Package 'pwrFDR'

August 22, 2018

Version 1.95

Title FDR Power

Author Grant Izmirlian Jr <izmirlig@mail.nih.gov>

Maintainer Grant Izmirlian Jr <izmirlig@mail.nih.gov>

Imports methods, stats

Description This is a package for calculating power and sample size in multiple testing situations using the Benjamini-Hochberg (BH) false discovery rate (FDR) procedure. The package computes power and sample size in one of either two ways, using the average power or the lambda-power. See Izmirlian, G. (2018) <arXiv:1801.03989>.

License GPL (>= 2)

LazyLoad yes

NeedsCompilation yes

Repository CRAN

Date/Publication 2018-08-22 21:30:03 UTC

R topics documented:

`cCDF.SoM` *Complementary CDF for the True Positive Fraction*

Description

Computes the complementary CDF for the True Positive Fraction, S_m/M_m, via approximation based upon the asymptotic distribution.

Usage

```
cCDF.SoM(lambda, x = NULL, groups, effect.size, n.sample, r.1, FDR, N.tests,
         control)
```

Arguments

`lambda`	Argument of the complementary CDF. Result will be Pr(S_m/M_m > lambda)
`x`	In the abbreviated call sequence, the user only needs to specify `lambda`, `N.tests`, and `x`, an object of class `pwr` returned by the function `pwrFDR`.
`groups`	The number of experimental groups to compare. Default value is 2.
`effect.size`	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
`n.sample`	The number of experimental replicates.
`r.1`	The proportion of all test statistics that are distributed under HA.
`FDR`	The false discovery rate.
`N.tests`	The number of simultaneous hypothesis tests.
`control`	Optionally, a list with components with the following components: 'groups', used when distop=3 (F-dist), specifying number of groups. 'version', used only in the 'JL' method, choice 0 gives the 'JL' version as published, whereas choice 1 replaces the FDR with r.0*FDR resulting in the infinite simultaneous tests limiting average power, which is the 'Iz' version, but this is redundant because you can specify the 'Iz' method to use this option. 'tol' is a convergence criterion used in iterative methods which is set to 1e-8 by default 'max.iter' is an iteration limit, set to 1000 by default 'distop', specifying the distribution family of the central and non-centrally located sub-populations. =1 gives normal (2 groups) =2 gives t- (2 groups) and =3 gives F- (2+ groups) 'CS', correlation structure, for use only with 'method="simulation"' which will simulate m simulatenous tests with correlations 'rho' in blocks of size 'n.WC'. Specify as list CS = list(rho=0.80, n.WC=50) for example

Details

The complementary CDF for the True Positive Fraction, S_m/M_m, is approximated using its asymptotic distribution. Since

m^0.5 (S_m/M_m - average.power) –D–> N(0, sigma^2)

then

P(S_m/M_m > lambda)

~ 1 - Phi(m^0.5 (lambda - average.power)/sigma)

The approximation is reasonable as long as m sigma^2 is large enough A formula for the asymptotic variance is given in the cited manuscript. There is a user level function, var.rtm.SoM, which computes the asymptotic variance.

NOTE: the capabilities of this function are available in the function, `pwrFDR` by requesting the lambda power by specifying the argument, `lambda`

Value

An object of class "vvv" which is a list having components

`cCDF.SoM`	The result
`average.power`	The average power at the supplied arguments
`c.g`	The per test threshold that is equivalent to the BH-FDR, on the test statistic scale
`gamma`	The limiting proportion of tests that were called significant
`objective`	Result of optimization producing the average power, should be close to zero.
`err.III`	The probability mass on side of the oppositely signed alternative in two sided tests
`sigma.rtm.SoM`	The square root of the asymptotic variance of the root-m scaled true positive fraction, m^0.5 * S_m/M_m
`call`	The call which produced the result

Author(s)

Grant Izmirlian <izmirlian at nih dot gov>

References

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

See Also

`pwrFDR` `cCDF.ToJ`

Examples

```
## Example 1: Explicit call
   ccdf <- cCDF.SoM(lambda=0.8201087, effect.size=0.79, n.sample=46, r.1=2000/54675,
                    FDR=0.15, N.tests=1000)
   ccdf

## Example 2: Abbreviated call using result of pwrFDR
   rslt.avgp <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
   ccdf <- cCDF.SoM(lambda=0.8201087, x=rslt.avgp, N.tests=1000)
   ccdf
```

`cCDF.ToJ` *Complementary CDF for the False Discovery Fraction*

Description

Computes the complementary CDF for the False Discovery Fraction, T_m/J_m, via approximation based upon the asymptotic distribution.

