Testing dependence between the failure time and failure modes: An application of enlarged filtration

Dario Gasbarra[a,*], **S. B. Kulathinal**[b], **Isha Dewan**[c], **Aulikki Nissinen**[b]

[a] *Rolf Nevanlinna Institute, University of Helsinki, P.L. 4, 00014 Helsinki, Finland,*

[*] *E-mail: dag@rni.helsinki.fi, Tel.: +358 9 191 22765, Fax: +358 9 191 22779.*

[b] *Department of Epidemiology and Health Promotion, National Public Health Institute, Mannerheimintie 166, 00300 Helsinki, Finland*

[c] *Indian Statistical Institute, New Delhi, India*

Abstract

The model of independent competing risks provides no information for the assessment of competing failure modes if the failure mechanisms underlying these modes are coupled. Models for dependent competing risks in the literature can be distinguished on the basis of the functional behaviour of the conditional probability of failure due to a particular failure mode given that the failure time exceeds a fixed time, as a function of time. There is an interesting link between monotonicity of such conditional probability and dependence between failure time and failure mode, via crude hazard rates. In this paper, we propose tests for testing the dependence between failure time and failure mode using the crude hazards and using the conditional probabilities mentioned above. We establish the equivalence between the two approaches and provide an asymptotically efficient weight function under a sequence of local alternatives. The tests are applied to simulated data and to mortality follow-up data.

Keywords : Competing risks; Conditional probability; Crude hazards; Enlarged filtration; Functional delta method; Kolmogorov-Smirnov-type tests

1 Introduction

In a follow-up study of mortality, it is observed that the contribution of death due to causes cardiovascular diseases, cancer, accident and suicide combined to total death decreases with age. In such situations it is of interest to compare the probabilities of dying due to these causes combined given that a person has survived up to a certain age. It is also of interest to test whether such conditional probabilities increase or decrease with age. Dewan *et al.* (2004) give several examples where the conditional probabilities are of interest. In this paper, we study the relationship between crude hazards and these conditional probabilities in the case of two competing risks. We develop test procedures using the crude hazards and a weighted Kolmogorov-Smirnov-type test for testing dependence between failure mode and failure time. For a specific choice of local alternative, the two tests are shown to be equally efficient and an optimal kernel is given. A test based on crude hazards can then easily be extended to include more than two risks as well as censoring. The methods are illustrated by simulated data on failure time T and the failure mode δ and by real data from a mortality follow-up study conducted in Finland.

The competing risks data consist of the failure time, T and an indicator of failure mode, δ which can have one of the values $\{0, 1\}$.

Define the joint distribution of (T, δ) through the subsurvival functions

$$S_i(t) = P[T \geq t, \delta = i], i = 0, 1,$$

leading to the overall survival function of the failure time

$$S(t) = P[T \geq t] = S_0(t) + S_1(t).$$

Let $F_i(t) = P(\delta = i) - S_i(t)$ and $F(t) = 1 - S(t)$ denote the absolutely continuous subdistribution and distribution functions, respectively with $f_i(.)$ and $f(.)$ denoting the corresponding subdensity and density functions. Also, denote the conditional probability of failure due to the first risk given that there is no failure up to time t by

$$\Phi_1(t) = P[\delta = 1 \mid T \geq t] = \frac{S_1(t)}{S(t)}. \tag{1}$$

Equivalently, we can define $\Phi_0(t) = P[\delta = 0 \mid T \geq t] = 1 - \Phi_1(t) = S_0(t)/S(t)$. It is clear that independence of T and δ is equivalent to $\Phi_1(t) = P[\delta = 1] = \phi, \forall\ t > 0$. In a general dependence set-up, the analysis of competing risks data is usually carried out using the subsurvival functions $S_i(t), i = 0, 1$. If T and δ are independent then $S_i(t) = S(t)P[\delta = i]$ and hence, testing the hypothesis of equality of subsurvival functions $S_1(t) = S_0(t),\ \forall\ t > 0$ is equivalent to testing whether $P[\delta = 1] = P[\delta = 0]$. Note that the common value of these probabilities is $1/2$ since $P[\delta = 1] + P[\delta = 0] = 1$.

Let $\Lambda_i(t)$ and $\tilde{A}_i(t)$ be the cumulative cause-specific and cumulative crude hazards for failure mode i, and are given by

$$\Lambda_i(t) = \int_0^t \frac{dF_i(u)}{S(u)}, \quad \tilde{A}_i(t) = \int_0^t \frac{dF_i(u)}{S_i(u)}.$$

It is easy to verify that

$$\frac{d\Lambda_i(t)}{d\tilde{A}_i(t)} = \Phi_i(t). \tag{2}$$

We consider the testing problems $H_0 : \Phi_1(t) = \phi$ for some constant ϕ against
$H_1 : \Phi_1(t)$ is not constant, and $H_2 : \Phi_1(t)$ is increasing in t.
Dewan *et al.*(2004) have proposed tests based on U-statistics for H_0 versus H_1 and H_2. Their test U_3 is shown to be asymptotically equivalent to a weighted test proposed here for a special choice of the weight function. In terms of cause-specific hazards, the null hypothesis is

$$\frac{\Phi_1(t)}{\Phi_0(t)} = \frac{S_1(t)}{S_0(t)} = \frac{\phi}{1 - \phi} = \theta$$

and is equivalent to testing $d\Lambda_1(t)/d\Lambda_0(t) = f_1(t)/f_0(t) = \theta$ or $a_1(t) = a_0(t) = a(t)$, where $a_i(t) = d\tilde{A}_i(t)/dt, i = 0, 1$ are the crude hazards. The alternative hypothesis that $\Phi_1(t)$ is increasing in t is equivalent to $a_1(t) \leq a_0(t)$.

In section 2, we propose a test based on crude hazards and a weighted Kolmogorov-Smirnov-type of test for testing the above hypotheses. We also prove the equivalence between the optimal tests obtained in these two classes of tests. In section 3, the proposed tests are illustrated via simulated data and mortality follow-up data. We also compare the optimal weight function with the Harrington and Fleming (1982) type of weight function and propose its use in practice, since the form of the optimal weight function is generally not known.

2 Test of significance

Let $(T_j, \delta_j),\ j = 1, 2, ..., n$ be the competing risks data obtained from n independent and identical copies of the system. Define the counting processes

$$\begin{aligned} N_i(t) &= \sum_{j=1}^{n} I[T_j \leq t, \delta_j = i], i = 0, 1, \\ N_{.}(t) &= N_0(t) + N_1(t), Y_{.}(t) = \sum_{j=1}^{n} I[T_j \geq t]. \end{aligned}$$

Note that $N_i(t)$ counts the number of failures due to competing risk i up to time t and $Y_{.}(t)$ counts the number of units at risk just prior to time t. The natural estimates of the subsurvival functions are given by their empirical counterparts

$$\begin{aligned} \hat{F}_{in}(t) &= \frac{N_i(t-)}{n},\ \ \hat{S}_{in}(t) = \frac{N_i(\infty)}{n} - \hat{F}_{in}(t)\ \ and \\ \hat{S}_n(t) &= \frac{Y_{.}(t)}{n},\ \ \hat{\phi}_n = \frac{N_1(\infty)}{n}. \end{aligned}$$

