Marek Gągolewski
Konstancja Bobecka-Wesołowska
Przemysław Grzegorzewski

Computer Statistics with R

8. Nonparametric Tests

Faculty of Mathematics and Information Science
Warsaw University of Technology
[Last update: December 9, 2012]

Contents

Info

These tutorials are likely to contain bugs and typos. In case you find any don't hesitate to *contact us*! Thanks in advance!

8.1. Goodness-of-fit tests

8.1.1. Tests for normality

To verify H_0: *given sample is normally distributed* versus K: *sample is not normally distributed*, we may use the well-known Shapiro-Wilk test (`shapiro.test()`).

Other tests for normality, included in the package `nortest` (loaded via `library(nortest);`), are also available: the Cramer-von Mises test (`cvm.test()`), Anderson-Darling (`ad.test()`), Lilliefors (`lillie.test()`), Pearson chi-squared (`pearson.test()`), and Shapiro-Francia test (`sf.test()`).

Moreover, there exists a quite intuitive graphical method called a normal quantile-quantile plot, implemented in the function `qqnorm()` (see also: `?qqline`). It may give us a general intuition on whether a given sample *may* be normally distributed or not. For example:

```
x <- rnorm(100)
qqnorm(x)
qqline(x, col = "red")
```

8.1.2. Pearson chi-squared test

Let $\mathbf{X} = (X_1, X_2, \ldots, X_{n_X})$ i.i.d. F, where F is an unknown c.d.f.

We would like to verify a null hypothesis: $H_0 : F = F_0$, where F_0 is a theoretical c.d.f. of interest against an alternative hypothesis: $K : F \neq F_0$.

Let us categorize all the observations into r disjoint classes (intervals) $S_1 = (-\infty, a_1], S_2 = (a_1, a_2], \ldots, S_r = (a_{r-1}, \infty)$. Person's chi-squared test statistic:

$$T = \sum_{i=1}^{r} d_i^2 \quad \overset{H_0}{\sim}_{n\to\infty} \; \chi^2_{r-1-t}, \tag{8.1}$$

where $d_i = \frac{N_i - np_i}{\sqrt{np_i}}$, N_i is the number of observations in the i-th class and $0 < p_i = \mathrm{P}_{F_0}(X \in S_i)$.

For large n, we may use the critical region: $R(\alpha) = \left[\chi^2_{1-\alpha, r-1}, \infty\right)$. In practice this method is appropriate for $np_i \geq 5$ and $n \geq 100$.

R implementation: see `?chisq.test`.

For example, we would like to check whether a coin is fair. We tossed the coin 1000 times and got 512 heads and 488 tails. The goodness-of-fit test for the Bernoulli distribution Bern(0.5) may be calculated by calling:

```
counts <- c(512,488)
probs <- (1/2, 1/2)
chisq.test(counts, p=probs)
```

8.1.3. Kolmogorov test

The test is implemented in R, see `?ks.test`. It assumes that a sample comes from a *continuous* distribution.

For example to test the goodness of fit to the uniform distribution $\mathrm{U}[0,1]$ we call:

```
x <- runif(100)
ks.test(x, "punif")
```

The normal distribution $\mathrm{N}(0{,}5,1)$:

```
ks.test(x, "pnorm", 0.5, 1)
```

The exponential distribution $\mathrm{Exp}(2)$:

```
ks.test(x, "pexp", 2)
```

8.2. Test for independence

8.2.1. Pearson chi-squared test

Given a two-way $r \times c$ contingency table (matrix) O consisting of observed values (counts) of discrete random variables X and Y, let $n = \sum_{i=1}^{r} \sum_{j=1}^{c} O_{ij}$. To test

$$H_0 : X \text{ and } Y \text{ are independent},$$

versus

$$K : X \text{ and } Y \text{ are not independent},$$

at an arbitrary significance level $\alpha \in (0,1)$, we can use the Pearson chi-squared test statistic

$$T = \sum_{i=1}^{r} \sum_{j=1}^{c} d_{ij}^2 \quad \overset{H_0}{\sim} \quad \chi^2_{(r-1)(c-1)}, \tag{8.2}$$

where $d_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}}$ and $E_{ij} = \frac{1}{n} \left(\sum_{k=1}^{c} O_{ik}\right) \left(\sum_{l=1}^{r} O_{lj}\right)$.

