FHtest: An R Package for the Comparison of Survival Curves with Censored Data

Ramon Oller
Universitat de Vic – UCC

Klaus Langohr
Universitat Politècnica de Catalunya

Abstract

The Fleming-Harrington class for right-censored data was first introduced by Harrington and Fleming (1982). This class is widely used in survival analysis studies and it is a subset of the so-called weighted logrank test statistics. Recently, Oller and Gómez (2012) proposed an extension of this class for interval-censored data. This paper introduces the R package **FHtest**, which implements the Fleming-Harrington class for right-censored and interval-censored survival data. It provides an integrated approach for performing two-sample, k-sample and trend tests based on either counting process theory, likelihood theory, or permutation distributions. In this paper, we summarize the main aspects of the theory framework and present several examples with R codes to illustrate the usage of the main functions of **FHtest**.

Keywords: interval-censored data, permutation tests, counting processes, R.

1. Introduction

Survival times are encountered in many areas of research: for example, the time from cancer diagnosis until death; the time until first failure of a new machine; the sensory shelf life of yogurt, or the time an unemployed worker needs to find a new job. One common feature of all these examples is the possible presence of censored data, that is, incomplete information on the times of interest. Several types of censored data are distinguished:

- Right-censored data: the (unknown) survival time lies above a certain value.

- Left-censored data: the survival time lies below a certain value.

- Interval-censored data: the survival time lies in an interval between two values.

- Double-censored data: survival times are both left- and right-censored.

As with other areas in statistics, common interests in survival analysis comprehend the estimation of distribution functions, hypothesis testing, or the fit of regression models. However, due to the presence of censored data, survival time data require specific methods that take into account that only partial information is available.

In R (R Core Team 2017), there are plenty of packages that provide functions for the analysis of survival times including the **survival** package (Therneau 2017), which is one of the recommended packages of R. The CRAN task view on survival analysis (Allignol and Latouche 2017) gives an exhaustive overview on all these packages available on CRAN (`https://CRAN.R-project.org`) or on Bioconductor (`http://www.bioconductor.org/`). The paragraph in the CRAN task view on hypothesis testing to compare survival functions lists several packages that offer functions for this purpose, such as function `survdiff` of the **survival** package. However, none of these packages included functions that implemented weighted logrank tests for interval-censored data based on the Fleming-Harrington class (Harrington and Fleming 1982). This lack of R functions for this specific type of statistical tests to compare survival functions was the principal motivation for the development of the package **FHtest** (Oller and Langohr 2017). The package implements the Fleming-Harrington class for right-censored and interval-censored survival data and provides an integrated approach for performing two-sample, *k*-sample and trend tests based on either counting process theory, likelihood theory, or permutation distributions.

Several differences exist between functions provided by the **FHtest** package and those of other packages. For example, function `survdiff` of the **survival** package implements the Fleming-Harrington class with right-censored data and the counting process approach. Function `survdiff`, however, is limited to the one-parameter subclass instead of the whole class of two parameters. Recently, package **survMisc** (Dardis 2016) contains function `comp`, which implements the whole class of two parameters. However, neither `survdiff` nor `comp` consider one-sided alternatives, perform the permutation approach, or work with interval-censored data. Function `ictest` of the **interval** package (Fay and Shaw 2010) is programmed for interval-censored data, implements both the permutation and the score vector approach, and allows for one-sided alternatives. However, only two test statistics from the Fleming-Harrington class, the logrank and the Wilcoxon-Peto statistics, can be used.

This paper introduces the **FHtest** package which is available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=FHtest` and it is organized around four sections: In Section 2, the theoretical background of the nonparametric tests of the Fleming-Harrington class for right and interval-censored data is set forth. Following, in Sections 3 and 4, the four main functions of the **FHtest** package are presented and applied to real data examples, respectively. Section 5 concludes the paper with some final remarks on the current state of the package and possible enhancements.

2. Comparison of survival curves with censored data

We denote by T the time from a well-defined start point until the occurrence of an event of interest, $\mathcal{E}$. T is, hence, a positive random variable. One of the functions characterizing its distribution is the survival function, S, which corresponds to the probability of surviving a time t: $S(t) = \mathsf{P}(T > t) = 1 - F(t)$.

Given the longitudinal nature of studies on survival times, a common feature of these are cen-

sored data. These arise whenever the occurrence of $\mathcal{E}$ is not observed exactly, but only partial information is given. If $\mathcal{E}$ does not occur during the follow-up of a study, the corresponding time is right-censored. That is, it is only known that T is larger than the study duration, but not by how much. Contrary to that, left censoring is present if T is below a certain value, but it is not known by how much. If $\mathcal{E}$ occurs during a time interval, T is said to be interval-censored. The general convention is to consider the intervals left-open, that is $T \in (L, R]$. Both left and right-censored data can be considered particular cases of interval-censored data with $L = 0$ and $R = \infty$, respectively. Alternatively, intervals can also be considered to be closed intervals, that is $T \in [L, R]$. Under this convention, exact observation can also be considered a particular case of interval-censored data, namely if $L = R$. In the current paper, we do not need to choose between both conventions because the **FHtest** package can handle both types of interval-censored data. For a more thorough introduction on survival times and censored data, see, e.g., Section 1.4 in Gómez, Calle, Oller, and Langohr (2009).

In R, the nonparametric estimation of $S(t)$ in the presence of right-censored data by means of the Kaplan-Meier estimator (Kaplan and Meier 1958) can be accomplished by using the function `survfit` of the **survival** package (Therneau 2017). For the estimation of the survival function with interval-censored data using the Turnbull estimator (Turnbull 1976), one may use the function `icfit` of the **interval** package (Fay and Shaw 2010) or different functions of the **Icens** package (Gentleman and Vandal 2017).

When dealing with samples from two or more different populations, one is generally interested in comparing the corresponding distributions by testing one of the following hypotheses of interest:

- Two-sample test:

$$\text{H}_0\colon S_1(t) = S_2(t)\ \forall t > 0 \quad \text{vs.} \quad \text{H}_1\colon S_1(t) \neq S_2(t) \text{ for some } t > 0. \tag{1}$$

  Alternatively, H_1 can also consider the case of a one-sided alternative.

- k-sample test:

$$\text{H}_0\colon S_1(t) = \cdots = S_k(t)\ \forall t > 0 \quad \text{vs.} \quad \text{H}_1\colon \exists i, j, t \text{ such that } S_i(t) \neq S_j(t). \tag{2}$$

- Increasing trend test:

$$\text{H}_0\colon S_1(t) = \cdots = S_k(t)\ \forall t > 0 \quad \text{vs.}$$
$$\text{H}_1\colon S_1(t) \leq \cdots \leq S_k(t)\ \forall t > 0 \text{ with, at least, one inequality.} \tag{3}$$

  Alternatively, H_1 can also consider the case of a decreasing alternative.

The R package **FHtest** provides functions for each of the three hypotheses above in the presence of right- or interval-censored data based on the Fleming-Harrington class. At this point, it is important to note that this family is geared to detect alternative hypotheses where the hazards between groups differ but do not cross. Under quite general conditions, the basic test statistic

$$\boldsymbol{U} = (U_1, \ldots, U_k)^\top$$

is asymptotically normal with zero mean and covariance matrix V. The explicit form of $\boldsymbol{U}$ and V is given below and depends on the type of censoring and the asymptotic approach. The test statistic for the hypothesis given in (1) is

$$W_1 = \frac{U_2}{\sqrt{v_{22}}}, \tag{4}$$

where v_{22} is the second value in the diagonal of V and the asymptotic distribution of W_1 is the standard normal distribution. The hypothesis in the k-sample problem (2) can be tested using the test statistic

$$W_2 = \boldsymbol{U}^\top V^- \boldsymbol{U}, \tag{5}$$

where V^- is the generalized inverse of V and the asymptotic distribution of W_2 is a χ^2_{k-1} distribution. The increasing trend alternative (3) can be tested with the statistic

$$W_3 = \frac{a^\top \boldsymbol{U}}{\sqrt{a^\top V a}}, \tag{6}$$

where $a = (a_1, \ldots, a_k)^\top$, $a_1 < \ldots < a_k$, are relevant covariate values for each group distribution. The asymptotic distribution of W_3 is the standard normal distribution. A one-tailed test in the left (or right) direction may show evidence of an increasing (or decreasing) trend.

