabilities, normalized to sum to 1.
log, log.p logical; If TRUE, probabilities are given as log(p).
n number of observations. If length(n) > 1, the length is taken to be the number required.
lower.tail logical; If TRUE (default), the probabilities are $P[X \leq x]$ otherwise, $P[X \geq x]$.

Details

These functions assume one-sided selection on the effects. alpha contains the selection thresholds and eta the vector of p-hacking probabilities. theta is the true effect, while sigma is the true standard deviation before selection.

Value

dphnorm gives the density, pphnorm gives the distribution function, and rphnorm generates random deviates.

References

Moss, Jonas and De Bin, Riccardo. "Modelling publication bias and p-hacking" Forthcoming (2019)

Examples

rphnorm(100, theta = 0, sigma = 0.1, eta = c(1, 0.5, 0.1))
Publication Selection Meta-analysis Model

Description

Density, distribution, quantile, random variate generation, and expectation calculation for the distribution for the publication selection meta-analysis model.

Usage

dpsnorm(x, theta, sigma, alpha = c(0, 0.025, 0.05, 1), eta, log = FALSE)

ppsnorm(
  q,
  theta,
  sigma,
  alpha = c(0, 0.025, 0.05, 1),
  eta,
  lower.tail = TRUE,
  log.p = FALSE
)

rpsnorm(n, theta, sigma, alpha = c(0, 0.025, 0.05, 1), eta)

Arguments

x, q          vector of quantiles.
theta         vector of means.
sigma         vector of study standard deviations.
alpha         vector of thresholds for publication bias.
eta           vector of publication probabilities, normalized to sum to 1.
log, log.p    logical; If TRUE, probabilities are given as \( \log(p) \).
lower.tail    logical; If TRUE (default), the probabilities are \( P[X \leq x] \) otherwise, \( P[X \geq x] \).
n             number of observations. If length(n) > 1, the length is taken to be the number required.

Details

The effect size distribution for the publication selection model is not normal, but has itself been selected for. These functions assume one-sided selection on the effects. These functions do not assume the existence of an underlying effect size distribution. For these, see \texttt{mpsnorm}.

Value

dpsnorm gives the density, ppsnorm gives the distribution function, and rpsnorm generates random deviates.
References


Moss, Jonas and De Bin, Riccardo. "Modelling publication bias and p-hacking" Forthcoming (2019)

Examples

```r
rpsnorm(100, theta = 0, sigma = 0.1, eta = c(1, 0.5, 0.1))
```

---

**snorm**  
*Selected Normal Effect Size Distribution*

**Description**

Density, random variate generation, and expectation calculation for the effect size distribution of the one-sided normal publication bias model.

**Usage**

```r
dsnorm(x, theta0, tau, sigma, alpha = c(0, 0.025, 0.05, 1), eta, log = FALSE)
rsnorm(n, theta0, tau, sigma, alpha = c(0, 0.025, 0.05, 1), eta)
esnorm(theta0, tau, sigma, alpha, eta)
```

**Arguments**

- `x`: vector of quantiles.
- `theta0`: vector of means.
- `tau`: vector of heterogeneity parameters.
- `sigma`: vector of study standard deviations.
- `alpha`: vector of thresholds for publication bias.
- `eta`: vector of publication probabilities, normalized to sum to 1.
- `log`: logical; If TRUE, probabilities are given as log(p).
- `n`: number of observations. If length(n) > 1, the length is taken to be the number required.

**Details**

The effect size distribution for the publication selection model is not normal, but has itself been selected for. These functions assume a normal underlying effect size distribution and one-sided selection on the effects.
Value

dsnorm gives the density, psnorm gives the distribution function, and rsnorm generates random deviates.

References


Examples

rsnorm(100, theta0 = 0, tau = 0.1, sigma = 0.1, eta = c(1, 0.5, 0.1))
Index

*Topic datasets
  dat.anderson2010, 2
  dat.baskerville2012, 3
  dat.cuddy2018, 4
  dat.dang2018, 4
  dat.motyl2017, 5

  allma (ma), 8
  cma (ma), 8

  dat.anderson2010, 2
  dat.baskerville2012, 3
  dat.cuddy2018, 4
  dat.dang2018, 4
  dat.motyl2017, 5
  dmpsnorm (mpsnorm), 10
  dphnorm (phnorm), 11
  dpsnorm (psnorm), 13
  dsnorm (snorm), 14

  esnorm (snorm), 14
  extract_eta (ExtractParameters), 6
  extract_isq (ExtractParameters), 6
  extract_tau (ExtractParameters), 6
  extract_theta (ExtractParameters), 6
  extract_theta0 (ExtractParameters), 6
  ExtractParameters, 6

  loo, 7
  loo, mafit-method, 7

  ma, 8
  mafit, 6
  mafit-class, 10
  mpsnorm, 10, 13

  phma (ma), 8
  phnorm, 11
  pmpsnorm (mpsnorm), 10
  pphnorm (phnorm), 11
  ppsnorm (psnorm), 13
  psma (ma), 8
  psnorm, 11, 13
  publipha (publipha-package), 2
  publipha-package, 2
  rmrsnorm (mpsnorm), 10
  rphnorm (phnorm), 11
  rpsnorm (psnorm), 13
  rsnorm (snorm), 14
  rstan::extract, 6
  snorm, 14