2.1 Test based on crude hazards

Let $\mathcal{F}_t^{N,Y}$ be the filtration generated by (N_0, N_1, Y). Consider the enlarged filtration $\mathcal{G}_t = \mathcal{F}_t^{N,Y} \vee \sigma(N_1(\infty))$. Note that $N_0(\infty)$ is implicitly included in the enlarged filtration since $N_0(\infty) = n - N_1(\infty)$. Using Bayes' formula, the compensator of $N_1(dt)$ with respect to the enlarged filtration $\mathcal{G}_t$ is given by

$$\begin{aligned} P(N_1(dt) = 1 | \mathcal{F}_{t-}^{N,Y}, N_1(\infty)) &= \frac{P(N_1(dt) = 1 | \mathcal{F}_{t-}^{N,Y}) P(N_1(\infty) | \mathcal{F}_{t-}^{N,Y}, N_1(dt) = 1)}{P(N_1(\infty) | \mathcal{F}_{t-}^{N,Y})} \\ &= \frac{D_1}{D_2} Y(t) \, \Lambda_1(dt) \\ &= \frac{N_1(\infty) - N_1(t-)}{\Phi_1(t)} \Lambda_1(dt), \end{aligned}$$

where $D_1 = \binom{Y(t) - 1}{N_1(\infty) - N_1(t-) - 1} \Phi_1(t)^{N_1(\infty) - N_1(t-) - 1} (1 - \Phi_1(t))^{Y(t) + N_1(t-) - N_1(\infty)}$ and $D_2 = \binom{Y(t)}{N_1(\infty) - N_1(t-)} \Phi_1(t)^{N_1(\infty) - N_1(t-)} (1 - \Phi_1(t))^{Y(t) + N_1(t-) - N_1(\infty)}$.

It is interesting to note that the conditional probability of interest, $\Phi_1(t)$, appears in the compensator. Similarly, $P(N_0(dt) = 1 | \mathcal{F}_{t-}^{N,Y}, N_1(\infty))$ can be obtained. Hence, for $i = 0, 1$,

$$\begin{aligned} \tilde{M}_i(t) &= N_i(t) - \int_0^t \frac{N_i(\infty) - N_i(s-)}{\Phi_i(t)} d\Lambda_i(s) \\ &= N_i(t) - \int_0^t Y_i(t) d\tilde{A}_i(s) \end{aligned} \tag{3}$$

is a $\{\mathcal{G}_t\}$-martingale where $Y_i(t) = N_i(\infty) - N_i(t-)$ and the second equality follows due to (2). The predictable variation process of $\tilde{M}_i(t)$ is

$$\langle \tilde{M}_i \rangle_t = \int_0^t Y_i(t) d\tilde{A}_i(s). \tag{4}$$

We can split the group of n individuals into a group of $N_1(\infty)$ individuals, those which will fail from cause 1, and a group of $N_0(\infty)$ individuals which will fail from cause 0.

The counting processes $N_0(t)$ and $N_1(t)$ are conditionally independent given $N_1(\infty)$. Testing the null hypothesis H_0 is equivalent to testing whether the $\{\mathcal{G}_t\}$-intensities of $N_0(t)$ and $N_1(t)$ are identical. This testing problem has been discussed by Andersen *et al.* (1993) on pages 345-348.

Let $K_n(t)$ be a $\{\mathcal{G}_t\}$-predictable weight process such that it is non-zero whenever the risk sets corresponding to the two groups are non-empty and is zero as soon as one of the two risk sets becomes empty. Define a test statistic

$$\begin{aligned} V_n &= \int_0^\infty K_n(s)\left(\frac{dN_1(s)}{Y_1(s)} - \frac{dN_0(s)}{Y_0(s)}\right) \\ &= \int_0^\infty K_n(s)\left(\frac{d\tilde{M}_1(s)}{Y_1(s)} - \frac{d\tilde{M}_0(s)}{Y_0(s)}\right) + \int_0^\infty K_n(s)\left(d\tilde{A}_1(s) - d\tilde{A}_0(s)\right), \end{aligned} \tag{5}$$

where $K_n(t)$ must be chosen in some efficient way and the second equality follows due to (3). Note that the second term in the last equality is zero under the null hypothesis. The definition of the weight function is natural since the estimates of the crude hazards need to be compared in the environment when there are subjects acting in both environments. A test based on V_n can be used to test H_0 against H_1 and H_2 since large values of V_n (either negative or positive) support H_1 while large negative values support H_2.

Theorem 2.1 *Assume that $n^{-1}Y_i(t)$ converges uniformly in probability to a deterministic function $y_i(t)$ for $i = 0, 1$. Further assume that $n^{-1}K_n(t)$ converges to $k(t)$ uniformly in probability with $k(.)$ bounded on $[0, \infty]$. Under the null hypothesis, as $n \to \infty$, $n^{-1/2}V_n$ converges in distribution to a Gaussian random variable with mean zero and variance σ^2 where*

$$\sigma^2 = \int_0^\infty k^2(t)\left(\frac{1}{y_1(t)} + \frac{1}{y_0(t)}\right)d\tilde{A}(t),$$

and $\tilde{A}(t) = \tilde{A}_1(t) = \tilde{A}_0(t)$, the common value of the crude hazards under the null hypothesis.

Proof: The result follows from (3), (4) and (5), and

$$< n^{-1/2}V_n, n^{-1/2}V_n >= \int_0^\infty n^{-1}K_n^2(t)(\frac{d\tilde{A}(t)}{Y_1(t)} + \frac{d\tilde{A}(t)}{Y_0(t)}). \qquad \square$$

σ^2 can be consistently estimated by

$$\begin{aligned} \hat{\sigma}^2 &= \int_0^\infty K_n^2(t)\left(\frac{1}{Y_1(t)Y.(t)} + \frac{1}{Y_0(t)Y.(t)}\right)dN.(t) \\ &= \int_0^\infty K_n^2(t)\{Y_1(t)Y_0(t)\}^{-1}dN.(t). \end{aligned} \tag{6}$$

Consider the sequence of local alternatives $\{P^{(n)}(\theta)\}$ for the crude hazards of the form

$$a_i^{(n)}(t) = a(t)(1 + \varepsilon_n\gamma(t)\theta_i),\ i = 0, 1 \tag{7}$$

where $\theta = (\theta_0, \theta_1) \in R^2$ is a local parameter, $\gamma(t), t > 0$ is a fixed function and $\epsilon_n = O(n^{-1/2})$. Under these local alternatives, the asymptotic mean and the variance of $n^{-1/2}V_n$ are

$$\begin{aligned} \mu &= (\theta_1 - \theta_0)\int_0^\infty k(t)a(t)\gamma(t)dt, \\ \sigma^2 &= \{\phi(1-\phi)\}^{-1}\int_0^\infty k^2(t)a(t)S^{-1}(t)dt. \end{aligned} \tag{8}$$

The asymptotic distribution of the test statistic under these local alternatives is a Gaussian distribution with mean μ and variance σ^2. To maximise the local asymptotic power is equivalent to maximising the noncentrality parameter μ^2/σ^2 which is given by

$$(\theta_1 - \theta_0)^2\phi(1-\phi)(\int_0^\infty k(t)a(t)\gamma(t)dt)^2(\int_0^\infty k^2(t)a(t)S^{-1}(t)dt)^{-1}.$$

It can be easily seen that the kernel $k(t)$ which maximises this noncentrality parameter is proportional to $\gamma(t)S(t)$ and hence this choice of kernel gives the most efficient test (see Andersen *et al.*, 1993, pages 372-375 for details). The maximum value of the noncentrality parameter is

$$(\theta_1 - \theta_0)^2\phi(1-\phi)\int_0^\infty \gamma^2(t)S(t)a(t)dt. \tag{9}$$

The above derivation is applicable to a more general sequence of local alternatives

$$a_i^{(n)}(t) = a(t)(1 + \varepsilon_n \gamma_i(t)),\ i = 0, 1, \tag{10}$$

with $\gamma(t) = \gamma_1(t) - \gamma_0(t)$.