Usage

```
cCDF.ToJ(lambda, x=NULL, groups, effect.size, n.sample, r.1, FDR, N.tests,
         control)
```

Arguments

`lambda`	Argument of the complementary CDF. Result will be Pr(S_m/M_m > lambda)
`x`	In the abbreviated call sequence, the user only needs to specify `lambda`, `N.tests`, and `x`, an object of class `pwr` returned by the function `pwrFDR`.
`groups`	The number of experimental groups to compare. Default value is 2.
`effect.size`	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
`n.sample`	The number of experimental replicates.
`r.1`	The proportion of all test statistics that are distributed under HA.
`FDR`	The false discovery rate.
`N.tests`	The number of simultaneous hypothesis tests.
`control`	Optionally, a list with components with the following components: 'groups', used when distop=3 (F-dist), specifying number of groups. 'version', used only in the 'JL' method, choice 0 gives the 'JL' version as published, whereas choice 1 replaces the FDR with r.0*FDR resulting in the infinite simultaneous tests limiting average power, which is the 'Iz' version, but this is redundant because you can specify the 'Iz' method to use this option. 'tol' is a convergence criterion used in iterative methods which is set to 1e-8 by default 'max.iter' is an

iteration limit, set to 1000 by default 'distop', specifying the distribution family of the central and non-centrally located sub-populations. =1 gives normal (2 groups) =2 gives t- (2 groups) and =3 gives F- (2+ groups) 'CS', correlation structure, for use only with 'method="simulation"' which will simulate m simulatenous tests with correlations 'rho' in blocks of size 'n.WC'. Specify as list CS = list(rho=0.80, n.WC=50) for example

Details

The complementary CDF for the False Discovery Fraction, T_m/J_m, is approximated using its asymptotic distribution. Since

m^0.5 (T_m/J_m - (1 - r.1) FDR) –D–> N(0, tau^2)

then

P(T_m/J_m > lambda)

~ 1 - Phi(m^0.5 (lambda - (1 - r.1) FDR)/tau)

The approximation is reasonable as long as m tau^2 is large enough A formula for the asymptotic variance is given in the cited manuscript. There is a user level function, var.rtm.ToJ, which computes the asymptotic variance.

Value

An object of class "vvv" which is a list having components

`cCDF.ToJ`	The result
`average.power`	The average power at the supplied arguments
`c.g`	The per test threshold that is equivalent to the BH-FDR, on the test statistic scale
`gamma`	The limiting proportion of tests that were called significant
`objective`	Result of optimization producing the average power, should be close to zero.
`err.III`	The probability mass on side of the oppositely signed alternative in two sided tests
`sigma.rtm.ToJ`	The square root of the asymptotic variance of the root-m scaled false discovery fraction, m^0.5 * T_m/J_m
`call`	The call which produced the result

Author(s)

Grant Izmirlian <izmirlian at nih dot gov>

References

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

See Also

`controlFDF` `cCDF.SoM`

Examples

```
## Example 1: Explicit call
  ccdf <- cCDF.ToJ(lambda=(1-2000/54675)*0.15, effect.size=0.79, n.sample=46, r.1=2000/54675,
                   FDR=0.15, N.tests=1000)
   ccdf

## Example 2: Abbreviated call using result of pwrFDR
   rslt.avgp <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
   ccdf <- cCDF.ToJ(lambda=(1-2000/54675)*0.15, x=rslt.avgp, N.tests=1000)
   ccdf
```

`CDF.Pval` *CDF of pooled (H0 and HA) population p-values*

Description

Computes the CDF of the pooled population p-values under the mixture model, e.g. the p-values are i.i.d. with CDF a mixture between a uniform (CDF in the null distributed population) and a concave function (CDF in the non-null distributed population).

Usage

```
CDF.Pval(u, groups, r.1, effect.size, n.sample, control)
```

Arguments

`u`	Argument of the CDF. Result will be Pr(P_i <= u)
`groups`	The number of experimental groups to compare. Default value is 2.
`r.1`	The proportion of all test statistics that are distributed under HA.
`effect.size`	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
`n.sample`	The number of experimental replicates.
`control`	Optionally, a list with components with the following components: 'groups', used when distop=3 (F-dist), specifying number of groups. 'version', used only in the 'JL' method, choice 0 gives the 'JL' version as published, whereas choice 1 replaces the FDR with r.0*FDR resulting in the infinite simultaneous tests limiting average power, which is the 'Iz' version, but this is redundant because you can specify the 'Iz' method to use this option. 'tol' is a convergence criterion used in iterative methods which is set to 1e-8 by default 'max.iter' is an iteration limit, set to 1000 by default 'distop', specifying the distribution family of the central and non-centrally located sub-populations. =1 gives normal (2 groups) =2 gives t- (2 groups) and =3 gives F- (2+ groups) 'CS', correlation structure, for use only with 'method="simulation"' which will simulate m simulatenous tests with correlations 'rho' in blocks of size 'n.WC'. Specify as list CS = list(rho=0.80, n.WC=50) for example

Details

Computes the CDF of the pooled population p-values under the mixture model, e.g. the p-values are i.i.d. with CDF a mixture between a uniform (CDF in the null distributed population) and a concave function (CDF in the non-null distributed population). If Fc_0 is the cCDF of a test statistic under H0 and Fc_A is the cCDF of a test statistic under HA then the CDF of the P-values is

G(u) = (1-r) u + r Fc_A(Fc_0^-1(u))

The limiting positve call fraction, lim_m J_m/m = gamma (a.s.) is the solution to the equation

G(gamma f) = gamma

where f is the FDR

Value

A list with components

`call`	The call which produced the result
`u`	The argument that was passed to the function
`CDF.Pval`	The value of the CDF

Author(s)

Grant Izmirlian <izmirlian at nih dot gov>

References

Genovese, C. and L. Wasserman. (2004) A stochastic process approach to false discovery control. Annals of Statistics. 32 (3), 1035-1061.

