Package 'rankrate'

June 9, 2022

Title Statistical Tools for Preference Learning with Rankings and Ratings

Version 1.0.0

Description An implementation of the statistical methodology proposed by Pearce and Erosheva, ``A Unified Statistical Learning Model for Rankings and Scores with Application to Grant Panel Review" (2022), which at time of release has been accepted in the Journal of Machine Learning Research. The package provides tools for estimating parameters of a Mallows-Binomial model, the first joint statistical preference learning model for rankings and ratings. The package includes functions for simulating rankings and ratings from the model, calculating the density of Mallows-Binomial data, estimating parameters using various exact and approximate algorithms, and for obtaining approximate confidence intervals based on the nonparametric bootstrap.

License GPL-3

Encoding UTF-8

RoxygenNote 7.1.2

Imports stats, nloptr, gtools, lpSolve

NeedsCompilation no

Author Michael Pearce [aut, cre, cph] (<https://orcid.org/0000-0002-9313-271X>)

Maintainer Michael Pearce <pearce790@gmail.com>

Repository CRAN

Date/Publication 2022-06-09 08:20:05 UTC

R topics documented:

ASTAR 2
ASTAR_LP 3
ci_mb 4
dmall 5
dmb 5

fit_mb 6
FV 7
getQ 8
Greedy 9
GreedyLocal 9
kendall 10
lp_heuristic 11
phat_conditional 11
psi 12
rmall 13
rmb 13
theta_conditional 14
totalcostheuristic_MB 15
totalcostheuristic_MB_LP 16

Index **17**

ASTAR *Calculate the exact MLE of a Mallows-Binomial distribution using an A* search algorithm*

Description

This function estimates the exact MLE of a Mallows-Binomial distribution using an A* tree search algorithm proposed in Pearce and Erosheva (2022). The algorithm may be very slow when number of objects, J, exceeds 15, but is often still tractable for larger J when ranking and rating consensus among judges is strong.

Usage

```
ASTAR(Pi, X, M)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
X	Matrix of ratings, one row per judge and one column per object.
M	Numeric specifying maximum (=worst quality) integer rating.

Value

List with elements pi0 (consensus ranking MLE), p (object quality parameter MLE), theta (scale parameter MLE), and numnodes (number of nodes traversed during algorithm, a measure of computational complexity).

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
ASTAR(Pi=Pi,X=X,M=5)
```

ASTAR_LP *Calculate the exact MLE of a Mallows-Binomial distribution using an A* search algorithm based on the LP total cost heuristic*

Description

This function estimates the exact MLE of a Mallows-Binomial distribution using an A* tree search algorithm based on the LP total cost heuristic, proposed in Pearce and Erosheva (2022). This algorithm is not recommended over the faster algorithm codified in the ASTAR function of this package, but is included for replicability of the paper. The algorithm may be very slow when number of objects, J, exceeds 15, but is often still tractable for larger J when ranking and rating consensus among judges is strong.

Usage

```
ASTAR_LP(Pi, X, M)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
X	Matrix of ratings, one row per judge and one column per object.
M	Numeric specifying maximum (=worst quality) integer rating.

Value

List with elements pi0 (consensus ranking MLE), p (object quality parameter MLE), theta (scale parameter MLE), and numnodes (number of nodes traversed during algorithm, a measure of computational complexity).

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
ASTAR_LP(Pi=Pi,X=X,M=5)
```

ci_mb *Calculate confidence intervals for Mallows-Binomial parameters based on the nonparametric bootstrap*

Description

This function calculates confidence intervals for parameters in a Mallows-Binomial model using the nonparametric bootstrap.

Usage

```
ci_mb(
  Pi,
  X,
  M,
  interval = 0.9,
  nsamples = 100,
  all = FALSE,
  method = c("ASTAR", "ASTAR_LP", "Greedy", "GreedyLocal", "FV"),
  localsearch = 0
)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
X	Matrix of ratings, one row per judge and one column per object.
M	Numeric specifying maximum (=worst quality) integer rating.
interval	Numeric between 0 and 1 specifying the confidence interval (e.g., .90 indicates a 90% confidence interval). Defaults to 0.90.
nsamples	Numeric indicating desired number of bootstrap samples to be used when calculating confidence intervals. Defaults to 100.
all	Boolean indicating if estimated parameters from all bootstrap samples should be returned. Defaults to FALSE.
method	String indicating which estimation method to use when estimating parameters.
localsearch	Numeric for use with the Greedy and FV methods; ignored otherwise. See documentation of those estimation functions for details. Defaults to 0, indicating no local search.

Value

List with elements ci (matrix of confidence intervals for Mallows-Binomial parameters; always returned), bootstrap_pi0 (matrix of bootstrap consensus rankings; returned only if all=TRUE), and bootstrap_ptheta (matrix of bootstrap estimates of (p,theta); returned only if all=TRUE).

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
ci_mb(Pi=Pi,X=X,M=5,method="FV",interval=0.95,nsamples=20,all=TRUE,localsearch=1)
```

---

dmall *Mallows density function*

Description

This function calculates the density of observation(s) under a Mallows distribution.