In the following, we briefly describe the test statistics of the Fleming-Harrington class for right-censored (RC) and interval-censored (IC) survival data and derive the distribution approaches.

2.1. Right-censored data: Counting process approach

The Fleming-Harrington class for right-censored data was first introduced by Harrington and Fleming (1982). This class of test statistics is a subset of the so-called weighted logrank test statistics. Fleming and Harrington (2005) give a theoretical framework for these statistics based on martingale and counting process methods. The Fleming-Harrington class has received much attention in the literature and many extensions have been proposed. For instance, Gillen and Emerson (2007) showed that these statistics are non-transitive under general alternatives and proposed an extension to avoid this drawback. More recently, Garès, Andrieu, Dupuy, and Savy (2015) proposed another extension which is designed to have good power against both late effects and proportional hazards alternatives.

Consider a random sample of possibly right-censored data $[L_i, R_i]$, $i = 1, \ldots, n$, with either $L_i = R_i$ (uncensored data) or $R_i = +\infty$ (right-censored data). Let $t_1 < \ldots < t_m$ be the m distinct uncensored times and suppose that there are $n - m$ right-censored times. Then, the components of the weighted logrank test statistic $\boldsymbol{U}^{\mathrm{RC}}$ are of the form

$$U_j^{\mathrm{RC}} = \sum_{r=1}^{m} w(t_r)\Big(d_{jr} - \frac{n_{jr}}{n_r} d_r\Big), \quad j = 1, \ldots, k, \tag{7}$$

where d_r stands for the number of deaths at time t_r and n_r gives the number of individuals at risk at time t_r, that is, the number of individuals alive just prior to time t_r. Both d_{rj} and n_{rj} correspond to the same definitions but restricted to observations in the jth group.

The weights suggested by Fleming-Harrington are

$$w(t_r) = \hat{S}(t_r-)^\rho (1 - \hat{S}(t_r-))^\lambda, \tag{8}$$

where $\hat{S}(t)$ is the nonparametric estimation of the survival function obtained with the so-called Kaplan-Meier estimator. When $(\rho, \lambda) = (0, 0)$ and $(\rho, \lambda) = (1, 0)$, the statistic $\boldsymbol{U}^{\mathrm{RC}}$ corresponds to the logrank and Prentice-Wilcoxon test statistics, respectively. The choice of the weights has to be made prior to the examination of the data taking into account that they should provide the greatest statistical power, which in turns depends on how the null is believed to be violated. For instance, in a given clinical trial, if one wanted to assess whether the effect of a treatment or therapy on the survival is stronger at the earlier phases of the therapy, $\lambda = 0$ should be chosen, with increasing values of ρ emphasizing stronger early differences. If there were clinical reasons to believe that the effect of the therapy would be more pronounced towards the middle or the end of the follow-up period, it would make sense to choose $\rho = \lambda > 0$ or $\rho = 0$, respectively, with increasing values of λ emphasizing stronger middle or late differences. Statistical power may not be the only consideration in determining the weighting scheme, clinical aspects should also be contemplated. For instance, in cases where an intervention is thought to not alter the quality of the patient's life, upweighting later differences in the survival curves might be important.

The asymptotic null distribution of (7) is derived through a martingale central limit theorem. Under the null hypothesis, $\boldsymbol{U}^{\mathrm{RC}}$ is asymptotically normal with zero mean and variance-covariance matrix V^{RCC} given by

$$\mathsf{VAR}(U_j) = \sum_{r=1}^{m} w(t_r)^2 \, \frac{n_{jr} \, (n_r - n_{jr})}{n_r^2} \, \frac{(n_r - d_r) \, d_r}{n_r - 1}$$

and

$$\mathsf{COV}(U_j, U_{j'}) = -\sum_{r=1}^{m} w(t_r)^2 \, \frac{n_{jr} \, n_{j'r}}{n_r^2} \, \frac{(n_r - d_r) \, d_r}{n_r - 1}, \quad j, j' = 1, \ldots, k.$$

2.2. Right-censored data: Permutation approach

As shown in Abd-Elfattah and Butler (2007), the statistic (7) can also be written in a linear form

$$\boldsymbol{U}^{\mathrm{RC}} = \sum_{i=1}^{n} \mathbf{z}_i c_i, \tag{9}$$

where $\mathbf{z}_i = \left(Z_{1i}, \ldots, Z_{Ki}\right)^{\top}$ is a covariate vector of group indicators (Z_{ji} equals 1 or zero according to whether or not the ith individual is in the jth group) and c_i is a score value associated with each individual:

$$c_i = \begin{cases} w(t_r) - \displaystyle\sum_{s=1}^{r} w(t_s) \, \dfrac{d_s}{n_s} & \text{if } L_i = R_i = t_r, \\[2ex] -\displaystyle\sum_{s=1}^{r} w(t_s) \, \dfrac{d_s}{n_s} & \text{if } R_i = +\infty \text{ and } t_r \leq L_i < t_{(r+1)}. \end{cases} \tag{10}$$

The permutation approach can be applied easily to this linear form. The idea is that if the null hypothesis is true and the underlying censoring process is identical across groups, the labels on the scores c_i are exchangeable. The permutation distribution of $\boldsymbol{U}^{\mathrm{RC}}$ is then obtained by permuting the labels and recomputing the test statistic for all the possible rearranged

labels. The p values from the permutation distribution may be calculated either exactly using complete enumeration or a network algorithm, or by an approximation method like Monte Carlo resampling, or asymptotically using a version of the central limit theorem for exchangeable random variables (Sen 2006). This last method yields a normal approximation with zero permutation expectation and variance-covariance matrix

$$V^{\mathrm{RCP}} = \frac{1}{n-1}\Big(\sum_{i=1}^{n} c_i^2\Big)\Big(\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i^{\top} - \bar{\mathbf{z}}\bar{\mathbf{z}}^{\top})\Big),$$

where $\bar{\mathbf{z}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i$ is the sample mean.

The permutation approach is, in fact, a conditional approach since the distribution of the test statistic is computed conditioned on the observed data. It is not obvious whether the permutation approach gives power properties similar to an unconditional approach. With right-censored data, Heimann and Neuhaus (1998) show that the permutation version of the logrank test and the unconditional version are asymptotically equivalent even under unequal censoring. A more recent contribution concerning this subject is given in Wang, Lagakos, and Gray (2010).

2.3. Interval-censored data: Permutation approach

Oller and Gómez (2012) proposed a new class of test statistics for interval-censored data that extends the Fleming-Harrington class for right-censored data given in (7). The asymptotic behavior of the proposed tests was derived using a permutation approach and an observed Fisher information approach.

