

Package ‘ldDesign’

January 2, 2012

Version 2.0-0

Title Design of experiments for detection of linkage disequilibrium

Author Rod Ball <rod.ball@scionresearch.com>

Maintainer Rod Ball <rod.ball@scionresearch.com>

Description R package for design of experiments for design of genome-wide association studies. Version 2 incorporating quantitative traits and case-control studies. The Bayes factor should be chosen large enough to give respectable posterior odds. This requires Bayes factors of the order of 10^6 in genome-wide association studies where prior odds are low. Sample sizes needed to get this strength of evidence are substantially higher than those from traditional power calculations. The corresponding threshold for p-values is substantially lower than commonly used. For quantitative traits ldDesign uses an existing deterministic power calculation for detection of linkage disequilibrium between a bi-allelic QTL and a bi-allelic marker, together with the Spiegelhalter and Smith Bayes factor to generate designs with power to detect effects with a given Bayes factor. For case-control studies an asymptotic approximate Bayes factor is used to derive an analytical power calculation in dominant, recessive, additive and general genetic models.

Depends

License GPL (>= 2)

URL <mailto:rod.ball@scionresearch.com> www.scionresearch.com/

Repository CRAN

Date/Publication 2011-11-18 07:10:46

R topics documented:

calc.B.ABF	2
cc.design	3
gpc.power	6
ld.design	7

ld.sim	9
luo.ld.populations	12
luo.ld.power	12
oneway.bf.alpha	14
SS.oneway.bf	16

Index	18
--------------	-----------

calc.B.ABF	<i>Calculate Z-values or alpha-values from approximate Bayes factors and vice versa.</i>
------------	--

Description

These functions find the critical values of Z and alpha corresponding to a given approximate Bayes factor, or vice versa for given sample size (n) and prior equivalent sample size (a), where Z is a statistic with sampling distribution $N(0, 1/n)$ for sample size n. The critical value Z_c is used by [cc.power](#) to calculate the power, for an experiment to obtain a given Bayes factor, and the corresponding alpha value α_{hac} may be used by the Genetic power calculator function [gpc.power](#).

Usage

```
calc.B.ABF(Z,n,a=1)
calc.Zc.ABF(B,n,a=1)
calc.Zalpha.ABF(alpha,n)
calc.Balpha.ABF(alpha,n,a)
calc.alphaB.ABF(B,n,a=1,alpha.start=1e-5,reduction.factor=1.5,niter=20,verbose=FALSE,show.progress)
```

Arguments

B	Bayes factor
Z	a sample statistic with sampling distribution $N(0,1/n)$ estimating a multiple of a quantity of interest
n	sample size
a	equivalent sample size for information in prior
alpha	significance level
alpha.start	initial guess (upper bound) for significance level
reduction.factor	factor to decrease alpha.start by each iteration in the search for an alpha value corresponding to B
niter	number of iterations in search
show.progress	if TRUE print dots to show progress
verbose	if TRUE print array of alpha and B values calculated

Details

These functions calculate the approximate Bayes factor and threshold calculations used in the power calculations for case-control studies ([cc.power](#), [cc.design](#)), but can be used more widely. For given Z , n , a , the approximate Bayes factor is given by the formula:

$$(1) B = \sqrt{a/(n+a)} * \exp(n^2 * Z^2 / (2 * (n+a)))$$

where n is the sample size and the sampling distribution of Z is $N(0, 1/n)$, and the prior for Z is $N(0, 1/a)$, (so that a is the equivalent sample size for information in the prior). In applications a subjective prior for the quantity of interest can be given and the corresponding value of a determined.

Value

For `calc.B.ABF` the Bayes factor from (1); for `calc.Zc.ABF` the value of Z corresponding to B in (1);

Author(s)

Rod Ball <rod.ball@scionresearch.com> www.scionresearch.com

References

- Ball, R.D. 2005: Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. *Genetics* 170: 859–873.
- Ball, R.D. 2007 (Jan.): “Statistical analysis and experimental design”, Chapter 8, In: *Association mapping in plants*. N.C. Oraguzie et al. editors, Springer Verlag, ISBN 0387358447. (69pp) (Approximate Bayes factors for S-TDT test: pp166-167)
- Wakefield, J. 2007 (Jul.): A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* 81: 208–227.
- Ball, R.D. 2011: Experimental designs for robust detection of effects in genome-wide case-control studies (submitted).

