

# Package ‘HardyWeinberg’

November 23, 2009

**Type** Package

**Title** Graphical tests for Hardy-Weinberg equilibrium

**Version** 1.4

**Date** 2009-11-23

**Author** Jan Graffelman <[jan.graffelman@upc.edu](mailto:jan.graffelman@upc.edu)>

**Maintainer** Jan Graffelman <[jan.graffelman@upc.edu](mailto:jan.graffelman@upc.edu)>

**Depends** R (>= 1.8.0), graphics, stats

**Description** Package HardyWeinberg is a package for exploring bi-allelic marker data. It focuses on the graphical representation of the results of tests for Hardy-Weinberg equilibrium in a ternary plot. Routines for several tests for Hardy-Weinberg equilibrium are included in the package.

**License** GPL (>= 2)

**URL** <http://www.r-project.org>, <http://www-eio.upc.edu/~jan>

**Repository** CRAN

**Date/Publication** 2009-11-23 19:19:13

## R topics documented:

HardyWeinberg-package	2
af	3
GenerateSamples	4
HWChisq	4
HWChisqMat	5
HWCondProbAB	6
HWData	7
HWExact	8
HWExactMat	9
HWLratio	10
HWternaryPlot	11
maf	14
UniqueGenotypeCounts	14

---

HardyWeinberg-package

*Graphical tests for Hardy-Weinberg equilibrium*

---

## Description

The HardyWeinberg-package provides graphical tests for Hardy-Weinberg equilibrium (HWE) based on the ternary plot (de Finetti diagram). The package constructs ternary plots for genotypic compositions for bi-allelic marker data. The acceptance region for several statistical tests of HWE (Chisquare test, Chisquare test with continuity correction, Haldane's exact test) can be depicted inside the ternary plot with the routines of the package. Large numbers of bi-allelic markers (e.g. SNPs) can be represented in a single ternary diagram and the statistical (non)significance of a test for HWE can be inferred from the position of the marker in the plot.

## Details

Package: HardyWeinberg  
Type: Package  
Version: 1.3  
Date: 2008-04-22  
License: GPL Version 2 or later.

The most important function of the package is `HWTernaryPlot` that can be used to create ternary plots with acceptance regions for HWE. Other routines implement statistical tests for HWE such as `HWChisq` and `HWLratio`.

## Author(s)

Jan Graffelman

Maintainer: Jan Graffelman <jan.graffelman@upc.edu>

## References

Weir, B.S. (1996) *Genetic Data Analysis II*. Sinauer Associates, Massachusetts.

Graffelman, J. and Morales, J. (2008) Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Human Heredity* 65(2):77-84.

## Examples

```
library(HardyWeinberg)

# make random compositions that are in HWE
```

```
set.seed(123)

m <- 100 # number of markers
n <- 100 # sample size

out <- HWDData(n,m)
Xc <- out$Xc
out <- HWTernaryPlot(Xc,100,region=1,vertex.cex=2,signifcolour=TRUE)
```

---

af

*Function to compute allele frequencies*

---

### Description

Function `af` computes the allele frequencies for a matrix or a vector containing genotypic compositions.

### Usage

```
af(x)
```

### Arguments

`x` a vector or matrix with compositions

### Value

a vector with allele frequencies

### Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

### See Also

[maf](#)

### Examples

```
X <- as.vector(rmultinom(1,100,c(0.5,0.4,0.1)))
X <- X/sum(X)
print(X)
print(af(X))
```

---

GenerateSamples      *Generate genotypic compositions*

---

**Description**

GenerateSamples generates all possible genotypic compositions (AA,AB,BB) for a given sample size n.

**Usage**

```
GenerateSamples(n = 5)
```

**Arguments**

n                      the desired sample size

**Value**

returns a matrix with in each row a possible genotypic composition for the given sample size.

