

Package ‘GenBinomApps’

March 6, 2024

Type Package

Title Clopper-Pearson Confidence Interval and Generalized Binomial Distribution

Version 1.2.1

Date 2024-03-06

Depends stats

Author Horst Lewitschnig, David Lenzi

Maintainer Horst Lewitschnig <Horst.Lewitschnig@infineon.com>

Description Density, distribution function, quantile function and random generation for the Generalized Binomial Distribution. Functions to compute the Clopper-Pearson Confidence Interval and the required sample size. Enhanced model for burn-in studies, where failures are tackled by countermeasures.

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2024-03-06 10:00:02 UTC

R topics documented:

GenBinomApps-package	2
clopper.pearson.ci	3
cm.clopper.pearson.ci	5
cm.n.clopper.pearson	6
Generalized Binomial	7
n.clopper.pearson	9

Index	11
--------------	-----------

Description

Density, distribution function, quantile function, and random generation for the Generalized Binomial Distribution. Also included are functions to compute the Clopper-Pearson confidence interval limits for the standard case, for an enhanced model, and the required sample size for a given target probability for both models.

Details

This package originates from semiconductor manufacturing but can also be used for other purposes. The functions are based on the paper *Decision-Theoretical Model for Failures which are Tackled by Countermeasures*, Kurz et al. (2014).

The generalized binomial distribution is defined as the sum of independent, not identically binomial distributed random variables. That means, they have different success probabilities, and they can have different sample sizes.

Example: A person has to drive 3 routes at each working day. The probabilities for a radar control on these routes are 0.1% for the first route, 0.5% for the second route and 1% for the third route. The person has to drive route 1 and route 2 one time per day and route 3 two times per day. What are the probabilities to have 0, 1, 2, more than 2 controls at 100 working days?

Knowing that the number of controls is binomially distributed for each route:

$$R_1 \sim \text{binom}(100, 0.001), R_2 \sim \text{binom}(100, 0.005), R_3 \sim \text{binom}(200, 0.01)$$

Thus the sum of these binomially distributed random variables has a generalized binomial distribution with parameters $n_1 = 100, n_2 = 100, n_3 = 200, p_1 = 0.001, p_2 = 0.005, p_3 = 0.01$.

$$R = R_1 + R_2 + R_3, R \sim \text{gbinom}(100, 100, 200, 0.001, 0.005, 0.01)$$

In this example the probabilities $P(R = 0), P(R = 1), P(R = 2), P(R > 2)$ can be computed straightforward.

See the examples for the results.

Consider now a burn-in study in which k failures are observed. The number of failures is binomially distributed. Thus, the Clopper-Pearson confidence interval limits can be used to obtain a confidence interval for the failure probability. If failures occur, countermeasures should be implemented with a type specific effectivity. Consider the case of different failure types. That leads to more than one countermeasure. Each countermeasure can have a different effectivity. The probability for solving a certain number of failures can be computed with the generalized binomial distribution. It gives the likelihoods for various possible outcome scenarios, if the countermeasures would have been introduced from the beginning on. Based on the model in Kurz et al. (2014), confidence intervals can be computed.

Note

The generalized binomial distribution described here is also known as Poisson-binomial distribution.

Author(s)

Horst Lewitschnig, David Lenzi.

Maintainer: Horst Lewitschnig <Horst.Lewitschnig@infineon.com>

References

D.Kurz, H.Lewitschnig, J.Pilz, *Decision-Theoretical Model for Failures which are Tackled by Countermeasures*, IEEE Transactions on Reliability, Vol. 63, No. 2, June 2014.

K.J. Klauer, *Kriteriumsorientierte Tests*, Verlag fuer Psychologie, Hogrefe, 1987, Goettingen, p. 208 ff.

M.Fisz, *Wahrscheinlichkeitsrechnung und mathematische Statistik*, VEB Deutscher Verlag der Wissenschaften, 1973, p. 164 ff.

C.J.Clopper and E.S. Pearson, *The use of confidence or fiducial limits illustrated in the case of the binomial*, Biometrika, vol. 26, 404-413, 1934.