In practice, the form of the optimal weight function is generally not known. An alternative is to use Harrington and Fleming (1982) type weight function which is given by $[\hat{S}_n(t)]^\rho$, where ρ is a fixed constant between 0 and 1 and $\hat{S}_n(t)$ is an estimate of the overall survival function.

2.2 Kolmogorov-Smirnov-type test

The hypothesis H_2 is identical to $\Phi(t_1) \le \Phi(t_2)$, whenever $t_1 \le t_2$ which can equivalently be written as

$$\begin{aligned} S_1(t_1)/S(t_1) &\le S_1(t_2)/S(t_2) \\ S_1(t_1)S(t_2) &\le S_1(t_2)S(t_1). \end{aligned}$$

This gives $\Psi(t_1, t_2) = S_1(t_2)S(t_1) - S_1(t_1)S(t_2) = S_1(t_2)S_0(t_1) - S_1(t_1)S_0(t_2) \ge 0$ whenever $t_1 \le t_2$ with strict inequality for some (t_1, t_2). Let $\hat{\Psi}_n(t_1, t_2)$ be obtained by replacing the functions by their empirical counterparts

$$\begin{aligned} \hat{\Psi}_n(t_1, t_2) &= \hat{S}_{1n}(t_2)\hat{S}_n(t_1) - \hat{S}_{1n}(t_1)\hat{S}_n(t_2) \\ &= \hat{F}_{1n}(\infty)(\hat{F}_n(t_2) - \hat{F}_n(t_1)) + \hat{F}_{1n}(t_1)(1 - \hat{F}_n(t_2)) - \hat{F}_{1n}(t_2)(1 - \hat{F}_n(t_1)). \end{aligned}$$

A Kolmogorov-Smirnov-type statistic to test H_0 against H_1 can be defined as $\sqrt{n}\hat{D}_n = \sqrt{n}\sup_{u \le v} \mid \hat{\Psi}_n(u, v) \mid$. Large values of the test statistic support H_1. A one-sided test to test H_0 against H_2 can be based on the statistic $\sqrt{n}\hat{D}_{1n} = \sqrt{n}\sup_{u \le v}(\hat{\Psi}_n(u, v))$ and large positive values support H_2. Similarly, a hypothesis that $\Phi_1(t)$ is decreasing can be tested and large negative values support the hypothesis.

The following theorem is proved in the Appendix I by the functional delta method.

Theorem 2.2 *As n tends to ∞, $\sqrt{n}(\hat{\Psi}_n(u,v) - \Psi(u,v))$ converges to a zero-mean Gaussian random field $Z(u,v)$ with covariance structure*

$$\begin{aligned} cov(Z(u_1,v_1), Z(u_2,v_2)) &= 0 \textit{ if } u_1 \leq v_1 \leq u_2 \leq v_2 \textit{ or } u_2 \leq v_2 \leq u_1 \leq v_1 \\ &= \phi(1-\phi)(1-F(max(v_1,v_2)) \\ &\quad (1-F(min(u_1,u_2))(F(min(v_1,v_2)-F(max(u_1,u_2)) \\ &\qquad otherwise \qquad (11) \\ var(Z(u,v)) &= \phi(1-\phi)(1-F(v))(1-F(u))(F(v)-F(u)) \\ &= \phi(1-\phi)(1-t)(1-s)(t-s), \qquad (12) \end{aligned}$$

where $F(u) = s$ and $F(v) = t$.

The above theorem can be used directly to test the hypothesis H_0 against H_1 and H_2 but this point is not elaborated here since our interest in the distribution of $Z(u,v)$ is in defining a more general class of tests by weighing the values of $\hat{\Psi}_n(u,v)$ rather than restricting attention to its extrema.

2.3 A class of weighted Kolmogorov-Smirnov-type tests

For some deterministic weight function $K(u,v)$, we consider

$$\begin{aligned} \Delta &= \iint_{0<u\leq v<\infty} K(u,v)(S_1(v)S_0(u) - S_1(u)S_0(v))dudv \\ &= \iint_{u\leq v} K(u,v)\Psi(u,v)dudv. \qquad (13) \end{aligned}$$

A weighted Kolmogorov-Smirnov-type test statistic for testing H_0 against H_1 and H_2 is defined via a stochastic kernel K_n as

$$\begin{aligned} \hat{\Delta}_n &= \iint_{0<u\leq v<\infty} K_n(u,v)(\hat{S}_{1n}(v)\hat{S}_{0n}(u) - \hat{S}_{1n}(u)\hat{S}_{0n}(v))dudv \\ &= \iint_{u\leq v} K_n(u,v)\hat{\Psi}_n(u,v)dudv, \qquad (14) \end{aligned}$$

such that

$$\sqrt{n}\iint\limits_{u\le v}(K_n(u,v)-K(u,v))\Psi(u,v)dudv \xrightarrow{\mathbf{P}} 0 \text{ as } n\to\infty. \tag{15}$$

The asymptotic distribution can be obtained by using the covariance structure (11) and (12) of the Kolmogorov-Smirnov-type test.

Theorem 2.3 *Under the assumption (15), as n tends to ∞, $\sqrt{n}(\hat{\Delta}_n-\Delta)$ converges in distribution to a Gaussian random variable with mean zero and variance*

$$\sigma^2=\iint\limits_{u\le v}\iint\limits_{u'\le v'}K(u,v)K(u',v')cov(Z(u,v),Z(u',v'))dudvdu'dv'.$$

The U-statistic, U_3 in Dewan *et al.*(2004) can be obtained from (14) by selecting $K_n(u,v)$ such that the limit $K(u,v)dudv=24dF_1(u)dF_1(v)$, and has asymptotic variance $\sigma^2=(96/35)\phi^5(1-\phi)$.

To check the efficiency of this test and also to compare it with the test based on V_n given by (5), we consider the same local alternatives $\{P^{(n)}(\theta)\}$ as in (7).

We define

$$\begin{aligned} A(t) &= \int_0^t a(s)ds,\ \Gamma(t)=A^{-1}(t)\int_0^t a(s)\gamma(s)ds,\\ S(t) &= \exp(-A(t)),\ V(t)=\exp(-A(t)\Gamma(t)). \end{aligned}$$

Then we let the corresponding cumulative hazards be given by

$$A_i^{(n)}(t)=A(t)(1+\varepsilon_n\Gamma(t)\theta_i),\ i=0,1 \tag{16}$$

and the corresponding sequence of subsurvival functions is

$$\begin{aligned} S_1^{(n)}(t) &= \phi\exp(-A_1^{(n)}(t))=\phi S(t)V(t)^{\varepsilon_n\theta_1},\\ S_0^{(n)}(t) &= (1-\phi)\exp(-A_0^{(n)}(t))=(1-\phi)S(t)V(t)^{\varepsilon_n\theta_0},\\ S^{(n)}(t) &= S_1^{(n)}(t)+S_0^{(n)}(t). \end{aligned}$$

To obtain the asymptotic mean μ of $\sqrt{n}\hat{\Delta}_n$ under the local alternatives, consider

$$\sqrt{n}\iint\limits_{u\le v} K(u,v)(S_1^{(n)}(v)S_0^{(n)}(u) - S_1^{(n)}(u)S_0^{(n)}(v))dudv =$$
$$\sqrt{n}\phi(1-\phi)\iint\limits_{u\le v} K(u,v)S(v)S(u)\Big(V(v)^{\theta_1\varepsilon_n}V(u)^{\theta_0\varepsilon_n} - V(v)^{\theta_0\varepsilon_n}V(u)^{\theta_1\varepsilon_n}\Big)dudv.$$

As n tends to ∞, the above expression converges to the limit

$$\begin{aligned}\mu &= \phi(1-\phi)\iint\limits_{u\le v} K(u,v)S(v)S(u)(\log V(v) - \log V(u))(\theta_1-\theta_0)dudv \\ &= \phi(1-\phi)(\theta_0-\theta_1)\iint\limits_{u\le v} K(u,v)S(v)S(u)(A(v)\Gamma(v) - A(u)\Gamma(u))dudv. \quad (17)\end{aligned}$$

To obtain the noncentrality parameter of the limiting test, we must square (17) and divide by the asymptotic variance of $\sqrt{n}\hat{\Delta}_n$

$$\mathrm{Var}\left(\iint\limits_{u\le v} K(u,v)Z(u,v)dudv\right)$$
$$= \iint\limits_{u\le v}\ \iint\limits_{u'\le v'} K(u,v)K(u',v')\ \mathrm{Cov}\ (Z(u,v),Z(u',v'))du'dv'dudv = \| K \|_{\mathcal{H}}^2$$

where $\|\cdot\|_{\mathcal{H}}$ is the norm in the corresponding reproducing kernel Hilbert space.

