Package 'HWEBayes'

February 19, 2015

Title Bayesian investigation of Hardy-Weinberg Equilibrium via estimation and testing.

Version 1.4

Date 2012-12-17

Imports MCMCpack, mvtnorm

Description Estimation and testing of HWE using Bayesian methods. Three models are currently considered: HWE, a model parameterized in terms of the allele frequencies and a single inbreeding coefficient f, and the saturated model. Testing is based on Bayes factors.

License GPL-2

LazyLoad yes

Author Jon Wakefield [aut, cre], Martyn Plummer [ctb]

Maintainer Jon Wakefield <jonno@u.washington.edu>

NeedsCompilation no

Repository CRAN

Date/Publication 2013-12-18 12:03:46

R topics documented:

`baselogit` *Calculates a set of baseline logits from a set of probabilities*

Description

Calculates a set of $k-1$ baseline logits $\log(p_1/p_k), ..., \log(p_{k-1}/p_k)$, from a set of probabilities $p_1, ..., p_k$.

Usage

```
baselogit(probs)
```

Arguments

`probs`	A set of probabilities, $p_1, ..., p_k$, where k is the number of alleles.

Details

This function is used by a number of other functions in the package, for example, to provide a parameterization for maximization and for importance sampling in the single f model.

Value

`baselogit`	Returns the set of $k-1$ baseline logits, where k is the number of alleles.

Author(s)

Jon Wakefield (jonno@u.washington).

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

`invbaselogit`

Examples

```
baselogit(probs=c(0.5,0.4,0.1))
```

DiabRecess *Data on diabetes patients antigen classes*

Description

Data are from Thomson et al. (1986) and describe the counts of combinations of four different antigen classes in 45 French type I diabetes patients. HWE is equivalent to a recessive model, see Thomson (1983) and Wakefield (2009) for more details.

Usage

```
data(DiabRecess)
```

Format

A vector with 10 observations in the order $n_{11}, n_{11}, n_{13}, n_{14}, n_{22}, n_{23}, n_{24}, n_{33}, n_{34}, n_{44}$.

Source

Thomson, G. et al (1986). HLA and IDDM predisposition: new aspects. Genetic Epidemiology, 1, 262-368.

References

Thomson, G. (1983). Investigation of the mode of inheritance of the HLA associated diseases by the method of antigen genotype frequencies among diseased individuals. Tissue Antigens, 21, 81-104.

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

Examples

```
data(DiabRecess)
```

DirichNormHWE *Evaluates the normalizing constant under the HWE model, with a conjugate prior*

Description

Function to evaluate the normalizing constant given a conjugate Dirichlet prior and the HWE model.

Usage

```
DirichNormHWE(nvec, bvec0)
```

Arguments

nvec	vector of genotype frequencies in the order $n_{11}, n_{12}, ..., n_{1k}, n_{22}..., n_{2k}, ..., n_{kk}$.
bvec0	vector of length k Dirichlet prior parameters, where k is the number of alleles.

Value

The normalizing constant.

Author(s)

Jon Wakefield (jonno@u.washington).

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

DirichNormSat

Examples

```
data(DiabRecess)
DirichNormHWE(nvec=DiabRecess,bvec0=rep(1,4))
```

`DirichNormSat` *Evaluates the normalizing constant (as used in the denominator of a Bayes factor) for a conjugate prior*

Description

Function to evaluate the normalizing constant given a conjugate Dirichlet prior and a saturated model.

Usage

```
DirichNormSat(nvec, bvec)
```

Arguments

`nvec`	vector of genotype frequencies in the order $n_{11}, n_{12}, ..., n_{1k}, n_{22}..., n_{2k}, ..., n_{kk}$.
`bvec`	vector of length $k(k+1)/2$ Dirichlet prior parameters, where k is the number of alleles.

Value

The normalizing constant.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

`DirichSampSat`, `DirichSampHWE`, `DirichNormSat`, `DirichNormHWE`, `HWEDirichBF2`

Examples

```
data(DiabRecess)
DirichNormSat(nvec=DiabRecess,bvec=rep(1,10))
```

DirichSampHWE *Simulate samples from a Dirichlet prior or posterior under HWE*

Description

Function to simulate samples from the HWE Dirichlet model. Can be used for samples from the prior or the (conjugate) Dirichlet posterior, both in the k allele case. Samples are generated for the allele frequencies in the order $p_1, p_2, ..., p_k$.

