Package 'harmonicmeanp'

October 13, 2022

Type Package

Title Harmonic Mean p-Values and Model Averaging by Mean Maximum
Likelihood

Version 3.0

Date 2019-08-17

Author Daniel J. Wilson

Maintainer Daniel Wilson <hmp.R.package@gmail.com>

Description The harmonic mean p-value (HMP) test combines p-values and corrects for multiple test-
ing while controlling the strong-sense family-wise error rate. It is more powerful than com-
mon alternatives including Bonferroni and Simes procedures when combining large propor-
tions of all the p-values, at the cost of slightly lower power when combining small propor-
tions of all the p-values. It is more stringent than controlling the false discovery rate, and pos-
sesses theoretical robustness to positive correlations between tests and un-
equal weights. It is a multi-level test in the sense that a superset of one or more signifi-
cant tests is certain to be significant and conversely when the superset is non-significant, the con-
stituent tests are certain to be non-significant. It is based on MAMML (model averag-
ing by mean maximum likelihood), a frequentist analogue to Bayesian model averag-
ing, and is theoretically grounded in generalized central limit theorem. For detailed exam-
ples type vignette(``harmonicmeanp'') after installation. Version 3.0 addresses errors in ver-
sions 1.0 and 2.0 that led function p.hmp to control the familywise er-
ror rate only in the weak sense, rather than the strong sense as intended.

License GPL-3

Depends FMStable

Suggests knitr, ape, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2019-08-19 18:00:05 UTC

R topics documented:

```
Asymptotically Exact Harmonic Mean p-Value
```
Asymptotically Exact Harmonic Mean p-Value

Description

Compute a combined *p*-value via the asymptotically exact harmonic mean *p*-value. The harmonic mean *p*-value (HMP) test combines *p*-values and corrects for multiple testing while controlling the strong-sense family-wise error rate. It is more powerful than common alternatives including Bonferroni and Simes procedures when combining large proportions of all the *p*-values, at the cost of slightly lower power when combining small proportions of all the *p*-values. It is more stringent than controlling the false discovery rate, and possesses theoretical robustness to positive correlations between tests and unequal weights. It is a multi-level test in the sense that a superset of one or more significant tests is certain to be significant and conversely when the superset is non-significant, the constituent tests are certain to be non-significant. It is based on MAMML (model averaging by mean maximum likelihood), a frequentist analogue to Bayesian model averaging, and is theoretically grounded in generalized central limit theorem.

Usage

```
p.hmp(p, w = NULL, L = NULL, w.sum.tolerance = 1e-6, multilevel = TRUE)
```

Arguments

p	A numeric vector of one or more *p*-values to be combined. Missing values (NAs) will cause a missing value to be returned.
w	An optional numeric vector of weights that can be interpreted as incorporating information about prior model probabilities and power of the tests represented by the individual *p*-values. The sum of the weights cannot exceed one but may be less than one, which is interpreted as meaning that some of the L *p*-values have been excluded.
L	The number of constituent *p*-values. This is mandatory when using p.hmp as part of a multilevel test procedure and needs to be set equal to the total number of individual *p*-values investigated, which may be (much) larger than the length of the argument p. If ignored, it defaults to the length of argument p, with a warning.
w.sum.tolerance	
	Tolerance for checking that the weights do not exceed 1.

multilevel	Logical, indicating whether the test is part of a multilevel procedure involving a total of L *p*-values intended to control the strong-sense familywise error rate (TRUE, default), or whether a stand-alone test is to be produced (FALSE), in which case L is ignored.

Value

When `multilevel==TRUE` an asymptotically exact combined *p*-value is returned, equivalent to `sum(w)*pharmonicmeanp(hmp.stat(p,w)/sum(w),L)`. This *p*-value can be compared against threshold `sum(w)*alpha` to control the strong-sense familywise error rate at level `alpha`. A test based on this asymptotically exact harmonic mean *p*-value is equivalent to comparison of the 'raw' harmonic mean *p*-value calculated by `hmp.stat(p,w)` to a 'harmonic mean *p*-value threshold' `sum(w)*qharmonicmeanp(alpha,L)`.