Consider a random sample of interval-censored data $(L_i, R_i]$, let $t_1 < \ldots < t_m$ denote the unique ordered elements of $\{L_i, R_i;\ i = 1, \ldots, n\}$, and let $\hat{S}(t)$ be the nonparametric estimation of the survival function obtained with the so-called Turnbull estimator. Then, the components of $\boldsymbol{U}^{\mathrm{IC}}$ are of the form

$$U_j^{\mathrm{IC}} = \sum_{r=1}^{m} w^{\mathrm{IC}}(t_r)\Big(d_{jr}^{\mathrm{IC}} - \frac{n_{jr}^{\mathrm{IC}}}{n_r^{\mathrm{IC}}} d_r^{\mathrm{IC}}\Big), \quad j = 1, \ldots, k, \tag{11}$$

where $d_r^{\mathrm{IC}} = n\left(\hat{S}(t_r-) - \hat{S}(t_r)\right)$ stands for the estimated number of deaths at time t_r and $n_r^{\mathrm{IC}} = n\,\hat{S}(t_r-)$ gives the estimated number of individuals at risk at time t_r. The computation of d_{rj}^{IC} and n_{rj}^{IC} is equivalent to d_r^{IC} and n_r^{IC} replacing $\hat{S}$ by $\hat{S}_j = \frac{1}{N_j}\sum_{i=1}^{N} Z_{ji}\hat{S}_i$ where N_j is the sample size of the jth group and $\hat{S}_i$ is an estimate of the individual survival function as defined by Fay and Shih (1998).

The weights proposed by the authors are

$$w^{\mathrm{IC}}(t_r) = \hat{S}(t_r-)\,\frac{B(1 - \hat{S}(t_r); \lambda + 1, \rho) - B(1 - \hat{S}(t_r-); \lambda + 1, \rho)}{\hat{S}(t_r-) - \hat{S}(t_r)},$$

where $B(t; a, b) = \int_0^t x^{a-1}(1-x)^{b-1}dx$ is an incomplete beta function. There are two reasons for these unusual weights. First, these weights follow from the score vector approach given in Section 2.4 for interval-censored data. Second, the equivalence between the weighted logrank form given in (11) and the linear form given in (12) is based on these weights. Further insights on these weights can be obtained from the following results. When $(\rho, \lambda) =$

$(0,0)$ and $(\rho, \lambda) = (1,0)$, the statistic $\boldsymbol{U}^{\mathrm{IC}}$ reduces to the logrank and Wilcoxon-Peto test statistics originally proposed in Peto and Peto (1972); for further details, see also Fay (1996, 1999). Asymptotically, as $\hat{S}$ is nearly continuous at the point t_r, the weight function $w^{\mathrm{IC}}(t_r)$ resembles $[\hat{S}(t_r-)]^{\rho}[1 - \hat{S}(t_r-)]^{\lambda}$ and, consequently, the interpretation of the parameters ρ and λ reproduces the interpretation in the original right-censored family.

This class $\boldsymbol{U}^{\mathrm{IC}}$ can also be written in a linear form (9) where

$$
c_i^{\mathrm{IC}} = \begin{cases} \frac{\hat{S}(R_i)B(1-\hat{S}(R_i);\lambda+1,\rho) - \hat{S}(L_i)B(1-\hat{S}(L_i);\lambda+1,\rho)}{\hat{S}(L_i)-\hat{S}(R_i)} & \text{if } \hat{S}(R_i) \neq 0, \\ -B(1 - \hat{S}(L_i); \lambda+1, \rho) & \text{if } \hat{S}(R_i) = 0. \end{cases}
\tag{12}
$$

Under the assumption that the underlying censoring process is identical across groups, this linear form of $\boldsymbol{U}^{\mathrm{IC}}$ allows the use of the permutation distribution (with zero expectation and variance-covariance matrix V^{ICP}). The approach is the same as explained in Section 2.2.

Herein, it is considered that the observed intervals are half open intervals, however the definitions above are easily modifiable if we observe closed or open intervals. It is also worth mentioning that the methodology for interval-censored data can also be used for right-censored data as a particular case. In this case, both (10) and (12) would give similar but not identical results.

2.4. Interval-censored data: Score vector approach

When we restrict the Fleming-Harrington class $\boldsymbol{U}^{\mathrm{IC}}$ to the subclass where $\lambda = 0$ (the so called G^{ρ} family), each test statistic (for a given value of ρ) can be seen as an efficient score test in a linear transformation model. We have to assume discrete or grouped continuous interval-censored data: $L, R \in \{t_0, t_1, \ldots, t_m\}$ where $0 = t_0 < t_1 < \ldots < t_m < t_{m+1} = +\infty$. We also assume a linear transformation model $g(T_i) = -\mathbf{z}_i^{\top}\boldsymbol{\beta} + \epsilon_i$ where g is an unknown increasing function and ϵ_i has survival function $S_{\epsilon}(t) = [1 + \rho\exp(t)]^{-\frac{1}{\rho}}$. Then,

$$
\boldsymbol{U}^{\mathrm{IC}} = \left[\frac{\partial \log(\mathrm{Lik}(\boldsymbol{\beta}, \boldsymbol{\theta}))}{\partial \boldsymbol{\beta}}\right]_{\boldsymbol{\beta}=\mathbf{0},\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0},
\tag{13}
$$

where $\mathrm{Lik}(\boldsymbol{\beta}, \theta)$ is the likelihood function and $\boldsymbol{\theta} = (\theta_r)_{r=1}^m$ are nuisance parameters such that $\theta_r = g(t_r)$.

Under discrete interval-censored data and the null hypothesis, the asymptotic behavior of the score vector $\boldsymbol{U}^{\mathrm{IC}}$ follows from maximum likelihood theory. This approach needs that none of the estimated $\boldsymbol{\theta}$ approach the boundary of the parameter space, i.e., $1 > \hat{S}(t_1) > \ldots > \hat{S}(t_m) > 0$. The distribution is asymptotically normal with zero mean and variance-covariance matrix V^{ICS}. The explicit formula for V^{ICS} is not presented here. Instead, see Gómez and Oller (2008) for details.

We note that $\boldsymbol{U}^{\mathrm{RC}}$ and V^{RCC} can also be derived by the score vector approach when $\lambda = 0$. In this case, however, a partial likelihood is used instead of a full likelihood. For this reason, though $\boldsymbol{U}^{\mathrm{RC}}$ and V^{RCC} or alternatively $\boldsymbol{U}^{\mathrm{IC}}$ and V^{ICS} may be used to analyze right-censored data, both approaches do not exactly coincide.

3. The FHtest package

3.1. Installation and dependencies

The **FHtest** package can easily be installed from any CRAN repository by executing the R command

```
R> install.packages("FHtest")
```

To use **FHtest**, several R packages need to be available. As shown on the corresponding web page on CRAN (https://CRAN.R-project.org/package=FHtest), **FHtest** depends on the packages **interval** (Fay and Shaw 2010) and **KMsurv** (Yan, Klein, and Moeschberger 2012), and it imports functions from **survival** (Therneau 2017), **perm** (Fay and Shaw 2010), and **MASS** (Venables and Ripley 2002). The **FHtest** package is partially constructed over the **interval** package, which itself depends on the packages **Icens** (Gentleman and Vandal 2017) and **MLEcens** (Maathuis 2013): functions in both packages use similar structures in the input arguments and output values; some functions in **FHtest** are modifications of functions in **interval** and use the same internal procedures. The **KMsurv** package gives right-censored data sets that are used for illustration purposes. Both the **survival** and the **MASS** packages are recommended packages of R and are, hence, included in all binary distributions of R. Finally, the package **perm** is from the same authors as **interval** and it is installed together with this package.

