Journal de la Société Française de Statistique
Vol. 155 No. 3 (2014)

Testing an "Exponential Delay Time model" against a "Random Sign Censoring model" in Reliability

Titre: Test du modèle "Delay Time" exponentiel contre le modèle "Random Sign Censoring" en Fiabilité

Jean-Yves Dauxois[1], Sarah Jomhoori[2] and Fatemeh Yousefzadeh[2]

Abstract: In this paper we consider an industrial system subject to different causes of failure and different types of maintenance: a corrective maintenance is performed after a critical failure and a preventive maintenance can be performed in order to decrease the risk of critical failure. The recurrence of these types of maintenance has been often modeled in a competing risks framework.

However rather few statistical inference has been carried out in these models. In particular, there is a need to introduce statistical tests in order to help the engineers to select the model which better fits their data. Thus, in this paper, we introduce a nonparametric test with aim to decide between a Delay Time model with exponential distribution and a Random Sign model. We prove the asymptotic normality of our test statistic and we carry out a Monte Carlo simulation to learn how works our test on finite sample sizes. An application on a real dataset is also given.

Résumé : Nous nous intéressons dans cet article à un système industriel sujet à différentes causes de pannes et sur lequel deux types de maintenance peuvent être effectuées : soit une maintenance corrective dans le cas d'une panne critique, soit une maintenance préventive afin de réduire le risque d'une panne critique. Des modélisations basées sur la notion de risques concurrents ont été proposées dans la littérature. Mais peu d'inférence statistique a été menée sur ces modèles. En particulier, il existe très peu de tests statistiques permettant de décider quel modèle pourrait le mieux s'ajuster à un jeu de données précis.

L'objectif de cet article est justement d'introduire un test non-paramétrique permettant de décider entre deux de ces modèles de fiabilité : le modèle "Delay Time" avec loi exponentielle et le modèle "Random Sign Censoring". Nous introduisons une statistique de test et prouvons sa normalité asymptotique. Nous terminons l'article en étudiant par simulation le comportement de notre procédure dans le cas de petits échantillons et en présentant une application du test sur un jeu de données réelles.

Keywords: Asymptotic distribution, Censoring, Competing risks, Corrective Maintenance, Delay Time model, Non-parametric test, Preventive Maintenance, Random Sign Censoring Model
Mots-clés : Censure, Loi asymptotique, Maintenance Corrective, Maintenance Préventive, Modèle "Delay Time", Modèle "Random Sign Censoring", Risques concurrents, Test non-paramétrique
AMS 2000 subject classifications: 62G10, 62N03, 62N05, 90B25

1. Introduction

This paper considers the modelling of the time to failure of an industrial system. We suppose that many types of failure can occur, some are seen as critical, other not. As it is often the case in practice, these causes of failure are not assumed to be independent. We also suppose that different types of maintenance can be carried out. A Corrective Maintenance (CM) is generally performed

[1] Université de Toulouse-INSA, IMT, UMR CNRS 5219, 135, Avenue de Rangueil, 31077 Toulouse Cedex 4, France. E-mail: jean-yves.dauxois@insa-toulouse.fr
[2] Department of Statistics, Faculty of Sciences, University of Birjand, Iran. P.O. BOX: 97175/615 E-mail: sjomhoori@birjand.ac.ir

after a critical failure. A condition-based Preventive Maintenance (PM) is also performed at random times, for example after a casual observation of a degradation. Although the subject of the efficiency of the repair is of interest and largely studied in the literature, we assume in our case that all the repairs are perfect, i.e. the system is considered as good as new after maintenance.

The matter of the modelling of different types of failure and/or maintenance has been often addressed in the literature through the use of competing risks. The notion of Competing risks has been introduced by Bernoulli in the 18th century to separate the risk to die from smallpox from the other causes. This is now a well established theory and the literature concerned with competing risks is huge, in particular in the field of Survival Analysis. One can not list the papers published in this area and we only refer to Crowder (2001) for a very readable introduction to competing risks and to Andersen et al. (1993) for a more mathematical presentation of this subject and of the theoretical results available in this field.

In Reliability, the competing risks approach has been considered by many authors, like Cooke (1996) and Li and Pham (2005) to model the different types of failure, or Cooke and Paulsen (1997) who have been the first to introduce this concept to model the dependence between preventive and corrective maintenance. One can mention also papers from Cooke (1993), Hokstad and Jensen (1998), Bunea and Bedford (2002), Christer (2002), Bunea et al. (2003), Langseth and Lindqvist (2003), Langseth and Lindqvist (2006), Lindqvist et al. (2006), Doyen and Gaudoin (2006), Dijoux et al. (2008), Dijoux and Gaudoin (2009), Deloux et al. (2012), Dijoux and Gaudoin (2014), among others...

Many of the above papers have introduced interesting models of maintenance which are often used by engineers. One can mention the "Independent Competing Risks with Mixture of Exponentials" model of Bunea et al. (2003), the "Delay-Time" model of Hokstad and Jensen (1998), the "Random Sign censoring" model of Cooke (1993, 1996) and finally the "(Intensity Proportional) Repair Alert" model of Langseth and Lindqvist (2003) and Lindqvist et al. (2006). Note that all the above papers consider models with only two competing risks. Either the different causes of failure are split into two groups or only the two types of maintenance (PM or CM) are considered.

But, the available results on these models are almost exclusively probabilistic. In particular, only graphical criteria are used to decide the model to use on a given dataset. They are based on empirical estimation of some functions known to have a specific behavior under these models. Some interesting exceptions are Dewan et al. (2004) who combined the concept of concordance and discordance with U-statistic approach to derive some tests for model selection and Langseth and Lindqvist (2006) who proposed to use parametric bootstrap to test IPRA model under perfect repair. Nevertheless, there is a need to introduce other statistical procedure.

This is the aim of our paper to introduce such a goodness-of-fit test. Our test attempts to detect characteristic properties of the function $\Phi(\cdot)$ which gives the probability of Preventive Maintenance beyond time t (this function is introduced precisely in next section).