When `multilevel==FALSE` an asymptotically exact combined *p*-value is returned, equivalent to `pharmonicmeanp(hmp.stat(p,w),length(p))`. L is ignored and w is normalized to sum to 1. This *p*-value can be compared against threshold `alpha` to control the per-test error rate at level `alpha`. A test based on this asymptotically exact harmonic mean *p*-value is equivalent to comparison of the 'raw' harmonic mean *p*-value calculated by `hmp.stat(p,w)` to a 'harmonic mean *p*-value threshold' `qharmonicmeanp(alpha,L)`.

Author(s)

Daniel J. Wilson

References

Daniel J. Wilson (2019) The harmonic mean *p*-value for combining dependent tests. *Proceedings of the National Academy of Sciences USA* 116: 1195-1200.

See Also

hmp.stat

Examples

```
# For detailed examples type vignette("harmonicmeanp")
# Example: simulate from a non-uniform distribution mildly enriched for small \emph{p}-values.
# Compare the significance of the combined p-value for Bonferroni, Benjamini-Hochberg (i.e. Simes),
# HMP and (equivalently) MAMML with 2 degrees of freedom.
L = 1000
p = rbeta(L,1/1.8,1)
min(p.adjust(p,"bonferroni"))
min(p.adjust(p,"BH"))
p.hmp(p,L=L)
p.mamml(1/p,2,L=L)

# Multilevel test: find significant subsets of the 1000 p-values by comparing overlapping
# subsets of size 100 and 10 in addition to the top-level test of all 1000, while maintaining
# the strong-sense family-wise error rate.
p.100 = sapply(seq(1,L-100,by=10),function(beg) p.hmp(p[beg+0:99],L=L))
plot(-log10(p.100),xlab="Test index",main="Moving average",
  ylab="Asymptotically exact combined -log10 p-value",type="o")
```

```
# The appropriate threshold is alpha (e.g. 0.05) times the proportion of tests in each p.100
abline(h=-log10(0.05*100/1000),col=2,lty=2)

p.10 = sapply(seq(1,L-10,by=5),function(beg) p.hmp(p[beg+0:9],L=L))
plot(-log10(p.10),xlab="Test index",main="Moving average",
  ylab="Asymptotically exact combined -log10 p-value",type="o")
# The appropriate threshold is alpha (e.g. 0.05) times the proportion of tests in each p.10
abline(h=-log10(0.05*10/1000),col=2,lty=2)
```

Compute the Harmonic Mean p-Value
                  Compute the Harmonic Mean p-*Value*

Description

The harmonic mean p-value (HMP) is defined as the inverse of the (possibly weighted) arithmetic mean of the inverse p-values. When the HMP is small (e.g. less than 0.05), it is approximately well-calibrated, meaning that it can be directly interpreted. However, the function p.hmp calculates an asymptotically exact p-value from the HMP and is preferred.

Usage

```
hmp.stat(p, w = NULL)
```

Arguments

p	A numeric vector of one or more p-values. Missing values (NAs) will cause a missing value to be returned.
w	An optional numeric vector of weights that can be interpreted as prior model probabilities for each of the alternative hypotheses represented by the individual p-values. The sum of the weights cannot exceed one but may be less than one, which is interpreted as meaning that some p-values have been excluded.

Value

The harmonic mean p-value is returned.

Author(s)

Daniel J. Wilson

References

Daniel J. Wilson (2019) The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences USA* 116: 1195-1200.

See Also

p.hmp

Examples

```
# For detailed examples type vignette("harmonicmeanp")
p = rbeta(1000,1/1.5,1)
hmp.stat(p)
p.hmp(p,L=1000)
```

Compute the Model-Averaged Mean Maximized Likelihood

Compute the Model-Averaged Mean Maximized Likelihood

Description

The model-averaged mean maximized likelihood (MAMML) is defined as the (possibly weighted) arithmetic mean of the maximized likelihood ratios from a series of likelihood ratio tests comparing mutually exclusive alternative hypotheses with the same nested null hypothesis based on the exact same data.