Critical region: $R(\alpha) = \left[\chi^2_{1-\alpha,(r-1)(c-1)}, \infty\right)$.

See: `?chisq.test` and also `?fisher.test` for small-sized samples.

8.3. R two-sample tests decision-tree

Assume we are given two i.i.d. random samples $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$, and we would like to verify a hypothesis about *average* values of $\mathbf{X}$ and $\mathbf{Y}$. Here is a suggested algorithm for choosing an appropriate statistical test in R.

1. If $n = m$ and observations are paired, i.e. (X_i, Y_i) are pairwise dependent for all $i = 1, 2, \ldots, n$,
   - a.1. Test for normality (if it is not assumed explicitly).
     - a.1.1. Both samples are normally distributed ⇒ Paired t-test (`t.test(x,y,paired=T)`).
     - a.1.2. At least one sample is not normally distributed ⇒ Wilcoxon signed-rank test (`wilcox.test(x,y,paired=T)`).
2. Otherwise,
   - b.1. Test for normality (if it is not assumed explicitly).
     - b.1.1. Both samples are normally distributed ⇒ Test for equality of variances.
       - b.1.1.1. Equal variances ⇒ Two-sample t-test Model I (`t.test(x,y,var.equal=T)`)
       - b.1.1.2. Unequal variances ⇒ Two-sample t-test Model II (`t.test(x,y)`)
     - b.1.2. At least one sample is not normally distributed ⇒ Wilcoxon rank-sum test (`wilcox.test(x,y)`).

8.3.1. Wilcoxon rank-sum test

Also known as: Mann-Whitney-Wilcoxon U test.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. F_X, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$ i.i.d. F_Y be two independent samples, where F_X, F_Y are unknown continuous distributions.

Null hypothesis: $H_0 : F_X = F_Y$. Let

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}, \quad \text{where } U_{ij} = \begin{cases} 1 & \text{for } X_i > Y_j, \\ 0 & \text{for } X_i \le Y_j. \end{cases} \tag{8.3}$$

This test statistic is called Mann-Whithey's U.

Alternatively, let $r(X_i)$ denote the rank of X_i in the joint sequence of observations $(X_1, X_2, \ldots, X_n, Y_1, Y_2, Y_3, \ldots, Y_m)$. The sum of ranks of all the elements from $\mathbf{X}$, i.e.

$$R_X = \sum_{i=1}^{n} r(X_i), \tag{8.4}$$

is called the Wilcoxon statistic. It may be shown that

$$U = R_X - \frac{n(n+1)}{2}. \tag{8.5}$$

Also, the following holds.

$$U = nm + \frac{m(m+1)}{2} - R_Y, \tag{8.6}$$

where $R_Y = \sum_{i=j}^{m} r(Y_j)$.
For an alternative hypothesis: $K_1 : F_X \neq F_Y$ we use the following (approximate) critical region: $R_1(\alpha) = (-\infty, -z_{1-0.5\alpha}] \cup [z_{1-0.5\alpha}, +\infty)$ for a modified test statistic

$$Z = \frac{U - 0.5nm}{\sqrt{nm(n+m+1)}}\sqrt{12} \tag{8.7}$$

R implementation: `wilcox.test(x,y)`.

8.3.2. Wilcoxon signed-ranks test

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. F_X, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ i.i.d. F_Y be two independent random samples, where F_X, F_Y are unknown continuous distributions. Assume that observations are matched in pairs, i.e. (X_i, Y_i) are dependent for $i = 1, 2, \ldots, n$. Let $\mathbf{Z} = \mathbf{X} - \mathbf{Y}$ i.i.d. F_Z. Null hypothesis: $H_0 : Z = -Z$, (that is, the distribution of Z is symmetric around 0). Test statistic:

$$W = 0.5\sum_{i=1}^{n} \overline{r}(Z_i) + \frac{n(n+1)}{4}, \tag{8.8}$$

where $\overline{r}(Z_i)$ is the *signed rank* of Z_i, i.e. the rank of $|Z_i|$ in $(|Z_1|, |Z_2|, \ldots, |Z_n|)$ multiplied by the sign of Z_i.

