Package 'LEAPFrOG'

February 19, 2015

Title Likelihood Estimation of Admixture in Parents From Offspring Genotypes

Version 1.0.7

Date 2014-08-11

Author Daniel JM Crouch & Michael E Weale

Maintainer Daniel JM Crouch <daniel.crouch@oncology.ox.ac.uk>

Description Contains LEAPFrOG Gradient Optimisation and Expectation Maximisation functions for inferring parental admixture proportions from an offspring with SNP genotypes.

URL http://sites.google.com/site/mikeweale

Depends alabama, MASS

Suggests rjags

License GPL

NeedsCompilation no

Repository CRAN

Date/Publication 2014-08-13 16:42:15

R topics documented:

LEAPFrOG-package *Likelihood Estimation of Admixture in Parents From Offspring Genotypes*

Description

Takes genotype data for a single individual, and allele frequencies for multiple populations, and returns estimations of admixture proportions for the individual, as well as the admixture proportion of their two parents.

Details

Package:	LEAPFROG
Type:	Package
Version:	1.0
Date:	2011-06-08
License:	GPL
LazyLoad:	yes

Use LEAPFrOG when regular genotype data is available (0,1 or 2 alleles at each SNP). Use LEAPFrOG_EM when phased haplotypes are available. Plot results using LEAPFrOG_plot

Author(s)

Daniel Crouch and Michael Weale

Maintainer: Daniel Crouch - daniel.crouch@oncology.ox.ac.uk

References

Crouch & Weale (2011), European Journal of Human Genetics

See Also

LEAPFrOG_plot,LEAPFrOG_EM,LEAPFrOG,BEAPFrOG

BEAPFrOG *Bayesian Estimation of Admixture in Parental From Offspring Genotype*

Description

Provides estimates of admixture proportions in offspring and ungenotyped parents, using genotype data

Usage

```
BEAPFrOG(data,p,nchains=1,iterations=1000,alpha=0.05,prior=1,burn=2000,SampSizes)
```

Arguments

data	Vector of allele counts: each element either 0,1,2 or NA.
p	Matrix of allele frequencies. Each row corresponds with a SNP. Number of rows must equal length of data. Each column is a population
nchains	Number of Markov Chains to perform gibbs sampling
iterations	Number of samples for each Markov Chain
alpha	1 minus the width of credible interval taken around the posterior mode.
prior	Concentration parameter for the dirichlet distribution. 1 is uninformative, small values suggest first-generation admixture (uninformative as to which source populations involved), and higher values suggest low parental divergence.
burn	Burn- in period. The number of MCMC samples to discard.
SampSizes	Vector of the same length as the number of source populations. Each element is the number of individuals used to calculate allele frequencies in that population. Parameterises the prior on allele frequencies

Details

BEAPFrOG requires jags and rjags to be installed. http://mcmc-jags.sourceforge.net/ Credible intervals are centered around the mode.

Values for parameter vectors m1 and m2 can be exchanged to give identical likelihoods. This may give rise to bimodal posterior distributions, particularly with first-generation admixture, and the resulting credible intervals are not useful. Therefore, for all MCMC samples, we redefine m1 as the admixture proportions for the parent with admixture from population 1 less than 0.5, and vice-versa for m2.

Value

A list including elements

P1est	A vector: Posterior mode for admixture proportions in parent 'A'.
P2	A vector: Posterior mode for admixture proportions in parent 'B'.
P1i	A matrix of two columns and number of rows equal to number of populations: upper and lower credible intervals of width defined by the argument alpha, for admixture proportions in parent 'A'.
P2i	A matrix of two columns and number of rows equal to number of populations: upper and lower credible intervals of width defined by the argument alpha, for admixture proportions in parent 'B'.
Monitor	All stored MCMC samples. Use plot() on this object for visualisation of posterior distribution and MCMC trace

Author(s)

Daniel Crouch & Michael Weale, Department of Medical and Molecular Genetics, King's College London

See Also

LEAPFrOG,LEAPFrOG_plot

---

constrOptim2 *constrOptim2*

Description

By Ravi Varadhan. Constrained optimisation for likelihood function, returning hessian matrix used for parameter standard errors.

LEAPFrOG *LEAPFrOG*

Description

Provides estimates of admixture proportions and parental divergence of these admixture proportions

Usage

```
LEAPFrOG(data,p,Nudge=0.001,NonLinCon=TRUE)
```

Arguments

data	Vector of allele counts: each element either 0,1,2 or NA.
p	Matrix of allele frequencies. Each row corresponds with a SNP. Number of rows must equal length of data. Each column is a population
Nudge	D for population 1 will be initialised at 0.5+Nudge. Nudge must be greater than 0. In theory the value for Nudge shouldn't affect the final optimum, but may influence the time to convergence. Default is 0.001.
NonLinCon	If TRUE (default), the auglag optimisation function is invoked with a nonlinear constraint imposed on D*m, preventing impossible admixture totals of >1 in the parents. We strongly advise this option

Details

Standard errors returned in the order P-1 m parameters followed by P-1 D parameters. m and D for the Pth population are not estimated directly and have no standard error.