Usage

```
mamml.stat(R, w = NULL)
```

Arguments

R	A numeric vector of one or more maximized likelihood ratios. Missing values (NAs) will cause a missing value to be returned.
w	An optional numeric vector of weights that can be interpreted as prior model probabilities for each of the alternative hypotheses represented by the individual *p*-values. The sum of the weights cannot exceed one but may be less than one, which is interpreted as meaning that some *p*-values have been excluded.

Value

The model-averaged mean maximized likelihood ratio is returned.

Author(s)

Daniel J. Wilson

References

Daniel J. Wilson (2019) The harmonic mean *p*-value for combining dependent tests. *Proceedings of the National Academy of Sciences USA* 116: 1195-1200.

See Also

p.mamml

Examples

```
# For detailed examples type vignette("harmonicmeanp")
nu = 3
R = exp(0.5*rchisq(1000,nu))
mamml.stat(R)
p.mamml(R,nu,L=1000)
```

Harmonic Mean p-Value Asymptotic Distribution

The Harmonic Mean p-Value Asymptotic Distribution

Description

Density, distribution function, quantile function and random number generation for the harmonic mean of *L* *p*-values under their null hypotheses, i.e. the harmonic mean of *L* standard uniform random variables, assuming *L* is large.

Usage

```
dharmonicmeanp(x, L, log=FALSE)
pharmonicmeanp(x, L, log=FALSE, lower.tail=TRUE)
qharmonicmeanp(p, L, log=FALSE, lower.tail=TRUE)
rharmonicmeanp(n, L)
```

Arguments

x	The value or vector of values of the harmonic mean *p*-value, for example calculated from data using function `hmp.stat`.
L	The number of constituent *p*-values used in calculating each value of x.
log	If true the log probability is output.
lower.tail	If true (the default) the lower tail probability is returned. Otherwise the upper tail probability.
p	The value or vector of values, between 0 and 1, of the probability specifying the quantile for which to return the harmonic mean *p*-value.
n	The number of values to simulate.

Value

`dharmonicmeanp` produces the density, `pharmonicmeanp` the tail probability, `qharmonicmeanp` the quantile and `rharmonicmeanp` random variates for the harmonic mean of *L* *p*-values when their null hypotheses are true.

Use `qharmonicmeanp(alpha,L)` to calculate α_L, the 'harmonic mean *p*-value threshold', as in Table 1 of Wilson (2019, corrected), where L is the total number of *p*-values under consideration and `alpha` is the intended strong-sense familywise error rate.

Author(s)

Daniel J. Wilson

References

Daniel J. Wilson (2019) The harmonic mean *p*-value for combining dependent tests. *Proceedings of the National Academy of Sciences USA* 116: 1195-1200.

See Also

p.hmp

Examples

```
# For a detailed tutorial type vignette("harmonicmeanp")
# Example: simulate from a non-uniform distribution mildly enriched for small \emph{p}-values.
# Compare the significance of the combined p-value for Bonferroni, Benjamini-Hochberg (i.e. Simes),
# HMP and (equivalently) MAMML with 2 degrees of freedom.
L = 1000
p = rbeta(L,1/1.5,1)
min(p.adjust(p,"bonferroni"))
min(p.adjust(p,"BH"))
x = hmp.stat(p)
pharmonicmeanp(x,length(p))
p.hmp(p,L=L)
p.mamml(1/p,2,L=L)

# Compute critical values for the HMP from asymptotic theory and compare to direct simulations
L = 100
alpha = 0.05
(hmp.crit = qharmonicmeanp(alpha,L))
nsim = 100000
p.direct = matrix(runif(L*nsim),nsim,L)
hmp.direct = apply(p.direct,1,hmp.stat)
(hmp.crit.sim = quantile(hmp.direct,alpha))

# Compare HMP of p-values simulated directly, and via the asymptotic distribution,
# to the asymptotic density
L = 30
nsim = 10000
p.direct = matrix(runif(L*nsim),nsim,L)
hmp.direct = apply(p.direct,1,hmp.stat)
hmp.asympt = rharmonicmeanp(nsim,L)
h = hist(hmp.direct,60,col="green3",prob=TRUE,main="Distributions of harmonic mean p-values")
hist(hmp.asympt,c(-Inf,h$breaks,Inf),col="yellow2",prob=TRUE,add=TRUE)
hist(hmp.direct,60,col=NULL,prob=TRUE,add=TRUE)
curve(dharmonicmeanp(x,L),lwd=2,col="red3",add=TRUE)
legend("topright",c("Direct simulation","Asymptotic simulation","Asymptotic density"),
  fill=c("green3","yellow2",NA),col=c(NA,NA,"red3"),lwd=c(NA,NA,2),bty="n",border=c(1,1,NA))
```