**Author(s)**

Jan Graffelman <jan.graffelman@upc.edu>

**Examples**

```
GenerateSamples(5)
```

---

HWChisq                      *Chi square tests for Hardy Weinberg equilibrium*

---

**Description**

HWChisq performs the chi-square test for Hardy Weinberg equilibrium with or without continuity correction.

**Usage**

```
HWChisq(X, cc = 0.5, alpha = 0.05, verbose = FALSE)
```

**Arguments**

X                      X a vector containing the genotypic counts (AA,AB,BB).  
cc                      cc the continuity correction parameter (default cc = 0).  
alpha                      significance level (0.05 by default).  
verbose                      verbose = 1 prints results, verbose = 0 is silent.

**Value**

HWChisq returns a list with the components:

chisq	value of the chi-square statistic. NA is returned if the marker is monomorphic.
pval	p-value of the chi-square test for Hardy-Weinberg equilibrium.
D	Half the deviation from Hardy-Weinberg equilibrium for the AB genotype.
p	allele frequency of A.

**Author(s)**

Jan Graffelman <jan.graffelman@upc.edu>

**See Also**

[HWLratio](#)

**Examples**

```
x <- c(298, 489, 213)
names(x) <- c("MM", "MN", "NN")
HW.test <- HWChisq(x, verbose=TRUE)
```

---

HWChisqMat

*Matrix version of HWChisq*

---

**Description**

HWChisqMat executes the Chisquare test for HWE for each row in a matrix.

**Usage**

```
HWChisqMat(X, ...)
```

**Arguments**

X	A n times 3 matrix of genotypic counts (AA,AB,BB)
...	extra arguments that are passed on to HWChisq

**Value**

pvalvec	Vector with the p-values of each test
chisqvec	Vector with the chi-square statistics
Dvec	Vector with deviations from independence

**Author(s)**

Jan Graffelman <jan.graffelman@upc.edu>

**See Also**[HWChisq](#)**Examples**

```
X <- HWData(100,10)$Xt
colnames(X) <- c("MM", "MN", "NN")
Results <- HWChisqMat(X)
Output <- cbind(X, Results$chisqvec, Results$pvalvec)
print(Output)
```

---

HWCondProbAB

*Compute probability of a genotypic sample*


---

**Description**

Computes the probability of a particular genotypic sample given the allele count, sample size and number of heterozygotes.

**Usage**

```
HWCondProbAB(n, nA, nAB)
```

**Arguments**

n	n is the total sample size (total number of individuals)
nA	nA is the number of A alleles in the sample
nAB	nAB is the number of heterozygotes in the sample

**Value**

p	probability of the particular sample
---	--------------------------------------

**Author(s)**

Jan Graffelman (jan.graffelman@upc.edu)

**See Also**[HWExact](#)**Examples**

```
x <- c(298, 489, 213)
names(x) <- c("MM", "MN", "NN")
n <- sum(x)
nM <- 2*x[1]+x[2]
nMN <- x[2]
p <- HWCondProbAB(n, nM, nMN)
```

**Description**

HWDData generates samples of genotypic counts under various schemes. It mainly uses sampling from the multinomial distribution given Hardy-Weinberg allele frequencies.

**Usage**

```
HWDData(n = 100, nm = 100, f = 0, p = NULL, pfixed = FALSE, exactequilibrium = FALSE)
```

**Arguments**

n	the sample size.
nm	the number of markers (or samples).
f	the inbreeding coefficient
p	the allele frequency
pfixed	if TRUE Haldane's distribution is used for sampling, if FALSE a multinomial distribution is used
exactequilibrium	generates data in exact HWE if set to TRUE
pdist	take a random allele frequency from a uniform or beta distribution of pfixed = FALSE and p is not given.
...	specific parameters for the uniform or beta

**Value**

Xt	the genotypic counts.
Xc	the genotypic compositions.