The rest of this paper is outlined as follows. In Section 2, we introduce the competing risks framework and we present a short review of reliability models with different types of maintenance based on competing risks. We recall precisely the properties fulfilled under these models by the function $\Phi(\cdot)$ and the conditional survival functions given the type of maintenance. We also recall the available results on nonparametric inference with competing risks, those which will be used in the following sections. Section 3 is concerned with the construction of a family of test with aim to decide *e.g.* between the “Delay Time” model and the “Random Sign” model. The statistic of test is introduced and its large sample behavior established. In Section 4 we present a numerical application of our testing procedure. First, we carry out a simulation study of the empirical level and power of our test on small sample sizes. The paper ends with a Conclusion where some prospects of further developments are given.

2. Framework and notations

2.1. Competing risks formulation

First, let us introduce precisely the competing risks framework used to modelise the lifetime of a repairable system with different types of failure and different types of maintenance. Let us denote by X the failure time associated to the failure mode(s) of interest. In this case a Corrective Maintenance is performed on the system. Let us denote by Y the termination time of observation due to other causes, like preventive maintenance or a non-critical failure. For abbreviation we will say that Y is the time where a Preventive Maintenance is performed on the system. Except when explicitly mentioned, we don’t assume in the sequel that X and Y are independent.

Thus, one can only observe the r.v. (T,δ) where :

$$\left\{\begin{array}{lcl} T & = & X\wedge Y \\ \delta & = & 1+I\{X<Y\} \end{array}\right. ,$$

that is $\delta=1$ when a PM is performed and $\delta=2$ when it is a CM.

From Tsiatis (1975), one knows that without any other assumption or information, the distribution of the random variables X and Y (e.g. their survival functions $S_X(\cdot)$ and $S_Y(\cdot)$) are not identifiable and thus not estimable. In general, only the cumulative incidence functions or sub-survival functions defined as below are identifiable. The cumulative incidence functions (CIF) are given by

$$F_j(t)=P(T\leq t,\delta=j),\ \text{for}\ j=1,2\ \text{and}\ t>0 \tag{1}$$

whereas the sub-survival functions are

$$\begin{array}{lcl} S_2(t) & = & P(T>t,\delta=2)=P(X>t,Y>X), \\ S_1(t) & = & P(T>t,\delta=1)=P(Y>t,X>Y). \end{array}$$

Of course, the cumulative distribution function of T is equal to

$$F(t)=P(T\leq t)=F_1(t)+F_2(t) \tag{2}$$

and let us write $S(t)=1-F(t)$ the survival function.

The conditional survival functions

$$\begin{array}{lcl} CS_2(t) & = & P(T>t|\delta=2)=P(X>t|Y>X), \\ CS_1(t) & = & P(T>t|\delta=1)=P(Y>t|X>Y) \end{array}$$

are also estimable.

Finally, we will also consider the function which, for all t, gives the probability of PM beyond time t:

$$\Phi(t) = P(\delta = 1 | T > t) = P(Y < X | T > t).$$

In the sequel we will use the notation $\gamma = \Phi(0) = P(\delta = 1)$.

As we will see in the following subsection, the functions $\Phi(\cdot)$, $CS_1(\cdot)$ and $CS_2(\cdot)$ have interesting properties under classical Reliability models of different types of maintenance and could be used to build goodness-of-fit tests of these models.

2.2. Review of some Reliability models with two types of maintenance

In this subsection we will recall briefly the definition of some competing-risks-based models for the occurrence of the PM and CM times. Presentations of these models have already been done in Bunea et al. (2003) or Dijoux and Gaudoin (2009). In particular, more details on these models can be found in Bunea et al. (2003). The aim of this paragraph is only to list the properties of the above functions under such models.

2.2.1. Independent Competing Risks

In this model the r.v. X and Y are supposed to be independent. This is a strong and untestable hypothesis. There is no general result on the behavior of the functions $\Phi(\cdot)$, $CS_2(\cdot)$ and $CS_1(\cdot)$. All depends on the distribution given to X and Y. One can note that if X and Y have an exponential distribution, the function $\Phi(\cdot)$ is constant, as it has been noted by Bunea et al. (2003).

2.2.2. Independent Competing Risks with Mixture of Exponentials

Bunea et al. (2003) introduced a model which assumes that the r.v. X and Y are independent with respective survival functions:

$$\begin{aligned} S_X(t) &= p\exp(-\lambda_1 t) + (1-p)\exp(-\lambda_2 t) \\ S_Y(t) &= \exp(-\lambda_y t). \end{aligned}$$

The reals $\lambda_1, \lambda_2, \lambda_y$ and p are the parameters of the model.

Bunea et al. (2003) proved that under this model the function $\Phi(\cdot)$ is strictly increasing when $\lambda_1 \neq \lambda_2$ and that we have $CS_2(t) \leq CS_1(t)$, for all $t > 0$.

2.2.3. Delay-Time Model

The Delay-Time model introduced by Hokstad and Jensen (1998) assumed that the r.v. X and Y are such that:

$$\begin{aligned} X &= U + V \\ Y &= U + W, \end{aligned}$$

where U, V and W are independent r.v. Thus, the r.v. X and Y are dependent, but independent given U. The r.v. U can represent the time where the degradation of the system crosses a threshold. In this case, the r.v. V (resp. W) represents the remaining time before CM (resp. PM).

If U, V and W are supposed to be exponentially distributed, Hokstad and Jensen (1998) established that the function $\Phi(\cdot)$ is constant and that $CS_1(t) = CS_2(t)$, for all $t > 0$.

2.2.4. *Random Signs Censoring Model*

This model has been introduced by Cooke (1993). Here the r.v. δ and X are supposed to be independent. This means that the type of maintenance is independent of the failure time. It is easy to see that this is equivalent to say that the sign of $Y - X$ is independent from X which explains the name of this model.