harmonicmeanp	*Harmonic mean p-values and model averaging by mean maximum likelihood*

Description

The harmonic mean p-value (HMP) test combines p-values and corrects for multiple testing while controlling the strong-sense family-wise error rate. It is more powerful than common alternatives including Bonferroni and Simes procedures when combining large proportions of all the p-values, at the cost of slightly lower power when combining small proportions of all the p-values. It is more stringent than controlling the false discovery rate, and possesses theoretical robustness to positive correlations between tests and unequal weights. It is a multi-level test in the sense that a superset of one or more significant tests is certain to be significant and conversely when the superset is non-significant, the constituent tests are certain to be non-significant. It is based on MAMML (model averaging by mean maximum likelihood), a frequentist analogue to Bayesian model averaging, and is theoretically grounded in generalized central limit theorem.

Details

Package:	harmonicmeanp
Type:	Package
Version:	3.0
Date:	2019-08-17
License:	GPL-3

The key function is p.hmp for combining p-values using the HMP. Type vignette("harmonicmeanp") for detailed examples.

Author(s)

Daniel J. Wilson

Maintainer: Daniel Wilson <hmp.R.package@gmail.com>

References

Daniel J. Wilson (2019) The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences USA* 116: 1195-1200.

See Also

Package FMStable

Examples

```
# For detailed examples type vignette("harmonicmeanp")
```

```
# Example: simulate from a non-uniform distribution mildly enriched for small \emph{p}-values.
# Compare the significance of the combined p-value for Bonferroni, Benjamini-Hochberg (i.e. Simes),
# HMP and (equivalently) MAMML with 2 degrees of freedom.
p = rbeta(1000,1/1.5,1)
min(p.adjust(p,"bonferroni"))
min(p.adjust(p,"BH"))
p.hmp(p,L=1000)
p.mamml(1/p,2,L=1000)
```

Landau Distribution   *The Landau Distribution*

Description

Density, distribution function, quantile function and random number generation for the Landau distribution with location parameter mu and scale parameter sigma.

Usage

```
dLandau(x, mu=log(pi/2), sigma=pi/2, log=FALSE)
pLandau(x, mu=log(pi/2), sigma=pi/2, log=FALSE, lower.tail=TRUE)
qLandau(p, mu=log(pi/2), sigma=pi/2, log=FALSE, lower.tail=TRUE)
rLandau(n, mu=log(pi/2), sigma=pi/2)
```

Arguments

x	The value or vector of values of the Landau-distributed random variable.
mu	The location parameter of the Landau distribution. Defaults to log(pi/2) to give Landau's original distribution.
sigma	The scale parameter of the Landau distribution. Defaults to pi/2 to give Landau's original distribution.
log	If true the log probability is output.
lower.tail	If true (the default) the lower tail probability is returned. Otherwise the upper tail probability.
p	The value or vector of values, between 0 and 1, of the probability specifying the quantile for which to return the Landau random variable x.
n	The number of values to simulate.

