Package: XNomial (via r-universe)

November 6, 2024

Type Package

Title Exact Goodness-of-Fit Test for Multinomial Data with Fixed Probabilities

Version 1.0.4.1

Date 2015-12-22

Description Tests whether a set of counts fit a given expected ratio. For example, a genetic cross might be expected to produce four types in the relative frequencies of 9:3:3:1. To see whether a set of observed counts fits this expectation, one can examine all possible outcomes with xmulti() or a random sample of them with xmonte() and find the probability of an observation deviating from the expectation by at least as much as the observed. As a measure of deviation from the expected, one can use the log-likelihood ratio, the multinomial probability, or the classic chi-square statistic. A histogram of the test statistic can also be plotted and compared with the asymptotic curve.

VignetteBuilder knitr

Depends R (>= 2.14)

Suggests knitr

NeedsCompilation yes

License GPL

RoxygenNote 5.0.0

Date/Publication 2024-11-05 05:38:55 UTC

Author Bill Engels [aut, cre]

Maintainer Bill Engels <wrengels@wisc.edu>

Repository https://wrengels.r-universe.dev

RemoteUrl https://github.com/cran/XNomial

RemoteRef HEAD

RemoteSha ab0c49bfdf18d3255939fff7392b350f3435dbb0

Contents

---

xmonte	*Perform Multinomial Goodness-Of-Fit Test By Monte-Carlo Simulations*

Description

Use xmonte to compute a P value to test whether a set of counts fits a specific multinomial distribution. It does this by examining a large number of random outcomes and finding the probability of those cases which deviate from the expectation by at least as much as the observed.

Usage

```
xmonte(obs, expr, ntrials = 1e+05, statName = "LLR", histobins = F,
  histobounds = c(0, 0), showCurve = T, detail = 1, safety = 1e+08)
```

Arguments

obs	vector containing the observed numbers. All are non-negative integers summing to > 0.
expr	vector containing expectation. The length should be the same as that of obs and they should be non-negative summing to > 0. They need not be integers or sum to one.
ntrials	the number of random trials to look at, such as ntrials=100000
statName	name of the test statistic to use as a measure of how deviant an observation is from the expectation. The choices are: "LLR" for the log-likelihood ratio, "Prob" for the probability, "Chisq" for the chisquare statistic.
histobins	specifies histogram plot. If set to 0, F or FALSE no histogram is plotted. If set to 1 or T or TRUE a histogram with 500 bins will be plotted. If set to a number > 1 a histogram with that number of bins is plotted.
histobounds	vector of length 2 indicating the bounds for the histogram, if any. If unspecified, bounds will be determined to include about 99.9 percent of the distribution.
showCurve	should an asymptotic curve be drawn over the histogram?
detail	how much detail should be reported concerning the P value. If 0, nothing is printed for cases where the function is used programmatically. Minimal information is printed if detail is set to 1, and additional information if it is set to 2.
safety	a large number, such as one billion, to set a limit on how many samples will be examined. This limit is there to avoid long computations.

Value

xmonte returns a list with the following components:

$ obs	the observed numbers used as imput
$ expr	expected ratios, arbitrary scale
$ ntrials	the number of random tables examined
$ statType	which test statistic was used
$ pLLR/pProb/pChi	
	the P value computed for the given test statistic
$ standard.error	
	the binomial standard error of the estimated P value
$ observedLLR	the value of LLR statistic for these data
$ observedProb	the multinomial probability of the observed data under the null hypothesis
$ observedChi	observed value of the chi square statistic
$ histobins	number of bins in the histogram (suppressed if zero)
$ histobounds	range in histogram (suppressed if not used)
$ histoData	data for histogram (suppressed if not used) Length is histobins
$ asymptotoc.p.value	
	the P value obtained from the classical asymptotic test – use for comparison only