**Author(s)**

Jan Graffelman (jan.graffelman@upc.edu)

**See Also**

HWternaryPlot

**Examples**

```
n <- 100
nm <- 100
out <- HWDData(n, nm)
```

---

 HWExact

*Exact test for Hardy-Weinberg equilibrium*


---

**Description**

HWExact performs an exact test for Hardy-Weinberg equilibrium

**Usage**

```
HWExact(X, alternative = "two.sided", pvaluetype = "dost", verbose = FALSE)
```

**Arguments**

X	vector with the genotype counts AA, AB, BB
alternative	two.sided (default) will perform a two-sided test where both an excess and a dearth of heterozygotes count as evidence against HWE. less is a one-sided test where only dearth of heterozygotes counts a evidence against HWE, greater is a one-sided test where only excess of heterozygotes counts as evidence against HWE.
pvaluetype	if pvaluetype is set to dost then the p-value of a two-sided test is computed as twice the tail area of a one-sided test. When set to selome, the p-value is computed as the sum of the probabilities of all samples less or equally likely as the current sample.
verbose	print results or not.

**Details**

HWExact use the recursion equations described by Wigginton et. al.

**Value**

pval	p-value of the exact test
pofthesample	probability of the observed sample

**Note**

HWExact is designed for a fast analysis of a large set of markers. HWExact is slower, but provides more detailed statistics for the sample under study.

**Author(s)**

Jan Graffelman (jan.graffelman@upc.edu)

**References**

Weir, B.S. (1996) Genetic data analysis II. Sinauer Associates, Massachusetts. See Chapter3.  
 Wigginton, J.E., Cutler, D.J. and Abecasis, G.R. (2005) A note on exact tests of Hardy-Weinberg equilibrium, American Journal of Human Genetics (76) pp. 887-893.

**See Also**

[HWLratio](#), [HWChisq](#)

**Examples**

```
x <- c(298,489,213)
names(x) <- c("MM", "MN", "NN")
HW.test <- HWExact(x)
print(HW.test)
```

---

HWExactMat

*Matrix version of HWExact*

---

**Description**

HWExactMat executes a fast Exact test for HWE for each row in a matrix.

**Usage**

```
HWExactMat(X, ...)
```

**Arguments**

X	A n times 3 matrix of genotypic counts (AA,AB,BB)
...	extra arguments that are passed on to HWExact

**Value**

pvalvec	Vector with the p-values of each test
---------	---------------------------------------

**Author(s)**

Jan Graffelman <[jan.graffelman@upc.edu](mailto:jan.graffelman@upc.edu)>

**See Also**

[HWExact](#)

**Examples**

```
X <- HWData(100,10)$Xt
colnames(X) <- c("MM", "MN", "NN")
Results <- HWExactMat(X)
Output <- cbind(X,Results$pvalvec)
print(Output)
```

---

`HWLratio`*Likelihood ratio test for Hardy Weinberg equilibrium*

---

**Description**

`HWLratio` performs the Likelihood ratio test for Hardy Weinberg equilibrium.

**Usage**

```
HWLratio(X, verbose = FALSE)
```

**Arguments**

`X`                    X a vector containing the genotypic counts (AA,AB,BB).  
`verbose`            `verbose = 1` prints results, `verbose = 0` is silent.

**Value**

`HWLratio` returns a list with the components:

`Lambda`            the likelihood ratio  
`G2`                  $-2 \cdot \log(\text{Lambda})$   
`pval`              the p-value

**Author(s)**

Jan Graffelman <jan.graffelman@upc.edu>

**References**

Weir, B.S. (1996) Genetic data analysis II. Sinauer Associates, Massachusetts. See Chapter 3.

**See Also**

[HWChisq](#)

**Examples**

```
x <- c(298, 489, 213)
names(x) <- c("MM", "MN", "NN")
HW.test <- HWLratio(x, verbose=TRUE)
```

**Description**

HWTernaryPlot is a routine that draws a ternary plot for three-way genotypic compositions (AA,AB,BB), and represents the acceptance region for different tests for Hardy-Weinberg equilibrium (HWE) in the plot. This allows for graphical testing of a large set of markers (e.g. SNPs) for HWE. The (non) significance of the test for HWE can be inferred from the position of the marker in the ternary plot. Different statistical tests for HWE can be done graphically with this routine: the ordinary chisquare test, the chisquare test with continuity correction and the Haldane's exact test.