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

See Also

`CDF.Pval.HA`

Examples

```
## First calculate an average power for a given set of parameters
rslt.avgp <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)

## Now verify that G( gamma f ) = gamma

gamma <- rslt.avgp$gamma
f <- rslt.avgp$call$FDR

G.gma.f <- CDF.Pval(u=gamma*f, r.1=2000/54675, effect.size=0.79, n.sample=46)

c(G.of.gamma.f=G.gma.f$CDF.Pval, gamma=gamma)
```

`CDF.Pval.HA` *CDF of p-values for test statistics distribted under HA.*

Description

Computes the CDF of p-values for test statistics distribted under HA.

Usage

```
CDF.Pval.HA(u, groups = 2, r.1, effect.size, n.sample, control)
```

Arguments

`u`	Argument of the CDF. Result will be Pr(P_i <= u)
`groups`	The number of experimental groups to compare. Default value is 2.
`r.1`	The proportion of all test statistics that are distributed under HA.
`effect.size`	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
`n.sample`	The number of experimental replicates.
`control`	Optionally, a list with components with the following components: 'groups', used when distop=3 (F-dist), specifying number of groups. 'version', used only in the 'JL' method, choice 0 gives the 'JL' version as published, whereas choice 1 replaces the FDR with r.0*FDR resulting in the infinite simultaneous tests limiting average power, which is the 'Iz' version, but this is redundant because you can specify the 'Iz' method to use this option. 'tol' is a convergence criterion used in iterative methods which is set to 1e-8 by default 'max.iter' is an iteration limit, set to 1000 by default 'distop', specifying the distribution family of the central and non-centrally located sub-populations. =1 gives normal (2 groups) =2 gives t- (2 groups) and =3 gives F- (2+ groups) 'CS', correlation structure, for use only with 'method="simulation"' which will simulate m simulatenous tests with correlations 'rho' in blocks of size 'n.WC'. Specify as list CS = list(rho=0.80, n.WC=50) for example

Details

Computes the CDF of p-values for test statistics distribted under HA. If Fc_0 is the cCDF of a test statistic under H0 and Fc_A is the cCDF of a test statistic under HA then the CDF of a P-value for a test statistic distributed under HA is

G_A(u) = Fc_A(Fc_0^-1(u))

The limiting true positive fraction is the infinite simultaneous tests average power,

lim_m S_m/M_m = average.power (a.s.),

which is used to approximate the average power for finite 'm', is G_1 at gamma f:

G_1(gamma f) = average.pwer

where f is the FDR and gamma = lim_m J_m/m (a.s.) is the limiting positive call fraction.

Value

A list with components

call	The call which produced the result
u	The argument that was passed to the function
CDF.Pval.HA	The value of the CDF

Author(s)

Grant Izmirlian <izmirlian at nih dot gov>

References

Genovese, C. and L. Wasserman. (2004) A stochastic process approach to false discovery control. Annals of Statistics. 32 (3), 1035-1061.

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

See Also

CDF.Pval

Examples

```
## First calculate an average power for a given set of parameters
rslt.avgp <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)

## Now verify that G_A( gamma f ) = average.power

gamma <- rslt.avgp$gamma
f <- rslt.avgp$call$FDR

GA.gma.f <- CDF.Pval.HA(u=gamma*f, r.1=2000/54675, effect.size=0.79, n.sample=46)

c(G.A.of.gamma.f=GA.gma.f$CDF.Pval.HA, average.power=rslt.avgp$average.power)
```

---

controlFDF *Calculates a reduced FDR required to control the FDF*

Description

Calculates a reduced FDR required to bound the the false discovery rate in probability.

Usage

```
controlFDF(groups=2, FDR, r.1, N.tests, effect.size, n.sample,
           use.prob=c("f*","f","user"), prob=NULL)
```

Arguments

groups	The number of experimental groups to compare. Default value is 2.
FDR	The false discovery rate.
r.1	The proportion of simultaneous tests that are non-centrally located
N.tests	The number of simultaneous hypothesis tests.
effect.size	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
n.sample	The number of experimental replicates.
use.prob	This sets the value of the probability that the FDF exceeds lambda*. Set this to the character string 'f*' (default) and the routine uses the value of f*. Set this to 'f' and the routine uses the value of the FDR. Set this to 'user' and then you can specify any desired probability in the argument 'prob'.
prob	The desired probability that the FDF exceeds lambda* that is user specified when the argument 'use.prob' is set to 'user'.

Details

Calculates a reduced FDR required to bound the the false discovery rate in probability...e.g. finds f* so that when the BH-FDR procedure at FDR, f*, is used, we ensure that

Pr(T/J > (1-r) f) < (1-r) f*

where 'f' is the original false discovery rate and 'r' is the proportion of non-null distributed test statistics.

Value

f.star	The reduced FDR required to bound the FDF in probability
obj	Result of optimization yielding 'f.star' should be close to 0
L.star	The bound on the FDF, should be (1-r) f. See above.
P.star	The probability that the FDF is greater than L.star. See above.
average.power	Resulting average power.
c.g	The BH-FDR threshold on the scale of the test statistics.
gamma	The proportion of all 'm' tests declared significant.
objective	Result of optimization yielding the 'average.power'.
err.III	Mass on the wrong side of the threshold.
sigma.rtm.SoM	Asymptotic variance of the true positive fraction.