Examples

```
#One of Gregor Mendel's crosses produced four types of pea seeds in the numbers as follows:
peas <- c(315, 108, 101, 32)
#and he expected them to appear in the ratio of 9:3:3:1 according to his genetic model.
expected <- c(9, 3, 3, 1)
#Test Mendels theory using
xmonte(peas, expected)
#To see a histogram of the likelihood ratio statistic, use:
xmonte(peas, expected, histobins = TRUE)
#The red areas of the histogram represent those outcomes deviating from the expected 9:3:3:1 ratio
#at least as much as the observed numbers. (Much has been made of the tendency for Mendel's data
#to fit the expectations better than expected!)
#If you wish to use the standard chisquare statistic as a measure of goodness-of-fit instead
#of the LLR, use:
xmonte(peas, expected, statName="Chisq", histobins=TRUE)
```

---

xmulti	*Perform Multinomial Goodness-Of-Fit Test By Full Enumeration*

Description

Use xmulti to compute a P value to test whether a set of counts fits a specific multinomial distribution. It does this by examining all possible outcomes with the same total count and determining the total (multinomial) probability of those cases which deviate from the expectation by at least as much as the observed. Please see the vignette for more.

Usage

```
xmulti(obs, expr, statName = "LLR", histobins = F, histobounds = c(0, 0),
  showCurve = T, detail = 1, safety = 1e+09)
```

Arguments

obs	vector containing the observed numbers. All are non-negative integers summing to > 0.
expr	vector containing expectation. The length should be the same as that of obs and they should be non-negative summing to > 0. They need not be integers or sum to one.
statName	name of the test statistic to use as a measure of how deviant an observation is from the expectation. The choices are: "LLR" for the log-likelihood ratio, "Prob" for the probability, "Chisq" for the chisquare statistic.
histobins	specifies histogram plot. If set to 0, F or FALSE no histogram is plotted. If set to 1 or T or TRUE a histogram with 500 bins will be plotted. If set to a number > 1 a histogram with that number of bins is plotted.
histobounds	vector of length 2 indicating the bounds for the histogram, if any. If unspecified, bounds will be determined to include about 99.9 percent of the distribution.
showCurve	should an asymptotic curve be drawn over the histogram?
detail	how much detail should be reported concerning the P value. If 0, nothing is printed for cases where the function is used programmatically. Minimal information is printed if detail is set to 1, and additional information if it is set to 2.
safety	a large number, such as one billion, to set a limit on how many samples will be examined. This limit is there to avoid long computations.

Value

xmulti returns a list with the following components:

$ obs	the observed numbers used as imput
$ expr	expected ratios
$ statType	which test statistic was used
$ pLLR	the P value with LLR as the test statistic
$ pProb	the P value with the multinomial probability as test statistic
$ pChi	the P value with the chisquare as test statistic
$ observedLLR	the value of LLR statistic for these data
$ observedProb	the multinomial probability of the observed data under the null hypothesis
$ observedChi	observed value of the chi square statistic
$ histobins	number of bins in the histogram (suppressed if zero)
$ histobounds	range in histogram (suppressed if not used)
$ histoData	data for histogram (suppressed if not used) Length is histobins

`$ asymptotoc.p.value`
> the P value obtained from the classical asymptotic test

`$ cases.examined`
> the total number of possible tables

Examples

```
#
#One of Gregor Mendel's crosses produced four types of pea seeds in the numbers:
#
peas <- c(315, 108, 101, 32)
#
#and he expected them to appear in the ratio of 9:3:3:1 according to his genetic model.
#
expected <- c(9, 3, 3, 1)
#
#Test Mendel's theory using
#
xmulti(peas, expected)
#
#In this example, the number of cases examined was 28956759,
#and it probably took your computer less than half a second.
#To see a histogram of the likelihood ratio statistic, use:
#
xmulti(peas, expected, histobins = TRUE)
#
#The red areas of the histogram represent those outcomes deviating from the expected 9:3:3:1 ratio
#at least as much as the observed numbers. (Much has been made of the tendency for Mendel's data
#to fit the expectations better than expected!)
#If you wish to use the standard chisquare statistic as a measure of goodness-of-fit instead
#of the LLR, use:
#
xmulti(peas, expected, statName="Chisq", histobins=TRUE)
```

Index