A kernel $K(u,v)$ is efficient under the sequence of local alternatives (16) (or equivalently (7)) when it maximises

$$\| K \|_{\mathcal{H}}^{-2}\ \phi^2(1-\phi)^2(\theta_0-\theta_1)^2(\iint\limits_{u\le v} K(u,v)S(v)S(u)(A(v)\Gamma(v) - A(u)\Gamma(u))dudv)^2. \quad (18)$$

If we denote $L(u,v) = S(u)S(v)(A(v)\Gamma(v) - A(u)\Gamma(u))$ and for a generic function $G(u,v)$ we define the convolution operator

$$(\mathcal{R}G)(u,v) = \iint\limits_{u'\le v'} G(u',v')\ \mathrm{Cov}\ (Z(u,v),Z(u',v'))du'dv',$$

then we can rewrite expression (18) without terms with ϕ and $\theta's$ as

$$\frac{(K,\mathcal{R}^{-1}L)_{\mathcal{H}}^2}{(K,K)_{\mathcal{H}}},$$

where $(,)_{\mathcal{H}}$ is the inner product in Hilbert space. It is clear that this is maximal when $K(u,v) = (\mathcal{R}^{-1}L)(u,v)$.

In other words, the asymptotically efficient kernel $K(u,v)$ satisfies

$$(\mathcal{R}K)(u,v) = \iint\limits_{u' \le v'} K(u',v') \text{ Cov } (Z(u,v), Z(u',v'))du'dv' = L(u,v). \tag{19}$$

It is shown in Appendix II that an optimal weight function which satisfies (19) is of the form

$$K(u,v) = k(u)\dot{\delta}_u(v) = \gamma(u)S^{-1}(u)\dot{\delta}_u(v) \tag{20}$$

where $\dot{\delta}_u$ is the derivative of the delta function in the sense of distributions for a smooth function f,

$$\int\limits_{-\infty}^{\infty} \dot{\delta}_u(v)f(v)dv = -f'(u)$$

where $f'(u)$ is the derivative of $f(u)$ with respect to u. The kernel in (20) can be approximated by a sequence of smooth kernels, and for such sequences the weighted Kolmogorov-Smirnov-type test (14) approximates the asymptotically efficient test (5) based on crude hazards.

3 Illustrations

3.1 Simulation study

Consider a bivariate exponential distribution with density function

$$f(x,y) = \lambda_1\lambda_2 exp(-\lambda_1 x - \lambda_2 y)[1 + \alpha(2exp(-\lambda_1 x) - 1)(2exp(-\lambda_2 y) - 1)]$$

and survival function

$$S(x,y) = exp(-\lambda_1 x - \lambda_2 y)[1 + \alpha(1 - exp(-\lambda_1 x))(1 - exp(-\lambda_2 y))],$$

where $-1 \leq \alpha \leq 1$ and $\lambda_1,\ \lambda_2 > 0$. The survival function of $T = \min(X, Y)$ is

$$S(t) = S(t,t) = exp(-\lambda_1 t - \lambda_2 t)[1 + \alpha(1 - exp(-\lambda_1 t))(1 - exp(-\lambda_2 t))].$$

It is clear that, for $\alpha = 0$, $S(t) = S(t,t) = exp(-\lambda_1 t - \lambda_2 t)$ and corresponds to independence of X and Y.

We fix λ_1 and λ_2 such that $\lambda_1 \neq \lambda_2$ and vary α. Consider the crude hazards

$$a_1(t) = a_1(t,\alpha) = a(t,\lambda_1,\lambda_2,\alpha) = \frac{d\tilde{A}_1(t)}{dt}$$
$$= \frac{\Big(1 + \alpha(1 - exp(-\lambda_1 t))(1 - exp(-\lambda_2 t))\Big)\left(1 - \frac{\alpha e^{-\lambda_1 t}(1-e^{-\lambda_2 t})}{1+\alpha(1-e^{-\lambda_1 t})(1-e^{-\lambda_2 t})}\right)}{\left\{\frac{1+\alpha(1+e^{-(\lambda_1+\lambda_2)t})}{\lambda_1+\lambda_2} - \frac{2\alpha e^{-\lambda_1 t}}{2\lambda_1+\lambda_2} - \frac{\alpha e^{-\lambda_2 t}}{\lambda_1+2\lambda_2}\right\}}$$

and $a_0(t) = a_0(t,\alpha) = a(t,\lambda_2,\lambda_1,\alpha)$ defined analogously by interchanging the role of λ_1 and λ_2. When $\alpha = 0$, $a(t,\lambda_1,\lambda_2,0) = a(t,\lambda_2,\lambda_1,0) = \lambda_1 + \lambda_2$ and a is continuous in its arguments.

The sequence of local alternatives obtained by expanding the crude hazards $a(t,\lambda_1,\lambda_2,\alpha_n)$ and $a(t,\lambda_2,\lambda_1,\alpha_n)$ around the point $\alpha = 0$ is

$$a_1^{(n)}(t) = (\lambda_1+\lambda_2)(1+\alpha_n\gamma_1(t)), \quad a_0^{(n)}(t) = (\lambda_1+\lambda_2)(1+\alpha_n\gamma_2(t))$$

where $\alpha_n = cn^{-1/2}$ and c is a constant such that $-1 \leq \alpha_n \leq 1$ and

$$\begin{aligned}
\gamma_1(t) &= \frac{\partial}{\partial\alpha} a(t,\lambda_1,\lambda_2,\alpha)_{\alpha=0} \\
&= e^{-(\lambda_1+\lambda_2)t} - \frac{2\lambda_1}{2\lambda_1+\lambda_2} e^{-\lambda_1 t} - \frac{\lambda_2}{\lambda_1+2\lambda_2} e^{-\lambda_2 t} \\
\gamma_2(t) &= \frac{\partial}{\partial\alpha} a(t,\lambda_2,\lambda_1,\alpha)_{\alpha=0} \\
&= e^{-(\lambda_1+\lambda_2)t} - \frac{2\lambda_2}{2\lambda_2+\lambda_1} e^{-\lambda_2 t} - \frac{\lambda_1}{\lambda_2+2\lambda_1} e^{-\lambda_1 t}
\end{aligned}$$

Let

$$\gamma(t) = \gamma_1(t) - \gamma_2(t) = \frac{\lambda_2}{\lambda_1+2\lambda_2} e^{-\lambda_2 t} - \frac{\lambda_1}{2\lambda_1+\lambda_2} e^{-\lambda_1 t}.$$

The optimal kernel for testing $a_1(t) = a_0(t)$ is proportional to $S(t)\gamma(t)$.

