Competing Risks Models: Problems of Modelling and of Identification

by

Michel Mouchart*, **Jean-Marie Rolin***

April 1995

Abstract

Competing risks models are presented in the framework of single transition-multiple causes models. Particular attention is paid to general distributions of latent durations and to model identification problems, which are different from parameter identification problems. Implications for modelling and interpreting empirical findings are considered. The exposition is fairly elementary and the techniques in use, although not the most usual ones, are deemed to be more powerful and, hopefully, more illuminating than the more traditional ones. In particular, continuous and discrete distributions are treated in a unified framework.

Acknowledgement:

The authors are indebted to G. Wünsch and I. Gijbels for many useful comments on an earlier version of this paper and retain full responsability for any remaining errors. Research support from the contract "Projet d'Actions de Recherche Concertées" (PARC No. 93/98-164) of the Belgian Government is gratefully acknowledged.

* CORE and Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve
Contact address: Voie du Roman Pays 34
B-1348 Louvain-la-Neuve
Belgium
e-mail: MOUCHART@STAT.UCL.AC.BE

1 INTRODUCTION

In this paper we consider problems met in modelling data of the following type. For a given individual, we observe the age at the time of death, say T, and the cause of the death, say A, among a finite set of possible causes, say $E = \{1, 2, \ldots, p\}$. The finite character of A may raise difficulties in applications where, for instance, the number of causes may increase with time without a priori given limits; think for instance to mutating viruses. In such situations, A might be conceptually infinite; this feature would call for modifying the models to be presented. For statistical purposes, models typically consider a finite number of precisely defined causes along with a "residual" cause that gathers all other possible causes; this residual cause is often treated as a censoring state.

Thus the data have the form $(T, A) \in I\!R_+ \times E$ if A is an element of E, i.e. a simple cause is given for each observed individual, or has the form $(T, A) \in I\!R_+ \times 2^E$ if A is a subset of E, which is the case when multiple causes are allowed. Multiple causes arise in situations where several causes may actually act simultaneously or in situations where the available information does not allow one to distinguish one from several possible causes.

The data density, or the likelihood function, of the data (T, A) is based on the specification of

$$P(T > t, A = a) = S(t, a) \tag{1.1}$$

which is called a *sub-distribution* (in the form of a survivor function) because:

$$S(0, a) = P(A = a) < 1 \tag{1.2}$$

whereas the survivor function of T, i.e. $S_T(t)$, would satisfy:

$$S_T(0) = P(T > 0) = 1 \tag{1.3}$$

as soon as $P(T = 0) = 0$, which we always assume from now on. We shall also assume: $P(T < \infty) = 1$. Clearly:

$$S_T(t) = \sum_a S(t, a) \tag{1.4}$$

Furthermore,

$$S(t, a) = P(T > t | A = a)P(A = a) \tag{1.5}$$

which shows that a sub-distribution is the product of a (conditional) distribution and a probability, but most models are based on the dual decomposition, namely:

$$\begin{aligned} S(t, a) &= E[P(A = a|T)\mathbb{1}_{\{T>t\}}] \\ &= \int_t^\infty P(A = a|T = s)dF_T(s) \end{aligned} \tag{1.6}$$

where $F_T(s) = 1 - S_T(s) = P(T \leq s)$ is the distribution function of the observed duration T

Remark

Integrating with respect fo dF_T, the so-called Lebesgue-Stieltjes integral, may be viewed, heuristically, as a short-hand notation encompassing the discrete, the absolutely continuous and the mixed case, in view of the following formula :

$$\int g(t)dF_T(t) = \int g(t)f_T(t)dt + \sum_t g(t)P(T = t) \tag{1.7}$$

where $f_T(t)$ is the (weighted) density of the absolutely continuous part of F_T and the summation is made with respect to the (at most countable) jumps of F_T. ∎

When $S(t,a)$ is differentiable, in t, the data density $L_{T,A}(t,a)$ is obtained as:

$$L_{T,A}(t,a) = -\frac{d}{dt}S(t,a) \tag{1.8}$$

If furthermore the marginal distribution of T admits a density, say $f_T(t)$, i.e.: $dF_T(t) = f_T(t)dt$, we obtain:

$$S(t,a) = \int_t^\infty P(A=a|T=s)f_T(s)ds \tag{1.9}$$

and therefore:

$$L_{T,A}(t,a) = P(A=a|T=t)f_T(t) \tag{1.10}$$

It is accordingly natural that most specifications of the sub-distribution $S(t,a)$, and consequently of the likelihood function, are based on the specification of the probability of the cause conditionally on the observed durations and on the specification of the (marginal) law of the observed duration.