Examples

```
calc.Zc.ABF(B=1e6,n=10000,a=1)
calc.B.ABF(Z=0.0607,n=10000,a=1)
calc.Balpha.ABF(alpha=0.01,n=100,a=1)
calc.alphaB.ABF(B=1e6,n=10000,a=1,alpha.start=1e-7)
calc.Zalpha.ABF(alpha=1.28e-9,n=10000)
```

cc.design

Functions for design of experiments to detect linkage disequilibrium in genome-wide case-control studies

Description

Find the sample sizes (number of cases and number of controls) required to detect linkage disequilibrium with a given Bayes factor, with a given power, or find the power of experimental designs to detect linkage equilibrium with a given Bayes factor.

Usage

```
cc.design(B, OR, D, p, q, power, baseline.risk, Dprime=NULL, R=NULL,
  prevalence=NULL, n.cases, n.controls, model=c("additive",
  "dominant","recessive","general"), a=1, sigma2.eta=NULL, verbose=FALSE,
  amalgamate.cells=FALSE, pmin=0.1, pmax=0.99, ninterp=20, print.power.curve=TRUE)
cc.power(B, OR, D, p, q, baseline.risk, Dprime=NULL, R=NULL, prevalence=NULL,
  n.cases, n.controls, model=c("additive","dominant","recessive","general"),
  a=1, sigma2.eta=NULL, verbose=FALSE, amalgamate.cells=FALSE, show.attributes=FALSE)
```

Arguments

B	Bayes factor
OR	Odds ratio
D	Linkage disequilibrium coefficient
p	Bi-allelic marker allele frequency
q	Bi-allelic QTL allele frequency (for risk allele)
power	cc.design: Power, or probability of detecting an effect with Bayes factor greater than B
baseline.risk	Baseline risk, i.e. the probability of being a case for a genotype with no risk alleles
Dprime	D': i.e. linkage disequilibrium as proportion of the maximum (or minimum, if negative); need to give D_or_ Dprime
R	Relative risk: in the additive model: relative risk per copy of the risk allele, in dominant or recessive models relative risk for the high risk genotype(s) compared with the low risk genotype(s) in the 2x2 contingency table; need to give R_or_ OR.
prevalence	disease prevalence, i.e. the probability of being a case in the population; need to give prevalence_or_ baseline.risk
n.cases	number of cases (can be vectorised), for cc.design only used for determining the proportion of cases and controls
n.controls	number of controls (can be vectorised) for cc.design only used for determining the proportion of cases and controls
model	genetic model assumed for the locus, one of: "additive", "dominant", "recessive" or "general"
a	equivalent sample size for information in prior, (alternative to specifying sigma2.eta); a=1 is a conservative 'default' value
sigma2.eta	prior variance for the log odds ratio
verbose	if TRUE display intermediate information
amalgamate.cells	if TRUE, amalgamate the genotypes with 1 or 2 copies of the risk allele (additive model only); this may be advantageous when the risk allele frequency (q) is low
pmin	cc.design: Lower bound for power — power curve calculated from pmin to pmax

`pmax` `cc.design`: Upper bound for power — power curve calculated from `pmin` to `pmax`
`ninterp` `cc.design`: Number of sample sizes to calculate for interpolation
`print.power.curve`
 `cc.design`: If TRUE print power curve for power from `pmin` to `pmax`
`show.attributes`
 `cc.power`: If TRUE print attributes of the returned object, otherwise just print the power.

Details

These functions implement the method described in Ball (2011) for obtaining the power of designs for detecting linkage disequilibrium with a given Bayes factor in genome-wide case-control studies. The method uses an asymptotic approximation to the Bayes factor (Ball 2007, 2011; Wakefield 2007) together with an estimator for the log-odds ratio and its standard error. This extends the method from Ball (2005) to case-control studies used for studying human diseases.