Usage

```
dmall(Pi, pi0, theta, log = FALSE)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
pi0	Vector specifying the consensus (modal probability) ranking.
theta	Numeric specifying the Mallows scale parameter.
log	Boolean indicating if loglikelihood should be returned.

Value

(Log) likelihood of rankings under a Mallows distribution.

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
dmall(Pi=Pi,pi0=c(2,1,3,4),theta=1,log=TRUE)
```

---

dmb *Mallows-Binomial density function*

Description

This function calculates the density of observation(s) under a Mallows-Binomial distribution.

Usage

```
dmb(Pi, X, p, pi0 = NULL, theta, M, log = FALSE)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
X	Matrix of ratings, one row per judge and one column per object.
p	Vector of object qualities.
pi0	Vector specifying the consensus (modal probability) ranking (useful to break ties in p).
theta	Numeric specifying the Mallows scale parameter.
M	Numeric specifying maximum (=worst quality) integer rating.
log	Boolean indicating if loglikelihood should be returned.

Value

(Log) likelihood of rankings under a Mallows-Binomial distribution.

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
dmb(Pi=Pi,X=X,p=c(.1,.2,.5,.9),theta=1.1,M=5,log=TRUE)
dmb(Pi=Pi,X=X,p=c(.1,.2,.5,.9),theta=2,M=5,log=TRUE)
```

---

fit_mb *Calculate the exact or approximate MLE of a Mallows-Binomial distribution using various methods*

Description

This function calculates the exact or approximate MLE of a Mallows-Binomial distribution using a user-specified method.

Usage

```
fit_mb(
  Pi,
  X,
  M,
  method = c("ASTAR", "ASTAR_LP", "Greedy", "GreedyLocal", "FV"),
  localsearch = 0
)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
X	Matrix of ratings, one row per judge and one column per object.
M	Numeric specifying maximum (=worst quality) integer rating.
method	String specifying method, with allowable options "ASTAR","ASTAR_LP","Greedy","GreedyLocal",and "FV".
localsearch	Numeric for use with the Greedy and FV methods; ignored otherwise. See documentation of those estimation functions for details. Defaults to 0, indicating no local search.

Value

List with elements pi0 (estimated consensus ranking MLE), p (estimated object quality parameter MLE), theta (estimated scale parameter MLE), and numnodes (number of nodes traversed during algorithm, a measure of computational complexity).

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
fit_mb(Pi=Pi,X=X,M=5,method="ASTAR")
fit_mb(Pi=Pi,X=X,M=5,method="Greedy",localsearch=0)
fit_mb(Pi=Pi,X=X,M=5,method="GreedyLocal")
fit_mb(Pi=Pi,X=X,M=5,method="FV",localsearch=3)
```

FV *Estimate the MLE of a Mallows-Binomial distribution using the FV method*

Description

This function estimates the MLE of a Mallows-Binomial distribution using the FV method.

Usage

```
FV(Pi, X, M, localsearch = 0)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
X	Matrix of ratings, one row per judge and one column per object.
M	Numeric specifying maximum (=worst quality) integer rating.
localsearch	Numberic specifying the maximum Kendall distance to the first-estimated consensus ranking of rankings which should be considered during a post-hoc local search; defaults to 0 (indicating no local search).

Value

List with elements pi0 (estimated consensus ranking MLE), p (estimated object quality parameter MLE), theta (estimated scale parameter MLE), and numnodes (number of nodes traversed during algorithm, a measure of computational complexity).

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
FV(Pi=Pi,X=X,M=5)
FV(Pi=Pi,X=X,M=5,localsearch=2)
```

---

getQ *Calculate Q Matrix*

Description

This function calculates the Q matrix given a collection of (partial) rankings. For use in estimation functions, but not likely on its own.

Usage

```
getQ(Pi, J)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
J	Total number of objects assessed.

Value

Matrix Q of dimensions J x J.

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
getQ(Pi=Pi,J=4)
```

`Greedy` *Estimate the MLE of a Mallows-Binomial distribution using the Greedy method*

Description

This function estimates the MLE of a Mallows-Binomial distribution using the Greedy method, as described in Pearce and Erosheva (2022).