**Usage**

```
HWTernaryPlot(X, n = NA, addmarkers = TRUE, newframe = TRUE, hwcurve = TRUE,
vbounds = TRUE, mafbounds = FALSE, mafvalue = 0.05, axis = 0, region = 1,
vertexlab = colnames(X), alpha = 0.05, vertex.cex = 1, pch = 19, cc = 0.5,
markercol = "black", markerbgcol = "black", cex = 0.75, axislab = "",
verbose = FALSE, markerlab = NULL, mcex = 1, connect = FALSE, curvecols =
rep("black",5), signifcolour = TRUE, curtyp = "solid", ssf = "max", pvaluetype = "c
```

**Arguments**

X	a matrix of n genotypic compositions or counts. If it is a matrix of compositions, X should have (n rows that sum 1, and 3 columns, with the relative frequencies of AA, AB and BB respectively. Argument n should be supplied as well. If X is a matrix of raw genotypic counts, it should have 3 columns with the absolute counts of AA, AB and BB respectively. Argument n may be supplied and will be used for painting acceptance regions. If not supplied n is computed from the data in X.
n	the samples size (for a complete composition with no missing data).
addmarkers	represent markers by dots in the triangle (addmarkers=TRUE) or not (addmarkers=FALSE).
newframe	allows for plotting additional markers in an already existing ternary plot. Overplotting is achieved by setting newframe to FALSE. Setting newframe = TRUE (default) will create a new ternary plot.
hwcurve	draw the HW parabola in the plot (hwcurve=TRUE) or not (hwcurve=FALSE).
vbounds	indicate the area corresponding to expected counts > 5 (vbounds=TRUE) or not (vbounds=FALSE).
mafbounds	indicate the area corresponding to MAF < mafvalue.
mafvalue	a critical value for the minor allele frequency (MAF).
axis	draw a vertex axis 0 = no axis is drawn 1 = draw the AA axis

	2 = draw the AB axis 3 = draw the BB axis
region	the type of acceptance region to be delimited in the triangle 0 = no acceptance region is drawn 1 = draw the acceptance region corresponding to a Chi-square test 2 = draw the acceptance region corresponding to a Chi-square test with continuity correction 3 = draw the acceptance region corresponding to a Chi-square test with continuity correction for $D > 0$ 4 = draw the acceptance region corresponding to a Chi-square test with continuity correction for $D < 0$ 5 = draw the acceptance regions for all preceding tests simultaneously 6 = draw the acceptance region corresponding to a Chi-square test with continuity correction with the upper limit for $D > 0$ and the lower limit for $D < 0$ 7 = draw the acceptance region corresponding to a two-sided exact test
vertexlab	labels for the three vertices of the triangle
alpha	significance level (0.05 by default)
vertex.cex	character expansion factor for the labels of the vertices of the triangle.
pch	the plotting character used to represent the markers.
cc	value for the continuity correction parameter (0.5 by default).
markercol	vector with colours for the marker points in the triangle.
markerbgcol	vector with background colours for the marker points in the triangle.
cex	expansion factor for the marker points in the triangle.
axislab	a label to be put under the horizontal axis.
verbose	print information on the numerically found cut-points between curves of the acceptance region and the edges of the triangle.
markerlab	labels for the markers in the triangle.
mcex	character expansion factor for the labels of the markers in the ternary plot.
connect	connect the represented markers by a line in the ternary plot.
curvecols	a vector with four colour specifications for the different curves that can be used to delimit the HW acceptance region. E.g. <code>curvecols=c("red", "green", "blue", "black", "purple")</code> will paint the Hardy-Weinberg curve red, the limits of the acceptance region for an ordinary chi-square test for HWE green, the limits of the acceptance region for a chi-square test with continuity correction when $D > 0$ blue and the limits of the acceptance region for a chi-square test with continuity correction when $D < 0$ black, and the limits of the exact acceptance region purple.
signifcolour	colour the marker points automatically according to the result of a significance test (green markers non-significant, red markers significant). <code>signifcolour</code> only takes effect if <code>region</code> is set to 1, 2 or 7.
curtyp	style of the drawn curves ("dashed", "solid", "dotted", ...)