Value

For `cc.power`, the power for the given sample size(s) (`n.cases`, `n.controls`). The returned object also includes attributes including Bayes factor, model, sample sizes, odds ratios, relative risks, non-centrality parameter (`ncp`), expected marker genotype frequencies for cases and controls (`ps`), and equivalent frequentist threshold (`alphac`). For `cc.design`, the sample sizes (`n`, `n.cases`, `n.controls`) required to obtain the specified power.

Author(s)

Rod Ball <rod.ball@scionresearch.com> www.scionresearch.com

References

Ball, R.D. 2005: Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. *Genetics* 170: 859–873.

Ball, R.D. 2007: “Statistical analysis and experimental design”, Chapter 8, In: Association mapping in plants. N.C. Oraguzie et al. editors, Springer Verlag, ISBN 0387358447. (69pp)

Ball, R.D. 2011: Experimental designs for robust detection of effects in genome-wide case-control studies (submitted).

Examples

```

cc.power(B=1e6, OR=1.6, D=0.1, p=0.3, q=0.2, baseline.risk=0.1,
         n.cases=1000*seq(2,12,by=2), n.controls=1000*seq(2,12,by=2),
         model="additive", a=1)
cc.design(B=1e6, OR=2.0, D=0.1, p=0.3, q=0.2, power=0.9, baseline.risk=0.1,
         n.cases=2000, n.controls=3000, model="additive", a=1,
         pmin=0.1, pmax=0.99, ninterp=20, print.power.curve=TRUE)
cc.design(B=1e6, OR=c(2.0,2.0), D=0.1, p=0.3, q=0.2, power=0.9, baseline.risk=0.1,
         n.cases=2000, n.controls=3000, model="general", a=1,
         pmin=0.1, pmax=0.99, ninterp=20, print.power.curve=TRUE)

```

```
cc.design(B=1e6, OR=3.0, Dprime=0.5, p=0.03, q=0.02, power=0.9, baseline.risk=0.1,
          n.cases=5000, n.controls=5000, model="recessive",
          pmin=0.1, pmax=0.9, ninterp=20, print.power.curve=TRUE)
cc.design(B=1e6,OR=c(1.2,1.2),D=0.25,p=0.5,q=0.5,prevalence=0.13,power=0.8,
          n.cases=5000,n.controls=5000,model="general",a=1)
```

gpc.power	<i>Power calculations for case-control studies using the Genetic Power calculator</i>
-----------	---

Description

Find the power of tests and the sample sizes required for case-control for given frequentist significance level (α). Uses the Genetic Power Calculator web site (Purcell 2008). Can be used to check the Bayesian calculations from [cc.design](#), [cc.power](#), if called with the appropriate alpha level (α attribute from the output of [cc.power](#)).

Usage

```
gpc.power(Dprime, p, q, prevalence, R1, R2, n.cases, n.controls, alpha, power,
          show.html=TRUE)
```

Arguments

Dprime	Dprime, i.e. linkage disequilibrium as a proportion of the maximum (or minimum, if negative)
p	Bi-allelic marker allele frequency
q	Bi-allelic QTL allele frequency (for risk allele)
prevalence	disease prevalence, i.e. the probability of being a case in the population
R1	Relative risk for the genotype with one copy of the risk allele, compared with the genotype with no copies of the risk allele
R2	Relative risk for the genotype with two copies of the risk allele
n.cases	number of cases
n.controls	number of controls
alpha	significance level or threshold
power	Power, or probability of detecting an effect with the given relative risks, with significance level α (i.e. $p < \alpha$)
show.html	if TRUE, try to show the html web page returned by the query

Details

This function generates a query to the Genetic Power Calculator website, and extracts the power and sample sizes corresponding to the input values of α and power. Requires `wget`, `sed`, and files in the `./bin` subdirectory.