We consider the optimal weight function $S(t)\gamma(t)$ and also the weight function $S_n(t)^\rho$ with $\rho = 1$. The level of significance used throughout is 0.05. The parameters used for the simulation are $\lambda_1 = 1$, $\lambda_2 = 3$ and $\alpha = 0$ for the null hypothesis. A sample of size 500 was generated with 1000 repetitions. Figure 1 gives the empirical distribution of the standardised test statistic $(n^{-1/2}V_n/\hat{\sigma})$ under the null hypothesis and also when α takes values -0.22, -0.44, -0.67 and -0.89 along with the standard normal distribution. The empirical distribution corresponding to $\alpha = 0$ is quite close to the standard normal distribution and as α goes away from 0, the distributions look like shifted normal and the curves move away from the standard normal distribution as expected.

To compare the two weight functions and the U-statistic U_3 proposed in Dewan *et al.*(2004), empirical distributions of the three statistics are computed using $\alpha = -0.894$. Figure 2 shows these three empirical distributions and also the standard normal distribution. It is clear that the test based on the Harrington and Fleming type weight function with $\rho = 1$ has power similar to the test based on optimal weight function. The noncentrality parameter of the U-statistic is smaller than that of the test based on crude hazards. In practice, when one does not want to make assumptions about the structure of the alternative hypothesis, the Harrington and Fleming type of weight function is a good choice.

3.2 Mortality follow-up study

We analyse the mortality follow-up data from the Finnish cohorts which was a part of the Seven Countries Study in which men in the age-group of 40-59 were examined during 1958-1964 (see Keys *et al.*, 1966 and Karvonen *et al.*, 1970 for the details of the study). There were two Finnish cohorts: one from Ilomantsi in the eastern Finland

and one from Pyty and Mellil in the south-western Finland, consisting mainly of rural agricultural populations. The original cohort consists of 823 men from the eastern Finland and 888 men from the south-western Finland. Here, we analyse 40-years of mortality follow-up data of 1560 men who died during the follow-up. The mortality follow-up data give the date of death and underlying cause of death. A death due to causes cardiovascular diseases, cancer, accidents and suicide, is defined as cause 1 and a death due to any other causes is defined as cause 0. The number of deaths due to cause 1 is 621 and that due to cause 0 is 939. Figure 3 shows the empirical conditional probability functions $\Phi_1(t)$ and $\Phi_0(t)$ and Figure 4 shows the corresponding estimates of the crude hazards. It can be seen from Figure 3 that the probability of dying due to cause 1 given that a person has survived up to certain age is a decreasing function of age, although after the age of 85 there is no clear trend, and hence the probability of dying due to cause 0 is increasing with age. In fact, there are several ages when the rate of change in the Φ function changes. It can be seen that $\Phi_1(t) \leq \Phi_1(0)$. Here the hypothesis of interest is whether $\Phi_1(t)$ is decreasing, *i.e.*, $a_0(t) \leq a_1(t)$ for all t. The value of the test statistic using Harrington and Fleming type of weight function is 5.2411. We reject the null hypothesis at 5% level of significance and hence it may be concluded that the probability of dying due to cause 1 given survival up to a certain age decreases with age.

4 Discussion

It is shown that the most efficient test based on crude hazards (5), is equivalent to the most efficient test in the class of the weighted Kolmogorov-Smirnov-type tests (14), for a specific choice of local alternatives. A simple well-known test for comparing hazards of two counting processes can be efficiently applied in the present situation. This allows a straightforward extension of testing hazards from two samples to k samples, in case of k failure modes. A $k-$sample test for comparing hazards given on pages 345-348 in Andersen *et al.* (1993) can be used in case of $k-$failure modes or competing risks.

It is demonstrated using the simulated data that Harrington and Fleming type of weight function performs satisfactorily when compared to the optimal weight function. In general when the form of the optimal weight function is not known, Harrington and Fleming type of weight function can be used.

It is easy to check that equality of crude hazards in the absence of censoring yields equality of crude hazard in the presence of independent censoring. Hence, in case of right-censored competing risks data with independent censoring, the above methods can be applied without any changes. In this case, the enlarged filtration explicitly includes both $N_1(\infty)$ and $N_2(\infty)$. In other words, the proposed methods can be applied to right-censored competing risks data by dropping the censored observations which do not contain any information about δ and using the uncensored observations only. We refer to Example V.2.1, Chapter V, Andersen *et al.* (1993) for a discussion regarding censored survival data.

Acknowledgements

We would like to thank Professor J. V. Deshpande, Department of Statistics, University of Pune, India for useful discussion and comments. The first author was supported by the Academy of Finland. The research of the second author was supported by the GenomEUtwin Project grant from the European Commission under the programme 'Quality of Life and Management of the Living Resources' of 5th Framework Programme (no. QLG2-CT-2002-01254) and by the Academy of Finland under grant number 53646. The authors would like to thank the reviewers for their helpful comments.

References

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

Breslow,N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Annals of statistics* **2**: 437-453.

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

Harrington, D. P. and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika*, **69**: 133-144.

Karvonen, M. J., Orma, E., Punsar, S. et al. (1970). Coronary heart disease in seven countries. Five-year experience in Finland. *Circulation*, **41** (Suppl. I): I-52-I-62.

Keys A., Aravanis, C., Blackburn, H. W. et al. (1966). Epidemiological studies related to coronary heart disease: characteristics of men aged 40-59 in seven countries. *Acta Med Scand* (Suppl. 460): 1-392.

van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge University Press, Cambridge.

Appendix I: Proof of Theorem 2.2

It is shown in Breslow *et al.* (1974) that as n tends to ∞, $\sqrt{n}(\hat{F}_{1n} - F_1, \hat{F}_n - F)$ converges jointly in $D([0,\infty)) \times D([0,\infty))$ with the Skorokhod topology to zero mean

Gaussian processes (X_1, X) with covariance structure, for $s \leq t$,

$$\begin{aligned} cov(X_1(s), X_1(t)) &= F_1(s)(1 - F_1(t)), \\ cov(X(s), X(t)) &= F(s)(1 - F(t)), \\ cov(X_1(s), X(t)) &= F_1(s)(1 - F(t)), \\ cov(X_1(t), X(s)) &= F_1(s) - F_1(t)F(s). \end{aligned}$$

Note that X is a time changed Brownian bridge that is $X(\infty) = 0$, but X_1 is not a Brownian bridge, as the limit $X_1(\infty)$ is random. The martingale decomposition for (X_1, X) can be written as

$$dX = dM - \frac{X}{(1-F)}dF, \quad dX_1 = dM_1 - \frac{X}{(1-F)}dF_1,$$

where M and M_1 are Gaussian martingales with

$$d\langle M, M\rangle = (1-F)dF \text{ and } d\langle M, M_1\rangle = d\langle M_1, M_1\rangle = (1-F)dF_1.$$

If we denote $X_0 = X - X_1$, $M_0 = M - M_1$, $F_0 = F - F_1$, we get the linear system of stochastic differential equations

$$\begin{aligned} dX_0 &= dM_0 - \frac{(X_0 + X_1)}{(1-F)}dF_0 \\ dX_1 &= dM_1 - \frac{(X_0 + X_1)}{(1-F)}dF_1 \end{aligned}$$

where M_0 and M_1 are orthogonal Gaussian martingales with $d\langle M_i\rangle = (1-F)dF_i$. The solution can be given explicitely in terms of (M_0, M_1) and matrix exponentials. Note that $X_0(\infty) + X_1(\infty) = 0$.

