The HardyWeinberg Package

version 1.2

Jan Graffelman

Department of Statistics and Operations Research
Universitat Politècnica de Catalunya
Avinguda Diagonal 647, 08028 Barcelona, Spain.
email: jan.graffelman@upc.edu

April 2008

1 Introduction

This guide gives some instructions on how to perform graphical significance tests for Hardy-Weinberg equilibrium (HWE) by depicting the acceptance region for HWE in a ternary plot with routines from the package `HardyWeinberg`. The outline of this guide is as follows. Section 2 describes how the R package `HardyWeinberg` can be installed. Section 3 shows how to perform some of the classical tests for Hardy-Weinberg equilibrium with routines from the package. Finally, Section 4 shows how to contruct ternary plots with the HW acceptance region and how to perform graphical tests for HWE. We refer to Graffelman & Morales (2007) for the theoretical foundation of the graphical tests. If you appreciate this software then please cite the following paper in your work:

Graffelman, J. & Morales-Camarena, J. (2008) Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Human Heredity* **65**(2): 77-84. (clic here to access the paper)

2 Installation

Packages in R can be installed inside the program with the option "Packages" in the main menu and then choosing "Install package" and picking the package "HardyWeinberg". Typing:

```
> library(HardyWeinberg)
```

will make the functions `HWChisq, HWData, HWLratio` and `HWTernaryPlot` available.

3 Classical tests for Hardy-Weinberg equilibrium

We show how to perform several classical tests for Hardy-Weinberg equilibrium. As an example we use a sample of 1000 individuals genotyped for the MN blood group locus described by Hedrick (2005, Table 2.4). We store the genotypic counts (298, 489 and 213 for MM, MN and NN respectively) in a vector `x`:

```
> x <- c(298, 489, 213)
> HW.test <- HWChisq(x, verbose = TRUE)

Chi2 =  0.2214896 p-value =  0.6379073 D =  -3.69375
```

This shows that the χ^2-statistic has value 0.2215, and that the corresponding p-value for the test is 0.6379. Taking a significance level of $\alpha = 0.05$, we do not reject HWE for the MN locus.

When `verbose` is set to `FALSE` (default) the test is silent, and `HW.test` is a list containg the results of the test (χ^2-statistic, the p-value of the test, half the deviation from HWE (`D`) for the heterozygote ($D = \frac{1}{2}(f_{AB} - e_{AB})$) and the allele frequency (`p`) of M).

```
> HW.test <- HWChisq(x)
> print(HW.test)

$chisq
[1] 0.2214896

$pval
[1] 0.6379073

$D
[1] -3.69375

$p
[1] 0.5425
```

The χ^2-test can also be performed with Yates' continuity correction by setting the `cc` parameter:

```
> HW.test <- HWChisq(x, cc = 0.5, verbose = TRUE)

Chi2 =  0.1789563 p-value =  0.6722717 D =  -3.69375
```

This gives a smaller χ^2-statistic and a larger p-value in comparison with the previous test. The likelihood ratio test (Weir, 1996, Chapter 3) for HWE can be performed by typing

```
> HW.lrtest <- HWLratio(x, verbose = TRUE)

G2 = 0.2214663 p-value = 0.637925
```

Note that the G^2-statistic and the p-value obtained are very close to the χ^2-statistic and its p-value. The Fisher exact test for HWE can be performed by using routine `fisher.test` from the `stats` package. The genotypic counts are re-organized into a 2×2 table:

```
> m <- matrix(c(x[1], x[2]/2, x[2]/2, x[3]), ncol = 2)
> colnames(m) <- c("M", "N")
> rownames(m) <- c("M", "N")
> print(m)

      M     N
M 298.0 244.5
N 244.5 213.0

> fisher.test(m, alternative = "two.sided")

        Fisher's Exact Test for Count Data

data:  m
p-value = 0.6555
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.8238894 1.3794978
sample estimates:
odds ratio
  1.066071
```

The Fisher exact test leads to the same conclusion, we do not reject HWE (`p=0.6555`). `HWChisq` and `HWLratio` assume that the data are supplied as a vector of genotypic counts listed in order (AA,AB,BB). Additional test for HWE may be added to the package in the near future.