Value

A table showing the values of alphac, power, and number of cases. The power is the power to obtain the specified significance level alphac, with the specified numbers of cases and controls. The number of cases required is the number of cases required to obtain the specified power with the specified significance level

Author(s)

Rod Ball <rod.ball@scionresearch.com> www.scionresearch.com

References

Purcell, S., Cherny, S.S., and Sham, P.C. 2003: Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19: 149–150. <http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html>

Purcell, S. 2008: Genetic Power Calculator: <http://pngu.mgh.harvard.edu/~purcell/cgi-bin/cc2k.cgi>

Examples

```
# Not run
power.res <- cc.power(B=1e6, R=1.2, Dprime=0.5, p=0.5, q=0.5, prevalence=0.13,
  n.cases=147546/2, n.controls=147546/2, model="dominant", a=1,
  verbose=TRUE)
n <- 147546
alphac <- attr(power.res, "alphac")
## use R1 = R2 for the dominant model, R1=1.0 for the recessive model
#gpc.res <- gpc.power(Dprime=0.5, p=0.5, q=0.5, prevalence=0.13, R1=1.2, R2=1.2, n.cases=n/2,
# n.controls=n/2, alpha=alphac, power=0.8)
##
## gpc.res
##          alphac  power ncases(pwr=80%)
## dominant    8.31e-10 0.80030          73752
## recessive   8.31e-10 0.02266          210022
## general (2df) 8.31e-10 0.80260          73580
## allelic     8.31e-10 0.74160          78035
```

Description

Find the sample size required to detect linkage disequilibrium with a given Bayes factor, with a given power, or find the power of experimental designs to detect linkage equilibrium with a given Bayes factor.

Usage

```
ld.design(p, q, D, h2, phi, Bf, power, nmin = 50, nmax = 1e+05, ninterp = 50,
          missclass.rate = 0, print.it = FALSE)
ld.power(n, p, q, D, h2, phi, Bf, missclass.rate = 0)
```

Arguments

n	ld.power: vector of sample sizes
p	Bi-allelic marker allele frequency
q	Bi-allelic QTL allele frequency
D	Linkage disequilibrium coefficient
h2	QTL ‘heritability’, i.e. proportion of total or phenotypic variance explained by the QTL
phi	Dominance ratio: phi = 0 denotes purely additive, phi = 1 denotes purely dominant allele effects
Bf	Bayes factor
power	ld.design: Power, or probability of detecting an effect with Bayes factor greater than Bf
nmin	ld.design: Lower bound for sample size
nmax	ld.design: Upper bound for sample size
ninterp	ld.design: Number of sample sizes to try
missclass.rate	Proportion of marker values which are missclassified, i.e. incorrect (to allow for genotyping errors)
print.it	If TRUE print results for sample sizes tried

Details

These functions implement the method described in Ball (2005) for obtaining the power of designs for detecting linkage disequilibrium with a given Bayes factor. The F values, (and hence significance levels) corresponding to the given Bayes factors, sample sizes, and marker genotype frequencies, are calculated using the method of Spiegelhalter and Smith (1982) (R functions [oneway.bf.alpha.required](#), [SS.oneway.bf](#)). The power is obtained using a corrected version of the classical deterministic power calculation from Luo (1988) (R function [luo.ld.power](#)).

Value

For `ld.power`, a matrix with columns:

n	Sample sizes
power	Power of the design with the given sample sizes

Additionally the return value has attributes indicating the linkage disequilibrium parameters used. For `ld.design` the sample size is returned.

Author(s)

Rod Ball <rod.ball@scionresearch.com> www.scionresearch.com

References

Ball, R.D. 2005: Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. *Genetics* 170: 859–873.

Ball, R.D. 2007: Statistical analysis and experimental design. Chapter 8, pp133–196 In: *Association Mapping in Plants*, N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. DeSilva (Editors), Springer, New York.

Luo, Z.W. 1988: Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Heredity* 80, 198–208

Spiegelhalter, D. and A.F.M. Smith 1982: Bayes factors for linear and log-linear models with vague prior information *J. Royal Statist Soc. B* 44: 377–387.