Three of the packages – **interval**, **MLEcens**, **KMsurv** – will be automatically installed together with **FHtest**, if this has not been already done before. By contrast, package **Icens** is not available at the CRAN repositories, but at the Bioconductor web site (http://www.bioconductor.org/). For this reason, it has to be installed separately, for example, by executing the following instructions in R:

```
R> source("http://bioconductor.org/biocLite.R")
R> biocLite("Icens")
```

Once all packages are installed, package **FHtest** can be loaded and used:

```
R> library("FHtest")
```

```
Loading required package: interval
Loading required package: survival
Loading required package: perm
Loading required package: Icens
Loading required package: MLEcens
Loading required package: KMsurv
```

3.2. The main functions

The four main functions of the **FHtest** package correspond to the censoring schemes and approaches exposed in Sections 2.1 through 2.4:

- `FHtestrcc`: The Fleming-Harrington test for right-censored data based on counting processes.

- `FHtestrcp`: The Fleming-Harrington test for right-censored data based on permutations.

- `FHtesticp`: The Fleming-Harrington test for interval-censored data based on permutations.

- `FHtestics`: The Fleming-Harrington test for interval-censored data based on a score vector distribution.

There are several features that are common to all four functions. First, the (censored) data and the group covariate can be specified in two different ways. One way is to assign two vectors, `L` and `R`, that define the censoring interval, and a `group` vector that gives the group covariate. The additional arguments `Lin` and `Rin` determine whether the endpoints are included or not. Another way is to use the common formula notation by including a 'Surv' object of the **survival** package; see the help page of function `Surv` for more details.

Second, all functions allow for two-sample, k-sample and trend tests with hypotheses and alternatives as discussed in Section 2. The `group` vector determines the type of test: a two-sample test when `group` has two unique values, otherwise a k-sample test when `group` is a character or factor vector, and a trend test when `group` is numeric. By default, the common argument `alternative` is set to `"different"`, which corresponds to a two-sided alternative. For a one-sided alternative as in (3), the user can choose `alternative = "increasing"` for an increasing trend test or `alternative = "decreasing"` for a decreasing trend test.

Third, the four functions have two arguments – `rho` and `lambda` – corresponding to the parameters ρ and λ of the weights of the Fleming-Harrington test statistic shown in (8). The only exception is function `FHtestics`, which does not allow values of `lambda` different from 0 as explained in Section 2.4. The default values are in all cases `rho = 0` and `lambda = 0`.

Finally, the outputs of all functions have the same structure indicating, among others, the type of test, the censoring scheme, the values of `rho` and `lambda`, and the alternative. In addition, the four functions, actually, return lists that can be stored as R objects each with its own class according to the function in use.

The arguments shared by the functions `FHtestrcp` and `FHtesticp`, but not available for `FHtestrcc` and `FHtestics`, have to do with the permutation approach. For example, argument `method` can be used to choose the way the p values are computed. There are four methods, the last two of which are only available for the two-sample test: `"pclt"` to use the permutational central limit theorem (Sen 2006); `"exact.mc"` to use an exact method via Monte Carlo; `"exact.network"` for an exact method using a network algorithm (see, e.g., Agresti, Mehta, and Patel 1990), and `"exact.ce"` for an exact method using complete enumeration, which is appropriate for very small sample sizes. The default value is `"pclt"`.

More details on all these aspects will be shown and explained in Section 4.

3.3. Data sets provided by package FHtest

Two data sets with interval-censored data are provided by package **FHtest**: `duser` and `illust3`.

The data frame `duser` contains the data of 940 injection drug users on five variables that come from the data base of the patients of the detoxification unit of the university hospital *Germans Trias i Pujol* in Badalona (Spain). The following output shows the first six observations:

```
R> data("duser", package = "FHtest")
R> head(duser)

  left right zper zgen age
1    0    44    2    0  24
2    0    83    1    0  15
3    0   128    1    0  16
4    0    59    2    1  21
5    0    83    2    0  20
6    0    37    2    0  17
```

Variables `left` and `right` are, respectively, the left and right endpoints of the interval-censored time (in months) from the start of injection drug use until seroconversion, which can be used as a proxy of HIV infection. Whenever `left` is 0, a left-censored time until seroconversion is given, whereas right-censored times are identified by a right endpoint of 9999. If both endpoints coincide, the time to seroconversion can be considered exact. Their frequencies are as follows:

```
R> sum(duser$left == 0)

[1] 530

R> sum(duser$right == 9999)

[1] 343

R> sum(duser$left == duser$right)

[1] 2
```

Hence, $940 - 530 - 343 - 2 = 65$ observations are strictly interval-censored.

The three remaining variables of the data frame are calendar period (`zper`; a value between 1 and 4), sex (`zgen`; 0: male; 1: female), and age at first injection drug use. For more information on the data set, see, e.g., Gómez, Calle, Egea, and Muga (2000) and Oller and Gómez (2012).

The second data set, `illust3`, is a data frame with 1607 observations on 3 variables. These data come from an AIDS clinical trial that started enrollment in 1987 and was designed to study the benefits of Zidovudine therapy in patients in the early stage of HIV infection (Volberding, Lagakos, Grimes, Stein, Rooney, Meng, Fischl, Collier, Phair, Hirsch, Hardy, Balfour, and Reichman 1995; Oller, Gómez, and Calle 2004). Herein, the interval-censored variable is the time (in months) from treatment start until either AIDS onset or death. The third variable is the treatment group (1, deferred therapy; 2, 500 mg/day dosage; 3, 1500 mg/day dosage). The first six observations of `illust3` are the following:

```
R> data("illust3", package = "FHtest")
R> head(illust3)
```

```
  left right group
1    4   999     1
2   74   999     1
3    8   999     1
4   58   999     1
5   11    11     1
6    4    11     1
```

Herein, the code for right-censored observations is where `right` equals 999 and identical values of `left` and `right` like in observation 5 indicate uncensored times. Contrary to data frame `duser`, there are no left-censored times, but there are right-censored times with a zero left endpoint, which, actually, can be considered missing since they do not contribute any information. The numbers of exact times, right-censored times, and missing data are:

```
R> sum(illust3$left == illust3$right)
```

```
[1] 326
```

```
R> sum(illust3$left > 0 & illust3$right == 999)
```

```
[1] 672
```

```
R> sum(illust3$left == 0 & illust3$right == 999)
```

```
[1] 149
```

Hence, $1607 - 326 - 672 - 149 = 460$ observations are strictly interval-censored.

4. Real data examples

4.1. Examples with right-censored data

The data sets `ovarian` and `bmt` of the R packages **survival** (Therneau 2017) and **KMsurv** (Yan *et al.* 2012), respectively, will be used as examples for the use of the functions `FHtestrcc` and `FHtestrcp` with right-censored data.

The data frame `ovarian` contains the survival times of 26 patients with ovarian cancer and comprehends six variables including survival time (`futime`), the censoring status (`fustat`), and treatment (`rx`):

```
R> library("survival")
R> head(ovarian)
```

```
   futime fustat      age resid.ds rx ecog.ps
1      59      1 72.3315        2  1       1
2     115      1 74.4932        2  1       1
3     156      1 66.4658        2  1       2
4     421      0 53.3644        2  2       1
5     431      1 50.3397        2  1       1
6     448      0 56.4301        1  1       2
```

The data come from a randomized clinical trial that compared two different treatments, in the following called Treatment 1 (`rx = 1`) and Treatment 2 (`rx = 2`), whose survival functions are shown in the left panel of Figure 1. According to that figure, there was a better survival at short term among patients with Treatment 2, but differences diminished after 1.5 years.