Examples

```
## n1=100, n2=100, n3=200, p1=0.001, p2=0.005, p3=0.01
dgbinom(c(0:2),size=c(100,100,200),prob=c(0.001,0.005,0.01))
# 0.07343377 0.19260317 0.25173556
pgbinom(2,size=c(100,100,200),prob=c(0.001,0.005,0.01),lower.tail=FALSE)
# 0.4822275
```

```
## n=110000 tested devices, 2 failures divided in 2 failure types k1=1, k2=1.
## 2 countermeasures with effectivities p1=0.5, p2=0.8
cm.clopper.pearson.ci(110000,size=c(1,1), cm.effect=c(0.5,0.8))
# Confidence.Interval = upper
# Lower.limit = 0
# Upper.limit = 3.32087e-05
# alpha = 0.1
```

```
## target failure probability p=0.00001, 2 failures divided in 2 failure types k1=1, k2=1.
## 2 countermeasures with effectivities p1=0.5, p2=0.8
cm.n.clopper.pearson(0.00001,size=c(1,1), cm.effect=c(0.5,0.8))
# 365299
```

clopper.pearson.ci *Clopper-Pearson Confidence Interval*

Description

Computing upper, lower or two-sided Clopper-Pearson confidence limits for a given confidence level.

Usage

```
clopper.pearson.ci(k, n, alpha = 0.1, CI = "upper")
```

Arguments

k	number of failures/successes.
n	number of trials.
alpha	significance level for the $(1 - \alpha) \cdot 100\%$ confidence level (default $\alpha = 0.1$).
CI	indicates the kind of the confidence interval, options: "upper" (default), "lower", "two.sided".

Details

Computes the confidence limits for the p of a binomial distribution. Confidence intervals are obtained by the definition of Clopper and Pearson. The two-sided interval for $k = 0$ is $(0, 1 - (\alpha/2)^{1/n})$, for $k = n$ it is $((\alpha/2)^{1/n}, 1)$.

Value

A data frame containing the kind of the confidence interval, upper and lower limits and the used significance level alpha.

References

D.Kurz, H.Lewitschnig, J.Pilz, *Decision-Theoretical Model for Failures which are Tackled by Countermeasures*, IEEE Transactions on Reliability, Vol. 63, No. 2, June 2014.

Thulin, Mans, *The cost of using exact confidence intervals for a binomial proportion*, Electronic Journal of Statistics, vol. 8, pp. 817-840, 2014.

C.J.Clopper and E.S. Pearson, *The use of confidence or fiducial limits illustrated in the case of the binomial*, Biometrika, vol. 26, pp. 404-413, 1934.

Examples

```
clopper.pearson.ci(5,100000,alpha=0.05)
# Confidence.Interval = upper
# Lower.limit = 0
# Upper.limit = 0.0001051275
# alpha = 0.05

clopper.pearson.ci(5,100000,CI="two.sided")
# Confidence.Interval = two.sided
# Lower.limit = 1.97017e-05
# Upper.limit = 0.0001051275
# alpha = 0.1
```

cm.clopper.pearson.ci *Clopper-Pearson Confidence Interval for Failures Which are Tackled by Countermeasures*

Description

Provides the extended Clopper-Pearson confidence limits for a failure model, where countermeasures are introduced.

Usage

```
cm.clopper.pearson.ci(n, size, cm.effect, alpha = 0.1, CI = "upper", uniroot.lower = 0,
  uniroot.upper = 1, uniroot.maxiter = 1e+05, uniroot.tol = 1e-10)
```

Arguments

n	sample size.
size	vector of the number of failures for each type.
cm.effect	vector of the success probabilities to solve a failure for each type. Corresponds to the probabilities p_i of a generalized binomial distribution.
alpha	significance level for the $(1 - \alpha) \cdot 100\%$ confidence level (default $\alpha = 0.1$).
CI	indicates the kind of the confidence interval, options: "upper" (default), "lower", "two.sided".
uniroot.lower	The value of the lower parameter sent to uniroot. Lower bound of the interval to be searched. See uniroot for more details.
uniroot.upper	The value of the upper parameter sent to uniroot. Upper bound of the interval to be searched. See uniroot for more details.
uniroot.maxiter	The value of the maxiter parameter sent to uniroot. Maximum number of iterations. See uniroot for more details.
uniroot.tol	The value of the tol parameter sent to uniroot. Convergence tolerance. See uniroot for more details.

Details

This is an extension of the Clopper-Pearson confidence interval, where different outcome scenarios of the random sampling are weighted by generalized binomial probabilities. The weights are the probabilities for observing $0, \dots, k$ failures after the introduction of countermeasures. Computes the confidence limits for the p of a binomial distribution, where p is the failure probability. The failures are tackled by countermeasures for specific failure types with different effectivity. See the references for further information.

Value

A data frame containing the kind of the confidence interval, upper and lower limits and the used significance level alpha.