4 Graphical tests for Hardy-Weinberg equilibrium

This section shows how to create ternary plots for a database of marker data (e.g. SNPs) and shows how the depict the acceptance region for HWE in the ternary plot, using different tests. An example with simulated data follows below. We obtain $m = 100$ markers for $n = 100$ individuals by taking random samples from a multinomial distribution with $\theta_{AA} = p^2$, $\theta_{AB} = 2pq$, and $\theta_{BB} = q^2$. This is done by routine `HWData`, which can generate data sets that are in Hardy-Weinberg equilibrium. Routine `HWData` can generate data that are in exact equilibrium (`exactequilibrium = TRUE`) or that are generated from a multinomial distribution (default). `HWData` returns a list with both the matrix of genotypic counts `Xt` and the matrix with genotypic compositions `Xc` with the relative frequencies of AA, AB and BB.

```
> set.seed(123)
> m <- 100
> n <- 100
> Xc <- HWData(n, m)$Xc
```

We create four different ternary plots for the simulated marker data shown in Figure 1. Panel (a) simply depicts the 100 genotypic compositions in a ternary plot. Note the marked curvature in the cloud of points. Panel (b) shows a nicer ternary plot with the HWE curve and the acceptance region for HWE according to an ordinary χ^2-test. Green markers are not significant, red markers significant ($\alpha = 0.05$). 6 markers show up significant. Panel (c) shows the same data, but the acceptance region represented corresponds to a χ^2-test with continuity correction (`cc = 0.5`), with separate curves for $D > 0$ and $D < 0$. Some markers previously significant markers now turn up insignificant. Panel (d) shows the acceptance region for Fisher's exact test. This option takes more computer time. The significant markers are the same ones that are significant in the χ^2-test with continuity correction. More examples with human SNP data are given in Graffelman & Morales (2008).

```
> plot.new()
> opar <- par(mfrow = c(2, 2), mar = c(3, 5, 3, 1) + 0.1, mex = 0.75,
+     oma = c(2, 0, 2, 0), new = TRUE)
> par(mfg = c(1, 1))
> Res <- HWTernaryPlot(Xc, 100, region = 0, hwcurve = FALSE, vbounds = FALSE,
+     vertex.cex = 1.25, main = "(a)")
> par(mfg = c(1, 2))
> Res <- HWTernaryPlot(Xc, 100, region = 1, vertex.cex = 1.25,
+     signifcolour = TRUE, main = "(b)")
> par(mfg = c(2, 1))
> Res <- HWTernaryPlot(Xc, 100, region = 2, vertex.cex = 1.25,
+     signifcolour = TRUE, main = "(c)")
> par(mfg = c(2, 2))
> Res <- HWTernaryPlot(Xc, 100, region = 7, vertex.cex = 1.25,
+     signifcolour = TRUE, main = "(d)")
> par(opar)
```