The data of the data frame `bmt` come from a study on the survival of 137 patients with leukemia. The complete data frame has 22 variables, among which are the disease-free survival time, that is, the time to either relapse or death (`t2`), the corresponding censoring indicator (`d3`; 1: relapse or death, 0: censored survival time), and the disease group (`group`): acute lymphoblastic leukemia (ALL), low risk acute myeloid leukemia (AML), and high risk AML. See the first six observations of these variables below:

```
R> data("bmt", package = "KMsurv")
R> dim(bmt)

[1] 137  22

R> head(bmt[, c("group", "t2", "d3")])

  group   t2 d3
1     1 2081  0
2     1 1602  0
3     1 1496  0
4     1 1462  0
5     1 1433  0
6     1 1377  0
```

The estimated survival functions of the three groups are presented in the right panel of Figure 1. On average, the disease-free survival times were largest among low risk AML patients and smallest among high risk AML patients.

For further information on both data sets, see the corresponding help pages in R and the articles cited therein.

In the remainder of Section 4.1, the use of functions `FHtestrcc` and `FHtestrcp` for two-sample, k-sample, and trend tests will be illustrated. Notice that for all tests used in this section, either function could be used.

Two-sample tests

For the first example, consider the logrank test with null and alternative hypothesis as shown in (1) in order to compare both treatments of the `ovarian` data set with respect to survival
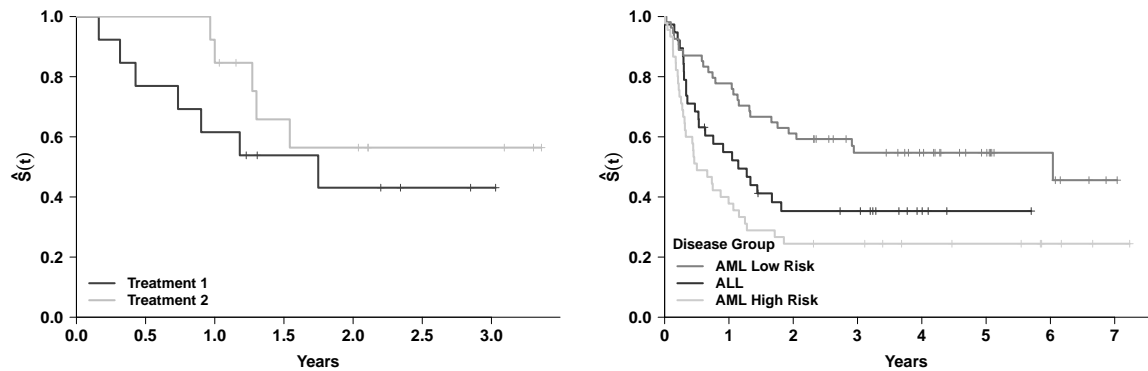
Figure 1: Left panel: Survival curves of patients with ovarian cancer. Right panel: Survival curves of patients with leukemia.

time. This test can be accomplished with function `survdiff` of the **survival** package with the following command syntax:

```
R> survdiff(Surv(futime, fustat) ~ rx, data = ovarian)


Call:
survdiff(formula = Surv(futime, fustat) ~ rx, data = ovarian)

      N Observed Expected (O-E)^2/E (O-E)^2/V
rx=1 13        7     5.23     0.596      1.06
rx=2 13        5     6.77     0.461      1.06

 Chisq= 1.1  on 1 degrees of freedom, p= 0.303
```

The same result ($p = 0.303$) is obtained with function `FHtestrcc` using, as well, the default values of its arguments:

```
R> FHtestrcc(Surv(futime, fustat) ~ rx, data = ovarian)


        Two-sample test for right-censored data

Parameters: rho=0, lambda=0
Distribution: counting process approach

Data: Surv(futime, fustat) by rx

      N Observed Expected    O-E (O-E)^2/E (O-E)^2/V
rx=1 13        7     5.23   1.77     0.596      1.06
rx=2 13        5     6.77  -1.77     0.461      1.06

Statistic Z= -1, p-value= 0.303
Alternative hypothesis: survival functions not equal
```

Compared with the output of function `survdiff`, the information on the observed and (under H$_0$) expected number of events provided by function `FHtestrcc` is the same. What differs between both is that `FHtestrecc`'s output is somewhat more detailed. For example, information is given on the values of the parameters (ρ and λ) of the weights of the Fleming-Harrington test statistic in (8) or on the alternative hypothesis H$_1$. The test statistic does also differ since `FHtestrecc` provides the one presented in (4), which follows (asymptotically) a standard normal distribution, whereas `survdiff` uses its quadratic value, whose asymptotic distribution is the χ^2 distribution. The advantage of using the former is that the alternative hypothesis may be one-sided.

As explained in Section 3.2, with all four functions of the **FHtest** package, the (censored) survival times can also be specified with two vectors, `L` and `R`, instead of using a 'Surv' object. In the case of right and uncensored data, vector `L` must contain the observed times, whereas vector `R` must contain either the uncensored survival times or infinity in the case of right-censored survival times. Hence, the following instruction returns the same result as before:

```
R> with(ovarian, FHtestrcc(futime, futime / fustat, rx))

        Two-sample test for right-censored data


Parameters: rho=0, lambda=0
Distribution: counting process approach


Data: {futime,futime/fustat} by rx

      N Observed Expected    O-E (O-E)^2/E (O-E)^2/V
rx=1 13        7     5.23   1.77     0.596      1.06
rx=2 13        5     6.77  -1.77     0.461      1.06


Statistic Z= -1, p-value= 0.303
Alternative hypothesis: survival functions not equal
```

The logrank test used in the previous example puts the same weights on all addends in (7), but a researcher's interest could actually be to detect either short or long term differences between survival functions. In the second case, for example, more weight needs to be put on latter differences between observed and expected number of events. This is accomplished in the following example with $\lambda = 1$ while ρ remains 0. For this example, function `FHtestrcp` is used, which computes the *p* value using the permutation approach:

```
R> FHtestrcp(Surv(futime, fustat) ~ rx, data = ovarian, lambda = 1)

        Two-sample test for right-censored data


Parameters: rho=0, lambda=1
Distribution:
  permutation approach (asymptotic approximation using central limit theorem)
```

```
Data: Surv(futime, fustat) by rx

        N       O-E
rx=1 13 -0.00447
rx=2 13  0.00447


Statistic Z= 0, p-value= 0.992
Alternative hypothesis: survival functions not equal
```

The command syntax used is the same as with function `FHtestrcc` with the only difference that parameter λ was set to 1. The reasons for the very large p value is two-fold: on the one hand, differences between both survival functions at long term are rather small (see Figure 1), and, on the other hand, both samples are comprised by only 13 patients each.