References

D.Kurz, H.Lewitschnig, J.Pilz, *Decision-Theoretical Model for Failures which are Tackled by Countermeasures*, IEEE Transactions on Reliability, Vol. 63, No. 2, June 2014.

See Also

[uniroot](#), [dgbinom](#), [clopper.pearson.ci](#)

Examples

```
## n=110000 tested devices, 2 failures divided in 2 failure types k1=1, k2=1.
## 2 countermeasures with effectivities p1=0.5, p2=0.8
cm.clopper.pearson.ci(110000,size=c(1,1),cm.effect=c(0.5,0.8))
# Confidence.Interval = upper
# Lower.limit = 0
# Upper.limit = 3.32087e-05
# alpha = 0.1
```

cm.n.clopper.pearson *Required Sample Size - Countermeasure Model*

Description

Provides the required sample size with respect to the extended upper Clopper-Pearson limit for a failure model, where countermeasures are introduced.

Usage

```
cm.n.clopper.pearson(p, size, cm.effect, alpha = 0.1, uniroot.lower = k + 1,
  uniroot.upper = 1e+100, uniroot.tol = 1e-10, uniroot.maxiter = 1e+05)
```

Arguments

p	target probability.
size	vector of the number of failures for each type.
cm.effect	vector of the success probabilities to solve a failure for each type. Corresponds to the probabilities p_i of a generalized binomial distribution.
alpha	significance level for the $(1 - \alpha) \cdot 100\%$ confidence level (default $\alpha = 0.1$).
uniroot.lower	The value of the lower parameter sent to uniroot. Lower bound of the interval to be searched. See uniroot for more details.
uniroot.upper	The value of the upper parameter sent to uniroot. Upper bound of the interval to be searched. See uniroot for more details.
uniroot.maxiter	The value of the maxiter parameter sent to uniroot. Maximum number of iterations. See uniroot for more details.
uniroot.tol	The value of the tol parameter sent to uniroot. Convergence tolerance. See uniroot for more details.

Details

Provides the required sample size with respect to the extended upper Clopper-Pearson limit. It applies for the case that failures are tackled by countermeasures. That means countermeasures with different effectivities for each failure type are introduced. See the references for further information.

Value

The value for the required sample size.

References

D.Kurz, H.Lewitschnig, J.Pilz, *Decision-Theoretical Model for Failures which are Tackled by Countermeasures*, IEEE Transactions on Reliability, Vol. 63, No. 2, June 2014.

See Also

[uniroot](#), [dgbinom](#), [cm.clopper.pearson.ci](#), [n.clopper.pearson](#)

Examples

```
## target failure probability p=0.00001, 2 failures divided in 2 failure types k1=1, k2=1.
## 2 countermeasures with effectivities p1=0.5, p2=0.8
cm.n.clopper.pearson(0.00001, size=c(1,1), cm.effect=c(0.5,0.8))
# 365299
```

Generalized Binomial *The Generalized Binomial Distribution*

Description

Density, distribution function, quantile function and random generation for the generalized binomial distribution with parameter vectors size and prob.

Usage

```
dgbinom(x, size, prob, log = FALSE)
pgbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
qgbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
rgbinom(N, size, prob)
```

Arguments

x, q	vector of quantiles.
p	vector of probabilities.
N	number of observations.
size	vector of the number of trials for each type.
prob	vector of the success probabilities for each type.
log, log.p	logical; if TRUE probabilities p are given as log(p).
lower.tail	logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.

Details

The generalized binomial distribution with $\text{size}=c(n_1, \dots, n_r)$ and $\text{prob}=c(p_1, \dots, p_r)$ is the sum of r binomially distributed random variables with different p_i (and, in case, with different n_i):

$$Z = \sum_{i=1}^r Z_i, \quad Z \sim \text{gbinom}(\text{size}, \text{prob}), \quad \text{with } Z_i \sim \text{binom}(n_i, p_i), \quad i = 1, \dots, r.$$

Since the sum of Bernoulli distributed random variables is binomially distributed, Z can be also defined as:

$$Z = \sum_{i=1}^r \sum_{j=1}^{n_i} Z_{ij}, \quad \text{with } Z_{ij} \sim \text{binom}(1, p_i), \quad j = 1, \dots, n_i.$$

The pmf is obtained by an algorithm which is based on the convolution of Bernoulli distributions. See the references below for further information.

The quantile is defined as the smallest value x such that $F(x) \geq p$, where F is the cumulative distribution function.