Usage

```
DirichSampHWE(nvec, bvec0, nsim)
```

Arguments

nvec	vector of genotype frequencies in the order $n_{11}, n_{12}, ..., n_{1k}, n_{22}..., n_{2k}, ..., n_{kk}$.
bvec0	vector of length k Dirichlet prior parameters, where k is the number of alleles.
nsim	number of samples to simulate from the prior/posterior.

Details

Uses the rdirichlet function from the MCMCpack library.

Value

pvec	matrix of size nsim $\times k$ containing samples for the genotype frequencies, in the order $p_1, p_{12}, ..., p_k$.

Author(s)

Jon Wakefield (jonno@u.washington).

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

DirichSampSat, DirichNormSat, DirichNormHWE

Examples

```
# First sample from the prior
PriorSampHWE <- DirichSampHWE(nvec=rep(0,10),bvec0=rep(1,4),nsim=1000)
par(mfrow=c(1,1))
hist(PriorSampHWE$pvec[,1],xlab="p1",main="")
# Now sample from the posterior
data(DiabRecess)
PostSampHWE <- DirichSampHWE(nvec=DiabRecess,bvec0=rep(1,4),nsim=1000)
par(mfrow=c(1,1))
hist(PostSampHWE$pvec[,1],xlab="p1",main="")
```

`DirichSampSat` *Simulate samples from a Dirichlet prior or posterior under the saturated model*

Description

Function to simulate samples from the satuated Dirichlet model. Can be used for samples from the prior or the (conjugate) Dirichlet posterior, both in the k allele case. Samples are generated for the genotype frequencies in the order $p_{11}, p_{12}, ..., p_{1k}, p_{22}..., p_{2k}, ..., p_{kk}$, the allele frequencies, and the fixation indices.

Usage

```
DirichSampSat(nvec, bvec, nsim)
```

Arguments

`nvec`	vector of genotype frequencies in the order $n_{11}, n_{12}, ..., n_{1k}, n_{22}..., n_{2k}, ..., n_{kk}$.
`bvec`	vector of length $k(k+1)/2$ Dirichlet prior parameters, where k is the number of alleles.
`nsim`	number of samples to simulate from the prior/posterior.

Details

Uses the `rdirichlet` function from the `MCMCpack` library.

Value

`pvec`	matrix of size `nsim` $\times k(k+1)/2$ containing samples for the genotype frequencies, in the order $p_{11}, p_{12}, ..., p_{1k}, p_{22}..., p_{2k}, ..., p_{kk}$.
`pmat`	matrix of size `nsim` $\times k(k+1)/2 \times k(k+1)/2$ containing samples for the genotype probabilities.
`pmarg`	matrix of size `nsim` $\times k$ containing samples for the allele frequencies, in the order $p_1, ..., p_k$.
`fixind`	matrix of size `nsim` $\times k(k+1)/2 \times k(k+1)/2$ containing samples for the fixation indices, contained in the lower diagonal, i.e. `fixind[,i,j]` for `[i>j]`.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

`DirichSampHWE`, `DirichNormSat`, `DirichNormHWE`

Examples

```
# First sample from the prior
PriorSampSat <- DirichSampSat(nvec=rep(0,10),bvec=rep(1,10),nsim=1000)
par(mfrow=c(1,2))
hist(PriorSampSat$pvec[,1],xlab="p1",main="")
hist(PriorSampSat$fixind[,2,1],xlab="f21",main="")
# Now sample from the posterior
data(DiabRecess)
PostSampSat <- DirichSampSat(nvec=DiabRecess,bvec=rep(1,10),nsim=1000)
par(mfrow=c(1,2))
hist(PostSampSat$pvec[,1],xlab="p1",main="")
hist(PostSampSat$fixind[,2,1],xlab="f21",main="")
```

`HWEDirichBF2` *Evaluates the Bayes factor in the k=2 allele case under conjugate priors*

Description

Function to evaluate the Bayes factor $\Pr(n|\text{ HWE }) / \Pr(n|\text{ saturated model })$ in the $k = 2$ allele case and with conjugate (Dirichlet) priors under HWE and saturated models.

Usage

```
HWEDirichBF2(nvec, bvec0, bvec1)
```

Arguments

`nvec`	vector of genotype frequencies in the order n_{11}, n_{21}, n_{22}.
`bvec0`	vector of length $k = 2$ Dirichlet prior parameters for the prior under the null, where k is the number of alleles.
`bvec1`	vector of length $k(k+1)/2 = 3$ Dirichlet prior parameters for the prior under the saturated model, where k is the number of alleles.