See Also

[ld.power](#), [ld.sim](#), [oneway.bf.alpha](#), [oneway.bf.alpha.required](#), [SS.oneway.bf](#)

Examples

```
ld.power(n=seq(100,1000,by=100),p=0.5,q=0.5,D=0.1,h2=0.1,phi=0,Bf=20)
ld.design(p=0.5,q=0.5,D=0.1,h2=0.1,phi=0,Bf=20,power=0.9,print.it=TRUE,nmin=600,nmax=4000)
ld.design(p=0.5,q=0.5,D=0.1,h2=0.1,phi=0,Bf=20,power=0.9,print.it=FALSE,nmin=1700,nmax=1900)
```

ld.sim

Functions to simulate populations with a bi-allelic marker and QTL in linkage disequilibrium and test for association.

Description

For a bi-allelic marker and QTL, with given allele frequencies, linkage disequilibrium, and QTL heritability, multiple replicate populations with marker, QTL, and trait values are simulated and tested for a marker-trait association. Results can be used to estimate the power of an experimental design for detecting linkage disequilibrium.

Usage

```
ld.sim(nsim, n, p, q, D, h2, Vp, phi, missclass.rate = 0, sim.chunk = 100,
       method = 1, print.it = TRUE, data.only=FALSE)
ld.sim1(n, p, q, D, d, h, sig2.error, missclass.rate = 0, nreps = 1,
       method = 2, print.it = TRUE, data.only=FALSE)
```

Arguments

nsim	Number of replicate simulations to do
n	The sample size, i.e. number of individuals genotyped and tested for the trait of interest
p	Bi-allelic marker allele frequency
q	Bi-allelic QTL allele frequency
D	Linkage disequilibrium coefficient
h2	QTL ‘heritability’, i.e. proportion of total or phenotypic variance explained by the QTL
Vp	ld.sim: Total or phenotypic variance: an arbitrary value may be used
phi	ld.sim: Dominance ratio: $\phi = 0$ denotes purely additive gene action, $\phi = 1$ denotes completely dominant gene action
d	ld.sim1: Expected value for trait when QTL genotype is QQ,qq respectively is d,-d
h	ld.sim1: Expected value for trait when QTL genotype is Qq is h
sig2.error	ld.sim1: Error variance when QTL genotype known and modelled
missclass.rate	Proportion of marker values which are missclassified, i.e. incorrect
sim.chunk	ld.sim: Number of replicates to do in a ‘chunk’ in each call to ld.sim1
nreps	ld.sim1: Number of replicate simulations to do for the given set of marker genotypes
method	If method=1 simulate random QTL genotypes conditional on marker values in ld.sim1; if method=2 simulate markers and QTL directly from table of joint probabilities. With method=1, a common set of marker values are used for each of the nreps replicates per call to ld.sim1, enabling MANOVA to be used.
print.it	if TRUE, print results
data.only	if TRUE, just return the simulated trait and marker genotype data

Details

Marker, QTL, and trait values are simulated according to the genetic model with normal errors. In ld.sim, QTL parameters d, h are determined from the parameters h2, q, phi, and Vq, and the main simulation done for each chunk of replicates by a call to ld.sim1. Marker-trait association is tested by a one-way analysis of variance of trait values in terms of marker classes. The proportion of results with P-value over a given threshold gives a stochastic estimate of the power calculated by [luo.ld.power](#).

Value

If data.only=FALSE, an array with 1 row per simulation run, and 4 columns with values for each run:

MS.beta	Between marker classes mean square
MS.within	Within marker classes mean square

F.value F value
 P.value P value

otherwise, if `data.only=TRUE`, an array `nsim*n` rows and 3 columns with marker and trait values for simulated populations:

marker Marker genotype indicator with values {1,2,3}, corresponding to genotypes {MM,Mm,mm}
 y trait values
 replicate Replicate population indicator. Each blocks of rows with a given replicate number is a simulated population with the given parameters

Author(s)

Rod Ball <rod.ball@scionresearch.com> www.scionresearch.com

References

Luo, Z.W. 1988 Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Heredity* 80, 198–208.