Compared with function `FHtestrcc`'s output, the most remarkable difference is the table on the difference between observed and expected events in both samples. Here, columns `(O-E)^2/E` and `(O-E)^2/V` are not provided since the p value is computed based on permutations (see Equations 9 and 10) and not on the (asymptotic) counting process distribution of the test statistic in (7).

k-sample tests

For the illustration of the k-sample test in (2), we will use the `bmt` data set and for the sake of a clearer output, we first put labels to each of the three study groups:

```
R> bmt$group <- factor(bmt$group, labels = c("ALL", "AML low risk",
+    "AML high risk"))
R> table(bmt$group)


          ALL  AML low risk AML high risk
           38            54            45
```

The corresponding survival functions are presented in the right panel of Figure 1. The estimation of the median survival is as follows:

```
R> survfit(Surv(t2, d3) ~ group, data = bmt)

Call: survfit(formula = Surv(t2, d3) ~ group, data = bmt)


                     n events median 0.95LCL 0.95UCL
group=ALL           38     24    418     194      NA
group=AML low risk  54     25   2204     704      NA
group=AML high risk 45     34    183     115     456
```

In the following example, the k-sample test (being $k = 3$) is accomplished with function `FHtestrcc` and most weight is put on differences close to the overall median, which is equal to 481 days, by choosing $\rho = \lambda = 1$:

```
R> FHtestrcc(Surv(t2, d3) ~ group, data = bmt, rho = 1, lambda = 1)
```

```
        K-sample test for right-censored data

Parameters: rho=1, lambda=1
Distribution: counting process approach


Data: Surv(t2, d3) by group


                       N Observed Expected    O-E (O-E)^2/E (O-E)^2/V
group=ALL             38     4.55     3.79  0.769     0.156      1.02
group=AML high risk 45     5.41     3.54  1.864     0.981      6.28
group=AML low risk  54     4.87     7.50 -2.633     0.924      8.99


Chisq= 9.9 on 2 degrees of freedom, p-value= 0.00697
Alternative hypothesis: survival functions not equal
```

The use of function `FHtestrcc` for the *k*-sample test is apparently the same as before, but it is important to remark that the grouping variable must be either a factor or a character variable. Otherwise, a trend test is performed.

Concerning the output of the *k*-sample test, it can be seen that the test statistic used, the one in (5), follows a χ^2 distribution with $k - 1 = 2$ degrees of freedom. It also shows a very small p value, that is, the null hypothesis can be rejected and we can assume that there are differences between the three risk groups with respect to disease-free survival after a bone marrow transplant. Moreover, we can observe that, contrary to the output of function `survfit`, the three rows of the output's table are ordered alphabetically and not according to the internal coding of the factor `group`.

Trend tests

In the above situation or when dealing with an ordinal grouping variable, a trend test may be, actually, of greater relevance than the *k*-sample test. We reorder the factor `group` such that the risk group with best disease-free survival, low risk AML, is internally coded 1 and the one with worst disease-free survival, high risk AML, is coded 3:

```
R> bmt$group <- relevel(bmt$group, ref = "AML low risk")
R> levels(bmt$group)

[1] "AML low risk"  "ALL"           "AML high risk"
```

Hence, the null and alternative hypothesis of the trend test of interest are

$$\text{H}_0\colon S_1(t) = S_2(t) = S_3(t) \ \forall t > 0 \quad \text{vs.}$$
$$\text{H}_1\colon S_1(t) \geq S_2(t) \geq S_3(t) \ \forall t > 0 \text{ with, at least, one inequality.} \quad (14)$$

Contrary to the *k*-sample test before, in the case of the trend test, the grouping variable must be a numeric variable. For this reason, in the example below, function `as.numeric` is applied to factor `group` in order to use its internal numeric codes. It is also important to specify the value of argument `alternative`, the direction of the alternative hypothesis, which in (14) is decreasing:

```
R> FHtestrcp(Surv(t2, d3)~as.numeric(group), data = bmt, rho = 1, lambda = 1,
+    alternative = "decreasing", method = "exact.mc")

        Trend FH test for right-censored data

Parameters: rho=1, lambda=1
Distribution:
  permutation approach (exact method using Monte Carlo with 999 replications)

Data: Surv(t2, d3) by as.numeric(group)

                    N    O-E
as.numeric(group)=1 54 -2.633
as.numeric(group)=2 38  0.769
as.numeric(group)=3 45  1.864

Statistic= 3, p-value= 0.001
99 percent confidence interval on p-value:  0.00000 0.00529
Alternative hypothesis:
  decreasing survival functions (higher group implies earlier event times)
```

Among other aspects, the output shows in its first line that the trend test has been applied, and in its last line it indicates the direction of the alternative hypothesis. The function makes use of test statistic (6). The order of the rows of the output's table follows the internal coding of the grouping variable. Therein, a negative sign in column `O-E` indicates less events observed than expected (under the null hypothesis), i.e., better survival than expected. Hence, the increase of the differences `O-E` from the first row to the last is in agreement with the alternative hypothesis of decreasing survival functions.

This example, actually, not only illustrates the accomplishment of the trend test with right-censored data, but also the use of an exact method via Monte Carlo to compute the p value. For this purpose, the value of `FHtestrcp`'s argument `method` was set to `"exact.mc"`. As a result, in addition to the p value, the 99% confidence interval for the p value is also returned.

Both, the number of the Monte Carlo replications (999) and the confidence level for the estimation of the p value, can be modified by changing the default values of argument `permcontrol`. To see which other characteristics of the permutation tests can be controlled through this argument, see the help page of (auxiliary) function `permControl` of the **perm** package, which is called by function `FHtestrcp`.

Further examples for the functions `FHtestrcc` and `FHtestrcp` can be found and studied by calling `example` using the function names as `topic` argument:

```
R> example("FHtestrcc")
R> example("FHtestrcp")
```

4.2. Examples with interval-censored data

For the illustration of functions `FHtesticp` and `FHtestics`, the data sets `duser` and `illust3` of the **FHtest** package, which are presented in Section 3.3, will be used. As shown therein,
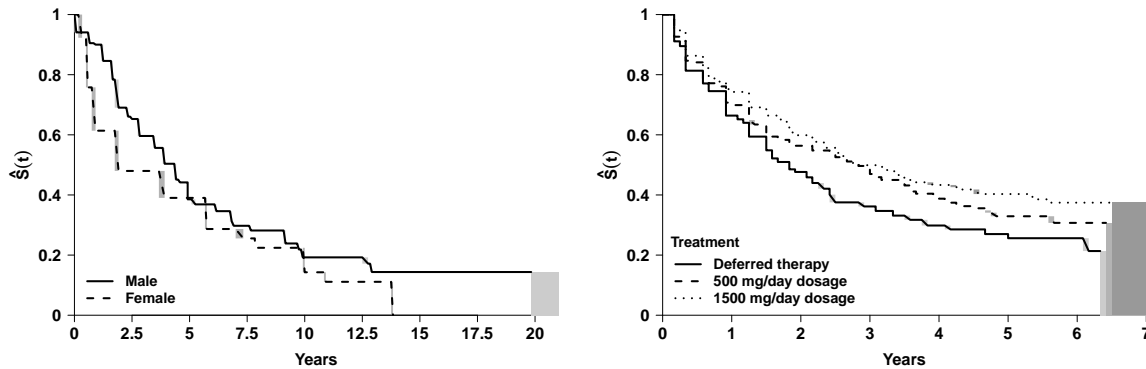
Figure 2: Left panel: Survival functions of time from first injection drug use until HIV seroconversion among injection drug users. Right panel: Survival functions of time from treatment start until either AIDS onset or death among HIV-infected patients. Shaded rectangles represent the intervals within which the Turnbull estimate of $S(t)$ is not identified.

the right-censored observations in both data sets are identified by extremely large values for the right-endpoints of the censoring intervals: `9999` and `999`, respectively. Even though not necessary for the application of `FHtesticp` and `FHtestics`, we first recodify these upper limits by `NA` values, in order to clearly distinguish right- from interval-censored data:

```
R> is.na(duser$right) <- duser$right == 9999
R> is.na(illust3$right) <- illust3$right == 999
```

This recodification is also preferable for the nonparametric estimation of the survival functions by means of the Turnbull estimator, since no upper limit for the survival times is assumed.

Two-sample tests

For the two-sample test, we will use the data set `duser`, which contains the times from first injection drug use until HIV seroconversion among 940 injection drug users (IDUs). The left panel of Figure 2 shows the estimation of the corresponding survival functions of both male and female IDUs. According to this graph, which is drawn with the `plot` method for 'icfit' objects from the **interval** package, the probability for seroconversion during the first three to four years is higher among female IDUs, after which the survival functions are nearly the same with the only difference that beyond 13 years, no more seroconversion is observed among male IDUs.