Details

The density of the Landau distribution can be written

$$f(x) = \frac{1}{\pi\,\sigma} \int_0^\infty \exp\left(-t\frac{(x-\mu)}{\sigma} - \frac{2}{\pi}t\log(t)\right) \sin(2t)\,\mathrm{d}t$$

Value

dLandau produces the density, pLandau the tail probability, qLandau the quantile and rLandau random variates for the Landau distribution.

Author(s)

Daniel J. Wilson

References

Landau LD (1944) On the energy loss of fast particles by ionization. *J Phys USSR* 8:201-205.

Daniel J. Wilson (2019) The harmonic mean *p*-value for combining dependent tests. *Proceedings of the National Academy of Sciences USA* 116: 1195-1200.

See Also

p.hmp

Examples

```
# For detailed examples type vignette("harmonicmeanp")
# Example: simulate from a non-uniform distribution mildly enriched for small \emph{p}-values.
# Compare the significance of the combined p-value for Bonferroni, Benjamini-Hochberg (i.e. Simes),
# HMP and (equivalently) MAMML with 2 degrees of freedom.
L = 1000
p = rbeta(L,1/1.5,1)
min(p.adjust(p,"bonferroni"))
min(p.adjust(p,"BH"))
x = hmp.stat(p)
pLandau(1/x,log(length(p))+(1 + digamma(1) - log(2/pi)),pi/2,lower.tail=FALSE)
p.hmp(p,L=L)
p.mamml(1/p,2,L=L)
```

Model-Averaged Mean Maximized Likelihood Ratio Asymptotic Distribution

Model-Averaged Mean Maximized Likelihood Ratio Asymptotic Distribution

---

Description

Density, distribution function, quantile function and random number generation for the mean of *L* maximized likelihood ratios under their null hypotheses, i.e. the mean of *L* Pareto(1,1) variables, assuming *L* is large.

Usage

```
dmamml(x, L, df, log=FALSE)
pmamml(x, L, df, log=FALSE, lower.tail=TRUE)
qmamml(p, L, df, log=FALSE, lower.tail=TRUE, xmin=1+1e-12, xmax=1e12)
rmamml(n, L, df)
```

Arguments

x	The value or vector of values of the mean maximized likelihood ratios, for example calculated from data using function `mamml.stat`.
L	The number of constituent likelihood ratios used in calculating each value of x.
df	Degrees of freedom of the constituent likelihood ratios, assumed all equal for each value of x.
log	If true the log probability is output.
lower.tail	If true (the default) the lower tail probability is returned. Otherwise the upper tail probability.
p	The value or vector of values, between 0 and 1, of the probability specifying the quantile for which to return the mean maximized likelihood ratio.
xmin, xmax	The range of values of the mean maximized likelihood ratio over which to search for the quantile specified by p, used by the numerical root finding algorithm. Must bracket the solution.
n	The number of values to simulate.

Value

dmamml produces the density, pmamml the tail probability, qmamml the quantile and rmamml random variates for the model-averaged mean maximized likelihood ratios when their null hypotheses are true.

Author(s)

Daniel J. Wilson

References

Daniel J. Wilson (2019) The harmonic mean *p*-value for combining dependent tests. *Proceedings of the National Academy of Sciences USA* 116: 1195-1200.