Competing risks models is a class of models originally developped for the case of a single observed cause i.e. $a \in E$. Formal generalizations may be obtained by redefining the state space as the power set of E but raise problems of specification out of the scope of this presentation.

The object of this paper is to give a brief and simple survey of some problems raising when modelling competing risks models. A particular

attention is paid to accommodate for general distributions rather than to stick to the absolutely continuous case. The rationale is twofold. Firstly we believe that (well chosen) general tools may be simpler to use than tools specific to the absolutely continuous case. Secondly, the upraise of semi-parametric and non-parametric models lead to construct procedures based, at the start, on the empirical process which is discrete in nature even when the theoretical distribution is assumedly continuous. Furthermore, data are typically observable on a discrete scale only and model-builders may want to take this feature into account explicitly, particularly when ties are apparent in the data. Bridging discrete and continuous distributions in a systematic way becomes, in these cases, of primary interest. A particular attention is also paid to identification problems, a crucial issue when modelling. Here also we endeavour to exposit the problem in a general framework, not only for the sake of generality but also for the simplicity of the basic argument. Thus the originality, if any, of this paper relies more on the mode of exposition than on the final results.

In the next section, we exposit the competing risks model in a form general enough for accommodating both discrete and continuous distributions, although earlier expositions considered only continuous distributions. In section 3 we consider a particular case where the competing risks are independent in probability. In section 4 we analyze identification problems. The main one comes from the following fact. In a competing risks model one is assumed to observe the minimum of a random vector along with the coordinate where the minimum is obtained. This feature impedes to distinguish cases where the coordinates are independent in probability or not.

2 Competing Risks Models

2.1 The general case

The basic idea of a competing risks model is to associate to each possible cause, j, a latent duration, τ_j, representing the date (assuming the process starts at time 0) of an event the realization of which determines the individual to die of cause j; thus the date of death is

interpreted as the date of the event first realized, more explicitly, T is modelled as:

$$T = \min_{1 \leq j \leq p} \tau_j \tag{2.1}$$

The observed cause of death, A, is defined as:

$$A = \min \{j | \tau_j = T\} \tag{2.2}$$

for the cases where ties among the $\tau_j's$ have (strictly) positive probabilities ($P(\tau_j = \tau_k) > 0$). This requires an implicit ordering of the states such that in case of ties for the minimum, i.e. $\tau_j = \tau_k < \tau_i \quad \forall i \notin \{j, k\}$, the observed cause corresponds to the one with the lowest label. When no tie is possible, i.e. $P(\tau_j = \tau_k) = 0 \quad \forall j \neq k$, A is also equal to:

$$A = \arg \min_{1 \leq j \leq p} \tau_j \tag{2.3}$$

Remark

When ties may occur with strictly positive probability but an adequate ordering of the causes does not seem possible, the model may be reformulated in terms of multiple causes, i.e. 2^E is taken as the set of causes instead of E; the reformulated model becomes accordingly a model without ties.

In a competing risks model, the sub-distribution $S(t, a)$ is completely determined by the joint survivor function of the latent durations, i.e. by:

$$S_*(t_1, t_2, \ldots t_p) = P(\tau_1 > t_1, \ldots, \tau_p > t_p) \tag{2.4}$$

more precisely, under a competing risks model, (1.4) becomes:

$$S_T(t) = S_*(t, t, \ldots t) \tag{2.5}$$

and (1.6) may be rewritten as follows:

$$\begin{aligned} S(t,a) &= P(T > t, A = a) \\ &= P(\tau_a > t, A = a) \\ &= E[P(A = a|\tau_a)\mathbb{1}_{\{\tau_a > t\}}] \\ &= \int_t^\infty P(A = a|\tau_a = s)dF_a(s) \end{aligned} \tag{2.6}$$

where $F_a(s)$ is the marginal distribution function of the latent duration τ_a, i.e.

$$F_a(s) = P(\tau_a \leq s) = 1 - S_*(0, \ldots 0, s, 0, \ldots 0) \tag{2.7}$$

Furthermore:

$$\begin{aligned} P(A &= a|\tau_a = s) \\ &= P[\cap_{1\leq i\leq a-1}\{\tau_i > \tau_a\} \cap \cap_{a+1\leq i\leq p}\{\tau_i \geq \tau_a\}|\tau_a = s] \\ &= P[\cap_{1\leq i\leq a-1}\{\tau_i > s\} \cap \cap_{a+1\leq i\leq p}\{\tau_i \geq s\}|\tau_a = s] \end{aligned} \tag{2.8}$$

Formulae (2.5) to (2.8) show how the data density is determined by the joint distribution of the latent durations.