`rgbinom` uses the inversion method (see Devroye, 1986).

Value

`dgbinom` gives the pmf, `pgbinom` gives the cdf, `qgbinom` gives the quantile function and `rgbinom` generates random deviates.

Note

If `size` contains just one trial number and `prob` one success probability, then the generalized binomial distribution results in the binomial distribution.

The generalized binomial distribution described here is also known as Poisson-binomial distribution. See the link below to the package `poibin` for further information.

References

D.Kurz, H.Lewitschnig, J.Pilz, *Decision-Theoretical Model for Failures which are Tackled by Countermeasures*, IEEE Transactions on Reliability, Vol. 63, No. 2, June 2014.

K.J. Klauer, *Kriteriumsorientierte Tests*, Verlag fuer Psychologie, Hogrefe, 1987, Goettingen, p. 208 ff.

M.Fisz, *Wahrscheinlichkeitsrechnung und mathematische Statistik*, VEB Deutscher Verlag der Wissenschaften, 1973, p. 164 ff.

L.Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, 1986, p. 85 ff.

See Also

[ppoibin](#), for another implementation of this distribution.
[dbinom](#)

Examples

```
## n=10 defect devices, divided in 3 failure types n1=2, n2=5, n3=3.
## 3 countermeasures with effectivities p1=0.8, p2=0.7, p3=0.3 are available.
## use dgbinom() to get the probabilities for x=0,...,10 failures solved.
```



```

dgbinom(x=c(0:10),size=c(2,5,3),prob=c(0.8,0.7,0.3))

## generation of N=100000 random values
rgbinom(100000,size=c(2,5,3),prob=c(0.8,0.7,0.3))

## n1=100, n2=100, n3=200, p1=0.001, p2=0.005, p3=0.01
dgbinom(c(0:2),size=c(100,100,200),prob=c(0.001,0.005,0.01))
# 0.07343377 0.19260317 0.25173556
pgbinom(2,size=c(100,100,200),prob=c(0.001,0.005,0.01),lower.tail=FALSE)
# 0.4822275

```

n.clopper.pearson	<i>Required Sample Size</i>
-------------------	-----------------------------

Description

Provides the required sample size with respect to the one-sided upper Clopper-Pearson limit.

Usage

```
n.clopper.pearson(k, p, alpha = 0.1, uniroot.lower = k + 1, uniroot.upper = 1e+100,
uniroot.maxiter = 1e+05, uniroot.tol = 1e-10)
```

Arguments

k	number of failures.
p	target probability.
alpha	significance level for the $(1 - \alpha) \cdot 100\%$ confidence level (default $\alpha = 0.1$).
uniroot.lower	The value of the lower parameter sent to uniroot. Lower bound of the interval to be searched. See uniroot for more details.
uniroot.upper	The value of the upper parameter sent to uniroot. Upper bound of the interval to be searched. See uniroot for more details.
uniroot.maxiter	The value of the maxiter parameter sent to uniroot. Maximum number of iterations. See uniroot for more details.
uniroot.tol	The value of the tol parameter sent to uniroot. Convergence tolerance. See uniroot for more details.

Details

Provides the required sample size with respect to the upper Clopper-Pearson limit for a given target failure probability at a certain confidence level.

Value

The value for the required sample size.

References

D.Kurz, H.Lewitschnig, J.Pilz, *Decision-Theoretical Model for Failures which are Tackled by Countermeasures*, IEEE Transactions on Reliability, Vol. 63, No. 2, June 2014.

See Also

[uniroot,clopper.pearson.ci](#)

Examples

```
## target failure probability p=0.0002, 8 failures
n.clopper.pearson(8,0.0002)
# 64972
```

Index

`clopper.pearson.ci`, [3](#), [6](#), [10](#)
`cm.clopper.pearson.ci`, [5](#), [7](#)
`cm.n.clopper.pearson`, [6](#)

`dbinom`, [8](#)
`dgbinom`, [6](#), [7](#)
`dgbinom (Generalized Binomial)`, [7](#)

`GenBinomApps (GenBinomApps-package)`, [2](#)
`GenBinomApps-package`, [2](#)
`Generalized Binomial`, [7](#)

`n.clopper.pearson`, [7](#), [9](#)

`pgbinom (Generalized Binomial)`, [7](#)
`ppoibin`, [8](#)

`qgbinom (Generalized Binomial)`, [7](#)
`rgbinom (Generalized Binomial)`, [7](#)

`uniroot`, [6](#), [7](#), [10](#)