Value

Bayes factor is returned.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

DirichNormHWE, DirichNormSat, DirichSampHWE, DirichSampSat, HWETriangBF2, TriangNormHWE

Examples

```
HWEDirichBF2(nvec=c(88,10,2),bvec0=c(1,1),bvec1=c(1,1,1))
```

HWEImportSamp — *Importance sampling to calculate the normalizing constant under the single f model*

Description

Importance sampling to calculate the normalizing constant under the single f model. Two proposals are available, either sampling from the prior or sampling from a normal distribution whose mean vector and variance-covariance matrix must be specified. The latter may be taken from an MCMC analysis using, for example, WinBUGS. In all cases the likelihood is multinomial and the prior is Dirichlet on the allele frequencies, and normal on λ where $\lambda = \log(f - f_{\min})/(1 - f))$. See Weir (1996) for a description of HWE and different models/parameterizations.

Usage

```
HWEImportSamp(nsim, nvec, ischoice, lambdamu, lambdasd, alpha,
gmu = rep(0, length(alpha)), gsigma = diag(0, nrow = length(alpha),
ncol = length(alpha)))
```

Arguments

nsim	the number of points to sample to calculate the estimate.
nvec	vector of genotype frequencies in the order $n_{11}, n_{12}, ..., n_{1k}, n_{22}..., n_{2k}, ..., n_{kk}$.
ischoice	choice of importance sampling proposal, =1 gives a normal distribution with mean and variance that must be specified (as gmu and gsigma) and =2 is from the prior.
lambdamu	the mean of the prior for λ.

lambdasd	the variance of the prior for λ.
alpha	the vector of k parameters for the Dirichlet prior on the allele frequencies.
gmu	the mean of the importance sampling proposal, of length k, where k is the number of alleles.
gsigma	the variance of the importance sampling proposal, a matrix of dimension $k \times k$, where k is the number of alleles.

Value

PrnH1	the estimate of the normalizing constant
varest	the variance of the estimate of the normalizing constant

Warning

As always with importance sampling the procedure can be very unstable, particularly for large k. Hence rerunning the function with different simulation sample sizes, and different gmu and gsigma is recommended

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

Weir, B.S. (1996). Genetic Data Analysis II. Sunderland MA: Sinauer.

See Also

LambdaOptim, DirichNormSat, DirichNormHWE, TriangNormHWE

Examples

```
alpha <- c(1,1,1,1) # prior on allele frequencies
# gmu and gsigma were obtained from a WinBUGS run
gmu <- c(-0.4633092,0.3391625,0.3397936,-3.5438008)
gsigma <- matrix(c(
0.07937341,0.02819656,0.02766583,0.04607996,
0.02819656,0.07091320,0.04023827,0.01657028,
0.02766583,0.04023827,0.07042278,0.01752266,
0.04607996,0.01657028,0.01752266,0.57273683),nrow=4,ncol=4)
data(DiabRecess)
HWEImportSamp(nsim=5000,nvec=DiabRecess,ischoice=1,lambdamu=-2.95,
   lambdasd=1.07,alpha=alpha,gmu,gsigma)
HWEImportSamp(nsim=5000,nvec=DiabRecess,ischoice=2,lambdamu=-2.95,
   lambdasd=1.07,alpha=alpha)
```

HWEmodelsMLE *Evaluates the maximum likelihood estimates of the parameters of various models in the k allele case*

Description

Function to obtain the MLEs of parameters under the HWE, single f and saturated models. For the single f model numerical maximization is required if $k > 2$, where k is the number of alleles.

Usage

```
HWEmodelsMLE(nvec)
```

Arguments

nvec	vector of genotype frequencies in the order $n_{11}, n_{21}, n_{22}, ..., n_{k1}, n_{k2}, ..., n_{kk}$.

Value

phat	matrix of $k \times k$ MLEs of genotype frequencies
qhat	MLEs of k allele frequencies under the HWE model
fqhat	MLEs of k allele frequencies under the single f model
fsingle	MLE of single f
fmaxloglik	maximized log-likelihood (without the normalizing constant), under the single f model
fmin	estimated lower bound of $f_{\min}$ in the single f model. Under the single f model $f_{\min} < f < 1$ where $f_{\min} = -p_{\min}/(1 - p_{\min})$ and $p_{\min}$ is the minimum of the allele frequencies.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

Weir, B.S. (1996). Genetic Data Analysis II. Sunderland MA: Sinauer.