Cooke (1993) showed that, in this case, the function $\Phi(\cdot)$ has its maximum at the origin:

$$\sup_t \Phi(t) = \Phi(0) = P(\delta = 1).$$

Moreover, Cooke (1996) proved that there exists a joint distribution on (X,Y) which satisfies the random signs censoring assumption if, and only if,

$$CS_2(t) > CS_1(t), \text{ for all } t > 0.$$

2.2.5. *(Intensity Proportional) Repair Alert Model*

Langseth and Lindqvist (2003) and Lindqvist et al. (2006) have developed the Repair Alert Model which is a special case of the Random Sign Model where it is also assumed that:

$$P(Y \leq y | X = x, Y < X) = \frac{G(y)}{G(x)},$$

with $G(\cdot)$ an increasing function such that $G(0) = 0$. The Intensity Proportional Repair Alert Model (IPRA) is obtained with the choice of $G(\cdot) = \Lambda_X(\cdot)$, the cumulative hazard rate function of the time to failure X. Thus, the IPRA model assumes that the conditional density of Y is proportional to the intensity of the underlying failure process. Note finally that Dijoux and Gaudoin (2009) mentioned that in this case the function $\Phi(\cdot)$ is decreasing.

2.3. *Testing the "Exponential Delay time model" against a "Random Sign Censoring model" using the function $\Phi(\cdot)$*

As we have seen in the previous subsection, the functions $\Phi(\cdot)$, $CS_1(\cdot)$ and $CS_2(\cdot)$ have different behaviors depending on the reliability models considered. These properties are often used empirically by the engineers in order to decide the model to use on a specific dataset. From the plot of the empirical estimates of these functions they choose the model which better fits the data. There is a need to have some statistical tests to help the engineers in their choice. And it appears that the literature is rather poor in developing goodness-of-fit tests for such models. As said in the introduction, some interesting exceptions are Dewan et al. (2004) and Langseth and Lindqvist (2006).

Our aim in this paper is to fill part of this gap. More precisely we introduce a family of tests of the hypothesis "the function $\Phi(\cdot)$ is constant" against "the function $\Phi(\cdot)$ is not constant and attains its maximum at the origin". With the notation $\gamma = \Phi(0) = P(\delta = 1)$ introduced in Section 2.1, the above two hypotheses can be rewritten as:

$$H_0 : \Phi(t) = \gamma, \text{ for all } t > 0,$$

and

$$H_1 : \Phi(t) < \gamma, \text{ for all } t > 0.$$

For example, deciding H_0 could give us indications that we are in presence of exponential distributions either under an "independent competing risks" model or a "Delay Time" model. Deciding H_1 would suggest that maybe a "Random Sign" model could better fit the data.

2.4. Nonparametric inference of the functions of interest

Let us close this section by a small reminder of the available results on the nonparametric estimation of the functions of interest in a competing risks model. These estimators will be used in the forthcoming test statistic. But, first of all and in order to consider the most general situation, let us also assume the presence of a random censoring mechanism acting on the observation of the lifetime T. Such a random censoring could be for example due to "end of study" or "loss to follow up" even if this latter case is less common in Reliability studies. But our work, exposed in the following sections, may also have applications in survival analysis where such a phenomenon is very usual.

So, to allow for possibly right censored data, let us now suppose that the observations are not a sample of (T, δ) but a sample (T_i^*, δ_i^*), for $i = 1, 2, ..., n$, of:

$$\begin{cases} T^* &= T \wedge C \\ \delta^* &= \delta I(T \leq C) \end{cases},$$

where the censoring random variable C, with continuous distribution function $H(\cdot)$, is supposed to be independent of T. Thus, the indicator δ is equal to 1 when a PM is performed, to 2 when it is a CM and to 0 when we have observed a censoring time.

Let us define, for all $t > 0$, the counting processes

$$N_j(t) = \sum_{i=1}^{n} I(T_i^* \leq t, \delta_i^* = j), \quad j = 1, 2.$$

The r.v. $N_1(t)$ (resp. $N_2(t)$) gives the number of PM (resp. of CM) observed in the interval $[0, t]$.

Finally, let us introduce the number at risk process

$$Y(t) = \sum_{i=1}^{n} I(T_i^* \geq t),$$

which gives, at time t, the number of times T^* not yet observed.

With the notations $N(\cdot) = \sum_{j=1}^{2} N_j(\cdot)$ and $\Delta N(t) = N(t) - N(t^-)$, the Kaplan-Meier estimator of the survival function $S(\cdot)$ of T is well-known (see Andersen et al., 1993, p. 256) to be given by:

$$\widehat{S}(t) = \prod_{i:t^*_{(i)} \leq t} \left(1 - \frac{\Delta N(t^*_{(i)})}{Y(t^*_{(i)})}\right), \tag{3}$$

where $T^*_{(1)} \leq T^*_{(2)} \leq \cdots \leq T^*_{(n)}$ are the ordered statistics associated to the observed sample. The Aalen-Johansen estimators of CIFs are given by:

$$\widehat{F}_j(t) = \int_0^t \widehat{S}(u^-) \frac{dN_j(u)}{Y(u)}, \text{ for } j = 1,2 \text{ and } t > 0.$$

One can find in Andersen et al. (1993, p. 288) and following, an introduction and a study of this estimator but in a very general setup. Dauxois and Kirmani (2004) showed that the above expression of this estimator applies in this specific case.

The sub-survival functions are therefore estimated by

$$\widehat{S}_j(t) = \widehat{F}_j(\tau) - \widehat{F}_j(t), \qquad j = 1,2,$$

where τ is the right endpoint of the support of F.