See Also

[luo.ld.power](#)

Examples

```
# Power from stochastic simulation for Luo's population 12
data(luo.ld.populations)
luo.pop12.sim <- ld.sim(nsim=3000,
                      n=luo.ld.populations[12,"n"],
                      p=luo.ld.populations[12,"p"],
                      q=luo.ld.populations[12,"q"],
                      D=luo.ld.populations[12,"D"],
                      h2=luo.ld.populations[12,"h2"],
                      phi=luo.ld.populations[12,"phi"],
                      Vp=100)

# power
table(luo.pop12.sim[,4] < 0.05)[2]/sum(table(luo.pop12.sim[,4] < 0.05))
# Cf power from deterministic calculation
luo.ld.power(n=luo.ld.populations[12,"n"],
            p=luo.ld.populations[12,"p"],
            q=luo.ld.populations[12,"q"],
            D=luo.ld.populations[12,"D"],
            h2=luo.ld.populations[12,"h2"],
            phi=luo.ld.populations[12,"phi"],
            Vp=100,
            alpha=0.05)
```

luo.ld.populations *Luo's Linkage disequilibrium example populations*

Description

Matrix with rows containing parameters (population size, allele frequencies, disequilibrium, dominance ratio) for example populations with a bi-allelic marker and QTL in linkage disequilibrium, from Luo (1998).

Usage

```
data(luo.ld.populations)
```

Format

The format is: num [1:12, 1:7] 1 2 3 4 5 6 7 8 9 10 ... - attr(*, "dimnames")=List of 2 ..\$: NULL
 ..\$: chr [1:7] "pop." "n" "p" "q" ...

Source

Luo, Z.W. 1988 Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Heredity* 80, 198–208.

Examples

```
data(luo.ld.populations)
luo.ld.populations
```

luo.ld.power *Classical deterministic power calculation for association studies to detect linkage disequilibrium*

Description

Classical deterministic power calculation for power to detect linkage disequilibrium between a bi-allelic QTL and a bi-allelic marker, at a given significance level in a population level association study.

Usage

```
luo.ld.power(n, p, q, D, h2, phi, Vp = 100, alpha, print.it = TRUE,
             missclass.rate = 0)
```

Arguments

n	The sample size, i.e. number of individuals genotyped and tested for the trait of interest
p	Bi-allelic marker allele frequency
q	Bi-allelic QTL allele frequency
D	Linkage disequilibrium coefficient
h ²	QTL 'heritability', i.e. proportion of total or phenotypic variance explained by the QTL
phi	Dominance ratio: phi = 0 denotes purely additive, phi = 1 denotes purely dominant allele effects
Vp	Total or phenotypic variance: and arbitrary value may be used
alpha	Significance level for hypothesis tests
print.it	If TRUE print summary of results
missclass.rate	Proportion of marker values which are missclassified, i.e. incorrect

Details

This is based on the 'fixed model' power calculation from Luo (1998, *Heredity* 80, 198–208), with corrections described in Ball (2005). This function, combined with [oneway.bf.alpha](#), [oneway.bf.alpha.required](#), is used in Ball (2005) to design experiments for detecting linkage disequilibrium with a given power to achieve a given Bayes factor.

Value

Returns the power, or probability of detecting an effect, with the given parameters, at the given significance level.

Author(s)

Rod Ball <rod.ball@scionresearch.co.nz> www.scionresearch.com

References

- Ball, R.D. 2005: Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. *Genetics* 170: 859–873.
- Ball, R.D. 2007: Statistical analysis and experimental design. Chapter 8, pp133–196 In: *Association Mapping in Plants*, N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. DeSilva (Editors), Springer, New York.
- Luo, Z.W. 1988: Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Heredity* 80, 198–208

See Also

[ld.sim](#), [oneway.bf.alpha](#), [oneway.bf.alpha.required](#), [SS.oneway.bf](#)