Author(s)

Grant Izmirlian <izmirlian at nih dot gov>

References

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

See Also

pwrFDR

Examples

```
## at FDR=0.15 and other parameters, it takes n.sample=46 replicates for
## average power > 80%
pwr.46.15 <- pwrFDR(FDR=0.15, r.1=0.03, N.tests=1000, effect.size=0.79, n.sample=46)

## when there are 'only' N.tests=1000 simultaneous tests, the distribution of the
## false discovery fraction, FDF, is not so highly spiked at the FDR=0.15
## You need to set the FDR down to FDR=0.0657 to ensure that  Pr( T/J > 0.145 ) < 0.0657
fstr <- controlFDF(FDR=0.15, r.1=0.03, N.tests=1000, effect.size=0.8, n.sample=46)

## at all the above settings, with FDR=0.0657 at an n.sample of 46, we only have 69%
## average power.
pwr.46.0657 <- pwrFDR(FDR=0.065747, r.1=0.03, N.tests=1000, effect.size=0.79, n.sample=46)

## it'll cost 7 more replicates to get the average power up over 80%.
pwr.53.0657 <- pwrFDR(FDR=0.065747, r.1=0.03, N.tests=1000, effect.size=0.8, n.sample=53)

## it costs only 8.75% more to get it right!
```

---

detail *The detail extraction function for simulated power objects*

Description

Objects created by the pwrFDR function with option method=="simulation" are returned with an attribute named detail. This is its extractor function

Usage

```
detail(obj)
```

Arguments

obj	An object created by the pwrFDR function with option method=="simulation".

Value

A list with components

reps	A data frame of n.sim rows containing the results of the simulations as columns: M1, the \# of non-null distributed statistics, J, the \# of statistical tests rejected under the BH-FDR procedure, and S, the \# of true positives.
CCDF	A data frame of dimension n.T, the number of test statistics with non-zero mean, by 2. The first column is s, taking values 1 to n.T, and the second column is the monte carlo estimated complementary CDF for the distribution of the number of true positives, S
X	A single simulation replicate of the m raw test statistics, included as a sanity check

Author(s)

Grant Izmirlian izmirlig at mail dot nih dot gov

---

dists *The Distribution family object*

Description

The pwrFDR package currently incorporates 3 distribution types, normal, t and F. The first two of these are strictly for statistics formed from two group comparison while the third is for statistics formed from the omnibus test of any difference among an arbitrary number of groups >=2. The structure is general and user expandable. One must specify the density, CDF and quantile function for a given distribution and its parameters under the null and under the alternative. These parameters must be expressions to be evaluated inside the kernel of the power program, functions of the arguments n.sample, groups and effect.size. This is not used directly by the user at all unless she (he) wants to add a distribution type.

Format

A data frame with 3 observations on the following 6 variables.

pars0 a list vector having components 'c(nd, p1, p2, ...)' where 'nd' is the distribution number starting with 0, and p1, p2, ..., are paramters of the distribution, which are functions of 'n.sample', 'groups' and 'effect.size' as mentioned above. These must be expressed as a call e.g. as.call(expression(c, nd, p1, p2, ...)) etc. 'pars0' are the parameters under the null.

pars1 a list vector. See directly above. Parameters under the alternative.

minv a list vector with components given the values -Inf or 0, which will be used to decide if the two sided corrections are used or not.

ddist a list vector with components set to functions, each one computing the probability density function corresponding to the particular distribution. A function of arguments 'x' and 'par'. See details below.

`pdist` a list vector with components set to the functions, each one computing the cumulative distribution function corresponding to the particular distribution. A function of arguments 'x' and 'par'. See details below.

`qdist` a list vector with components set to the functions, each one computing the quantile function (inverse cumulative distribution function) corresponding to the particular distribution. A function of arguments 'x' and 'par'. See details below.

Details

The instance shipped with the current version of the package was created using the following commands. Read this and you'll figure out how to add a distribution type yourself.

```
"dists" <-
as.data.frame(rbind(
### Normal with 2 groups ###
c(pars0=as.call(expression(c,0,ncp=0,sd=1)),
pars1=as.call(expression(c,0,ncp=(n.sample/2)^0.5*effect.size, sd=1)),
minv=-Inf,
ddist=function(x, par) dnorm(x, mean=par[2], sd=par[3]),
pdist=function(x, par) pnorm(x, mean=par[2], sd=par[3]),
qdist=function(x, par) qnorm(x, mean=par[2], sd=par[3])),

### t with 2 groups ###
c(pars0=as.call(expression(c,1,ncp=0, ndf=2*n.sample - 2)),
pars1=as.call(expression(c,1,ncp=(n.sample/2)^0.5*effect.size, ndf=2*n.sample - 2)),
minv=-Inf,
ddist=function(x, par) dt(x, ncp=par[2], df=par[3]),
pdist=function(x, par) pt(x, ncp=par[2], df=par[3]),
qdist=function(x, par) qt(x, ncp=par[2], df=par[3])),

### F with 'groups' groups, effect.size=theta*c(0, 0.5, 0.5, ..., 0.5, 1) ###
c(pars0=as.call(expression(c,2,ncp=0, ndf1=groups-1, ndf2=groups*(n.sample-1))),
pars1=as.call(expression(c,2,ncp=n.sample*effect.size^2/2, ndf1=groups-1, ndf2=groups*(n.sample-
1))),
minv=0,
ddist=function(x, par) df(x, ncp=par[2], df1=par[3], df2=par[4]),
pdist=function(x, par) pf(x, ncp=par[2], df1=par[3], df2=par[4]),
qdist=function(x, par) qf(x, ncp=par[2], df1=par[3], df2=par[4])))
```

Source

This isn't 'data' data, its a kind of a 'family' object.