3. Testing H_0 against H_1

3.1. Development of the test statistic

Recall that we consider the problem of testing

$$H_0 : \Phi(t) = \gamma, \text{ for all } t > 0,$$

against

$$H_1 : \Phi(t) < \gamma, \text{ for all } t > 0.$$

From $\Phi(\cdot) = S_1(\cdot)/S(\cdot)$ it appears clearly that under H_0 the functions $S_1(\cdot)$ and $S(\cdot)$ are proportional and that under H_1 the following equivalent properties are fulfilled:

$$\begin{aligned} & \gamma S(t) - S_1(t) > 0, \text{ for all } t > 0 \\ \Longleftrightarrow \quad & (1-\gamma)F_1(t) - \gamma F_2(t) > 0 \text{ for all } t > 0 \\ \Longleftrightarrow \quad & CS_2(t) > CS_1(t) \text{ for all } t > 0. \end{aligned}$$

Using the second equivalence and a positive weight function $w(\cdot)$, one can see that

$$\psi = \int_0^\tau w(t) \left[(1-\gamma)F_1(t) - \gamma F_2(t)\right] dt,$$

is a measure of non-proportionality between $S_1(t)$ and $S(t)$. It is null under H_0 and positive under H_1.

Thus a natural test statistic for detecting the alternative H_1 is given by

$$\widehat{\psi} = \int_0^\tau \widehat{w}(t)\left[(1-\widehat{\gamma})\widehat{F}_1(t) - \widehat{\gamma}\widehat{F}_2(t)\right]dt, \tag{4}$$

where $\widehat{w}(\cdot)$ is a consistent estimator of $w(\cdot)$, the estimators $\widehat{F}_1(\cdot)$ and $\widehat{F}_2(\cdot)$ have been introduced in Section 2.4 and

$$\widehat{\gamma} = \widehat{F}_1(\tau) = \widehat{P(\delta = 1)}.$$

Note that this test looks like one introduced by Dewan et al. (2004), but it has to be said that our test assumes the presence of censoring and uses a weight function, which was not the case in their paper. Our method of proof of the forthcoming theoretical results is also different and it is not sure that their approach based on U-statistics would work in this case.

It is clear from the definitions of ψ and $\hat{\psi}$ that the choice of $w(\cdot)$ essentially corresponds to the choice of $\hat{w}(\cdot)$ in $\hat{\psi}$. The key considerations in choosing $\hat{w}(\cdot)$ are consistency, asymptotic relative efficiency and computational convenience. At the same time, certain considerations specific to the actual Reliability or Survival analysis problem at hand may have to be taken into account. Thus, one may wish to put more weight on early departures from H_0, late departures from H_0 or on departures in the mid-range. Another argument in the choice of the weight function could be to put more weights in the area where there is more observation, that is to say to use a weight function which is a function of the number at risk process $Y(\cdot)$. Consequently, it appears difficult to make general recommendations. In any case, our procedure requires that $\hat{w}(\cdot)$ be uniformly convergent in probability (on $[0,\tau]$). An attractive family of weight functions for choosing $\hat{w}(\cdot)$ is a generalization of the Fleming-Harrington family (Fleming and Harrington, 1981; Harrington and Fleming, 1982) defined by

$$\hat{w}(t) = (Y(t))^\zeta \left(\widehat{S}(t)\right)^p \left(1-\widehat{S}(t)\right)^q,$$

where $\zeta \geq 0$, $p \geq 0$ and $q \geq 0$. The weight function of the log-rank test corresponds to $p = q = 0$ and $\zeta = 1$. Moreover, early (late) departures from H_0 receive more weight if p is chosen to be much larger (smaller) than q whereas departures in the mid-range are given more weight if $p = q > 0$.

3.2. Asymptotic behavior

The following theorem establishes the asymptotic distribution of our test statistic.

Theorem 1. *Let us suppose that*

$$\int_0^\tau \frac{dF(s)}{\bar{H}(s)} < \infty, \text{ where } \bar{H}(\cdot) = 1 - H(\cdot) \tag{5}$$

and

$$\sup_{s\in[0,\tau]} |\widehat{w}(s) - w(s)| \xrightarrow{p} 0, \text{ as } n \to \infty.$$

Then, $\sqrt{n}(\widehat{\psi}-\psi)$ converges weakly to a mean zero normal random variable $\mathbb{Z}$, with finite variance σ^2. Under H_0 the limiting variance can be expressed in the form of

$$\begin{aligned}\sigma_0^2 &= (1-\gamma)\int_0^\tau\int_0^\tau w(t)w(s)\int_0^{s\wedge t}\frac{dF_1(u)}{\bar{H}(u)}dtds\\ &\quad+(\frac{1-\gamma}{\gamma^2})\Big(\int_0^\tau w(t)F_1(t)dt\Big)^2\Big(\int_0^\tau\frac{dF_1(u)}{\bar{H}(u)}\Big)\\ &\quad-2(\frac{1-\gamma}{\gamma})\Big(\int_0^\tau w(t)F_1(t)dt\Big)\Big(\int_0^\tau w(t)\int_0^t\frac{dF_1(u)}{\bar{H}(u)}dt\Big).\end{aligned}$$

Proof. It has been shown by Dauxois and Guilloux (2008) that, under (5), the following weak convergence holds in the Skorohod space of *cadlag* functions $\mathbb{D}^3[0,+\infty]$:

$$\sqrt{n}\begin{pmatrix}\widehat{S}(\cdot)-S(\cdot)\\ \widehat{F}_1(\cdot)-F_1(\cdot)\\ \widehat{F}_2(\cdot)-F_2(\cdot)\end{pmatrix}\xrightarrow{\mathscr{L}}\begin{pmatrix}Z_0(\cdot)\\ Z_1(\cdot)\\ Z_2(\cdot)\end{pmatrix},\text{ as } n\to\infty, \tag{6}$$

where $Z_i(\cdot), i=0,1,2$, are mean zero Gaussian processes defined by

$$\begin{aligned}Z_0(\cdot) &= S(\cdot)U_0(\cdot),\\ Z_j(\cdot) &= \int_0^{\cdot}\big(F_j(\cdot)-F_j(u)\big)dU_0(u)+\int_0^{\cdot}S(u)dU_j(u), \quad j=1,2.\end{aligned}$$

Recall that, in this result, the processes $U_1(\cdot)$ and $U_2(\cdot)$ are mean zero Gaussian local square integrable orthogonal martingales with covariance function given by

$$\langle U_j(s),U_j(t)\rangle=\int_0^{s\wedge t}\frac{dF_j(u)}{S^2(u)\bar{H}(u)}, \quad j=1,2, \tag{7}$$

and $U_0(\cdot)=-\big(U_1(\cdot)+U_2(\cdot)\big)$.