Value

A list including elements

m	A vector of admixture proportions in the genotyped offspring, one proportion per population. These sum to 1.
D	A vector of parental divergence paramaters, one per population.
mse	A vector of length number of populations-. Standard errors for all m estimates save the last populaion
Dse	A vector of length number of populations-. Standard errors for all D estimates save the last populaion
P1	Admixture proportions for each population, for parent 'A', derived from the m and D estimates.
P2	Admixture proportions for each population, for parent 'B', derived from the m and D estimates
value	Value of the optimised likelihood function.
counts	Number of times the likelihood function and gradient function were called during optimisation.

Author(s)

Daniel Crouch & Michael Weale, Department of Medical and Molecular Genetics, King's College London

See Also

LEAPFrOG_plot,LEAPFrOG_EM,BEAPFrOG

Examples

```
#Example with nonsense data -
#10000 random SNP genotypes
#...and uniform, random allele frequencies from two populations.
library(LEAPFrOG)
z1=LEAPFrOG(sample(0:2,10000,replace=TRUE),cbind(runif(10000,0,1),runif(10000,0,1)))
z1
```

LEAPFrOG_EM *LEAPFrOG Expectation Maximisation*

Description

Provides estimates of admixture proportions in offspring and ungenotyped parents, using phased data

Usage

```
LEAPFrOG_EM(data,p,chr,alpha=1e-6)
```

Arguments

data	2 column matrix of allele counts, with each row as a SNP. Columns 1 and 2 refer to the 2 haplotypes. Each entry is either 1, 0 or NA.
p	Matrix of allele frequencies. Each row corresponds with a SNP. Number of rows must equal length of data. Each column is a population
chr	Vector of chromosome identifiers, one for each SNP. Each entry is an integer, 1-22 for the autosomes. If two X chromosomes for a female are being analysed, it should be identified by the number 23.
alpha	Convergence tolerance for the EM algorithm. The optimisation will stop when an interation fails to change the parental admixture proportions (total change across all parameters) by this amount

Details

LEAPFrOG_EM requires python to be installed. Only the parental admixture proportions are estimated directly (all except the last population), and therefore standard errors are only reported for these only.

Value

A list including elements

m	A vector: Admixture proportions (one per population) for the genotyped offspring.
P1	A vector: Admixture proportions (one per population) for parent 'A'.
P2	A vector: Admixture proportions (one per population) for parent 'B'.
P1se	A vector: Standard errors (All populations except the last) foradmixture proportions in parent 'A'.
P2se	A vector: Standard errors (All populations except the last) foradmixture proportions in parent 'A'.
iterations	Number of expectation steps performed.
value	Value of the likelihood function for the final maximisation step.

Author(s)

Daniel Crouch & Michael Weale, Department of Medical and Molecular Genetics, King's College London

See Also

LEAPFrOG,LEAPFrOG_plot,BEAPFrOG

LEAPFrOG_plot *LEAPFrOG plotting function*

Description

Plots offspring and parental admiture proportions in the style of STRUCTURE, the popular population genetic software.

Usage

```
LEAPFrOG_plot(Results,PopNames,SampNames=NULL)
```

Arguments

Results	Array of dimensions 3*Npopulations*Noffspring. The first row is for the genotyped offspring and the second two for the unobserved parents. Each cell contains an admixture proportion.
PopNames	Character vector of length J (number of reference populations), Eg. c("Africa","Asia","Europe"). Order of names should correspond with order of parameters in Results.
SampNames	Character vector of sample names, equal to number of rows in Results, or NULL (detault), which will be printed underneath the admixture bars. Most useful when dealing with a small number of samples with distinct identity e.g. c("Hair","Blood","Door Handle","Ballroom","Lead Piping"). If NULL then no labels are printed beneath the plot (more appropriate for simulations or large population samples)

Author(s)

Daniel Crouch & Michael Weale, Department of Medical and Molecular Genetics, King's College London

See Also

LEAPFrOG,LEAPFrOG_EM,BEAPFrOG

Examples

```
#Example with nonsense data -
#10000 random SNP genotypes
#...and uniform, random allele frequencies from two populations.
library(LEAPFrOG)
#Get LEAPFrOG parameter estimates  for 10 simulated individuals
Results=array(dim=c(3,2,10))
for(i in 1:10){
z1=LEAPFrOG(sample(0:2,10000,replace=TRUE),cbind(runif(10000,0,1),runif(10000,0,1)))
Results[1,,i]=z1$m    #Offspring
Results[2,,i]=z1$P1 #Parent 'A'
Results[3,,i]=z1$P2 #Parent 'B
}
```

```
#Now plot these 10 individuals
LEAPFrOG_plot(Results,PopNames=c("PopA","PopB"))
#With sample names:
names=c("Hair","Blood","Door Handle","Ballroom","Lead Piping")
names=c(names,"Briefcase","Toothbrush","Sock","Shirt","Skin")
LEAPFrOG_plot(Results,PopNames=c("PopA","PopB"),SampNames=names)
```

Index