By the functional delta method (see Chapter 20 in van der Vaart, 1998), it can be shown that

$$\sqrt{n}(\hat{\Psi}_n(u,v) - \Psi(u,v)) =$$
$$\sqrt{n}(\hat{F}_{1n}(\infty) - F_1(\infty))(F(v) - F(u)) +$$
$$\hat{F}_{1n}(\infty)\sqrt{n}[(\hat{F}_n(v) - F(v)) - (\hat{F}_n(u) - F(u))] +$$
$$\sqrt{n}(\hat{F}_{1n}(u) - F_1(u))(1 - F(v)) - \hat{F}_{1n}(u)\sqrt{n}(\hat{F}_n(v) - F(v))$$
$$-\sqrt{n}(\hat{F}_{1n}(v) - F_1(v))(1 - F(u)) + \hat{F}_{1n}(v)\sqrt{n}(\hat{F}_n(u) - F(u))$$

converges to

$$Z(u,v) = X_1(\infty)(F(v) - F(u)) + F_1(\infty)(X(v) - X(u)) +$$
$$X_1(u)(1 - F(v)) - F_1(u)X(v) - X_1(v)(1 - F(u)) + F_1(v)X(u).$$

We can express

$$Z(u,v) = \int_0^\infty f(u,v,t)dX_1(t) + \int_0^\infty g(u,v,t)dX(t),$$

so that

$$\mathrm{Cov}(Z(u_1,v_1), Z(u_2,v_2)) =$$
$$\mathrm{Cov}\Big(f(u_1,v_1,\infty)I(\eta=1) + g(u_1,v_1,\infty), f(u_2,v_2,\infty)I(\eta=1) + g(u_2,v_2,\infty)\Big),$$

where

$$\begin{aligned} f(u,v,t) &= (F(v) - F(u)) + I_{[0,u]}(t)(1 - F(v)) - I_{[0,v]}(t)(1 - F(u)), \\ g(u,v,t) &= I_{[0,u]}(t)F_1(v) - I_{[0,v]}(t)F_1(u) + F_1(\infty)I_{[u,v]}(t). \end{aligned}$$

Under H_0, $F_1(t) = F_1(\infty)F(t) = \phi F(t)$ and hence,

$$\begin{aligned} f(u,v,t) &= (F(v) - F(u)) + I_{[0,u]}(t)(1 - F(v)) - I_{[0,v]}(t)(1 - F(u)), \\ g(u,v,t) &= \phi(I_{[0,u]}(t)F(v) - I_{[0,v]}(t)F(u) + I_{[0,v]}(t) - I_{[0,u]}(t)) \\ &= \phi(-I_{[0,u]}(t)(1 - F(v)) + I_{[0,v]}(t)(1 - F(u))) \\ &= \phi(F(v) - F(u) - f(u,v,t)). \end{aligned}$$

Note that

$$\int_0^\infty f(u,v,t)dF_1(t) \;=\; \phi\int_0^\infty f(u,v,t)dF(t) = 0$$

and

$$\int_0^\infty g(u,v,t)dF(t) \;=\; \phi\int_0^\infty (F(v) - F(u) - f(u,v,t))dF(t) = \phi(F(v) - F(u)).$$

It can be verified using simple calculations that

$$\begin{aligned} cov(Z(u_1,v_1), Z(u_2,v_2)) &= 0 \text{ if } u_1 \leq v_1 \leq u_2 \leq v_2 \text{ or } u_2 \leq v_2 \leq u_1 \leq v_1 \\ &= \phi(1-\phi)(1-F(max(v_1,v_2)) \\ &\quad (1-F(min(u_1,u_2))(F(min(v_1,v_2) - F(max(u_1,u_2)), \\ &\qquad\qquad \text{otherwise,} \\ var(Z(u,v)) &= \phi(1-\phi)(1-F(v))(1-F(u))(F(v)-F(u)) \\ &= \phi(1-\phi)(1-t)(1-s)(t-s), \end{aligned}$$

where $F(u) = s$ and $F(v) = t$, and hence $0 \leq s \leq t \leq 1$.
Hence the Theorem 2.2.

Appendix II: Optimal weight function

To find the asymptotic noncentrality parameter under the sequence of local alternatives $\{P^{(n)}(\theta)\}$ given in (7), we need to compute the asymptotic mean and variance of the weighted Kolmogorov-Smirnov-type test (14) for a sequence of possibly random kernels $K_n(u,v)$ approximating $K(u,v) = k(u)\dot{\delta}_u(v)$, so that

$$\sqrt{n}\iint\limits_{u\leq v}(K_n(u,v) - k(u)\dot{\delta}_u(v))\Psi^{(n)}(u,v)dudv =$$

$$\sqrt{n}\Bigg(\iint\limits_{u\leq v} K^n(u,v)\Psi^{(n)}(u,v)dudv - \int\limits_0^\infty k(u)S_1^{(n)}(u)S_0^{(n)}(u)(dA_1^{(n)}(u) - dA_0^{(n)}(u))\Bigg)$$

$$\overset{\mathbf{P}^{(n)}(\theta)}{\rightarrow} 0.$$

Using (17) and (18), it is easy to verify that the asymptotic mean and variance of $\sqrt{n}\hat{\Delta}_n$ is

$$\mu = (\theta_1 - \theta_0)\phi(1-\phi)\int_0^\infty k(u)S^2(u)a(u)\gamma(u)du,$$

and

$$\mathrm{Var}\left(\iint_{u\le v} K^*(u,v)Z(u,v)dudv\right) =$$
$$\iint\iint I(u\le v)I(u'\le v')k(u)\dot{\delta}_u(v)k(u')\dot{\delta}_{u'}(v')\ \mathrm{Cov}\ (Z(u,v),Z(u',v'))du'dv'dudv.$$

Note that

$$-\int \dot{\delta}_{u'}(v')I(u'\le v')R(u,v,u',v')dv' \ = \ \frac{\partial}{\partial v'}I(u'\le v')R(u,v,u',v')|_{v'=u'},$$

where

$$\begin{aligned} R(u,v,u',v') &= cov(Z(u,v),Z(u',v')) \\ &= \phi(1-\phi)S(v')S(min(u,u'))(S(max(u,u'))-S(v))I(v\le v')I(u'\le v) \\ &\quad +\phi(1-\phi)S(v)S(min(u,u'))(S(max(u,u'))-S(v'))I(v'\le v)I(u\le v'). \end{aligned}$$

Finally,

$$\sigma^2 = \mathrm{Var}\left(\iint_{u\le v} K^*(u,v)Z(u,v)dudv\right) \ = \ \phi(1-\phi)\int k^2(u)S^3(u)a(u)du.$$

The value of the noncentrality parameter μ^2/σ^2 is

$$(\theta_1-\theta_0)^2\phi(1-\phi)(\int_0^\infty k(u)S^2(u)a(u)\gamma(u)du)^2(\int_0^\infty k^2(u)S^3(u)a(u)du)^{-1}$$

and is maximised when $k(u) = \gamma(u)S^{-1}(u)$ and the maximum value of the noncentrality parameter is

$$(\theta_1-\theta_0)^2\phi(1-\phi)\int_0^\infty \gamma^2(u)S(u)a(u)du$$

which is identical to the maximum value of the noncentrality parameter (9) of the test (5).