Considering the covariance structure of $Z_j(\cdot)$, for $j=1,2$, we can write:

$$\begin{aligned}\langle Z_i(s),Z_j(t)\rangle &= \int_0^s\int_0^t(F_i(s)-F_i(u))(F_j(t)-F_j(v))\,d\langle U_0(u),U_0(v)\rangle\\ &+\int_0^s\int_0^t(F_i(s)-F_i(u))\,S(v)d\langle U_0(u),U_j(v)\rangle\\ &+\int_0^s\int_0^t S(u)\,(F_j(t)-F_i(v))\,d\langle U_i(u),U_0(v)\rangle\\ &+\int_0^s\int_0^t S(u)S(v)d\langle U_i(u),U_j(v)\rangle.\end{aligned}$$

But, thanks to the orthogonality of $U_1(\cdot)$ and $U_2(\cdot)$ and the expression of their covariance functions given in (7), we have:

$$\begin{aligned}
\langle U_0(u),U_0(v)\rangle &= \langle U_1(u),U_1(v)\rangle + \langle U_2(u),U_2(v)\rangle = \int_0^{u\wedge v} \frac{dF(x)}{S^2(x)\bar{H}(x)};\\
\langle U_0(u),U_j(v)\rangle &= -\langle U_j(u),U_j(v)\rangle = -\int_0^{u\wedge v} \frac{dF_j(x)}{S^2(x)\bar{H}(x)} \text{ for } j=1,2;\\
\text{and } \langle U_i(u),U_j(v)\rangle &= \delta_{ij}\int_0^{u\wedge v} \frac{dF_i(x)}{S^2(x)\bar{H}(x)},
\end{aligned}$$

where δ_{ij} is the Kronecker delta. Hence,

$$\begin{aligned}
\langle Z_i(s),Z_j(t)\rangle &= \int_0^{s\wedge t} \big(F_i(s)-F_i(u)\big)\big(F_j(t)-F_j(u)\big)\frac{dF(u)}{S^2(u)\bar{H}(u)}\\
&- \int_0^{s\wedge t} \big(F_i(s)-F_i(u)\big)\frac{dF_j(u)}{S(u)\bar{H}(u)} - \int_0^{s\wedge t} \big(F_j(t)-F_j(u)\big)\frac{dF_i(u)}{S(u)\bar{H}(u)}\\
&+ \delta_{ij}\int_0^{s\wedge t}\frac{dF_i(u)}{\bar{H}(u)}.
\end{aligned}$$

Now, since $\hat{w}(\cdot)$ is assumed to be an uniformly consistent estimator of $w(\cdot)$, we can write:

$$\sqrt{n}(\widehat{\psi}-\psi) = \sqrt{n}\left(\Psi(\hat{F}_1,\hat{F}_2)-\Psi(F_1,F_2)\right)+o_P(1),$$

where

$$\Psi(F_1,F_2) = \int_0^{\tau} w(t)\big[(1-\gamma)F_1(t)-\gamma F_2(t)\big]dt,$$

and $\gamma = F_1(\tau)$. It is easily seen that this latter function is Hadamard-differentiable (see e.g. van der Vaart and Wellner, 1996) with derivative:

$$\begin{aligned}
D_{\Psi}^{F_1,F_2}(\alpha_1,\alpha_2) &= \int_0^{\tau} w(t)\big[(1-F_1(\tau))\alpha_1(t)-\alpha_1(\tau)F_1(t)-F_1(\tau)\alpha_2(t)-\alpha_1(\tau)F_2(t)\big]dt\\
&= \int_0^{\tau} w(t)\big[(1-\gamma)\alpha_1(t)-\gamma\alpha_2(t)\big]dt - \alpha_1(\tau)\int_0^{\tau} w(t)F(t)dt.
\end{aligned}$$

Using the functional δ-method in the version of Theorem 3.9.5 of van der Vaart and Wellner (1996) and the convergence in (6), one gets

$$\sqrt{n}(\widehat{\psi}-\psi) \xrightarrow{\mathscr{L}} D_{\Psi}^{F_1,F_2}(Z_1,Z_2) = \textstyle\int_0^{\tau} w(t)\Big((1-\gamma)Z_1(t)-\gamma Z_2(t)\Big)dt - Z_1(\tau)\Big(\int_0^{\tau} w(t)F(t)dt\Big),$$

as n tends to $+\infty$. The limiting random variable has a mean zero Gaussian distribution with variance function given by

$$\begin{aligned}
\sigma^2 &= \text{Var}\left(\int_0^{\tau} w(t)\Big((1-\gamma)Z_1(t)-\gamma Z_2(t)\Big)dt - Z_1(\tau)\Big(\int_0^{\tau} w(t)F(t)dt\Big)\right)\\
&= \text{Var}\left(\int_0^{\tau} w(t)\Big((1-\gamma)Z_1(t)-\gamma Z_2(t)\Big)dt\right) + \left(\int_0^{\tau} w(t)F(t)dt\right)^2 \text{Var}(Z_1(\tau))\\
&\quad -2\Big(\int_0^{\tau} w(t)F(t)dt\Big)\Big(\int_0^{\tau} w(t)\Big\{(1-\gamma)\langle Z_1(t),Z_1(\tau)\rangle - \gamma\langle Z_2(t),Z_1(\tau)\rangle\Big\}dt\Big).
\end{aligned}$$

