Package ‘BiasedUrn’

January 20, 2025

Type Package

Title Biased Urn Model Distributions

Version 2.0.12

Date 2024-06-16

Author Agner Fog

Maintainer Agner Fog <agner@agner.org>

Description Statistical models of biased sampling in the form of
univariate and multivariate noncentral hypergeometric distributions,
including Wallenius' noncentral hypergeometric distribution and
Fisher's noncentral hypergeometric distribution.
See vignette(``UrnTheory'') for explanation of these distributions.
Literature:
Fog, A. (2008a). Calculation Methods for Wallenius' Noncentral Hypergeometric Distribution, Communications in Statistics, Simulation and Computation, 37(2) <doi:10.1080/03610910701790269>.
Fog, A. (2008b). Sampling methods for Wallenius’ and Fisher’s noncentral hypergeometric distributions, Communications in Statistics—Simulation and Computation, 37(2) <doi:10.1080/03610910701790236>.

License GPL-3

Encoding UTF-8

URL https://www.agner.org/random/ https://www.r-project.org/

NeedsCompilation yes

Repository CRAN

Date/Publication 2024-06-16 14:50:01 UTC

Contents

`BiasedUrn-package` *Biased Urn Model Distributions*

Description

Statistical models of biased sampling in the form of univariate and multivariate noncentral hypergeometric distributions, including Wallenius' noncentral hypergeometric distribution and Fisher's noncentral hypergeometric distribution (also called extended hypergeometric distribution).

These are distributions that you can get when taking colored balls from an urn without replacement, with bias. The univariate distributions are used when there are two colors of balls. The multivariate distributions are used when there are more than two colors of balls.

The (central) univariate and multivariate hypergeometric distribution can be obtained by setting `odds` = 1.

Please see `vignette("UrnTheory")` for a definition of these distributions and how to decide which distribution to use in a specific case.

Details

Package:	BiasedUrn
Type:	Package
Version:	2.0.12
Date:	2024-06-16
License:	GPL-3

Univariate functions in this package

	Wallenius' noncentral hypergeometric	Fisher's noncentral hypergeometric
Probability mass function	dWNCHypergeo	dFNCHypergeo
Cumulative distribution function	pWNCHypergeo	pFNCHypergeo
Quantile function	qWNCHypergeo	qFNCHypergeo
Random variate generation function	rWNCHypergeo	rFNCHypergeo
Calculate mean	meanWNCHypergeo	meanFNCHypergeo
Calculate variance	varWNCHypergeo	varFNCHypergeo
Calculate mode	modeWNCHypergeo	modeFNCHypergeo
Estimate odds from mean	oddsWNCHypergeo	oddsFNCHypergeo
Estimate number from mean and odds	numWNCHypergeo	numFNCHypergeo
Minimum x	minHypergeo	minHypergeo
Maximum x	maxHypergeo	maxHypergeo

Multivariate functions in this package

	Wallenius' noncentral hypergeometric	Fisher's noncentral hypergeometric
Probability mass function	dMWNCHypergeo	dMFNCHypergeo

Random variate generation function	rMWNCHypergeo	rMFNCHypergeo
Calculate mean	meanMWNCHypergeo	meanMFNCHypergeo
Calculate variance	varMWNCHypergeo	varMFNCHypergeo
Calculate mean and variance	momentsMWNCHypergeo	momentsMFNCHypergeo
Estimate odds from mean	oddsMWNCHypergeo	oddsMFNCHypergeo
Estimage number from mean and odds	numMWNCHypergeo	numMFNCHypergeo
Minimum x	minMHypergeo	minMHypergeo
Maximum x	maxMHypergeo	maxMHypergeo

Note

The implementation cannot run safely in multiple threads simultaneously

Author(s)

Agner Fog

Maintainer: Agner Fog <agner@agner.org>

References

https://www.agner.org/random/

Fog, A. 2008a. Calculation methods for Wallenius' noncentral hypergeometric distribution. *Communications in Statistics—Simulation and Computation* **37**, 2 doi:10.1080/03610910701790269

Fog, A. 2008b. Sampling methods for Wallenius' and Fisher's noncentral hypergeometric distributions. *Communications in Statistics—Simulation and Computation* **37**, 2 doi:10.1080/03610910701790236