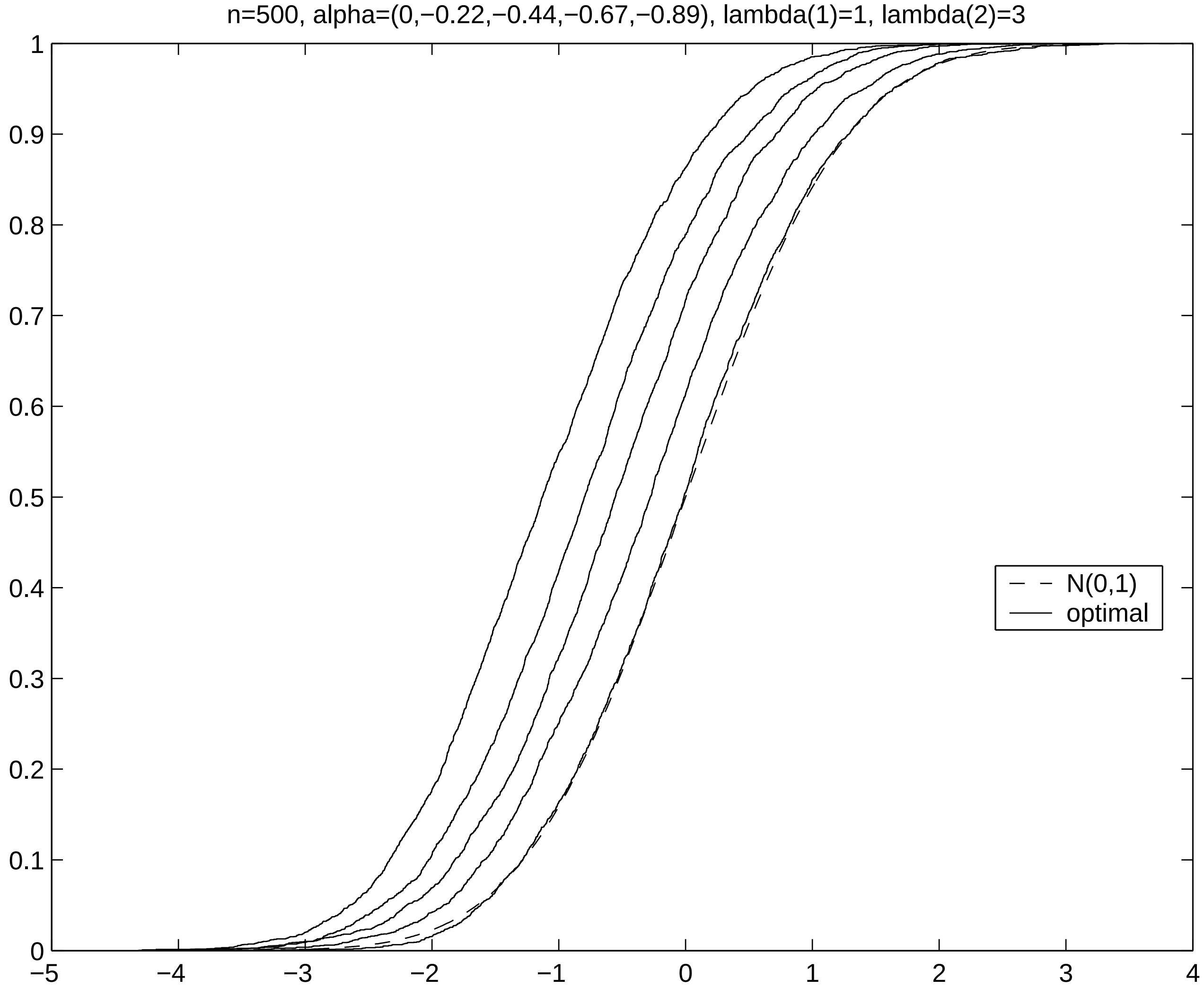

Figure 1: Empirical distributions of the standardised test statistic based on crude hazards for various values of α $n = 500$, $\alpha = (0, -0.22, -0.44, -0.67, -0.89)$, $\lambda_1 = 1, \lambda_2 = 3$

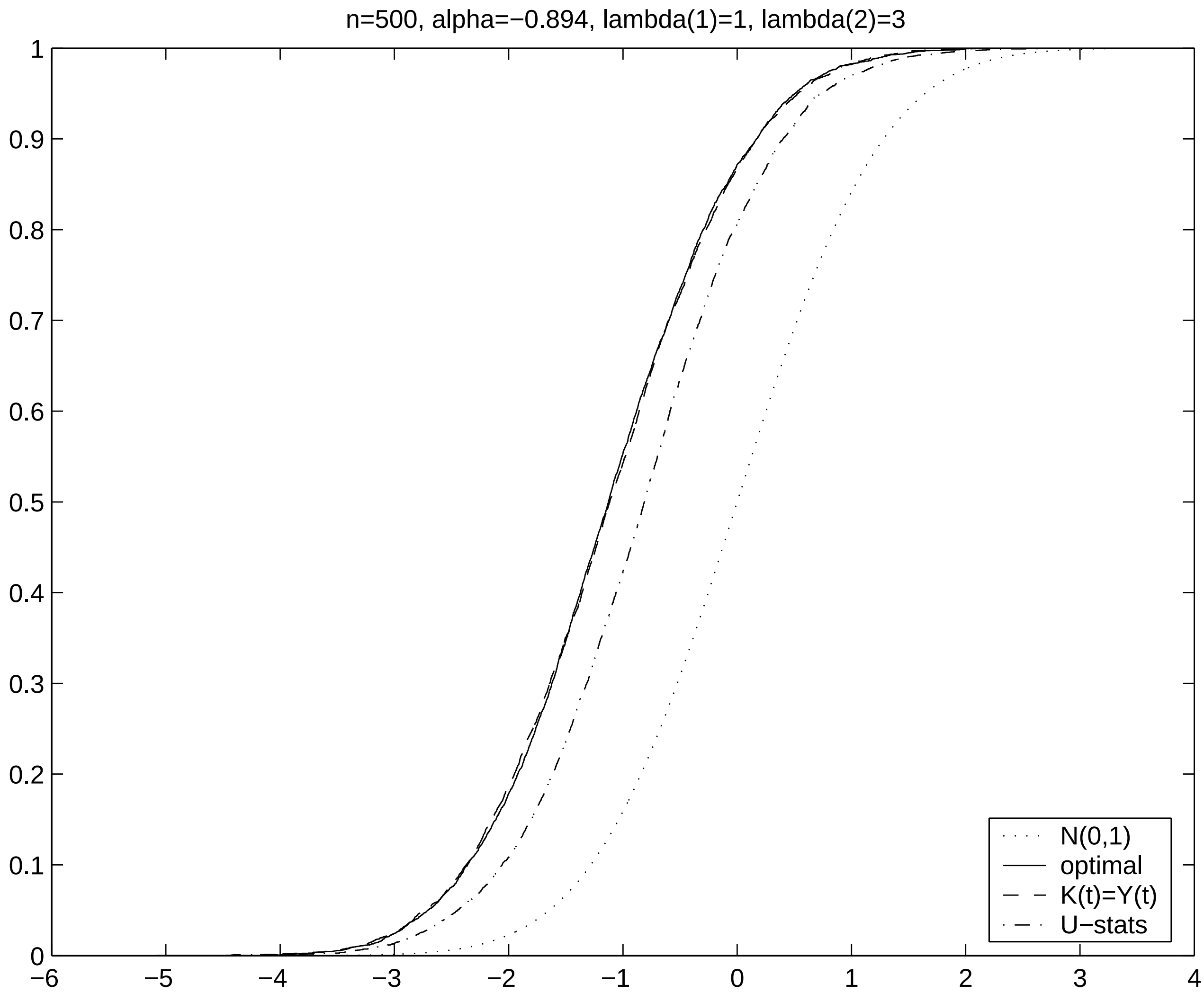

Figure 2: Empirical distributions of test statistics based on crude hazards using optimal kernel and Harrington-Fleming type kernel, and U_3 test $n = 500, \quad \alpha = -0.89, \quad \lambda_1 = 1, \lambda_2 = 3$