Since

$$\begin{aligned}
&\mathrm{Var}\left(\int_0^\tau w(t)\Big((1-\gamma)Z_1(t)-\gamma Z_2(t)\Big)dt\right)\\
&=\int_0^\tau\int_0^\tau w(t)w(s)\Big\{(1-\gamma)^2\langle Z_1(t),Z_1(s)\rangle-2\gamma(1-\gamma)\langle Z_1(t),Z_2(s)\rangle+\gamma^2\langle Z_2(t),Z_2(s)\rangle\Big\}dsdt,
\end{aligned}$$

it follows that:

$$\begin{aligned}
\sigma^2 &= \int_0^\tau\int_0^\tau w(t)w(s)\Big\{(1-\gamma)^2\langle Z_1(t),Z_1(s)\rangle-2\gamma(1-\gamma)\langle Z_1(t),Z_2(s)\rangle+\gamma^2\langle Z_2(t),Z_2(s)\rangle\Big\}dsdt\\
&+ \Big(\int_0^\tau w(t)F(t)dt\Big)^2\langle Z_1(\tau),Z_1(\tau)\rangle\\
&-2 \Big(\int_0^\tau w(t)F(t)dt\Big)\Big(\int_0^\tau w(t)\Big\{(1-\gamma)\langle Z_1(t),Z_1(\tau)\rangle-\gamma\langle Z_2(t),Z_1(\tau)\rangle\Big\}\Big)dt.
\end{aligned}$$

When H_0 is true, one can use the equation

$$F_2(\cdot)=\frac{1-\gamma}{\gamma}F_1(\cdot)$$

to simplify the expression of σ^2 and obtain (after some rather long but easy computations):

$$\begin{aligned}
\sigma_0^2 &= (1-\gamma)\int_0^\tau\int_0^\tau w(t)w(s)\int_0^{s\wedge t}\frac{dF_1(u)}{\bar{H}(u)}dtds\\
&+(\frac{1-\gamma}{\gamma^2})\Big(\int_0^\tau w(t)F_1(t)dt\Big)^2\Big(\int_0^\tau\frac{dF_1(u)}{\bar{H}(u)}\Big)\\
&-2(\frac{1-\gamma}{\gamma})\Big(\int_0^\tau w(t)F_1(t)dt\Big)\Big(\int_0^\tau w(t)\int_0^t\frac{dF_1(u)}{\bar{H}(u)}dt\Big),
\end{aligned}$$

which completes the proof. □

Remark. Now, thanks to this asymptotic result, it's straightforward to use our test statistic in order to decide between the two hypotheses. One has to reject (resp. do not reject) H_0, with level α, when the value of $\sqrt{n}\hat{\Psi}/\hat{\sigma}_0$ is higher (resp. smaller) than the (1-α)-quantile of the normal distribution, with $\hat{\sigma}_0$ a consistent estimator of σ_0.

4. Numerical study

4.1. Simulation study

We carried out a Monte Carlo simulation to study firstly the power of our test. In this order, we have simulated samples of the random couple (X,Y) from the bivariate distribution:

$$f(x,y)=\frac{1}{2x}e^{-x},\text{ with } 0<y<2x.$$

In this case, the hypothesis H_1 is satisfied since the distribution of X is exponential (of parameter 1) and the conditional distribution of Y given $X = x$ is uniform on the interval $[0, 2x]$, which shows that $P(Y - X > 0|X = x) = 1/2$, for all $x \in \mathbb{R}^+$. Thus, the sign of $Y - X$ is independent of X and the Random Sign assumption is fulfilled. In fact, one could even show that it is also the case for the IPRA assumption. The independent censoring time C is simulated according to the exponential distribution with values of the parameter in order to get different percentages of censoring: 0%, 10%, 30% and 50%. We have simulated samples of sizes 50, 100 and 200 and in each case two different levels have been considered 5% and 2%. Before considering the simulation results, let us recall that our test with no censoring (0% of censoring) and a weight function $w(\cdot) = 1$ coincides with the test of Dewan et al. (2004).

The simulation design described above was replicated 10000 times. Table 1 gives the resulting Monte Carlo estimates of the power of our test in the case of a constant weight function $w(\cdot)$ equal to 1. The results suggest that, at least when the sample size is not too small and the percentage of censoring not too high, our test has reasonably good power against the alternative considered. Moreover, one can see that the estimated values of the power respond in the expected manner to changes in the simulation parameters. The power increases with the sample size, decreases when the percentage of censoring increases or when the level decreases.

In order to have an idea of the influence of the choice of the weight function in our test statistic, we have tried a nonconstant weight function. Table 2 gives the simulation results with the same scenario as before but when the weight function is now $w(\cdot) = S(\cdot)$. Note that in this case the test statistic $\hat{\Psi}$ uses $\hat{w}(\cdot) = \hat{S}(\cdot)$, the Kaplan-Meier estimator of the survival function given in (4). This weight function is a member of the generalized Fleming-Harrington family where $\zeta = 0 = q$ and $p = 1$. One can see that the power is better with this weight function than before, showing that there is some interest of using such a weight function.

TABLE 1. *Simulation results. Monte Carlo estimates of the power of the test of H_0 against H_1 with weight function $w(\cdot) = 1$*

	Sample Size					
	50		100		200	
Level	5%	2%	5%	2%	5%	2%
Censoring						
0%	0.64	0.42	0.90	0.77	0.99	0.98
10%	0.54	0.34	0.81	0.65	0.98	0.94
30%	0.39	0.21	0.65	0.44	0.91	0.80
50%	0.29	0.15	0.46	0.28	0.74	0.55

TABLE 2. *Simulation results. Monte Carlo estimates of the power of the test of H_0 against H_1 with weight function $w(\cdot) = S(\cdot)$*

	Sample Size					
	50		100		200	
Level	5%	2%	5%	2%	5%	2%
Censoring						
0%	0.69	0.47	0.96	0.87	0.999	0.996
10%	0.61	0.41	0.89	0.74	0.997	0.98
30%	0.44	0.25	0.74	0.53	0.96	0.89
50%	0.31	0.16	0.54	0.35	0.84	0.68

We have also carried out a simulation study of the level of our test on small sample sizes. In this order, we have simulated samples of the random couple (X,Y) from the Block and Basu (1974) absolutely continuous bivariate exponential with probability density function:

$$f(x,y) = \begin{cases} \frac{\lambda_1\lambda(\lambda_2+\lambda_0)}{\lambda_1+\lambda_2} e^{-\lambda_1 x-(\lambda_2+\lambda_0)y}, & \text{for} \quad x<y \\ \frac{\lambda_2\lambda(\lambda_1+\lambda_0)}{\lambda_1+\lambda_2} e^{-\lambda_2 y-(\lambda_1+\lambda_0)x}, & \text{for} \quad x>y \end{cases},$$

where $\lambda = \lambda_0+\lambda_1+\lambda_2$. In this case, the hypothesis H_0 is satisfied since one can show that the function $\Phi(\cdot)$ is constant with value:

$$\Phi(t) = \frac{\lambda_1}{\lambda_1+\lambda_2},$$

for all $t>0$.

