

Package: discent (via r-universe)

June 16, 2024

Type Package

Title Estimation of Deme Inbreeding Spatial Coefficients with Gradient Descent

Version 0.5.0

Description In the early 1970s, Malécot described the relationship between genetic relatedness and physical distance, forming the framework of isolation by distance, or -- put simply -- pairs that are far apart are less likely to mate. Capitalizing on this framework by using measures of Identity by Descent, we produce a deme inbreeding spatial coefficient (DISC) using ``vanilla" gradient descent. For the mathematical formulation of the of DISC, see: <TODO>. Briefly, we assume that the relatedness between two locations (demes) in space is given by the average pairwise IBD between the two locations conditional on the distance that separates them. Further, we assume that geographic distance is scaled by a migration rate, which is a global parameter among all spatial locations.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports Rcpp, dplyr, magrittr, tidyr, purrr, utils, crayon

Suggests stats, broom, knitr, rmarkdown, testthat (>= 2.1.0), tibble, covr

VignetteBuilder knitr

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

LinkingTo Rcpp

SystemRequirements C++11

Depends R (>= 2.10)

Repository <https://plasmogenepi.r-universe.dev>

RemoteUrl <https://github.com/nickbrazeau/discent>

RemoteRef HEAD

RemoteSha 7f96210377482aa708033dfa4cf474aba38458d0

Contents

deme_inbreeding_spccoef	2
IBD_simulation_data	4
is.DISCresult	4
print.DISCresult	5
summary.DISCresult	5
tidyout	6
tidyout.DISCresult	6

Index 7

deme_inbreeding_spccoef

Identify Deme Inbreeding Spatial Coefficients in Continuous Space

Description

The purpose of this statistic is to identify an inbreeding coefficient, or degree of relatedness, for a given location in space. We assume that locations in spaces can be represented as "demes," such that multiple individuals live in the same deme (i.e. samples are sourced from the same location). The expected pairwise relationship between two individuals, or samples, is dependent on the each sample's deme's inbreeding coefficient and the geographic distance between the demes. The program assumes a symmetric distance matrix.

Usage

```
deme_inbreeding_spccoef(
  discdat,
  start_params = c(),
  f_learningrate = 0.001,
  m_learningrate = 1e-06,
  m_lowerbound = 0,
  m_upperbound = Inf,
  b1 = 0.9,
  b2 = 0.999,
  e = 1e-08,
  steps = 1000,
  thin = 1,
  normalize_geodist = TRUE,
  report_progress = TRUE,
  return_verbos = FALSE
)
```

Arguments

discdat	dataframe; The genetic-geographic data by deme (K)
start_params	named double vector; vector of start parameters.
f_learningrate	double; alpha parameter for how much each "step" is weighted in the gradient descent for inbreeding coefficients
m_learningrate	double; alpha parameter for how much each "step" is weighted in the gradient descent for the migration parameter
m_lowerbound	double; lower limit value for the global "m" parameter; will use a reflected normal within the gradient descent algorithm to adjust any aberrant values
m_upperbound	double; upper limit value for the global "m" parameter; will use a reflected normal within the gradient descent algorithm to adjust any aberrant values
b1	double; Exponential decay rates for the first moment estimate
b2	double; Exponential decay rates for the second moment estimate
e	double; Epsilon (error) for stability in the Adam optimization algorithm
steps	integer; the number of "steps" as we move down the gradient
thin	integer; the number of "steps" to keep as part of the output (i.e. if the user specifies 10, every 10th iteration will be kept)
normalize_geodist	boolean; whether geographic distances between demes should be normalized (i.e. rescaled to [0-1]). Helps increase model stability at the expense of complicating the interpretation of the migration rate parameter.
report_progress	boolean; whether or not a progress bar should be shown as you iterate through steps
return_verbose	boolean; whether the inbreeding coefficients and migration rate should be returned for every iteration or only for the final iteration. User will typically not want to store every iteration, which can be memory intensive

Details

The `gen.geo.dist` dataframe must be named with the following columns: "smp1"; "smp2"; "deme1"; "deme2"; "gendist"; "geodist"; which corresponds to: Sample 1 Name; Sample 2 Name; Sample 1 Location; Sample 2 Location; Pairwise Genetic Distance; Pairwise Geographic Distance. Note, the order of the columns do not matter but the names of the columns must match.

The `start_params` vector names must match the cluster names (i.e. clusters must have a name that we can match on for the starting relatedness parameters). In addition, you must provide a start parameter for "m".

Note: We have implemented coding decisions to not allow the "f" inbreeding coefficients to be negative by using a logit transformation internally in the code.

Gradient descent is performed using the Adam (adaptive moment estimation) optimization approach. Default values for moment decay rates, epsilon, and learning rates are taken from DP Kingma, 2014.

IBD_simulation_data *Simulated Identity by Descent from Isolation by Distance*

Description

Simulated Identity by Descent from Isolation by Distance

Usage

IBD_simulation_data

Format

A dataframe with 45 rows and 6 columns:

smp11, smp12 Placeholder sample names

deme1, deme2 Placeholder discrete demes

gendist Simulated genetic distances based on identity by descent

geodist Simulated geographic distances

Source

A toy dataset generated by basic simulation assuming an exponential relationship between relatedness and geographic distance. Data is not representative or generalizable but is simply meant to be used as input for various tests and function explanations

is.DISCresult *Check if DISCresult S3 Class*

Description

Overload is: function for determining if object is of class DISCresult

Usage

is.DISCresult(x)

Arguments

x DISC result from deme_inbreeding_spcoef function

`print.DISCresult` *print DISCresult S3 Class*

Description

overload `print()` function to print summary only

Usage

```
## S3 method for class 'DISCresult'  
print(x, ...)
```

Arguments

`x` DISC result from `deme_inbreeding_spccoef` function
`...` further arguments passed to or from other methods.

`summary.DISCresult` *Summary of DISCresult S3 Class*

Description

overload `summary()` function.

Usage

```
## S3 method for class 'DISCresult'  
summary(object, ...)
```

Arguments

`object` DISCresult Simulation
`...` further arguments passed to or from other methods.

tidyout	<i>Tidy Out Sim Method</i>
---------	----------------------------

Description

Method assignment

Usage

```
tidyout(x)
```

Arguments

x DISC result from deme_inbreeding_spcoef function

tidyout.DISCresult	<i>Tidy Out Sim</i>
--------------------	---------------------

Description

Function for taking output of SIR NE and lifting it over

Usage

```
## S3 method for class 'DISCresult'  
tidyout(x)
```

Arguments

x DISC result from deme_inbreeding_spcoef function

Index

* datasets

IBD_simulation_data, [4](#)

deme_inbreeding_spcoef, [2](#)

IBD_simulation_data, [4](#)

is.DISCresult, [4](#)

print.DISCresult, [5](#)

summary.DISCresult, [5](#)

tidyout, [6](#)

tidyout.DISCresult, [6](#)