

Package ‘BiG’

October 16, 2017

Type Package

Title Bayesian Aggregation in Genomic Applications

Version 0.1.0

Author Xue Li

Maintainer Xue Li <xuel@smu.edu>

Description

An implementation of Bayesian Aggregation in Genomic Applications (BiG), where BiG is a Bayesian latent variable approach to aggregation of partial and top ranked lists (Li et. al in preparation). It provides implementations for three different prior setups for variance/standard deviation parameters: diffuse inverse gamma (IG), diffuse uniform, half-t.

Depends R (>= 2.15.0)

License GPL-3

Encoding UTF-8

LazyData true

Imports truncnorm

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-10-16 13:23:04 UTC

R topics documented:

BiG_DA	2
BiG_diffuse	3
init_W	4
qtruncgamma	4
sim_lvm	5

Index	6
--------------	-------------------

BiG_DA

*BiG with half-t prior through a data augmentation approach***Description**

BiG implemented with half-t prior for the standard deviation parameters of the platform bias and diffuse uniform prior for the variance parameters of the study bias.

Usage

```
BiG_DA(r, n_T, n_p1, M = 20000, burnin, a = 0.0202, b = 98.5025, dp = 1,
       W = init_W(r), sigma_p10 = 0.5, sigma_p20 = 0.5, mu0 = numeric(G),
       xi10 = numeric(G), xi20 = numeric(G), sigma_s0 = rep(1, S))
```

Arguments

r	G*S matrix that contains the ranked lists to be aggregated, where G is the total number of items (genes) and S is the total number of ranked lists (studies). NA indicates non-inclusion of item. Note the matrix needs to be arranged such that all lists that belong to the same platform are next to each other, i.e. the first n_p1 columns are lists from platform 1.
n_T	vector of length S that contains number of top ranked items in each study.
n_p1	number of studies belong to platform 1.
M	number of MCMC iterations.
burnin	number of burn-in iterations.
a	hyperparameters for the prior distributions of standard deviation parameters. Used only when prior="uniform".
b	hyperparameters for the prior distributions of standard deviation parameters. Used only when prior="uniform".
dp	hyperparameter for the prior distributions of variance parameters for study bias and platform bias respectively. Used only when prior="IG".
W	G*S matrix that contains initial values for W. Each element of W is the local importance of the corresponding item in the corresponding study, i.e. the latent variable that determines the observed rank.
sigma_p10	initial values for the variance of the platform bias for platform 1 and platform 2 respectively.
sigma_p20	initial values for the variance of the platform bias for platform 1 and platform 2 respectively.
mu0	vector of length G that contains initial values for mu. Each element of mu is global importance of the corresponding item, i.e. the latent variable that determines the true rank.
xi10, xi20	vectors of length G that contain initial values for xi1 and xi2.
sigma_s0	vector of length S that contains initial values for the variances of the study bias.

Examples

```
set.seed(1234)
sim = sim_lvm(G=25, S=6, n_p1=3, rho=runif(6,min=0.3,max=0.9), p_p1=0.6, p_p2=0.8,
             lambda=runif(6,min=0.6,max=0.8), n_T=sample(c(5,10,15),6,replace=TRUE))
BiG_DA(r=sim$r, n_T=sim$n_T, n_p1=3, M=100, burnin=50)
```

BiG_diffuse

BiG with diffuse Inverse Gamma/Uniform prior

Description

BiG implemented with diffuse Inverse Gamma prior or diffuse Uniform prior for the variance/standard deviation parameters.

Usage

```
BiG_diffuse(r, n_T, n_p1, M = 20000, burnin, prior, ds = 1, dp = 1,
            W = init_W(r), sigma_p10 = 0.5, sigma_p20 = 0.5, mu0 = numeric(G),
            kappa10 = numeric(G), kappa20 = numeric(G), sigma_s0 = rep(1, S),
            a = 0.0202, b = 98.5025)
```