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

gentempfilenm *Generate a tempfile name*

Description

Generates a tempfile name with an optional user specified prefix and suffix Result is a character string

Usage

```
gentempfilenm(prfx = "temp", sfx = ".txt")
```

Arguments

prfx	prefix for the file name, e.g. "temp"
sfx	suffix (file extension) for the file name, e.g. ".txt"

Value

a character string containing the randomly generated name of the tempfile.

Author(s)

Grant Izmirlian izmirlig at mail dot nih dot gov

logit *Computes the logit transform*

Description

Computes the logit transform for objects of type numeric and objects of class "pwr".

Usage

```
logit(mu)
```

Arguments

mu	A real number on the interval [0, 1]

Value

A numeric equal to the logit of mu, a real number.

Author(s)

Grant Izmirlian <izmirlian at mail.nih.gov>

logitInv *Computes the inverse logit transform*

Description

Computes the inverse logit transform for objects of type numeric and objects of class "pwr".

Usage

```
logitInv(eta)
```

Arguments

eta	Any real number

Value

A numeric equal to the logit inverse of mu, a real number on the interval [0, 1]

Author(s)

Grant Izmirlian <izmirlian at mail.nih.gov>

paste *The paste operator*

Description

A binary operator shortcut for paste(x,y)

Usage

```
x %,% y
```

Arguments

x	a character string
y	a character string

Value

The concatenated character string

Author(s)

Grant Izmirlian <izmirlian@nih.gov>

Examples

```
library(pwrFDR)
"var" %,% (1:10)
```

---

pwrFDR *Power in the BH-FDR procedure.*

Description

Calculates average power or λ-power in the Benjamini-Hochberg FDR procedure, or sample size required for given average- or λ- power.

Usage

```
pwrFDR(groups = 2, effect.size, n.sample, r.1, FDR, use.LFD = FALSE,
       N.tests, average.power, L.power, lambda, method = c("approximate",
       "simulation", "JL", "Iz"), control = list(version = 0,
       tol = 1e-08, max.iter = 1000, distopt = 1, CS = list(NULL),
       verb = FALSE), n.sim = 1000, temp.file)
```

Arguments

groups	The number of experimental groups to compare. Default value is 2.
effect.size	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
n.sample	The number of experimental replicates. Required for calculation of power
r.1	The proportion of simultaneous tests that are non-centrally located
FDR	The false discovery rate.
use.LFD	Under the default value, 'FALSE', the procedure is controlled under the specified FDR. Specifying "use.LFD=TRUE" will instead first find an FDR, FDR.star, so that the probability that the FDF exceeds FDR.star is less than or equal to the originally specified FDR. Power (either average- or lambda-) or sample size under the average- or lambda- power is then calculated when the procedure is controlled at FDR.star. This is recommended when the distribution of the FDF is wide enough to matter but still well approximated by its asymptotic distribution, e.g. between several hundred and several thousand simultaneous tests. The argument 'N.tests' is required when "use.LFD=TRUE".
N.tests	The number of simultaneous hypothesis tests.
average.power	The desired average power. Sample size calculation requires specification of either 'average.power' or 'L.power'.
L.power	The desired λ-power (see details for explanation). Sample size calculation requires specification of either 'average.power' or 'L.power'.

`lambda` The λ-power threshold, required when calculating the λ-power (see details for explanation) or when calculating the sample size required for λ-power.

`method` Specify the method whereby the average power is calculated. The selection 'approximate' uses the approximation technique of Izmirlian (2016). To use direct simulation, specify 'simulation'. Specification of 'JL' uses the method of Jung (2005) and Liu (2007). Specification of 'Iz' uses the form which was shown (Izmirlian) to be the large number of tests limit of the average power. Valid selections are 'Iz' (default), 'JL', and any substring of 'approximate' or 'simulation'.

`control` Optionally, a list with components with the following components: 'groups', used when distop=3 (F-dist), specifying number of groups. 'version', used only in the 'JL' method, choice 0 gives the 'JL' version as published, whereas choice 1 replaces the FDR with r.0*FDR resulting in the infinite simultaneous tests limiting average power, which is the 'Iz' version, but this is redundant because you can specify the 'Iz' method to use this option. 'tol' is a convergence criterion used in iterative methods which is set to 1e-8 by default 'max.iter' is an iteration limit, set to 1000 by default 'distop', specifying the distribution family of the central and non-centrally located sub-populations. =1 gives normal (2 groups) =2 gives t- (2 groups) and =3 gives F- (2+ groups) 'CS', correlation structure, for use only with 'method="simulation"' which will simulate m simulatenous tests with correlations 'rho' in blocks of size 'n.WC'. Specify as list CS = list(rho=0.80, n.WC=50) for example

`n.sim` If 'simulation' method is chosen you may specify number of simulations. Default is 1000.