Examples

```
data(DiabRecess)
HWEmodelsMLE(nvec=DiabRecess)
```

`HWEsimdat` *Simulate data under the single f model with k alleles.*

Description

Simulate data under the single f model with k alleles (so $f = 0$ gives data under HWE).

Usage

```
HWEsimdat(npop, q, f)
```

Arguments

`npop`	population size.
`q`	vector of k allele frequencies.
`f`	value of inbreeding coefficient

Value

`nvec`	vector of genotype counts, in the order $n_{11}, n_{21}, n_{22}, ..., n_{k1}, n_{k2}, ..., n_{kk}$.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

Weir, B.S. (1996). Genetic Data Analysis II. Sunderland MA: Sinauer.

Examples

```
counts <- HWEsimdat(100,q=c(0.1,0.8,.1),f=0.1)
```

HWETriangBF2 *Evaluates the Bayes factor in the k=2 allele case with a "triangular" prior under the null*

Description

Function to evaluate the Bayes factor $\Pr(n|\text{ HWE }) / \Pr(n|\text{ saturated model })$ in the $k = 2$ allele case and with a conjugate (Dirichlet) priors under the saturated model and a "triangular" distribution under the null. the latter is the marginal prior distribution under the (1,1,1) Dirichlet prior under the saturated model.

Usage

```
HWETriangBF2(nvec)
```

Arguments

nvec	vector of genotype frequencies in the order n_{11}, n_{21}, n_{22}.

Value

Bayes factor is returned.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

TriangNormHWE

Examples

```
HWETriangBF2(nvec=c(88,10,2))
```

`invbaselogit` *Converts a set of k-1 baseline logits into a set of probabilities*

Description

Converts a set of $k-1$ baseline logits $\log(p_1/p_k), ..., \log(p_{k-1}/p_k)$ into a set of probabilities $p_1, ..., p_k$, where k is the number of alleles.

Usage

```
invbaselogit(baselogit)
```

Arguments

`baselogit`	A set of $k-1$ baseline logits, where k is the number of alleles.

Details

This is used by a number of other functions in the package, for example, to provide a parameterization for maximization and for importance sampling in the single f model.

Value

`invbaselogit`	the probability vector corresponding to the baseline logit

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

`baselogit`

Examples

```
invbaselogit(baselogit=c(0,0))
```

`LambdaOptim` *Obtains values for the prior specification for lambda*

Description

In the single f model we may parameterize in terms of the allele frequencies and $\lambda = \log((f - f_{\min})/(1 - f))$ where $f_{\min} = -p_{\min}/(1 - p_{\min})$ and $p_{\min}$ is the minimum allele frequency. The prior for λ is assumed normal and this function finds the mean and standard deviation of this normal, given two values for f, with associated probabilities.

Usage

```
LambdaOptim(nsim, bvec, f1, f2, p1, p2, init)
```

Arguments

`nsim`	the optimization is carried out by simulating from the joint prior on allele frequencies and λ, and this argument gives the number of simulations to take from the prior
`bvec`	vector of length k of prior specification for the HWE Dirichlet prior, where k is the number of alleles.
`f1`	first quantile for inbreeding coefficient f
`f2`	second quantile for inbreeding coefficient f
`p1`	probability associated with `f1`
`p2`	probability associated with `f2`
`init`	initial values for `lambdamu` and `lambdasd`

Value

`lambdamu`	prior mean for λ
`lambdasd`	prior standard deviation for λ

Warning

This function can be unstable and good starting values may be needed. It is also recommended to check the output by simulating from the given prior to see if the empirical quantiles match with those desired; the function `SinglefPrior` may be used for this

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

HWEImportSamp

Examples

```
bvec <- c(1,1,1,1)
init <- c(-3,log(1.1))
lampr <- LambdaOptim(nsim=10000,bvec=bvec,f1=0,f2=0.26,p1=0.5,p2=0.95,init)
```

MultLogLik *Evaluates the Multinomial likelihood under the single f model*

Description

Evaluates the Multinomial likelihood under the single f model. The normalizing constant is not included. This function is called by a number of other functions, and should not be needed.

Usage

```
MultLogLik(x, nvec, paramch = 1)
```

Arguments

x	a set of $k-1$ baseline logits, where k is the number of alleles), and a transformed version of f. Hence a vector of length k. The transformation adopted depends on the value of paramch.
nvec	vector of genotype frequencies in the order $n_{11}, n_{21}, ..., n_{k1}, n_{22}, ..., n_{k2}, ..., n_{kk}$.
paramch	a variable that if =1 assumes f is on the range (-1,+1) before transformation, and if =2 assumes on the range $(f_{\min}, +1)$.

