Package 'PMD'

September 2, 2021

Type Package

Title The Poisson Multinomial Distribtuion

Version 1.0

Date 2021-08-06

Maintainer Yili Hong <yilihong@vt.edu>

Description Applying exact, simulation and approximation methods for computing probability mass functions, cumulative distribution functions of Poisson-Multinomial distributions together with a random number generator to conduct Poisson-Multinomial distribution sampling. The exact method is based on fast Fourier transformation of the characteristic functions of Poisson-Multinomial distributions. Simulation method is built via naive sampling scheme of different multinomial distributions. Approximation method is designed by asymptotic distrbutions of Poisson-Multinomial distributions, which are normal.

License GPL (>= 2)

Encoding UTF-8

Imports mvtnorm, Rcpp

LinkingTo Rcpp, RcppArmadillo

SystemRequirements fftw3(>=3.3)

RoxygenNote 7.1.1

NeedsCompilation yes

Author Yili Hong [aut, ctb],
Zhengzhi Lin [aut, cre],
Yueyao Wang [aut, ctb]

R topics documented:

dpmd *Probability Mass Function of Poisson-Multinomial Distributions*

Description

Probability mass function of Poisson-Multinomial distributions specified by input matrix and computed through selected method. This function is capable for computation of the whole probability mass function as well as of one single probability mass point.

Usage

```
dpmd(pmat, xmat = NULL, method = "DFT-CF", B = 1000)
```

Arguments

pmat	An $\mathrm{n} \times \mathrm{m}$ matrix of probabilities. n is the number of independent trials. m is the number of categories. Also called success probability matrix. Each row of pmat describes the success probability for the corresponding trial and it should add up to 1.
xmat	Result matrix of column dimension m (probability mass point) specified by user. Each row of the matrix should has the form $\mathrm{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{m})$ which is used for computing $\mathrm{P}(\mathrm{X}_1 = \mathrm{x}_1, \ldots, \mathrm{X}_\mathrm{m} = \mathrm{x}_\mathrm{m})$, the values of x should sum up to n.
method	Character string stands for the method selected by user to compute the probability mass. The method can only be one of the following four: "DFT-CF", "NA", "SIM",
B	Number of repetitions in the simulation method. Will be ignored if users do not choose "SIM" method.

Details

Consider n independent trials and each trial leads to a success for exactly one of m categories. Each category has varying success probabilities from different trials. The Poisson multinomial distribution (PMD) gives the probability of any particular combination of numbers of successes for the m categories. The success probabilities form an $\mathrm{n} \times \mathrm{m}$ matrix, which is called the success probability matrix and denoted by pmat. For the methods we applied in dpmd, "DFT-CF" is an exact method to calculate all mass points of Poisson-Multinomial Distributions via FFT algorithm. When the dimension of pmat increases, the computation burden of "DFT-CF" might challenge the capability of a computer because the method automatically compute all probability mass points regardless the input of xmat.

"SIM" is a simulation method using a naive simulation scheme to calculate the whole probability mass function. Notice that the accuracy and running time will be affected by user choice of B. Usually B=1e5 or 1e6 will be accurate enough. Increasing B to larger than 1e8 will heavily aggravate computational burden of a CPU or GPU.

"NA" is an approximation method using Normal approximation to compute the probability mass function of xmat vector specified by user. This method requires an input of xmat.

Notice if xmat is not specified then it will be NULL, dpmd will compute the whole pmf if the selected method is "SIM" or "DFT-CF". Under the input of xmat, only the probability mass funtion limited by xmat will be computed.