Examples

```

data(luo.ld.populations)
options(digits=3)
powers <- numeric(nrow(luo.ld.populations))
for(ii in 1:nrow(luo.ld.populations)){
  cat("ii=", ii, "\n")
  powers[ii] <- luo.ld.power(n=luo.ld.populations[ii,"n"],
                           p=luo.ld.populations[ii,"p"],
                           q=luo.ld.populations[ii,"q"],
                           D=luo.ld.populations[ii,"D"],
                           h2=luo.ld.populations[ii,"h2"],
                           phi=luo.ld.populations[ii,"phi"],
                           Vp=100,
                           alpha=0.05)
}
cbind(luo.ld.populations,power=powers)

```

oneway.bf.alpha	<i>Correspondence between significance levels and Bayes factors for effects of marker genotype classes.</i>
-----------------	---

Description

Functions to calculate the correspondence between significance levels alpha and the Bayes factor, for association between a bi-allelic marker and QTL, for given sample sizes and marker genotype frequencies for bi-allelic marker.

Usage

```

oneway.bf.alpha(n, group.sizes = c(0.25, 0.5, 0.25) * n, alpha = 0.05)
oneway.bf.alpha.required(n, group.sizes = c(0.25, 0.5, 0.25) * n, Bf)

```

Arguments

n	Sample size, i.e. number of individuals genotyped and phenotyped for the trait
group.sizes	Number in each of the 3 possible marker genotype classes MM, Mm, mm
alpha	Significance level, i.e. threshold for ‘detection’
Bf	Bayes factor, used as threshold for detection

Details

These functions implement the correspondence between the significance levels and Bayes factors used in Ball (2005) to design experiments for detecting linkage disequilibrium with a given power to achieve a given Bayes factor. The function `SS.oneway.bf` is used to calculate the Bayes factor corresponding to a given F statistic (Spiegelhalter and Smith 1982). This is combined with a call to `qf`, for `oneway.bf.alpha` or calls to `pf` and interpolation for `oneway.bf.alpha.required`, to calculate the Bayes factor corresponding to a given alpha or alpha values for a given Bayes factor.

Value

oneway.bf.alpha returns the Bayes factor corresponding to a given significance level (alpha).
 oneway.bf.alpha.required returns the significance level (alpha) corresponding to a given Bayes factor.

Author(s)

Rod Ball <rod.ball@scionresearch.com> www.scionresearch.com

References

- Ball, R.D. 2005: Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. *Genetics* 170: 859–873.
- Ball, R.D. 2007: Statistical analysis and experimental design. Chapter 8, pp133–196 In: *Association Mapping in Plants*, N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. DeSilva (Editors), Springer, New York.
- Spiegelhalter, D. and A.F.M. Smith 1982 Bayes factors for linear and log-linear models with vague prior information *J. Royal Statist Soc. B* 44: 377–387.

See Also

[SS.oneway.bf,luo.ld.power](#)

Examples

```
# calculations for Table 4 in the manuscript
data(luo.ld.populations)
Bs <- numeric(nrow(luo.ld.populations))
n.Bf20s <- numeric(nrow(luo.ld.populations))
ns <- c(seq(200,400,by=25),450,seq(500,4000,by=100))
powers <- numeric(length(ns))
alphas <- numeric(length(ns))
P.Bf20s <- numeric(length(ns))
for(ii in 1:nrow(luo.ld.populations)){
  cat("ii=",ii,"\n")
  powers[ii] <- luo.ld.power(n=luo.ld.populations[ii,"n"],
                           p=luo.ld.populations[ii,"p"],
                           q=luo.ld.populations[ii,"q"],
                           D=luo.ld.populations[ii,"D"],
                           h2=luo.ld.populations[ii,"h2"],
                           phi=luo.ld.populations[ii,"phi"],
                           Vp=100,
                           alpha=0.05)

  p1 <- luo.ld.populations[ii,"p"]
  Bs[ii] <- oneway.bf.alpha(n=luo.ld.populations[ii,"n"],
                          group.sizes=c(p1^2,2*p1*(1-p1),(1-p1)^2)*
                          luo.ld.populations[ii,"n"])
  for(jj in seq(along=ns)){
    alphas[jj] <- oneway.bf.alpha.required(ns[jj],
```

```

      group.sizes=c(p1^2,2*p1*(1-p1),(1-p1)^2)*ns[jj],Bf=20)
P.Bf20s[jj] <- luo.ld.power(n=ns[jj],
      p=luo.ld.populations[ii,"p"],
      luo.ld.populations[ii,"q"],
      D=luo.ld.populations[ii,"D"],
      h2=luo.ld.populations[ii,"h2"],
      phi=luo.ld.populations[ii,"phi"],
      Vp=100,
      alpha=alphas[jj],
      print.it=FALSE)
}
n.Bf20s[ii] <- approx(P.Bf20s,ns,xout=0.9)$y
cat("n =",n.Bf20s[ii],"\n")
}
cbind(luo.ld.populations,powers,n.Bf20s)