In practice, for $n > 16$ we use an approximation with a statistic

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{n(n+1)(2n+1)}}\sqrt{24}, \tag{8.9}$$

which asymptotically has the standardized normal distribution. Therefore for an alternative hypothesis: $K_1 : Z \neq -Z$. we use the critical region: $R_1(\alpha) = (-\infty, -z_{1-0.5\alpha}] \cup [z_{1-0.5\alpha}, +\infty)$.

R implementation: `wilcox.test(x,y,paired=T)`.

8.4. Examples

Ex. 8.1. Lengths of 7 telephone calls were measured: 2.5, 1.8, 6.0, 0.5, 8.75, 1.2, 3.75 [min]. Verify whether the call lengths follow an exponential distribution with expectation of 4 mins. Use $\alpha = 0.01$.

Solution.
Here we may use the Kolmogorov goodness-of-fit test. We verify the null hypothesis H_0: sample has the exponential distribution Exp(1/4) against $K : \sim H_0$.

```
phones <- c(2.5, 1.8, 6, 0.5, 8.75, 1.2, 3.75)
ks.test(phones, "pexp", 0.25)
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  phones
## D = 0.1175, p-value = 0.9997
## alternative hypothesis: two-sided
```

Draw conclusions.

⊡

Ex. 8.2. A sample of 10000 eligible voters was randomly chosen for a study on the relationship between sex and attendance at the recent parliamentary election. The results are summarized in the following contingency table.

	Men	Women
Voted	2792	3591
Didn't vote	1486	2131

Check whether there exists any dependence between sex and the participation in the last election. Use significance level of $\alpha = 0.05$.

Solution.
Create the contingency matrix O:

```
counts <- c(2792, 3591, 1486, 2131)
labels <- list(c("Voted", "Didn't vote"), c("Men", "Women"))
O <- matrix(counts, byrow = T, nrow = 2, ncol = 2, dimnames = labels)
O

##             Men Women
## Voted       2792  3591
## Didn't vote 1486  2131
```

We will use the Pearson chi-squared test for independence. Null hypothesis: sex and participation are independent, alternative: they are not independent.

First let us create the matrix E, such that for $i, j = 1, 2$:

$$E_{ij} = \frac{1}{n}\left(\sum_{k=1}^{2} O_{ik}\right)\left(\sum_{l=1}^{2} O_{lj}\right),$$

where $n = \sum_{k=1}^{2}\sum_{l=1}^{2} O_{ij}$. We see that

$$E_{ij} = \frac{1}{n} r_i c_j,$$

where r_i is the sum of observations in the i-th row and c_j is the sum of observations in the j-th column. given $\mathbf{r} = (r_1, r_2)$ and $\mathbf{c} = (c_1, c_2)$, E may be easily calculated using the `outer()` function.

```
(n <- sum(O))
## [1] 10000
(r <- rowSums(O))
##       Voted Didn't vote
##        6383        3617
(c <- colSums(O))
##   Men Women
##  4278  5722
(E <- (1/n) * outer(r, c))
##             Men Women
## Voted       2731  3652
## Didn't vote 1547  2070
```

Calculate the test statistic

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}:$$

```
(T <- sum((O - E)^2/E))
## [1] 6.66
```

Determine the left bound of the rejection region:

```
dim(O)
## [1] 2 2
prod(dim(O) - 1)  # degrees of freedom
## [1] 1
qchisq(0.95, prod(dim(O) - 1))
## [1] 3.841
```

Therefore, at a given significance level of $\alpha = 0.05$ we reject the null hypothesis: sex and participation are dependent.

The test may also be calculated (this approach is recommended) using the `chisq.test()` function:

```
chisq.test(O)
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  O
## X-squared = 6.552, df = 1, p-value = 0.01047
```

Note that to disable the continuity correction we type:

```
chisq.test(O, correct = F)
##
##  Pearson's Chi-squared test
##
## data:  O
## X-squared = 6.66, df = 1, p-value = 0.009858
```

Compare the results and draw conclusions.