Value

For a given xmat, dpmd returns the probability mass function at points specified by xmat. If xmat is NULL, all mass points of a given pmat will be computed and stored in an output multi-dimensional array, say, res. Let the dimension of pmat is $n \times m$, res will be a $\mathrm{n} \times (\mathrm{m}-1)$ array. Then $\mathrm{P}(\mathrm{X}_1 = \mathrm{x}_1, \ldots, \mathrm{X}_\mathrm{m} = \mathrm{x}_\mathrm{m})$ will be $\mathrm{res}[\mathrm{x}_1 + 1, \ldots, \mathrm{x}_{\mathrm{m}-1} + 1]$. For instance, for the pmat matrix in the following example, the value of the array element res[1,2,1]=0.90 means the value of probability mass point $\mathrm{P}(\mathrm{X}_1 = 0, \mathrm{X}_2 = 1, \mathrm{X}_3 = 0, \mathrm{X}_4 = 2) = 0.90$.

Examples

```
pp <- matrix(c(.1, .1, .1, .7, .1, .3, .3, .3, .5, .2, .1, .2), nrow = 3, byrow = TRUE)
x <- matrix(c(0,0,1,2), nrow=1)
x1 <- matrix(c(0,0,1,2,2,1,0,0),nrow=2,byrow=TRUE)

dpmd(pmat = pp)
dpmd(pmat = pp, xmat = x1)
dpmd(pmat = pp, xmat = x)

dpmd(pmat = pp, xmat = x, method = "NA" )
dpmd(pmat = pp, xmat = x1, method = "NA" )

dpmd(pmat = pp, xmat = x, method = "SIM", B = 1e3)
dpmd(pmat = pp, xmat = x1, method = "SIM", B = 1e3)
```

---

ppmd *Cumulative Distribution Function of Poisson-Multinomial Distribution*

Description

This function computes cumulative distribution function of Poisson-Multinomial distributions that specified by input probability matrix via given method.

Usage

```
ppmd(pmat, x, method = "DFT-CF", B = 1000)
```

Arguments

pmat	An $n \times m$ matrix of probabilities. n is the number of independent trials. m is the number of categories. Each row of pmat describes the success probability for the corresponding trial and it should add up to 1.
method	Character string stands for the method selected by user to compute the probability mass. The method can only be one of the following three: "DFT-CF", "NA", "SIM-ALL".
B	Number of repetitions in the simulation method. Will be ignored if users do not choose "SIM-ALL" method.
xmat	A matrix with column number m and each row has the form $x = (x_1, \ldots, x_m)$ for computing $P(X_1 \leq x_1, \ldots, X_m \leq x_m)$.

Details

See Details in dpmd for the definition of the PMD and the introduction of notations. ppmd computes the cumulative distribution function by adding all probability mass points within hyper-dimensional space limited by x.

"DFT-CF" is an exact method to calculate all mass points of Poisson-Multinomial Distributions via FFT algorithm. "SIM-ALL" is a simulation method using a naive simulation scheme to calculate the whole probability mass function. "NA" is an approximation method using Normal approximation method.

Value

The value of $P(X_1 \leq x_1, \ldots, X_m \leq x_m)$ of given $x = (x_1, \ldots, x_m)$.

Examples

```
pp <- matrix(c(.1, .1, .1, .7, .1, .3, .3, .3, .5, .2, .1, .2), nrow = 3, byrow = TRUE)
x <- matrix(c(3,2,1,3),nrow=1)

ppmd(pmat = pp, x = x)
ppmd(pmat = pp, x = x, method = "NA")
ppmd(pmat = pp, x = x, method = "SIM-ALL", B = 1e3)
```

---

rpmd *Poisson-Multinomial Distribution Random Number Generator*

Description

Generating random samples from Poisson-Multinomial distribution based on a given success probability matrix.

Usage

```
rpmd(pmat, s = 1)
```

Arguments

pmat	The $n \times m$ success probability matrix, where n is the number of independent trials and m is the number of categories. Each row of pmat describes the success probability for the corresponding trial, which adds up to 1.
s	The number of samples to be generated.

Value

An $s \times m$ matrix of samples, each row stands for one sample from the PMD with success probability matrix pmat.

Examples

```
pp=matrix(c(.1, .1, .1, .7, .1, .3, .3, .3, .5, .2, .1, .2), nrow = 3, byrow = TRUE)

rpmd(pmat = pp, s = 5)
```

Index