See Also

BiasedUrn-Univariate.
BiasedUrn-Multivariate.
vignette("UrnTheory")
demo(CompareHypergeo)
demo(ApproxHypergeo)
demo(OddsPrecision)
demo(SampleWallenius)
dhyper
fisher.test

Examples

```
dWNCHypergeo(12, 25, 32, 20, 2.5)
```

BiasedUrn-Multivariate

Biased urn models: Multivariate distributions

Description

Statistical models of biased sampling in the form of multivariate noncentral hypergeometric distributions, including Wallenius' noncentral hypergeometric distribution and Fisher's noncentral hypergeometric distribution (also called extended hypergeometric distribution).

These are distributions that you can get when taking colored balls from an urn without replacement, with bias. The univariate distributions are used when there are two colors of balls. The multivariate distributions are used when there are more than two colors of balls.

Please see `vignette("UrnTheory")` for a definition of these distributions and how to decide which distribution to use in a specific case.

Usage

```
dMWNCHypergeo(x, m, n, odds, precision = 1E-7)
dMFNCHypergeo(x, m, n, odds, precision = 1E-7)
rMWNCHypergeo(nran, m, n, odds, precision = 1E-7)
rMFNCHypergeo(nran, m, n, odds, precision = 1E-7)
meanMWNCHypergeo(m, n, odds, precision = 0.1)
meanMFNCHypergeo(m, n, odds, precision = 0.1)
varMWNCHypergeo(m, n, odds, precision = 0.1)
varMFNCHypergeo(m, n, odds, precision = 0.1)
momentsMWNCHypergeo(m, n, odds, precision = 0.1)
momentsMFNCHypergeo(m, n, odds, precision = 0.1)
oddsMWNCHypergeo(mu, m, n, precision = 0.1)
oddsMFNCHypergeo(mu, m, n, precision = 0.1)
numMWNCHypergeo(mu, n, N, odds, precision = 0.1)
numMFNCHypergeo(mu, n, N, odds, precision = 0.1)
minMHypergeo(m, n)
maxMHypergeo(m, n)
```

Arguments

`x`	Number of balls of each color sampled. Vector with length = number of colors, or matrix with nrows = number of colors.
`m`	Initial number of balls of each color in the urn. Length of vector = number of colors.
`n`	Total number of balls sampled. Scalar.
`N`	Total number of balls in urn before sampling. Scalar.
`odds`	Odds or weight for each color, arbitrarily scaled. Length of vector = number of colors. Gives the (central) multivariate hypergeometric distribution if all odds are equal.

nran	Number of random variates to generate. Scalar.
mu	Mean x for each color. Length of vector = number of colors.
precision	Desired precision of calculation. Scalar.

Details

Allowed parameter values
x, m, odds and mu are all vectors with one element for each color. These vectors must have the same length. x can also be a matrix with one column for each observation. The number of rows in this matrix must be equal to the number of colors. The maximum number of colors is currently set to 32.

All parameters must be non-negative. n cannot exceed N = sum(m). The odds may be arbitrarily scaled. The code has been tested with odds ratios in the range $10^{-9} \ldots 10^{9}$ and zero. The code may work with odds ratios outside this range, but errors or NAN can occur for extreme values of odds. A ball with odds = 0 is equivalent to no ball. mu must be within the possible range of x.

Calculation time
The calculation time depends on the specified precision and the number of colors. The calculation time can be high for rMWNCHypergeo and rMFNCHypergeo when nran is high. The calculation time can be extremely high for dMFNCHypergeo when n is high and the number of colors is high. The calculation time can be extremely high for the mean... var... and moments... functions when precision < 0.1 and n is high and the number of colors is high.

