Package 'PMD'

August 14, 2021

Type Package

Title Computation of Poisson-Multinomial Distribtuions

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]

Archs i386, x64

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, x = c(0, 0, 0, 0), method = "DFT-CF", B = 1e3)
```

Arguments

pmat	An $n \times m$ 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.
x	Result vector of length m (probability mass point) specified by user when the selected method is "SIM" or "NA". The vector $x = (x_1, x_2, \ldots, x_m)$ is used for computing $P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m)$.
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", "SIM-ALL".
B	Number of repetitions in the simulation method. Will be ignored if users do not choose "SIM-ALL" or "SIM" method.

Details

Given pmat with dimension $n \times m$, the total number of outcomes is $(n+1)^{m-1}$. 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 users select "DFT-CF", dpmd will ignore vec and return the probability mass function for all outcomes.

"SIM-ALL" is a simulation method using a naive simulation scheme to calculate the whole probability mass function. Under this selection, the input of vec will be ignored. 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.

When the dimension of pmat increases, the computation burden of "DFT-CF" and "SIM-ALL" method might challenge the capability of a computer because both of the methods calculate all mass points of Poisson-Multinomial distributions.

"SIM" is as same as "SIM-ALL" except that it only computes the probability mass function at a single outcome specified by vec.

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

Value

For a single mass point, dpmd returns the probability mass function at that point.

For all mass points of a given pmat, it returns a multi-dimensional array. For instance, for the pmat matrix in the following example, the value of the array element $a_{1,2,1}$ = 0.90 means the value of probability mass point (0,1,0,2) is 0.90.

Examples

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

dpmd(pmat = pp)
dpmd(pmat = pp, method = "SIM-ALL", B = 1e3)
dpmd(pmat = pp, x = c(0,0,1,2), method = "NA" )
dpmd(pmat = pp, x = c(0,0,1,2), 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 = 1e3)
```

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.
x	A length m vector $x = (x_1, x_2, \ldots, x_m)$ for computing $P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_m \leq x_m)$.
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.

Details

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 $P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_m \leq x_m)$ of given $x = (x_1, x_2, \ldots, x_m)$.

Examples

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

ppmd(pmat = pp, x = c(3,2,1,3))
ppmd(pmat = pp, x = c(3,2,1,3), method = "NA")
ppmd(pmat = pp, x = c(3,2,1,3), method = "SIM-ALL", B = 1e3)
```

---

rpmd *Poisson-Multinomial Distribution Random Number Generator*

Description

Generating random samples of given a Poisson-Multinomial distribution.

Usage

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

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.
n	Number of samples to be generated.

Value

An $n \times m$ matrix of samples, each row stands for one draw from 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, n = 5)
```

Index