The independent censoring time C is still simulated according to the exponential distribution with values of the parameter in order to get different percentages of censoring: 0%, 10% and 30%. We have simulated samples of sizes 50 and 100 and in each case two different levels have been considered 5% and 2%. This simulation design has been replicated 10000 times and the results are listed in Table 3. One can see that the empirical levels are not far from the nominal ones and that, as expected, they tend to be closer when the sample size increases. The percentage of censoring seems to affect more the empirical levels when the sample size is small (n=50) than when n=100.

TABLE 3. *Simulation results. Monte Carlo estimates of the level of the test of H_0 against H_1 with weight function $w(\cdot)=1$*

	Sample Size			
	50		100	
Level	5%	2%	5%	2%
Censoring				
0%	5.34%	2.04%	4.34%	1.9%
10%	4.78%	2.17%	4.99%	2.09%
30%	4.49%	2.14%	5.22%	1.99%

4.2. Application on Norsk Hydro data set

Now let us consider a real dataset introduced by Bunea et al. (2003). This dataset gives failure times of two identical compressor units of the Norsk Hydro ammonia plant between the 2nd of October 1968 and the 25th of June 1989. The dataset contains a large part of the history of the compressor units like: the time of component failure, the failure modes (leakage, no start, unwanted start, vibration, warming, overhaul, little gas stream, great gas stream, others), the degree of failure (critical or non critical), the down times of the component, identification of the compressor unit where the failure occurs (first unit failed, second unit failed, both units failed), the action taken after a failure, system or subsystem failure, the action taken (immediate reparation, immediate replacement, adjustement, planned overhaul, modification, others) and finally the revision periods (18 revision periods with different lengths).

There are many different ways to use this dataset in order to illustrate our testing procedure on real problems. We have decided to follow Bunea et al. (2003) and Dijoux and Gaudoin (2009) in considering the lifetime T as the operation time since last failure. Like these authors, we have assumed that these operation times since last failure are i.i.d., even if the plausibility of such an assumption can be criticized. See our comments on this subject in Section 5. Since it’s natural to associate a CM to a critical failure and a PM to a non-critical failure, we follow Bunea et al. (2003) in considering that $\delta = 2$ in case of a critical failure and $\delta = 1$ in case of a non-critical failure. This yields 338 observed lifetimes, where 247 are with cause $\delta = 1$ and 91 with $\delta = 2$. There is no censoring time in this case.

Figure 1 plots the empirical estimate of the function $\Phi(\cdot)$ defined by

$$\hat{\Phi}(t) = \frac{\sum_{i=1}^{338} I\{T_i > t, \delta_i = 1\}}{\sum_{i=1}^{338} I\{T_i > t\}},$$

for all $t > 0$. One can see that this function appears to be almost constant. This is confirmed by our test. Indeed, its p-value is 36.55% which does not lead to reject H_0.

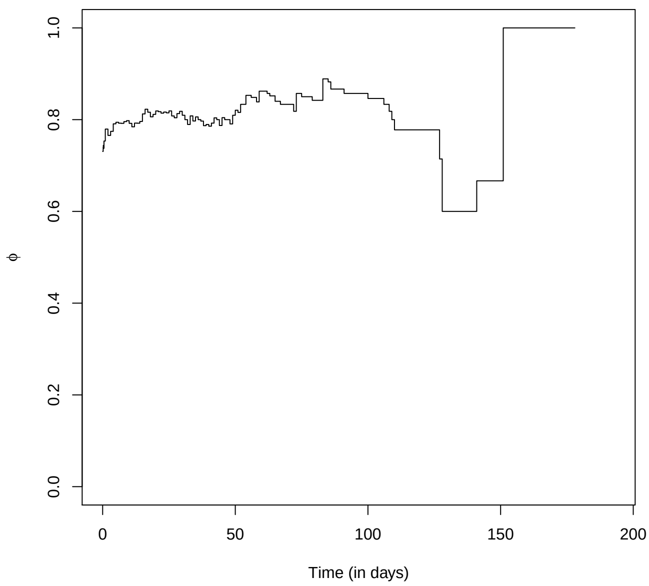

FIGURE 1. *Norsk Hydro dataset (Bunea et al., 2003). Empirical estimation of the function* $\Phi(t) = P(\delta = 1|T > t)$

5. Conclusion

In this paper we have considered, in a setup of competing risks, the problem of testing “the function $\Phi(\cdot)$ is constant” against “the function $\Phi(\cdot)$ is not constant and attains its maximum at the origin”, where $\Phi(t) = P(\delta = 1|T > t)$, for all $t > 0$. Our Section 2.2 devoted to recall the main models (based on competing risks) of maintenance analysis in Reliability shows that our test can be used to decide between “Exponential Delay Time model” and a “Random Sign Censoring model”.

We have obtained the asymptotic distribution of our test statistic and have shown, thanks to a simulation study, its rather good behavior even on small sample sizes. We have illustrated its use on the Norsk Hydro dataset where the question of the constancy of the $\Phi(\cdot)$ function has been addressed by many authors (see Bunea et al., 2003 and Dijoux and Gaudoin, 2009). The conclusion of our testing procedure is that the p-value is 36.55% and that we can't reject H_0 in favor of H_1. Thus, in this case it should be preferable to use an "Exponential Delay Time model" rather than a "Random Sign Censoring model". However, one has also to check that an other model (for example one of those introduced in Section 2.2) doesn't better fit the data.