<code>ssf</code>	sample size function ("max", "min", "mean", "median", ...). Indicates how the sample size for drawing acceptance regions is determined from the matrix of counts.
<code>pvalue</code>	method to compute p-values in an exact test ("dost" or "selome")
<code>...</code>	other arguments passed on to the plot function (e.g. <code>main</code> for a main title).

### Value

<code>minp</code>	minimum allele frequency above which testing for HWE is appropriate (expected counts exceeding 5).
<code>maxp</code>	maximum allele frequency below which testing for HWE is appropriate.
<code>inrange</code>	number of markers in the appropriate range.
<code>percinrange</code>	percentage of markers in the appropriate.
<code>nsignif</code>	number of significant markers (only if <code>region</code> equals 1,2 or 7.)

### Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

### References

Graffelman, J. and Morales, J. (2008) Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Human Heredity* 65(2):77-84.

### See Also

[HWChisq](#)

### Examples

```
n <- 100 # sample size
m <- 100 # number of markers

out <- HWData(n,m)
Xc <- out$Xc

HWTernaryPlot(Xc,100,region=1,hwcurve=TRUE,vbounds=FALSE,vertex.cex=2)
```

maf

*Function to compute minor allele frequencies*

---

**Description**

Function `maf` computes the minor allele frequency for a matrix or vector of compositions.

**Usage**

```
maf(x)
```

**Arguments**

`x` a vector or matrix of genotypic compositions

**Value**

a vector of minor allele frequencies.

**Author(s)**

Jan Graffelman ([jan.graffelman@upc.edu](mailto:jan.graffelman@upc.edu))

**Examples**

```
X <- as.vector(rmultinom(1,100,c(0.5,0.4,0.1)))
X <- X/sum(X)
print(X)
print(maf(X))
```

---

`UniqueGenotypeCounts`*Extract unique genotypic compositions from a matrix*

---

**Description**

Function `UniqueGenotypeCounts` creates a matrix containing only the unique rows in the given matrix, together with their frequency of occurrence

**Usage**

```
UniqueGenotypeCounts(X)
```

**Arguments**

`X` A  $n$  by 3 matrix with genotypic counts (AA,AB,BB)

**Value**

A matrix with 4 columns, AA, AB, BB, and frequency of occurrence

**Author(s)**

Jan Graffelman <jan.graffelman@upc.edu>

**See Also**

[GenerateSamples](#)

**Examples**

```
set.seed(123)
X <- HWData(n=100, nm=100)$Xt
print(nrow(X))
Y <- UniqueGenotypeCounts(X)
print(nrow(Y))
print(sum(Y$w))
```

# Index

## \*Topic **aplot**

HWternaryPlot, 11

## \*Topic **datagen**

GenerateSamples, 4

HWData, 7

## \*Topic **htest**

HWChisq, 4

HWChisqMat, 5

HWExact, 8

HWExactMat, 9

HWLratio, 10

## \*Topic **misc**

af, 3

HWCondProbAB, 6

maf, 14

UniqueGenotypeCounts, 14

## \*Topic **package**

HardyWeinberg-package, 2

af, 3

GenerateSamples, 4, 15

HardyWeinberg

(HardyWeinberg-package), 2

HardyWeinberg-package, 2

HWChisq, 4, 6, 9, 10, 13

HWChisqMat, 5

HWCondProbAB, 6

HWData, 7

HWExact, 6, 8, 9

HWExactMat, 9

HWLratio, 5, 9, 10

HWternaryPlot, 11

maf, 3, 14

UniqueGenotypeCounts, 14