In order to test the null hypothesis of equal survival functions among male and female IDUs, in the following two examples, we will use function `FHtesticp`, which implements the permutation approach. In the first example, we carry out the logrank test ($\rho = \lambda = 0$), use the permutational central limit theorem to compute the p value and choose the two-sided alternative hypothesis as shown in (1). In addition, we use arguments that are only available with functions `FHtesticp` and `FHtestics`: The arguments `Lin` and `Rin` are both set to `TRUE` to indicate that the censoring intervals are closed intervals, which is meaningful since time is measured in months. Moreover, the argument `icontrol` is used to increase the maximum

number of iterations used internally by the EM algorithm to estimate the survival functions. This is necessary here because convergence problems arise when using the default value of 10000 iterations.

```
R> FHtesticp(Surv(left, right, type = "interval2") ~ zgen, data = duser,
+     Lin = TRUE, Rin = TRUE, icontrol = icfitControl(maxit = 50000))


        Two-sample test for interval-censored data

Parameters: rho=0, lambda=0
Distribution:
  permutation approach (asymptotic approximation using central limit theorem)

Data: Surv(left, right, type = "interval2") by zgen


          N    O-E
zgen=0 759 -16.4
zgen=1 181  16.4


Statistic Z= 1.8, p-value= 0.0694
Alternative hypothesis: survival functions not equal
```

The output is very similar to the one of function `FHtestrcp` with the title being the only difference. Here, we can see that the data set consists of 759 men (`zgen = 0`) and 181 women and that the p value is 0.0694.

What the output does not show is that the computational cost for this example is high. The execution of the previous command, with 940 interval-censored data, takes more than 40 seconds under OS X Yosemite with an Intel processor of 2.5 GHz and the following session info:

```
R> sessionInfo()


R version 3.3.3 (2017-03-06)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X Yosemite 10.10.5
```

If we used function `FHtestics` instead, the computational cost would roughly be the same. That is, the high computation time is not due to the use of the permutational central limit theorem, but due to the use of interval-censored data, in particular due to the necessary nonparametric estimation of the common survival function.

For this reason and in order to reduce computation times, if we are interested in carrying out further hypothesis tests with the same data, it is recommended to save the 'FHtesticp' object and use its element `fit` a posteriori as the value of `FHtesticp`'s argument `icFIT`. This is done in the following example, where we put more weight on early differences and use a one-sided alternative that could be motivated by some prior knowledge:

```
R> icpexam <- FHtesticp(Surv(left, right, type = "interval2") ~ zgen,
+    data = duser, Lin = TRUE, Rin = TRUE,
+    icontrol = icfitControl(maxit = 50000))
R> FHtesticp(Surv(left, right, type = "interval2") ~ zgen, data = duser,
+    rho = 1, Lin = TRUE, Rin = TRUE, alternative = "decreasing",
+    icFIT = icpexam$fit)
```

```
        Two-sample test for interval-censored data

Parameters: rho=1, lambda=0
Distribution:
  permutation approach (asymptotic approximation using central limit theorem)

Data: Surv(left, right, type = "interval2") by zgen


          N    O-E
zgen=0 759 -12.4
zgen=1 181  12.4


Statistic Z= 2.3, p-value= 0.0109
Alternative hypothesis:
  decreasing survival functions (group=1 has earlier event times)
```

In this example, the use of the stored 'FHtesticp' object has greatly reduced the computational time to less than one second.

k-sample tests

The data set illust3, which contains the data from an AIDS clinical trial, will be used for the illustration of the *k*-sample test with interval-censored data. Since the group indicator is a numeric variable, we first put the corresponding labels to each of the three study groups, which are roughly of the same sample size:

```
R> illust3$group <- factor(illust3$group, labels = c("Deferred therapy",
+    "500 mg/day dosage", "1500 mg/day dosage"))
R> table(illust3$group)

  Deferred therapy  500 mg/day dosage 1500 mg/day dosage
               541                538                528
```

The Turnbull estimates of the survival functions for time from treatment start until either AIDS onset or death are shown in the right panel of Figure 2. Therein, we can see that the estimated survival functions are nearly identical throughout the first year under treatment and then start to split: The survival function of the deferred therapy lies below the other two. These remain the same for nearly four years, after which the probabilities for disease-free survival are somewhat larger among the 1500 mg/day dosage group.

To test, whether these difference allow us to assume that there are really differences between the three treatments with respect to disease-free survival, we now use function FHtestics,

which implements the score vector approach set forth in Section 2.4. The null and alternative hypotheses are the ones of the *k*-sample test as shown in (2) and for this reason, the grouping variable used in the following command syntax has to be a factor or a character vector. By setting `rho = 3`, we put more weight on early differences than to differences at long term, and since disease-free survival is given in months, arguments `Lin` and `Rin` are, again, set to `TRUE`.

```
R> FHtestics(Surv(left, right, type = "interval2") ~ group, data = illust3,
+     rho = 3, Lin = TRUE, Rin = TRUE)

        K-sample test for interval-censored data

Parameters: rho=3, lambda=0
Distribution: score vector approach

Data: Surv(left, right, type = "interval2") by group

                          N     O-E
group=1500 mg/day dosage 528 -17.27
group=500 mg/day dosage  538  -3.23
group=Deferred therapy   541  20.50

Chisq= 12.2 on 2 degrees of freedom, p-value= 0.00221
Alternative hypothesis: survival functions not equal
```

Like before, when we applied the *k*-sample test to right-censored data, the value of test statistic (5) is computed, which asymptotically follows a χ^2 with 2 degrees of freedom. This value is fairly large and, consequently, the *p* value is quite small.

Trend tests

We also apply the trend test to this data set using the same null but the opposite alternative hypotheses as in (14):

$$H_0\colon S_1(t) = S_2(t) = S_3(t) \; \forall t > 0 \quad \text{vs.}$$
$$H_1\colon S_1(t) \leq S_2(t) \leq S_3(t) \; \forall t > 0 \text{ with, at least, one inequality.} \quad (15)$$

Here, S_1 is the survival function that corresponds to the deferred therapy, S_3 the one of the 1500 mg/day dosage treatment, and since H_1 states increasing disease-free survival functions, the argument `alternative` will now be set to `"increasing"`. As with the trend test with right-censored data, the grouping variable must now be a numeric variable:

```
R> FHtestics(Surv(left, right, type = "interval2") ~ as.numeric(group),
+     data = illust3, Lin = TRUE, Rin = TRUE, alternative = "increasing")

        Trend FH test for interval-censored data
```

```
Parameters: rho=0, lambda=0
Distribution: score vector approach

Data: Surv(left, right, type = "interval2") by as.numeric(group)

                       N    O-E
as.numeric(group)=1 541   51.4
as.numeric(group)=2 538   -8.0
as.numeric(group)=3 528  -43.4

Statistic Z= -4.2, p-value= 1.19e-05
Alternative hypothesis:
  increasing survival functions (higher group implies later event times)
```

In both examples with function `FHtestics`, we either used equal weights for all differences between observed and expected events or put more weight on short-term differences. If, contrary to that, we wanted to put more weight on late differences, we would need to use function `FHtesticp`, because function `FHtestics` does not allow for values of `lambda` different from 0.

Further examples for the functions `FHtesticp` and `FHtestics` can be found and studied by executing the following instructions:

```
R> example("FHtesticp")
R> example("FHtestics")
```

5. Final remarks

In this paper we have presented the R package **FHtest**, whose functions perform tests for right- and interval-censored survival data based on the Fleming-Harrington class. We have shown that the **FHtest** package is a flexible and user-friendly tool that integrates and extends methods in other R packages and we have given several detailed examples of the use of each function.