Usage

```
Greedy(Pi, X, M, localsearch = 0)
```

Arguments

`Pi`	Matrix of partial or complete rankings, one row per ranking.
`X`	Matrix of ratings, one row per judge and one column per object.
`M`	Numeric specifying maximum (=worst quality) integer rating.
`localsearch`	Numberic specifying the maximum Kendall distance to the first-estimated consensus ranking of rankings which should be considered during a post-hoc local search; defaults to 0 (indicating no local search).

Value

List with elements pi0 (estimated consensus ranking MLE), p (estimated object quality parameter MLE), theta (estimated scale parameter MLE), and numnodes (number of nodes traversed during algorithm, a measure of computational complexity).

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
Greedy(Pi=Pi,X=X,M=5)
Greedy(Pi=Pi,X=X,M=5,localsearch=2)
```

`GreedyLocal` *Estimate the MLE of a Mallows-Binomial distribution using the GreedyLocal method*

Description

This function estimates the MLE of a Mallows-Binomial distribution using the GreedyLocal method, as described in Pearce and Erosheva (2022). The method is identical to the Greedy method but includes an automatic and targeted post-hoc local search.

Usage

```
GreedyLocal(Pi, X, M)
```

Arguments

Pi	Matrix of partial or complete rankings, one row per ranking.
X	Matrix of ratings, one row per judge and one column per object.
M	Numeric specifying maximum (=worst quality) integer rating.

Value

List with elements pi0 (estimated consensus ranking MLE), p (estimated object quality parameter MLE), theta (estimated scale parameter MLE), and numnodes (number of nodes traversed during algorithm, a measure of computational complexity).

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
GreedyLocal(Pi=Pi,X=X,M=5)
```

---

kendall *Kendall's tau function*

Description

This function calculates Kendall's tau distance between a (partial) ranking pi and a complete ranking pi0.

Usage

```
kendall(pi, pi0)
```

Arguments

pi	A partial or complete ranking.
pi0	A complete ranking.

Value

Numeric Kendall's tau distance between pi and pi0.

Examples

```
kendall(pi=c(1,2,3),pi0=c(4,3,2,1))
kendall(pi=c(1,2,3),pi0=c(1,3,2))
```

lp_heuristic *Calculate the LP heuristic of a Mallows-Binomial model*

Description

This function calculates the LP heuristic of a Mallows-Binomial model, for use during an A* tree search for the MLE of a Mallows-Binomial model.

Usage

```
lp_heuristic(Q, Pi, I, J, order)
```

Arguments

Q	Matrix of dimension J x J.
Pi	Matrix of partial or complete rankings, one row per ranking.
I	Numeric specifying number of judges
J	Numeric specifying number of objects
order	Vector specifying a top-r or complete ordering of the desired p vector.

Value

Numeric specifying the LP heuristic.

phat_conditional *Estimate phat in a Mallows-Binomial given an order constraint*

Description

This function calculates the MLE of p in a Mallows-Binomial(p,theta) distribution given Order(p).

Usage

```
phat_conditional(X, M, order)
```

Arguments

X	Matrix of ratings, one row per judge and one column per object.
M	Numeric specifying maximum (=worst quality) integer rating.
order	Vector specifying a top-r or complete ordering of the desired p vector.

Value

Vector of length J that is the MLE of p given Order(p) and X.

Examples

```
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
phat_conditional(X=X,M=5,order=c(1,2,3,4))
phat_conditional(X=X,M=5,order=c(2,1))
```

---

psi *Psi function*

Description

This function calculates the normalizing constant of a Mallows distribution under the Kendall distance.

Usage

```
psi(theta, J, R, log = FALSE)
```

Arguments

theta	Non-negative scale parameter.
J	Positive integer indicating total number of objects.
R	Positive integer <=J indicating size of partial ranking.
log	Boolean indicating if log(Psi) should be returned.

Value

Numeric representing normalizing constant of a Mallows distribution.

Examples

```
psi(theta=1,J=10,R=8)
psi(theta=2,J=3,R=3,log=TRUE)
```

rmall *Random Mallows generation.*

Description

This function randomly generates rankings from a Mallows distribution.

Usage

```
rmall(I, pi0, theta, R = length(pi0))
```

Arguments

I	Numeric indicating the number of observations to be drawn.
pi0	Vector specifying the consensus (modal probability) ranking.
theta	Numeric specifying the Mallows scale parameter.
R	Numeric specifying the length of the (partial) rankings to be drawn.

Value

Matrix of rankings, one row per ranking.

Examples

```
rmall(I=5,pi0=1:5,theta=1,R=3)
rmall(I=10,pi0=1:5,theta=1)
```

rmb *Random Mallows-Binomial generation.*

Description

This function randomly generates rankings and ratings from a Mallows-Binomial distribution.

