

# Package ‘bmmix’

February 19, 2015

**Version** 0.1-2

**Date** 2014/05/19

**Title** Bayesian multinomial mixture

**Author** Thibaut Jombart <t.jombart@imperial.ac.uk>

**Maintainer** Thibaut Jombart <t.jombart@imperial.ac.uk>

**Depends** R (>= 3.0.0)

**Imports** utils, ggplot2, reshape2

**Description** Bayesian multinomial mixture model

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2014-05-27 13:03:46

## R topics documented:

Bayesian multinomial mixture model . . . . .	1
Process bmmix outputs . . . . .	5
<b>Index</b>	<b>6</b>

---

Bayesian multinomial mixture model

*Bayesian multinomial mixture model estimation using MCMC*

---

## Description

This function and model are under development. Do not use them, contact the author if interested.

**Usage**

```
bmmix(x, y, n=5e4, sample.every=200,
      move.alpha=TRUE, move.phi=FALSE,
      sd.alpha=0.1, sd.phi=0.05, move.phi.every=10,
      model.unsampled=FALSE, prior.unsampled.contrib=0.1,
      min.ini.freq=0.01, file.out="mcmc.txt", quiet=FALSE)
```

**Arguments**

<code>x</code>	a matrix containing multinomial data in columns used to compose the mixture (i.e., each column is an 'origin').
<code>y</code>	a vector of the same length as the number of rows of <code>x</code> containing the response variable.
<code>n</code>	the length of the MCMC.
<code>sample.every</code>	an integer indicating the frequency at which to save MCMC samples.
<code>move.alpha</code>	a logical indicating whether the mixture coefficients (alpha) should be estimated.
<code>move.phi</code>	a logical indicating whether the frequencies in <code>x</code> (phi) should be estimated; see details.
<code>sd.alpha</code>	the standard deviation of the normal distribution used as proposal for alpha.
<code>sd.phi</code>	the standard deviation of the normal distribution used as proposal for phi.
<code>move.phi.every</code>	the frequency at which values of phi should be moved.
<code>model.unsampled</code>	a logical indicating whether an 'unsampled origin' should be allowed; if TRUE, then <code>move.phi</code> should be TRUE as well, to allow for frequencies in this group to be estimated.
<code>prior.unsampled.contrib</code>	the mean of the exponential distribution used as prior for the contribution of the unsampled origin in the mixture; all other origins have flat priors.
<code>min.ini.freq</code>	the default minimum frequency of unobserved items in <code>x</code> used for the initial frequency estimate.
<code>file.out</code>	the name of the file used to store the outputs.
<code>quiet</code>	a logical indicating whether output messages should be hidden.

**Details**

There are essentially 4 variants of the model implemented by `bmmix`:

- estimate mixture only (default) only the mixture coefficients are estimated; frequencies (phi) are fixed to their maximum likelihood estimate from the data; this model has 'K' parameters (where 'K' is the number of putative origins, i.e. the number of columns in 'x').
- estimate mixture and frequencies both mixture coefficients and frequencies for each group and origin are estimated; this model has (N+1)K parameters (N being the number of rows in 'x'); to use this model, specify `move.phi=TRUE`.

- estimate mixture, allowing unsampled origin mixture coefficients are estimated with an additional 'unsampled' origin whose frequencies are estimated; this model has  $K+N+1$  parameters ( $N$  being the number of rows in 'x'); to use this model, specify `move.phi=FALSE, model.unsigned=TRUE`; this is the only practical model allowing unsampled origin for medium-sized or large datasets..
- estimate mixture, frequencies, and allow unsampled origin this is the most complex model; in addition to the previous one, an unsampled origin is allowed, and its frequencies are estimated; this model therefore has  $(N+1)(K+1)$  parameters; to use this model, specify `move.phi=TRUE` and `model.unsigned=TRUE`; note that if frequencies are not estimated (`move.phi=FALSE`), the frequencies in the unsampled origin will be fixed to their initial value in which all groups have the same frequency; this model quickly becomes hard to fit for medium-sized to large datasets.