The **FHtest** package has some limitations. For example, it covers left- and double-censored data as particular cases of interval-censored data, however, it does not cover neither truncated data nor doubly-censored (i.e., both the time origin and the event of interest are censored) data. Another intrinsic drawback is that weighted logrank tests are not appropriate for crossing hazards alternatives.

In the future, we plan to perform a simulation study to compare the four **FHtest** methods. With right-censored data and for both small and large samples, we are interested in studying the differences between distinct approaches: (i) the RC counting process versus the IC score vector method; (ii) the RC permutation versus the IC permutation method; and (iii) the RC counting process versus the RC permutation method. With interval-censored data, we are also interested in studying the differences between both IC approaches: the IC permutation versus the IC score vector method.

Acknowledgments

References

Abd-Elfattah EF, Butler RW (2007). "The Weighted Log-Rank Class of Permutation Tests: p-Values and Confidence Intervals Using Saddlepoint Methods." *Biometrika*, **94**(3), 543–551. doi:10.1093/biomet/asm060.

Agresti A, Mehta CR, Patel NR (1990). "Exact Inference for Contingency Tables with Ordered Categories." *Journal of the American Statistical Association*, **85**(410), 453–458. doi:10.1080/01621459.1990.10476220.

Allignol A, Latouche A (2017). *CRAN Task View: Survival Analysis.* Version 2017-04-25, URL https://CRAN.R-project.org/view=Survival.

Dardis C (2016). **survMisc**: *Miscellaneous Functions for Survival Data.* R package version 0.5.4, URL https://CRAN.R-project.org/package=survMisc.

Fay MP (1996). "Rank Invariant Tests for Interval Censored Data Under the Grouped Continuous Model." *Biometrics*, **52**(3), 811–822. doi:10.2307/2533044.

Fay MP (1999). "Comparing Several Score Tests for Interval Censored Data." *Statistics in Medicine*, **18**(3), 273–285. doi:10.1002/(sici)1097-0258(19990215)18:3<273:: aid-sim19>3.0.co;2-7.

Fay MP, Shaw PA (2010). "Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The **interval** R Package." *Journal of Statistical Software*, **36**(2), 1–34. doi:10.18637/jss.v036.i02.

Fay MP, Shih JH (1998). "Permutation Tests Using Estimated Distribution Functions." *Journal of the American Statistical Association*, **93**(441), 387–396. doi:10.1080/01621459.1998.10474120.

Fleming TR, Harrington DP (2005). *Counting Processes and Survival Analysis.* John Wiley & Sons. doi:10.1002/9781118150672.

Garès V, Andrieu S, Dupuy JF, Savy N (2015). "An Omnibus Test for Several Hazard Alternatives in Prevention Randomized Controlled Clinical Trials." *Statistics in Medicine*, **34**(4), 541–557. doi:10.1002/sim.6366.

Gentleman R, Vandal A (2017). **Icens**: *NPMLE for Censored and Truncated Data.* doi:10.18129/B9.bioc.Icens. R package version 1.48.0.

Gillen DL, Emerson SS (2007). "Nontransitivity in a Class of Weighted Logrank Statistics Under Nonproportional Hazards." *Statistics & Probability Letters*, **77**(2), 123–130. doi:10.1016/j.spl.2006.06.001.

Gómez G, Calle ML, Egea JM, Muga R (2000). "Risk of HIV Infection as a Function of the Duration of Intravenous Drug Use: A Non-Parametric Bayesian Approach." *Statistics in Medicine*, **19**(19), 2641–2656. doi:10.1002/1097-0258(20001015)19:19<2641::aid-sim527>3.0.co;2-p.

Gómez G, Calle ML, Oller R, Langohr K (2009). "Tutorial on Methods for Interval-Censored Data and Their Implementation in R." *Statistical Modelling*, **9**(4), 259–297. doi:10.1177/1471082x0900900402.

Gómez G, Oller R (2008). "A New Class of Rank Tests for Interval-Censored Data." *Harvard University Biostatistics Working Paper Series*. Working Paper 93, URL http://biostats.bepress.com/harvardbiostat/paper93/.

Harrington DP, Fleming TR (1982). "A Class of Rank Test Procedures for Censored Survival Data." *Biometrika*, **69**(3), 553–566. doi:10.1093/biomet/69.3.553.

Heimann G, Neuhaus G (1998). "Permutational Distribution of the Log-Rank Statistic Under Random Censorship with Applications to Carcinogenicity Assays." *Biometrics*, **54**(1), 168–184. doi:10.2307/2534005.

Kaplan EL, Meier P (1958). "Nonparametric Estimation From Incomplete Observations." *Journal of the American Statistical Association*, **53**(282), 457–481. doi:10.2307/2281868.

Maathuis M (2013). *MLEcens: Computation of the MLE for Bivariate (Interval) Censored Data.* R package version 0.1-4, URL https://CRAN.R-project.org/package=MLEcens.

Oller R, Gómez G (2012). "A Generalized Fleming and Harrington's Class of Tests for Interval-Censored Data." *The Canadian Journal of Statistics*, **40**(3), 501–516. doi:10.1002/cjs.11139.

Oller R, Gómez G, Calle ML (2004). "Interval Censoring: Model Characterizations for the Validity of the Simplified Likelihood." *The Canadian Journal of Statistics*, **32**(3), 315–326. doi:10.2307/3315932.

Oller R, Langohr K (2017). *FHtest: Tests for Right and Interval-Censored Survival Data Based on the Fleming-Harrington Class.* R package version 1.4, URL https://CRAN.R-project.org/package=FHtest.

Peto R, Peto J (1972). "Asymptotically Efficient Rank Invariant Test Procedures." *Journal of the Royal Statistical Society A*, **135**(2), 185–207. doi:10.2307/2344317.

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sen PK (2006). "Permutational Central Limit Theorems." In *Encyclopedia of Statistical Sciences*. John Wiley & Sons.

Therneau TM (2017). **survival**: *Survival Analysis.* R package version 2.41-3, URL https://CRAN.R-project.org/package=survival.

Turnbull BW (1976). "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data." *Journal of the Royal Statistical Society B*, **38**(3), 290–295.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S.* 4th edition. Springer-Verlag, New York.

Volberding PA, Lagakos SW, Grimes JM, Stein DS, Rooney J, Meng T, Fischl MA, Collier AC, Phair JP, Hirsch MS, Hardy WD, Balfour HH, Reichman RC (1995). "A Comparison of Immediate with Deferred Zidovudine Therapy for Asymptomatic HIV-Infected Adults with CD4 Cell Counts of 500 or More per Cubic Millimeter." *New England Journal of Medicine*, **333**(7), 401–407. doi:10.1056/nejm199508173330701.

Wang R, Lagakos SW, Gray RJ (2010). "Testing and Interval Estimation for Two-Sample Survival Comparisons with Small Sample Sizes and Unequal Censoring." *Biostatistics*, **11**(4), 676–692. doi:10.1093/biostatistics/kxq021.

Yan J, Klein JP, Moeschberger ML (2012). **KMsurv**: *Data Sets From Klein and Moeschberger (1997), Survival Analysis.* R package version 0.1-5, URL https://CRAN.R-project.org/package=KMsurv.

Affiliation:

Ramon Oller
Departament d'Economia i Empresa
Facultat d'Empresa i Comunicació
Universitat de Vic – UCC
Vic, Spain
E-mail: ramon.oller@uvic.cat
URL: https://www.uvic.cat/en/resultdir/ramon.oller