Value

dMWNCHypergeo and dMFNCHypergeo return the probability mass function for the multivariate Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A single value is returned if x is a vector with length = number of colors. Multiple values are returned if x is a matrix with one column for each observation. The number of rows must be equal to the number of colors.

rMWNCHypergeo and rMFNCHypergeo return random vectors with the multivariate Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A vector is returned when nran = 1. A matrix with one column for each observation is returned when nran > 1.

meanMWNCHypergeo and meanMFNCHypergeo return the mean of the multivariate Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A simple and fast approximation is used when precision >= 0.1. A full calculation of all possible x combinations is used when precision < 0.1. This can take extremely long time when the number of colors is high.

varMWNCHypergeo and varMFNCHypergeo return the variance of the multivariate Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A simple and fast approximation is used when precision >= 0.1. A full calculation of all possible x combinations is used when precision < 0.1. This can take extremely long time when the number of colors is high.

momentsMWNCHypergeo and momentsMFNCHypergeo return a data frame with the mean and variance of the multivariate Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. Calculating the mean and variance in the same operation saves time when precision < 0.1.

oddsMWNCHypergeo and oddsMFNCHypergeo estimate the odds from an observed mean for the multivariate Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A vector of odds is returned if mu is a vector. A matrix is returned if mu is a matrix with one row for each color. A simple and fast approximation is used regardless of the specified precision. Exact calculation is not supported. See demo(OddsPrecision).

numMWNCHypergeo and numMFNCHypergeo estimate the number of balls of each color in the urn before sampling from experimental mean and known odds ratios for Wallenius' and Fisher's noncentral hypergeometric distributions. The returned m values are not integers. A vector of m is returned if mu is a vector. A matrix of m is returned if mu is a matrix with one row for each color. A simple and fast approximation is used regardless of the specified precision. Exact calculation is not supported. The precision of calculation is indicated by demo(OddsPrecision).

minMHypergeo and maxMHypergeo calculate the minimum and maximum value of x for the multivariate distributions. The values are valid for the multivariate Wallenius' and Fisher's noncentral hypergeometric distributions as well as for the multivariate (central) hypergeometric distribution.

References

https://www.agner.org/random/

Fog, A. 2008a. Calculation methods for Wallenius' noncentral hypergeometric distribution. *Communications in Statistics—Simulation and Computation* **37**, 2 doi:10.1080/03610910701790269

Fog, A. 2008b. Sampling methods for Wallenius' and Fisher's noncentral hypergeometric distributions. *Communications in Statistics—Simulation and Computation* **37**, 2 doi:10.1080/03610910701790236

See Also

vignette("UrnTheory")
BiasedUrn-Univariate.
BiasedUrn.

Examples

```
# get probability
dMWNCHypergeo(c(8,10,6), c(20,30,20), 24, c(1.,2.5,1.8))
```

BiasedUrn-Univariate *Biased urn models: Univariate distributions*

Description

Statistical models of biased sampling in the form of noncentral hypergeometric distributions, including Wallenius' noncentral hypergeometric distribution and Fisher's noncentral hypergeometric distribution (also called extended hypergeometric distribution).

These are distributions that you can get when taking colored balls from an urn without replacement, with bias. The univariate distributions are used when there are two colors of balls. The multivariate distributions are used when there are more than two colors of balls.

Please see vignette("UrnTheory") for a definition of these distributions and how to decide which distribution to use in a specific case.

Usage

```
dWNCHypergeo(x, m1, m2, n, odds, precision=1E-7)
dFNCHypergeo(x, m1, m2, n, odds, precision=1E-7)
pWNCHypergeo(x, m1, m2, n, odds, precision=1E-7, lower.tail=TRUE)
pFNCHypergeo(x, m1, m2, n, odds, precision=1E-7, lower.tail=TRUE)
qWNCHypergeo(p, m1, m2, n, odds, precision=1E-7, lower.tail=TRUE)
qFNCHypergeo(p, m1, m2, n, odds, precision=1E-7, lower.tail=TRUE)
rWNCHypergeo(nran, m1, m2, n, odds, precision=1E-7)
rFNCHypergeo(nran, m1, m2, n, odds, precision=1E-7)
meanWNCHypergeo(m1, m2, n, odds, precision=1E-7)
meanFNCHypergeo(m1, m2, n, odds, precision=1E-7)
varWNCHypergeo(m1, m2, n, odds, precision=1E-7)
varFNCHypergeo(m1, m2, n, odds, precision=1E-7)
modeWNCHypergeo(m1, m2, n, odds, precision=1E-7)
modeFNCHypergeo(m1, m2, n, odds, precision=0)
oddsWNCHypergeo(mu, m1, m2, n, precision=0.1)
oddsFNCHypergeo(mu, m1, m2, n, precision=0.1)
numWNCHypergeo(mu, n, N, odds, precision=0.1)
numFNCHypergeo(mu, n, N, odds, precision=0.1)
minHypergeo(m1, m2, n)
maxHypergeo(m1, m2, n)
```