`temp.file` If 'simulation' method is chosen you may specify a tempfile where the current simulation replicate is updated. Very usefull for batch runs. You can use the included utility 'gentempfilenm'

Details

Power for the BH-FDR procedure on 'm' simultaneous tests, at a given FDR, f, is computed under the following model. A-priori, a proportion, 'r.1', of the 'm' test statistics are given a non-centrally located distribution, the remaining proportion are given a centrally distribution. Each of the non-centrally located test statistics is given the same location parameter. Suppose that a total of 'M' are chosen to have a non-central distribution, that a total of 'J' of all 'm' statistics are declared significant, and that of these, 'S' come from the non-centrally located population. This results in the following table.

		rej H0	acc H0	row Total
1.		rej H0	acc H0	row Total
2.	H0 is FALSE	S	M-S	M
3.	H0 is TRUE	J-S	(m-M)-(J-S)	m-M.1
4.	col Total	J	m-J	m

By default, when 'n.sample' is specified, the function computes the the average power,

AVERAGE POWER: $E[S/M]$

This is approximated using the infinite tests limit.

When 'n.sample' and 'λ' are both specified, the function computes the λ-power, which is the probability:

λ-POWER: $P\{S/M > \lambda\}$

that the true positive fraction S/M exceeds the threshold, 'λ'. This is done using an asymptotic approximation. If the user desires instead the sample size required for given average- or λ- power, then leave the argument 'n.sample' unspecified and specify instead either the 'average.power' or the 'λ', the threshold for the λ-power. By default, the model uses the assumption that test statistics are distributed as t-distributed as in the two group comparison. If the user specifies a value of 'groups' larger than 2 then the model assumes the test statistics are F-distributed as in the omnibus F-test for any difference. The above approximations are done under the default method. If the user wants to compare with simulated answers to obtain simulation estimates of the 'exact' values, one need only specify 'method="sim"'.

Value

An object of class "pwr" with with components including:

`call`	The call which produced the result
`average.power`	Resulting average power.
`L.power`	When 'lambda' is specified, the lambda-power is also computed
`L.eq`	The lambda at which the lambda-power and average-power are equal.
`n.sample`	If 'n.sample' is missing from the argument list, then the sample size required for the specified average- or lambda- power.
`FDR.star`	If "use.LFD=TRUE" was specified, the FDR at which the probability that the FDF exceeds FDR.star is less than or equal to the originally specified FDR.
`c.g`	The BH-FDR threshold on the scale of the test statistics.
`gamma`	The proportion of all 'm' tests declared significant.
`objective`	Result of optimization yielding the 'average.power'.
`err.III`	Mass on the wrong side of the threshold.
`sigma.rtm.SoM`	Asymptotic variance of the true positive fraction.

Author(s)

Grant Izmirlian <izmirlian at nih dot gov>

References

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

Jung S-H. (2005) Sample size for FDR-control in microarray data analysis. Bioinformatics; 21:3097-3104.

Liu P. and Hwang J-T. G. (2007) Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. Bioinformatics; 23:739-746.

See Also

controlFDF

Examples

```
## Example 1a: average power
   rslt.avgp <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
   rslt.avgp

## Example 1b: lambda-power
   rslt.lpwr <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675,
                       FDR=0.15, lambda=0.80, N.tests=54675)
   rslt.lpwr

## Example 1c: sample size required for given average power
   rslt.ss.avgp <- pwrFDR(effect.size=0.79, average.power=0.82,
                          r.1=2000/54675, FDR=0.15)
   rslt.ss.avgp

## Example 1d: sample size required for given lambda-power
   rslt.ss.lpwr <- pwrFDR(effect.size=0.79, L.power=0.82, lambda=0.80,
                          r.1=2000/54675, FDR=0.15, N.tests=54675)
   rslt.ss.lpwr

## Example 1e: simulation
   rslt.sim <- update(rslt.avgp, method="sim", n.sim=500, N.tests=1000)
   rslt.sim

## Example 2: methods for adding, subtracting, multiplying, dividing, exp, log,
## logit and inverse logit
   rslt.avgp - rslt.sim
   logit(rslt.avgp)       ## etc

## Example 3: Compare the asymptotic distribution of S/M with kernel
##            density estimate from simulated data
   pdf <- with(detail(rslt.sim)$reps, density(S/M1))

   med <- with(detail(rslt.sim)$reps, median(S/M1))
   avg <- rslt.sim$average.power
   sd <- rslt.sim$v.SoM.emp^0.5

   rng.x <- range(pdf$x)
   rng.y <- range(c(pdf$y, dnorm(pdf$x, mean=avg, sd=sd)))

   plot(rng.x, rng.y, xlab="u", ylab="PDF for S/M", type="n")
   with(pdf, lines(x, y))
   lines(rep(rslt.sim$average.power, 2), rng.y, lty=2)
   lines(pdf$x, dnorm(pdf$x, mean=avg, sd=sd), lty=3)
```

`var.J.o.rtm` *Calculates asymptotic variance of significant call fraction, J_m/m in the BH-FDR procedure on m=N.tests simultaneous tests.*

Description

A function which computes the asymptotic variance of the proportion of significant calls, J_N/N.