Usage

```
rmb(I, p, pi0 = NULL, theta, M, R = length(p))
```

Arguments

I	Numeric indicating the number of observations to be drawn.
p	Vector specifying the underlying object qualities.
pi0	Vector specifying the consensus (modal probability) ranking; should be used only for tie-breaking equal values in p.
theta	Numeric specifying the Mallows scale parameter.
M	Numeric specifying the maximum integer rating.
R	Numeric specifying the length of the (partial) rankings to be drawn.

Value

List containing elements X (I x J matrix of ratings) and Pi (I x R matrix of rankings).

Examples

```
rmb(I=5,p=c(.1,.3,.4,.7,.9),theta=1,M=10)
rmb(I=10,p=c(.1,.3,.3,.7,.9),pi0=c(1,3,2,4,5),theta=5,M=40,R=3)
```

---

theta_conditional *Estimate thetahat in a Mallows-Binomial given a constraint*

Description

This function calculates the MLE of theta in a Mallows-Binomial(p,theta) distribution given a constraint. The constraint is a partial or complete central ranking.

Usage

```
theta_conditional(Pi, order = NULL, D = NULL, J = NULL)
```

Arguments

Pi	Matrix of rankings, with one row per ranking
order	Vector specifying a complete ordering of the desired p vector. Either order or D must be specified; if both, use the order.
D	Numeric specifying a minimum total kendall distance between the observed rankings and a supposed central ranking. Either order or D must be specified; if both, use the order.
J	Numeric specifying the total number of objects to be assessed. Not used if order is specified.

Value

Numeric specifying the MLE of theta under a constraint.

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
theta_conditional(Pi=Pi,order=c(1,2,4,3))
theta_conditional(Pi=Pi,D=3,J=4)
```

`totalcostheuristic_MB` *Calculate the naive total cost heuristic of a Mallows-Binomial model*

Description

This function calculates the total cost heuristic of a Mallows-Binomial model given Q, Pi, X, M, and an order, for use during an A* tree search for the MLE of a Mallows-Binomial model.

Usage

```
totalcostheuristic_MB(Q, Pi, X, M, order)
```

Arguments

`Q`	Matrix of dimension J x J.
`Pi`	Matrix of partial or complete rankings, one row per ranking.
`X`	Matrix of ratings, one row per judge and one column per object.
`M`	Numeric specifying maximum (=worst quality) integer rating.
`order`	Vector specifying a top-r or complete ordering of the desired p vector.

Value

Numeric specifying the total cost heuristic.

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
Q <- matrix(c(0,.5,0,0,.5,0,0,0,1,1,0,0,1,1,.5,0),nrow=4,ncol=4)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
totalcostheuristic_MB(Q=Q,Pi=Pi,X=X,M=5,order=c(1,2))
totalcostheuristic_MB(Q=Q,Pi=Pi,X=X,M=5,order=c(2,1,4,3))
```

totalcostheuristic_MB_LP *Calculate the LP total cost heuristic of a Mallows-Binomial model*

Description

This function calculates the LP total cost heuristic of a Mallows-Binomial model given Q, Pi, X, M, and an order, for use during an A* tree search for the MLE of a Mallows-Binomial model.

Usage

```
totalcostheuristic_MB_LP(Q, Pi, X, M, order)
```

Arguments

Q	Matrix of dimension J x J.
Pi	Matrix of partial or complete rankings, one row per ranking.
X	Matrix of ratings, one row per judge and one column per object.
M	Numeric specifying maximum (=worst quality) integer rating.
order	Vector specifying a top-r or complete ordering of the desired p vector.

Value

Numeric specifying the LP total cost heuristic.

Examples

```
Pi <- matrix(c(1,2,3,4,2,1,NA,NA),byrow=TRUE,nrow=2)
Q <- matrix(c(0,.5,0,0,.5,0,0,0,1,1,0,0,1,1,.5,0),nrow=4,ncol=4)
X <- matrix(c(0,1,2,3,1,2,2,5),byrow=TRUE,nrow=2)
totalcostheuristic_MB_LP(Q=Q,Pi=Pi,X=X,M=5,order=c(1,2))
totalcostheuristic_MB_LP(Q=Q,Pi=Pi,X=X,M=5,order=c(2,1,4,3))
```

Index

ASTAR, 2
ASTAR_LP, 3

ci_mb, 4

dmall, 5
dmb, 5

fit_mb, 6
FV, 7

getQ, 8
Greedy, 9
GreedyLocal, 9

kendall, 10

lp_heuristic, 11

phat_conditional, 11
psi, 12

rmall, 13
rmb, 13

theta_conditional, 14
totalcostheuristic_MB, 15
totalcostheuristic_MB_LP, 16