Arguments

x	Number of red balls sampled.
m1	Initial number of red balls in the urn.
m2	Initial number of white balls in the urn.
n	Total number of balls sampled.
N	Total number of balls in urn before sampling.
odds	Probability ratio of red over white balls.
p	Cumulative probability.
nran	Number of random variates to generate.
mu	Mean x.
precision	Desired precision of calculation.
lower.tail	if TRUE (default), probabilities are $P(X \le x)$, otherwise, $P(X > x)$.

Details

Allowed parameter values

All parameters must be non-negative. n cannot exceed N = m1 + m2. The code has been tested with odds in the range $10^{-9} \ldots 10^{9}$ and zero. The code may work with odds outside this range, but errors or NAN can occur for extreme values of odds. A ball with odds = 0 is equivalent to no ball. `mu` must be within the possible range of `x`.

Calculation time
The calculation time depends on the specified precision.

Value

`dWNCHypergeo` and `dFNCHypergeo` return the probability mass function for Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A single value is returned if `x` is a scalar. Multiple values are returned if `x` is a vector.

`pWNCHypergeo` and `pFNCHypergeo` return the cumulative probability function for Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A single value is returned if `x` is a scalar. Multiple values are returned if `x` is a vector.

`qWNCHypergeo` and `qFNCHypergeo` return the quantile function for Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A single value is returned if `p` is a scalar. Multiple values are returned if `p` is a vector.

`rWNCHypergeo` and `rFNCHypergeo` return random variates with Wallenius' and Fisher's noncentral hypergeometric distribution, respectively.

`meanWNCHypergeo` and `meanFNCHypergeo` calculate the mean of Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A simple and fast approximation is used when $precision \geq 0.1$.

`varWNCHypergeo` and `varFNCHypergeo` calculate the variance of Wallenius' and Fisher's noncentral hypergeometric distribution, respectively. A simple and fast approximation is used when $precision \geq 0.1$.

`modeWNCHypergeo` and `modeFNCHypergeo` calculate the mode of Wallenius' and Fisher's noncentral hypergeometric distribution, respectively.

`oddsWNCHypergeo` and `oddsFNCHypergeo` estimate the odds of Wallenius' and Fisher's noncentral hypergeometric distribution from a measured mean. A single value is returned if `mu` is a scalar. Multiple values are returned if `mu` is a vector. A simple and fast approximation is used regardless of the specified precision. Exact calculation is not supported. See `demo(OddsPrecision)`.

`numWNCHypergeo` and `numFNCHypergeo` estimate the number of balls of each color in the urn before sampling from an experimental mean and a known odds ratio for Wallenius' and Fisher's noncentral hypergeometric distributions. The returned numbers `m1` and `m2` are not integers. A vector of `m1` and `m2` is returned if `mu` is a scalar. A matrix is returned if `mu` is a vector. A simple approximation is used regardless of the specified precision. Exact calculation is not supported. The precision of calculation is indicated by `demo(OddsPrecision)`.

`minHypergeo` and `maxHypergeo` calculate the minimum and maximum value of `x`. The value is valid for Wallenius' and Fisher's noncentral hypergeometric distribution as well as for the (central) hypergeometric distribution.

References

https://www.agner.org/random/

Fog, A. 2008a. Calculation methods for Wallenius' noncentral hypergeometric distribution. *Communications in Statistics—Simulation and Computation* **37**, 2 doi:10.1080/03610910701790269

Fog, A. 2008b. Sampling methods for Wallenius' and Fisher's noncentral hypergeometric distributions. *Communications in Statistics—Simulation and Computation* **37**, 2 doi:10.1080/03610910701790236

See Also

`vignette("UrnTheory")`
`BiasedUrn-Multivariate`.
`BiasedUrn`.
`fisher.test`

Examples

```
# get probability
dWNCHypergeo(12, 25, 32, 20, 2.5)
```

Index