Usage

```
var.J.o.rtm(x, groups, effect.size, n.sample, r.1, FDR, N.tests, control)
```

Arguments

`x`	Calls to this function can be made either specifying the single argument, x, which is an object of class "pwr" returned from the function, `pwrFDR` or by a full call specification, by specifying all the arguments required by the function `pwrFDR` (see below)
`groups`	The number of experimental groups to compare. Default value is 2.
`effect.size`	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
`n.sample`	The number of experimental replicates. Required for calculation of average.power or L.power.
`r.1`	The proportion of simultaneous tests that are non-centrally located
`FDR`	the false discovery rate.
`N.tests`	The number of simultaneous tests.
`control`	Optionally, a list with components with the following components: 'groups', used when distop=3 (F-dist), specifying number of groups. 'version', used only in the 'JL' method, choice 0 gives the 'JL' version as published, whereas choice 1 replaces the FDR with r.0*FDR resulting in the infinite simultaneous tests limiting average power, which is the 'Iz' version, but this is redundant because you can specify the 'Iz' method to use this option. 'tol' is a convergence criterion used in iterative methods which is set to 1e-8 by default 'max.iter' is an iteration limit, set to 1000 by default 'distop', specifying the distribution family of the central and non-centrally located sub-populations. =1 gives normal (2 groups) =2 gives t- (2 groups) and =3 gives F- (2+ groups) 'CS', correlation structure, for use only with 'method="simulation"' which will simulate m simulatenous tests with correlations 'rho' in blocks of size 'n.WC'. Specify as list CS = list(rho=0.80, n.WC=50) for example

Details

The proportion of significant calls, J_m/m, i.e., the proportion of all hypothesis tests declared significant by the Benjamini-Hochberg procedure is shown in the cited publication to be root-m consistent and asymptotically normal i.e.

m^1/2 (J_m/m - gamma) –D–> N(0, tau^2)

which is to say that the distribution of J_m/m is spiked about its mean, gamma, and the width dies off as 1/m^1/2. This is of interest in its own right, and is also needed as an intermediate in the calculation of the asymptotic variance of the empirical average power, S_m/M_m. The normal approximation for this latter quantity can be used to power multiple testing experiments on a lower quantile of the empirical mean rather than on its expected value. The width of this distribution is non-negligible for as many as 200 simultaneous tests, where the average power could be 80% but the lower quantile of the empirical average power could be as low as 50%.

Value

Returns a value of class vvv, containing components

`var.J.o.rtm`	The computed asymptotic variance
`power`	The average power
`gamma`	The expected proportion of significant calls
`c.g`	The 'q-value', which is the value of the criterion on the scale of the statistic (t of given number of 2 n - 2 degrees of freedom) which can be used as a per test criterion resulting in the equivalent Benjamini-Hochberg procedure
`call`	The call which produced the result

Author(s)

Grant Izmirlian <izmirlig at mail dot nih dot gov>

References

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

See Also

`var.rtm.SoM` `var.rtm.ToJ`

Examples

```
## call using result of pwrFDR
rslt.Iz <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
vJ <- var.J.o.rtm(rslt.Iz)

## call via argument list specification
vJ <- var.J.o.rtm(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
```

`var.rtm.SoM` *Calculates asymptotic variance of true positive fraction, S_m/M_m.*

Description

A function which computes the asymptotic variance of the true positive fraction, S_m/M_m in the BH-FDR procedure on m=N.tests simultaneous tests.

Usage

```
var.rtm.SoM(x, groups, effect.size, n.sample, r.1, FDR, N.tests, control)
```

Arguments

`x`	Calls to this function can be made either specifying the single argument, x, which is an object of class "pwr" returned from the function, `pwrFDR` or by a full call specification, by specifying all the arguments required by the function `pwrFDR` (see below)
`groups`	The number of experimental groups to compare. Default value is 2.
`effect.size`	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
`n.sample`	The number of experimental replicates. Required for calculation of power
`r.1`	The proportion of simultaneous tests that are non-centrally located
`FDR`	The false discovery rate.
`N.tests`	Number of simultaneous tests.
`control`	Optionally, a list with components with the following components: 'groups', used when distop=3 (F-dist), specifying number of groups. 'version', used only in the 'JL' method, choice 0 gives the 'JL' version as published, whereas choice 1 replaces the FDR with r.0*FDR resulting in the infinite simultaneous tests limiting average power, which is the 'Iz' version, but this is redundant because you can specify the 'Iz' method to use this option. 'tol' is a convergence criterion used in iterative methods which is set to 1e-8 by default 'max.iter' is an iteration limit, set to 1000 by default 'distop', specifying the distribution family of the central and non-centrally located sub-populations. =1 gives normal (2 groups) =2 gives t- (2 groups) and =3 gives F- (2+ groups) 'CS', correlation structure, for use only with 'method="simulation"' which will simulate m simulatenous tests with correlations 'rho' in blocks of size 'n.WC'. Specify as list CS = list(rho=0.80, n.WC=50) for example

Details

The true positive fraction, S_m/M_m, i.e., is the proportion of all non-centrally located statistical tests that are declared significant by the Benjamini-Hochberg procedure. It is shown, in the cited publication, to be root-m consistent and asymptotically normal i.e.

m^1/2 (S_m/M_m - average.power) –D–> N(0, sigma^2)

which is to say that the distribution of S_m/M_m is spiked about its mean, the average power, and the width dies off as 1/m^1/2. This noraml approximation is of interest in its own right and is also useful for powering multiple testing experiments on a more conservative operating characteristic than the average power, E[S_m / M_m]. For example, we can power the experiment on the lower 10th percentile of the distribution of S_m/M_m, which is approximated as

average.power + qnorm(0.10)*(vS/N.tests)^0.5

While the width of this distribution is negligible for micro-array studies, e.g. when N.tests=54675, it is non-negligible for as many as 200 simultaneous tests, where the average power could be 80% but the lower quantile of the empirical average power could be as low as 50%.