This is why there is place for further work in the direction of our paper. Indeed, it would be of interest to derive statistical tests for the other models of maintenance introduced in Section 2.2. For example, it should be possible to consider the test of H_0 against

$$H_2 : \Phi(t) \text{ is a non-constant decreasing function of } t,$$

since hypothesis H_2 would suggest that an "Intensity Proportional Repair Alert" model (or an independent competing risk model with exponential mixture after an easy adaptation of our test) could be applied. Tests based on the functions $CS_1(\cdot)$ and $CS_2(\cdot)$ would also be of interest.

Finally, let us mention that all the models introduced in Section 2 (sometime called Usual Competing Risk models) assume a perfect maintenance. Of course, this assumption may be violated in real applications. To take into account such situations, some authors have built more general models, like the Generalized Competing Risks models developed in Doyen and Gaudoin (2006), Dijoux et al. (2008),Deloux et al. (2012) and Dijoux and Gaudoin (2014). More work is needed in order to construct goodness-of-fit test for these models.

Acknowledgments

The authors are grateful to the editor and the guest editor of this special issue to have invited them to submit an article. The authors also thank the two anonymous referees for their careful reading of our paper and their constructive comments and suggestions that have led to improve the paper. The work of the first author is supported by the French National Agency of Research (ANR), project AMMSI, number ANR-2011-BS01-021.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York.

Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential extension. *J. Amer. Statist. Assoc.*, 69:1031–1037.

Bunea, C. and Bedford, T. (2002). The effect of model uncertainty on maintenance optimization. *IEEE Transactions on Reliability*, 51(4):486–493.

Bunea, C., Cooke, R., and Lindqvist, B. H. (2003). Competing risk perspective on reliability databases. In *Mathematical and statistical methods in reliability (Trondheim, 2002)*, volume 7 of *Ser. Qual. Reliab. Eng. Stat.*, pages 355–370. World Sci. Publ., River Edge, NJ.

Christer, A. H. (2002). A review of delay time analysis for modelling plant maintenance. In *Stochastic models in reliability and maintenance*, pages 89–123. Springer, Berlin.

Cooke, R. M. (1993). The total time on test statistic and age-dependent censoring. *Statist. Probab. Lett.*, 18(4):307–312.

Cooke, R. M. (1996). The design of reliability databases, part i and ii. *Reliability Eng. System safety*, 51:137–146 and 209–223.

Cooke, R. M. and Paulsen, J. (1997). Concepts for measuring maintenance performance and methods for analysing competing failures modes. *Reliability Eng. System safety*, 55:135–141.

Crowder, M. (2001). *Classical Competing Risks*. Springer Series in Statistics. Chapman & Hall/CRC, Boca Raton, FL.

Dauxois, J.-Y. and Guilloux, A. (2008). Nonparametric inference under competing risks and selection-biased sampling. *J. Multivariate Anal.*, 99(4):589–605.

Dauxois, J.-Y. and Kirmani, S. N. U. A. (2004). On testing the proportionality of two cumulative incidence functions in a competing risks setup. *J. Nonparametr. Stat.*, 16(3-4):479–491.

Deloux, E., Dijoux, Y., and Fouladirad, M. (2012). Generalization of the proportional hazards model for maintenance modelling and optimization. *J. Risk Reliab.*, 226(5):439–447.

Dewan, I., Deshpande, J. V., and Kulathinal, S. B. (2004). On testing dependence between time to failure and cause of failure via conditional probabilities. *Scand. J. Statist.*, 31(1):79–91.

Dijoux, Y., Doyen, L., and Gaudoin, O. (2008). Conditionally independent generalized competing risks for maintenance analysis. In *Advances in mathematical modeling for reliability*, pages 88–95. IOS, Amsterdam.

Dijoux, Y. and Gaudoin, O. (2009). The alert-delay competing risks model for maintenance analysis. *J. Statist. Plann. Inference*, 139(5):1587–1603.

Dijoux, Y. and Gaudoin, O. (2014). Generalized random sign and alert delay models for imperfect maintenance. *Lifetime Data Anal.*, 20(2):185–209.

Doyen, L. and Gaudoin, O. (2006). Imperfect maintenance in a generalized competing risks framework. *J. Appl. Probab.*, 43(3):825–839.

Fleming, T. R. and Harrington, D. P. (1981). A class of hypothesis tests for one- and two-sample censored survival data. *Comm. Statist. A—Theory Methods*, 10(8):763–794.

Harrington, D. P. and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika*, 69(3):553–566.

Hokstad, P. and Jensen, R. (1998). Predicting the failure rate for components that go through a degradation state. In *Safety and Reliability, S. Lydersen, G.K. Hansen and H.A. Sandtorv (eds), Balkema, Rotterdam*, pages 389–396.

Langseth, H. and Lindqvist, B. H. (2003). A maintenance model for components exposed to several failure mechanisms and imperfect repair. In *Mathematical and statistical methods in reliability (Trondheim, 2002)*, volume 7 of *Ser. Qual. Reliab. Eng. Stat.*, pages 415–430. World Sci. Publ., River Edge, NJ.

Langseth, H. and Lindqvist, B. H. (2006). Competing risks for repairable systems: a data study. *J. Statist. Plann. Inference*, 136(5):1687–1700.

Li, W. and Pham, H. (2005). Reliability modeling of multi-state degraded systems with multi-competing failures and random shocks. *IEEE Transactions on Reliability*, 54(2):297–301.

Lindqvist, B. H., Støve, B., and Langseth, H. (2006). Modelling of dependence between critical failure and preventive maintenance: the repair alert model. *J. Statist. Plann. Inference*, 136(5):1701–1717.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. U.S.A.*, 72:20–22.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.