Arguments

<code>r</code>	G*S matrix that contains the ranked lists to be aggregated, where G is the total number of items (genes) and S is the total number of ranked lists (studies). NA indicates non-inclusion of item. Note the matrix needs to be arranged such that all lists that belong to the same platform are next to each other, i.e. the first <code>n_p1</code> columns are lists from platform 1.
<code>n_T</code>	vector of length S that contains number of top ranked items in each study.
<code>n_p1</code>	number of studies belong to platform 1.
<code>M</code>	number of MCMC iterations.
<code>burnin</code>	number of burn-in iterations.
<code>prior</code>	either "IG" or "uniform"
<code>ds, dp</code>	hyperparameter for the prior distributions of variance parameters for study bias and platform bias respectively. Used only when <code>prior="IG"</code> .
<code>W</code>	G*S matrix that contains initial values for W. Each element of W is the local importance of the corresponding item in the corresponding study, i.e. the latent variable that determines the observed rank.
<code>sigma_p10, sigma_p20</code>	initial values for the variance of the platform bias for platform 1 and platform 2 respectively.
<code>mu0</code>	vector of length G that contains initial values for mu. Each element of mu is global importance of the corresponding item, i.e. the latent variable that determines the true rank.

kappa10, kappa20
vectors of length G that contain initial values for kappa1 and kappa2. Each element of kappa1 and kappa2 is the platform bias for the corresponding item in platform 1 and 2 respectively

sigma_s0
vector of length S that contains initial values for the variances of the study bias.

a, b
hyperparameters for the prior distributions of standard deviation parameters. Used only when prior="uniform".

Examples

```
set.seed(1234)
sim = sim_lvm(G=25, S=6, n_p1=3, rho=runif(6,min=0.3,max=0.9), p_p1=0.6, p_p2=0.8,
             lambda=runif(6,min=0.6,max=0.8), n_T=sample(c(5,10,15),6,replace=TRUE))
rank(-BiG_diffuse(r=sim$r, n_T=sim$n_T, n_p1=3, M=100, burnin=50, prior="IG"))
#rank(-BiG_diffuse(r=sim$r, n_T=sim$n_T, n_p1=3, M=100, burnin=50, prior="uniform"))
```

init_W	Generate initial values for W
--------	-------------------------------

Description

Generate initial values for W.

Usage

```
init_W(r)
```

Arguments

r
matrix that contains the ranked lists to be aggregated. NA indicates non-inclusion of item.

qtruncgamma	Truncated Gamma distribution
-------------	------------------------------

Description

Quantile function and random generation for truncated Gamma distribution with parameters shape and rate.

Usage

```
qtruncgamma(p, a = -Inf, b = Inf, shape, rate = 1)

rtruncgamma(n, a = -Inf, b = Inf, shape, rate = 1)
```

Arguments

p	vector of probabilities.
a	vector of lower bounds. These may be $-\text{Inf}$.
b	vector of upper bounds. These may be Inf .
shape, rate	shape and rate parameters. Must be positive, rate strictly.
n	number of observations. If $\text{length}(n) > 1$, the length is taken to be the number required.

Examples

```
qtruncgamma(0.6,1,2,2,1)
rtruncgamma(5,1,2,2,1)
```

sim_lvm

*Simulate rank data from latent variable model***Description**

Simulate rank data from latent variable model

Usage

```
sim_lvm(G, S, n_p1, rho, p_p1, p_p2, lambda, n_T)
```

Arguments

G	total number of genes involved in all of the studies.
S	number of studies (ranked lists) to be aggregated.
n_p1	number of studies belong to platform 1.
rho	correlation between local importance (w) and global importance (mu) for each study, which determines the total variance of w.
p_p1, p_p2	percentage of total variance of w contributed by platform variance from platform 1 and platform 2 respectively for the study with the lowest total variance.
lambda	inclusion rate for each study.
n_T	vector of length S that contains number of top ranked items in each study.

Examples

```
set.seed(1234)
sim_lvm(G=25, S=6, n_p1=3, rho=runif(6,min=0.3,max=0.9), p_p1=0.6, p_p2=0.8,
        lambda=runif(6,min=0.6,max=0.8), n_T=sample(c(5,10,15),6,replace=TRUE))
```

Index

BiG_DA, [2](#)

BiG_diffuse, [3](#)

init_W, [4](#)

qtruncgamma, [4](#)

rtruncgamma (qtruncgamma), [4](#)

sim_lvm, [5](#)