Value

MultLoglik	The value of the (unnormalized) multinomial log-likelihood.

Note

MultLogLikP also calculates the mulinimial likelihood using a different parameterization.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

Weir, B.S. (1996). Genetic Data Analysis II. Sunderland MA: Sinauer.

See Also

SinglefReject, MultLogLikP

MultLogLikP *Evaluates the Multinomial likelihood under the single f model*

Description

Evaluates the Multinomial likelihood under the single f model. The normalizing constant is not included. This function is called by a number of other functions, and should not be needed.

Usage

```
MultLogLikP(p, f, nvec)
```

Arguments

p	A set of probabilities, $p_1, ..., p_k$, where k is the number of alleles
f	The f parameter
nvec	vector of genotype frequencies in the order $n_{11}, n_{21}, ..., n_{k1}, n_{22}, ..., n_{k2}, ..., n_{kk}$.

Value

The unnormalized multinomial log-likelihood.

Note

MultLogLik also calculates the multinomial likelihood using an alternate parameterization.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

Weir, B.S. (1996). Genetic Data Analysis II. Sunderland MA: Sinauer.

See Also

SinglefReject,MultLogLik

SinglefPrior *Samples from the single f prior.*

Description

Function to sample from the single f prior, that is the Dirichlet and normal on λ, where $\lambda = \log((f - f_{\min})/(1 - f))$.

Usage

```
SinglefPrior(nsim, alpha, lambdamu, lambdasd)
```

Arguments

nsim	number of simulations from prior
alpha	vector of k parameters for the Dirichlet prior on the k allele frequencies.
lambdamu	mean of the normal prior on λ.
lambdasd	standard deviation of the normal prior on λ.

Value

p	sample for vector of k allele frequencies
f	sample of f parameters
lgts	samples for logits of baseline logits
lambda	samples for λ

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

SinglefReject, HWEsimdat

Examples

```
SinglefSamp <- SinglefPrior(nsim=1000,alpha=c(1,1,1,1),
  lambdamu=-2.95,lambdasd=1.07)
```

SinglefReject *Samples from the posterior for the single f model*

Description

Function to generate samples from the posterior for allele frequencies and f, under the single f model. Samples are generated using a rejection algorithm that simulates from the prior.

Usage

```
SinglefReject(nsim, bvec, lambdamu, lambdasd, nvec)
```

Arguments

nsim	number of samples to generate from the prior.
bvec	vector of size k that is the specification for the Dirichlet prior on the allele frequencies.
lambdamu	prior mean for λ.
lambdasd	prior standard deviation for λ.
nvec	vector of genotype frequencies in the order $n_{11}, n_{21}, n_{22}, ..., n_{k1}, n_{k2}, ..., n_{kk}$.

Value

psamp	samples for k allele frequencies.
fsamp	samples for inbreeding coefficient f.
accrate	acceptance rate of the rejection algorithm.
PrnH1	estimate of normalizing constant (which may be used in Bayes factor calculations). Calculated by averaging the likelihood over the sampled points.
varest	estimated variance of the estimate of the normalizing constant.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

Examples

```
data(DiabRecess)
postsampf1 <- SinglefReject(nsim=100,bvec=rep(1,4),lambdamu=-2.95,
   lambdasd=1.07,nvec=DiabRecess)
```

`TriangNormHWE` *Evaluates the normalizing constant under the HWE model, for the "triangular" prior distribution*

Description

Function to evaluate the normalizing constant given a "triangular" prior and the HWE model, in the $k = 2$ allele case. This prior results from marginalizing the conjugate Dirichlet prior with parameters (1,1,1) on the genotype frequencies under the alternative.

Usage

```
TriangNormHWE(nvec)
```

Arguments

`nvec`	vector of genotype frequencies in the order n_{11}, n_{21}, n_{22}.

Value

Normalizing constant is returned.

Author(s)

Jon Wakefield (jonno@u.washington.edu)

References

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics; Vol 66:257-65

See Also

`DirichNormHWE`, `DirichNormSat`, `DirichSampHWE`, `DirichSampSat`, `HWEDirichBF2`, `TriangNormHWE`

Examples

```
nvec <- c(88,10,2)
TriangNormHWE(nvec)
```

Index