### Value

A data.frame with class 'bmmix', containing the MCMC outputs: step, log-posterior, log-likelihood, log-prior, alpha values (mixture coefficients), and optionally frequencies for each group and origin (phi).

### Author(s)

Thibaut Jombart <t.jombart@imperial.ac.uk>

### Examples

```
## GENERATE TOY DATA ##
## ST frequencies in 3 origins:
## dogs, cows, asymptotic cases in human
f.dogs <- c(.5, .3, .1, .1, 0)
f.cows <- c(.6, .1, .1, .1, .1)
f.asymp <- c(0, .1, .2, 0, .7)

## mixture (y would be symptomatic cases)
f.y <- .1*f.dogs + .1*f.cows + .8*f.asymp

set.seed(1)
dogs <- rmultinom(1, 30, f.dogs)
cows <- rmultinom(1, 50, f.cows)
asymp <- rmultinom(1, 35, f.asymp)
X <- data.frame(dogs, cows, asymp,
  row.names=paste("ST", letters[1:5]))
X
y <- rmultinom(1, 40, f.y)
y

cbind(X,y)

## RUN BMMIX ##

## BASIC MODEL
## note: small n for this example only!
```

```

set.seed(1)
res <- bmmix(X,y, n=3e4)
head(res)

## VISUALIZE RESULTS ##
if(require("ggplot2") && require("reshape2")){

## manually ##
## check log-posterior
ggplot(dat=res) + geom_line(aes(x=step, y=post)) +
  labs(title="Trace of log-posterior values")

## check mixture coefficients
fig.dat <- melt(res, id=1:4)

ggplot(dat=fig.dat, aes(x=step)) +
  geom_line(aes(y=value, colour=variable)) +
  labs(title="Trace of mixture coefficients")

## with process.bmmix ##
## mixture coefficients
temp <- process.bmmix(res, "alpha")
names(temp)
temp$alpha # values
temp$trace # graphics: trace
temp$hist # graphics: histograms
temp$dens # graphics: densities
temp$violin # graphics: violinplot

}

## Not run:
## MODEL WITH ESTIMATED FREQUENCIES
set.seed(1)
res <- bmmix(X,y, move.phi=TRUE)
head(res)

## VISUALIZE RESULTS
if(require("ggplot2") && require("reshape2")){

## check log-posterior
ggplot(dat=res) + geom_line(aes(x=step, y=post)) +
  labs(title="Trace of log-posterior values")

fig.dat <- melt(res[,1:7], id=1:4)

## check mixture coefficients
ggplot(dat=fig.dat, aes(x=step)) +
  geom_line(aes(y=value, colour=variable)) +
  labs(title="Trace of mixture coefficients")

```

```
## check ST frequencies, i.e. in dogs:
fig.dat <- melt(res[,c(1,grep("dogs", names(res))[-1])], id=1)

ggplot(dat=fig.dat) +
  geom_line(aes(x=step, y=value, colour=variable)) +
  labs(title="Estimate of ST frequencies in dogs")

ggplot(dat=fig.dat) +
  geom_density(aes(x=value, fill=variable),alpha=.2) +
  labs(title="Estimate of ST frequencies in dogs")
}

## End(Not run)
```

---

Process bmmix outputs *Extract results and make graphics from bmmix outputs*

---

## Description

This function and model are under development. Do not use them, contact the author if interested.

## Usage

```
process.bmmix(x, what="post", burnin=1e4, ggplot=TRUE)
```

## Arguments

x	a data.frame output by bmmix.
what	a character string indicating which result is sought, matched against the names of the columns of x.
burnin	an integer indicating the burnin, i.e. the number of MCMC iterations to be discarded.
ggplot	a logical indicating whether graphics using ggplot2 should be returned.

## Value

A list containing processed results, and optionally ggplot graphics.

## Author(s)

Thibaut Jombart <t.jombart@imperial.ac.uk>

# Index

Bayesian multinomial mixture model, [1](#)  
bmmix (Bayesian multinomial mixture  
model), [1](#)

Process bmmix outputs, [5](#)  
process.bmmix (Process bmmix outputs), [5](#)