⊡

Ex. 8.3. Consider an exemplary multidimensional contingency table `Titanic` from the `datasets` library, which gives information on the fate of passengers on the fatal maiden voyage of the ocean liner *Titanic*, summarized according to economic status (class), sex, age and survival. Check whether there exists any significant ($\alpha = 0.01$) dependence between survival and (a) economic status, (b) sex and (c) age.

Solution.

```
library("datasets")
Titanic
## , , Age = Child, Survived = No
##
##       Sex
## Class  Male Female
##   1st     0      0
##   2nd     0      0
##   3rd    35     17
##   Crew    0      0
##
## , , Age = Adult, Survived = No
```

```
##
##       Sex
## Class  Male Female
##   1st   118      4
##   2nd   154     13
##   3rd   387     89
##   Crew  670      3
##
## , , Age = Child, Survived = Yes
##
##       Sex
## Class  Male Female
##   1st     5      1
##   2nd    11     13
##   3rd    13     14
##   Crew    0      0
##
## , , Age = Adult, Survived = Yes
##
##       Sex
## Class  Male Female
##   1st    57    140
##   2nd    14     80
##   3rd    75     76
##   Crew  192     20
```

For example, let us examine the potential dependence between survival (`Survived`) and economic status (`Class`). We will use `ftable()` to select the values of interest from the given 4-dimensional contingency table.

```
(O <- ftable(Titanic, row.vars = "Survived", col.vars = "Class"))
```

```
##          Class 1st 2nd 3rd Crew
## Survived
## No             122 167 528  673
## Yes            203 118 178  212
```

```
chisq.test(O)
```

```
##
##  Pearson's Chi-squared test
##
## data:  O
## X-squared = 190.4, df = 3, p-value < 2.2e-16
```

Draw conclusions and check the other variables.

⊡

Ex. 8.4. Two different lots of fish pâté cans were considered. The quality controller wants to check whether the weight measuring machine is working properly. The weights of randomly selected cans produced by a faultless machine are as follows (in grams): 146.08, 150.97, 149.30, 148.48, 148.05, 144.33, 148.95, and 147.91. After 2 weeks of intensive usage the machine measured out weights of 152.28, 152.45, 153.32, 150.84, 153.34, 150.90, 153.84, 152.31, 152.26, 149.51, 151.32, 152.18, and 148.90. Test an appropriate hypothesis at significance level of 0.01.

Solution.
No information about the distribution of data is given. Therefore we shall apply goodness-of-fit tests for normality for each vector.

```
x <- c(146.08, 150.97, 149.3, 148.48, 148.05, 144.33, 148.95, 147.91)
y <- c(152.28, 152.45, 153.32, 150.84, 153.34, 150.9, 153.84, 152.31, 152.26,
    149.51, 151.32, 152.18, 148.9)
shapiro.test(x)
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.954, p-value = 0.7514
shapiro.test(y)
##
##  Shapiro-Wilk normality test
##
## data:  y
## W = 0.9323, p-value = 0.3648
```

We may thus assume that we are given two independent samples from normal distributions. In a moment we will then use the two-sample t-test to verify $H_0 : \mathbb{E}\,X = \mathbb{E}\,Y$ against $K : \mathbb{E}\,X \neq \mathbb{E}\,Y$. Let us check for the equality of variances.

```
var.test(x, y)
##
##  F test to compare two variances
##
## data:  x and y
## F = 1.919, num df = 7, denom df = 12, p-value = 0.3065
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5322 8.9563
## sample estimates:
## ratio of variances
##               1.92
```

```
t.test(x, y, var.equal = T)
##
##  Two Sample t-test
##
## data:  x and y
## t = -4.985, df = 19, p-value = 8.222e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.389 -2.202
## sample estimates:
## mean of x mean of y
##     148.0     151.8
```

Compare the results with a non-parametric test for medians (it has smaller power):

```
wilcox.test(x, y)
##
##  Wilcoxon rank sum test
##
## data:  x and y
## W = 6, p-value = 0.0002949
## alternative hypothesis: true location shift is not equal to 0
```

Therefore, the quality controller should ask his repairmen to fix the machine, because it is not functioning properly ($\alpha = 0.01$). ⊡