See Also

p.hmp

Examples

```
# For a detailed tutorial type vignette("harmonicmeanp")
# Example: simulate from a gamma distribution mildly enriched for large likelihood ratios.
# Compare the significance of the combined p-value for Bonferroni, Benjamini-Hochberg (i.e. Simes),
# MAMML with 4 degrees of freedom.
L = 100
df = 4
enrich = 1.5
y = exp(0.5*rgamma(L,shape=df/2,rate=1/2/enrich))
p = pchisq(2*log(y),df,lower.tail=FALSE)
min(p.adjust(p,"bonferroni"))
min(p.adjust(p,"BH"))
```

```
(x = mamml.stat(y))
pmamml(x,L,df,lower.tail=FALSE)
p.mamml(y,df,L=L)
# Compare to the HMP - MAMML may act more conservatively because of adjustments for its
# poorer asymptotic approximation
x = hmp.stat(p)
p.hmp(p,L=L)

# Compute critical values for MAMML from asymptotic theory and compare to direct simulations
L = 100
df = 1
alpha = 0.05
(mamml.crit = qmamml(1-alpha,L,df))
nsim = 10000
y.direct = matrix(exp(0.5*rchisq(L*nsim,df)),nsim,L)
mamml.direct = apply(y.direct,1,mamml.stat)
# mamml.crit may be more conservative than mamml.crit.sim because of adjustments for its
# poorer asymptotic approximation
(mamml.crit.sim = quantile(mamml.direct,1-alpha))

# Compare MAMML simulated directly, and via the asymptotic distribution, to the asymptotic density
# Works best for df = 2
L = 30
df = 3
nsim = 1000
y.direct = matrix(exp(0.5*rchisq(L*nsim,df)),nsim,L)
mamml.direct = apply(y.direct,1,mamml.stat)
xmax = quantile(mamml.direct,.95)
h = hist(mamml.direct,c(-Inf,seq(0,xmax,len=60),Inf),col="green3",prob=TRUE,
  main="Distributions of MAMML",xlim=c(1,xmax))
# Slow because rmamml calls qmamml which uses numerical root finding
# mamml.asympt = rmamml(nsim,L,df)
# hist(mamml.asympt,c(-Inf,seq(0,xmax,len=60),Inf),col="yellow2",prob=TRUE,add=TRUE)
hist(mamml.direct,c(-Inf,seq(0,xmax,len=60),Inf),col=NULL,prob=TRUE,add=TRUE)
curve(dmamml(x,L,df),lwd=2,col="red3",add=TRUE)
legend("topright",c("Direct simulation","Asymptotic density"),
  fill=c("green3",NA),col=c(NA,"red3"),lwd=c(NA,2),bty="n",border=c(1,NA))
```

---

p.mamml	*Compute a combined* p-*value via the model-averaged mean maximized likelihood ratio*

Description

The model averaging by mean maximum likelihood (MAMML) test combines likelihood ratio tests and corrects for multiple testing while controlling the weak-sense family-wise error rate in a way that is more powerful than common alternatives including Bonferroni and Simes procedures and possesses theoretical robustness to positive correlations between tests and unequal weights. It is a frequentist analogue to Bayesian model averaging, is theoretically grounded in generalized central limit theorem, and motivates the simpler and better-calibrated harmonic mean *p*-value (HMP) test.

The model-averaged mean maximized likelihood (MAMML) is defined as the (possibly weighted) arithmetic mean of the maximized likelihood ratios from a series of likelihood ratio tests comparing mutually exclusive alternative hypotheses with the same nested null hypothesis based on the exact same data.

Usage

```
p.mamml(R, nu, w = NULL, L = NULL)
```

Arguments

R	A numeric vector of one or more maximized likelihood ratios. Missing values (NAs) will cause a missing value to be returned.
nu	A numeric scalar or vector for the degrees of freedom corresponding to all or each of the maximized likelihood ratios respectively.
w	An optional numeric vector of weights that can be interpreted as prior model probabilities for each of the alternative hypotheses represented by the individual *p*-values. The sum of the weights cannot exceed one but may be less than one, which is interpreted as meaning that some *p*-values have been excluded.
L	The number of constituent maximized likelihood ratios. If ignored, it defaults to the length of argument R, with a warning.

Value

The model-averaged mean maximized likelihood ratio is returned.

Author(s)

Daniel J. Wilson

References

Daniel J. Wilson (2019) The harmonic mean *p*-value for combining dependent tests. *Proceedings of the National Academy of Sciences USA* 116: 1195-1200.

See Also

mamml.stat, hmp.stat, p.hmp

Examples

```
# For detailed examples type vignette("harmonicmeanp")
nu = 3
R = exp(0.5*rchisq(1000,nu))
mamml.stat(R)
p.mamml(R,nu,L=1000)
```

Index