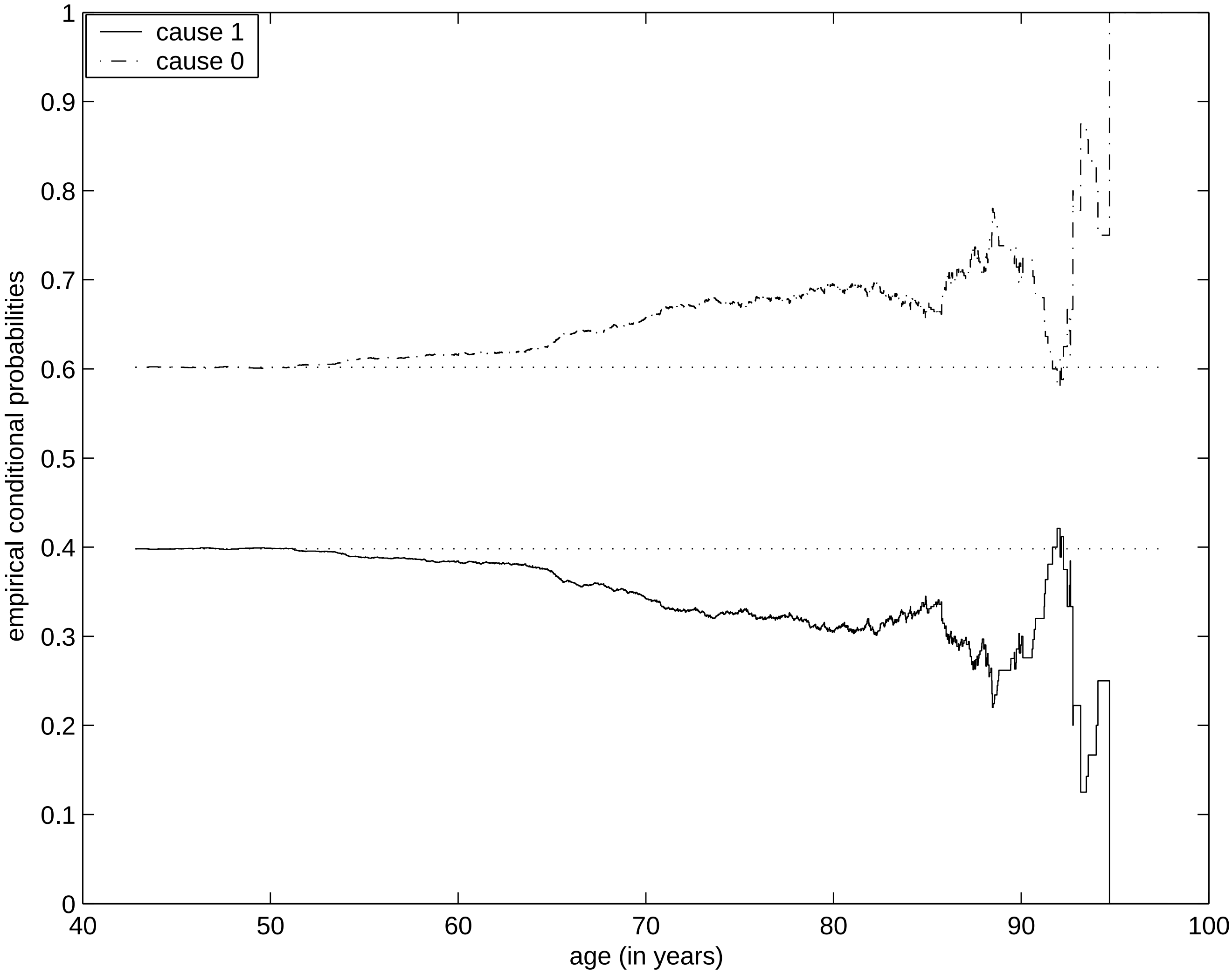

Figure 3: Empirical conditional probabilities for two competing causes of death

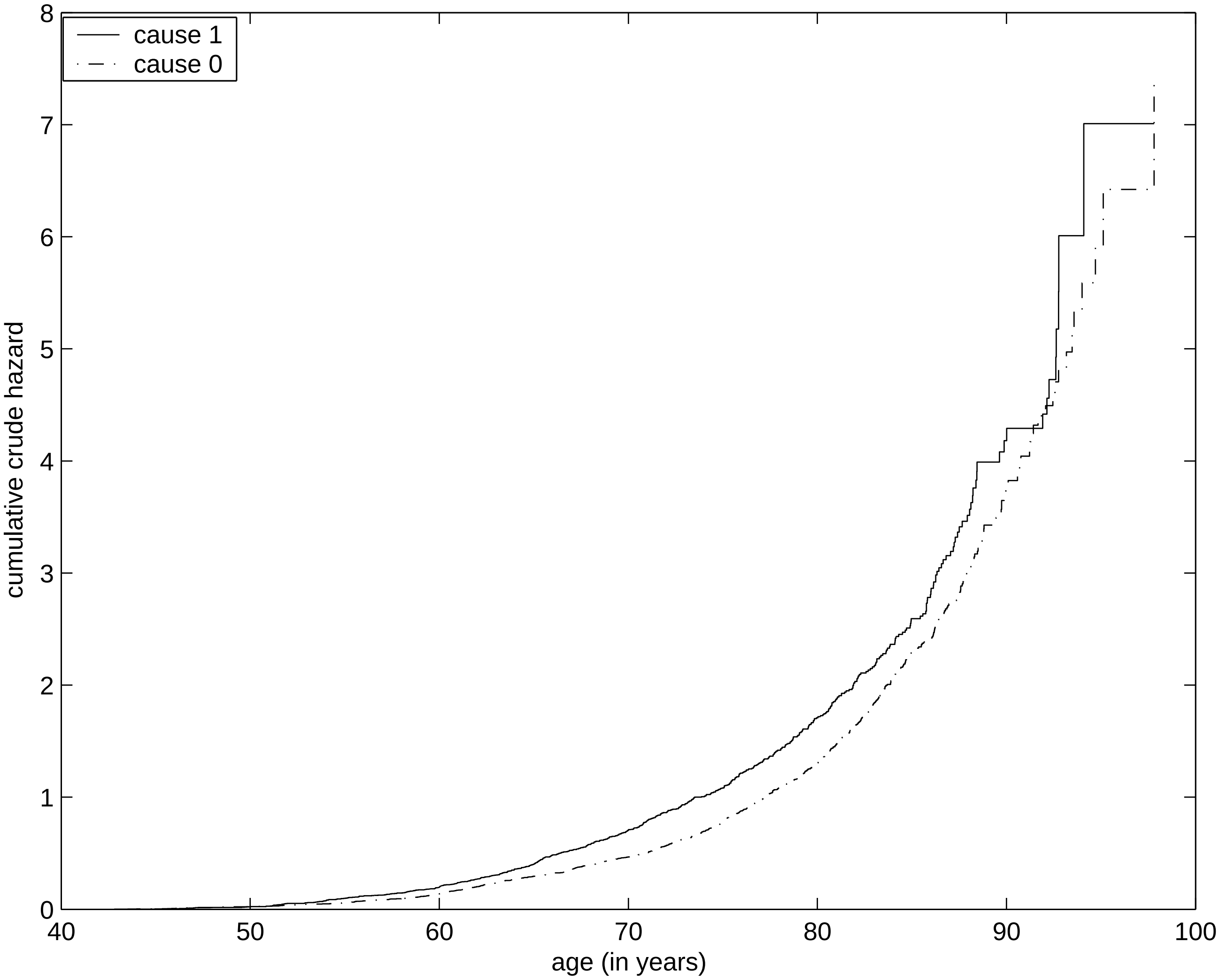

Figure 4: Nelson-Aalen estimates of cumulative crude hazards for two competing causes of death