Value

Returns a value of class vvv, containing components

`var.rtm.SoM`	The computed asymptotic variance
`power`	The average power
`gamma`	The expected proportion of significant calls
`c.g`	The 'q-value', which is the value of the criterion on the scale of the statistic (t of given number of 2 n - 2 degrees of freedom) which can be used as a per test criterion resulting in the equivalent Benjamini-Hochberg procedure
`call`	The call which produced the result

Author(s)

Grant Izmirlian <izmirlig at mail dot nih dot gov>

References

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

See Also

`var.J.o.rtm` `var.rtm.ToJ`

Examples

```
## call using result of pwrFDR
rslt.Iz <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
vS <- var.rtm.SoM(rslt.Iz)

## call via argument list specification
vS <- var.rtm.SoM(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
```

`var.rtm.ToJ` *Calculates asymptotic variance of false discovery fraction, T_N/J_N.*

Description

A function which computes the asymptotic variance of the false discovery fraction, T_N/J_N.

Usage

```
var.rtm.ToJ(x, groups, effect.size, n.sample, r.1, FDR, N.tests, control)
```

Arguments

`x`	Calls to this function can be made either specifying the single argument, x, which is an object of class "pwr" returned from the function, `pwrFDR` or by a full call specification, by specifying all the arguments required by the function `pwrFDR` (see below)
`groups`	The number of experimental groups to compare. Default value is 2.
`effect.size`	The effect size (mean over standard deviation) for test statistics having non-zero means. Assumed to be a constant (in magnitude) over non-zero mean test statistics.
`n.sample`	The number of experimental replicates. Required for calculation of power
`r.1`	The proportion of simultaneous tests that are non-centrally located
`FDR`	The false discovery rate.
`N.tests`	Number of simultaneous tests.
`control`	Optionally, a list with components with the following components: 'groups', used when distop=3 (F-dist), specifying number of groups. 'version', used only in the 'JL' method, choice 0 gives the 'JL' version as published, whereas choice 1 replaces the FDR with r.0*FDR resulting in the infinite simultaneous tests limiting average power, which is the 'Iz' version, but this is redundant because you can specify the 'Iz' method to use this option. 'tol' is a convergence criterion used in iterative methods which is set to 1e-8 by default 'max.iter' is an iteration limit, set to 1000 by default 'distop', specifying the distribution family of the central and non-centrally located sub-populations. =1 gives normal (2 groups) =2 gives t- (2 groups) and =3 gives F- (2+ groups) 'CS', correlation structure, for use only with 'method="simulation"' which will simulate m simulatenous tests with correlations 'rho' in blocks of size 'n.WC'. Specify as list CS = list(rho=0.80, n.WC=50) for example

Details

The false discovery fraction (FDF), T_m/J_m, i.e., is the proportion of all centrally located statistical tests that are declared significant by the Benjamini-Hochberg procedure. It is shown, in the cited publication, to be root-m consistent and asymptotically normal i.e.

m^1/2 (T_m/J_m - (1-r.1) f) –D–> N(0, sigma^2)

which is to say that the distribution of T_m/J_m is spiked about its mean, the false discovery rate, and the width dies off as 1/m^1/2. This noraml approximation is of interest in its own right, allowing tighter control over the false discovery fraction that afforded by controlling its mean, the false discovery rate. One can determine the FDR required to bound the FDF by a given value with a probability no less than a given value.

Value

Returns a value of class vvv, containing components

`var.J.o.rtm`	The computed asymptotic variance
`power`	The average power
`gamma`	The expected proportion of significant calls
`c.g`	The 'q-value', which is the value of the criterion on the scale of the statistic (t of given number of 2 n - 2 degrees of freedom) which can be used as a per test criterion resulting in the equivalent Benjamini-Hochberg procedure
`call`	The call which produced the result

Author(s)

Grant Izmirlian <izmirlig at mail dot nih dot gov>

References

Izmirlian G. (2017) Average Power and λ-power in Multiple Testing Scenarios when the Benjamini-Hochberg False Discovery Rate Procedure is Used. arXiv:1801.03989

See Also

`var.J.o.rtm` `var.rtm.SoM`

Examples

```
## call using result of pwrFDR
rslt.Iz <- pwrFDR(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
vJ <- var.J.o.rtm(rslt.Iz)

## call via argument list specification
vJ <- var.J.o.rtm(effect.size=0.79, n.sample=46, r.1=2000/54675, FDR=0.15)
```

`%over%` *Division operator with divide by zero clobbering*

Description

x %over% y = x/y when y!=0, equals 0 when y==0.

Usage

```
x %over% y
```

Arguments

`x,y`	Numeric or complex vectors or objects that can be coerced to such.

Value

x/y when y!=0, otherwise 0.

Author(s)

Grant Izmirlian izmirlig at mail dot nih dot gov

Index