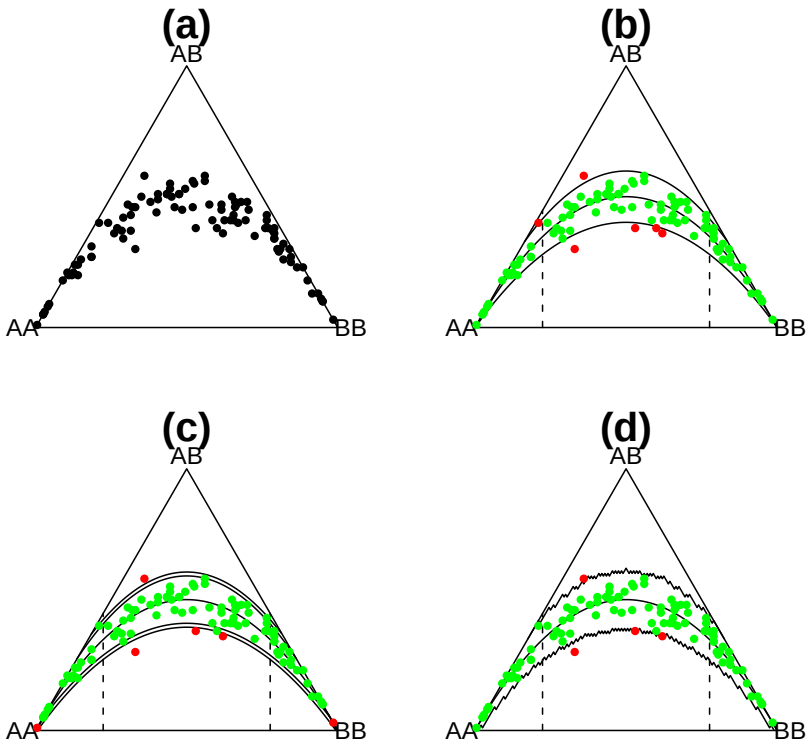

Figure 1: Ternary plot of 100 simulated SNPs for 100 individuals. (a): ordinary ternary plot, (b): with χ^2-acceptance region, (c) with acceptance region for χ^2-test with continuity correction, (d): with acceptance region for two-tailed FE test.

5 Online documentation

The online documentation for the most important routines of the package is included below.

`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, verbose = FALSE)
```

Arguments

`X`	`X` a vector containg the genotypic counts (AA,AB,BB).
`cc`	`cc` the continuity correction parameter (default `cc = 0`).
`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)
```

Description

HWData takes samples from the multinomial distribution given Hardy-Weinberg allele frequencies.

Usage

```
HWData(n = 100, nm = 100, pfixed = NULL, exactequilibrium = FALSE, pdist
= "runif",...)
```

Arguments

`n`	the sample size.
`nm`	the number of markers (or samples).
`pfixed`	take a fixed allele frequency with value `pfixed`.
`exactequilibrium`	generate data in exact HWE or use the multinomial distribution
`pdist`	take a random allele frequency from a uniform or beta distribution.
`...`	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 <- HWData(n,nm)
```

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*log(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)
```

`HWTernaryPlot` *Ternary plot with the Hardy-Weinberg acceptance region*

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 Fisher exact test.

Usage

```
HWTernaryPlot(X, n, 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, ...)
```

Arguments

`X`	a matrix of `n` genotypic compositions (`n` rows, rows summing 1, 3 columns, AA, AB and BB respectively).
`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 Fisher 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. `curvecols=c("red", "green","blue","black","purple")` 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 FE acceptance region purple.
`signifcolour`	colour the marker points automatically according to the result of a signifance test (green markers non-siginficant, red markers significant). `signifcolour` only takes effect if `region` is set to 1, 2 or 7.
`...`	other arguments passed on to the plot function (e.g. `main` for a main title).

Value

`minp`	minimum allele frequency above which testing for HWE is appropriate (expected counts exceeding 5).
`maxp`	maximum allele frequency below which testing for HWE is appropriate.
`inrange`	number of markers in the appropriate range.
`percinrange`	percentage of markers in the appropriate.
`nsignif`	number of significant markers (only if `region` 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

```
nm <- 100 # number of markers
n <- 100
Xc <- HWData(n,nm)$Xc

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

Acknowledgements

This work was partially supported by the Spanish grant SEJ2006-13537 This document was generated by Sweave (Leisch, 2002).

6 References

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

Hedrick, P. W. (2005) *Genetics of Populations.* Third edition. Jones and Bartlett Publishers, Sudbury, Massachusetts.

Leisch, F. (2002) Sweave: Dynamic generation of statistical reports using literate data analysis. *Compstat 2002, Proceedings in Computational Statistics.* pp. 575-580, Physica Verlag, Heidelberg. ISBN 3-7908-1517-9 URL http:/www.ci.tuwien.ac.at/ leisch/Sweave.

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