```

SS.oneway.bf

Bayes factors for one-way analysis of variance models.

Description

Function to calculate the Bayes factor for a one-way analysis of variance layout with vague or improper priors.

Usage

```
SS.oneway.bf(group.sizes, Fstat)
```

Arguments

group.sizes	Sizes of groups in the one-way layout
Fstat	F statistic obtained

Details

The function the Bayes factor corresponding to a given F statistic in a one-way analysis of variance model is calculated using the method of Spiegelhalter and Smith 1982. With improper priors the marginal probabilities of the data under each of the models (corresponding to the NULL and alternative hypotheses) is indeterminate. This is resolved by updating each prior with a small imaginary training sample, which is equivalent to normalising the Bayes factor to be 1 for the small training sample. Spiegelhalter and Smith obtain a formula for the Bayes factor in terms of the classical F value.

Value

Returns the Bayes factor corresponding to the given design and observed value of F statistic.

Author(s)

Rod Ball <rod.ball@scionresearch.com> www.scionresearch.com

References

Spiegelhalter, D. and A.F.M. Smith 1982 Bayes factors for linear and log-linear models with vague prior information J. Royal Statist Soc. B 44: 377–387.

See Also

[oneway.bf.alpha](#), [oneway.bf.alpha.required](#)

Examples

```
# Bayes factors corresponding to P-values 0.05,0.01,0.001,0.0001 for n=200
SS.oneway.bf(group.sizes=c(50,100,50),Fstat=qf(0.95,2,197))
SS.oneway.bf(group.sizes=c(50,100,50),Fstat=qf(0.99,2,197))
SS.oneway.bf(group.sizes=c(50,100,50),Fstat=qf(0.999,2,197))
SS.oneway.bf(group.sizes=c(50,100,50),Fstat=qf(0.9999,2,197))
```

Index

*Topic **datasets**

luo.ld.populations, [12](#)

*Topic **design, htest**

calc.B.ABF, [2](#)

*Topic **design**

cc.design, [3](#)

ld.design, [7](#)

ld.sim, [9](#)

luo.ld.power, [12](#)

*Topic **htest**

oneway.bf.alpha, [14](#)

SS.oneway.bf, [16](#)

*Topic **models**

oneway.bf.alpha, [14](#)

SS.oneway.bf, [16](#)

calc.alphaB.ABF (calc.B.ABF), [2](#)

calc.B.ABF, [2](#)

calc.Balpha.ABF (calc.B.ABF), [2](#)

calc.Zalpha.ABF (calc.B.ABF), [2](#)

calc.Zc.ABF (calc.B.ABF), [2](#)

cc.design, [3](#), [3](#), [6](#)

cc.power, [2](#), [3](#), [6](#)

cc.power (cc.design), [3](#)

gpc.power, [2](#), [6](#)

ld.design, [7](#)

ld.power (ld.design), [7](#)

ld.sim, [9](#), [9](#), [13](#)

ld.sim1 (ld.sim), [9](#)

luo.ld.populations, [12](#)

luo.ld.power, [8–11](#), [12](#), [15](#)

oneway.bf.alpha, [9](#), [13](#), [14](#), [17](#)

oneway.bf.alpha.required, [8](#), [9](#), [13](#), [17](#)

SS.oneway.bf, [8](#), [9](#), [13–15](#), [16](#)