2.2 The absolutely continuous case

Suppose now that $S_*(t_1 \ldots t_p)$ is continuously differentiable. This implies that the probability of ties in the latent durations is zero and that (2.8) may be rewritten as:

$$\begin{aligned} P(A=a|\tau_a=s) &= P[\cap_{i\neq a}\{\tau_i > s\}|\tau_a = s] \\ &= -\frac{(D_a S_*)(s,s,\ldots,s)}{f_a(s)} \end{aligned} \tag{2.9}$$

where $D_a S_*$ denotes the a-th partial derivative of S_*; the relationship (2.9) may indeed be obtained from the identity:

$$S_*(t,t,\ldots t) = \int_t^\infty P[\cap_{i\neq a}\{\tau_i > t\}|\tau_a = s] f_a(s) ds \tag{2.10}$$

where $f_a(s)$ is the marginal density of τ_a. Thus the data density (1.9) may be written as:

$$L_{T,A}(t,a) = -\frac{d}{dt} S(t,a) = -(D_a S_*)(t,t,\ldots t) \tag{2.11}$$

In other words, the data density is the partial derivative, corresponding to the cause of death, of the joint survivor function of the latent duration evaluated on the main diagonal at the observed duration $\{\tau_1 = \tau_2 = \ldots \tau_p = t\}$. It may be useful to rewrite that data density in terms of the conditional hazard function h_a^c defined as follows:

$$h_a^c(t) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P[t \leq \tau_a \leq t+\delta | T > t] \tag{2.12}$$

This is the hazard function of the a-th latent duration conditional on the observed duration being greater than t; this conditional hazard function may be evaluated as:

$$h_a^c(t) = -(D_a \; ln \; S_*)(t,t,\ldots t) \tag{2.13}$$

Using (2.5) and (2.12), the data density may also be written as:

$$L_{T,A}(t,a) = h_a^c(t) S_T(t) \tag{2.14}$$

The hazard function of the observed duration, h_T, is defined as:

$$h_T(t) = \frac{f_T(t)}{S_T(t)} = -\frac{d}{dt}\ ln\ S_T(t) \tag{2.15}$$

and may be evaluated as:

$$h_T(t) = \sum_a h_a^c(t) \tag{2.16}$$

which shows that in the smooth case the data density of the competing risks model rests on the specification of the conditional hazard functions h_a^c.

Remark

Let us define $m = \inf\{t | S_T(t) = 0\}$. When $m < \infty, S_T(m) = S_*(m, m, \ldots, m) = 0$ and formulae (2.12) to (2.16) are valid for $t < m$ but undefined and irrelevant for $t \geq m$. This situation arises, for instance, when one of the latent durations is bounded in probability, i.e. $S_j(m) = 0$, for some $m < \infty$, where S_j is the j-th marginal survivor function of S_*.

3 Independent Risks

3.1 The general case

That the latent durations τ_a, are independent in probability means that the joint survivor function is factorizable:

$$S_T(t_1 \dots t_p) = \prod_{1 \leq a \leq p} S_a(t_a) \tag{3.1}$$

where:

$$S_a(t_a) = P(\tau_a > t_a) = 1 - F_a(t_a) \tag{3.2}$$

This implies that:

$$S_T(t) = \prod_{1 \leq a \leq p} S_a(t) \tag{3.3}$$

In such a case, (2.8) may be simplified into:

$$P(A = a | \tau_a = s) = \prod_{1 \leq i \leq a-1} S_i(s). \prod_{a+1 \leq i \leq p} S_i(s-) \tag{3.4}$$

where

$$S_i(s-) = P(\tau_i \geq s) \tag{3.5}$$

As in section 2.1, writing a data density, or a likelihood function, requires either more assumptions or rather advanced mathematics. In the next two subsections we examine two particular cases that cover most useful applications and allow one to obtain the data density without too advanced mathematics. These particular cases aim at avoiding too complicated structures of ties among the latent durations.

3.2 The continuous case

Let us first suppose that the marginal survivor functions are continuous, i.e. $S_a(s) = S_a(s-) \quad \forall a \in A, \forall s \in I\!R_+$. This means that no value of s receives a strictly positive probability. Now (2.8) may be written as follows:

$$\begin{aligned} P(A = a|\tau_a = s) &= \prod_{i \neq a} S_i(s) \\ &= \frac{1}{S_a(s)} S_T(s) \end{aligned} \tag{3.6}$$

The sub-distribution (2.6) is accordingly rewritten as:

$$S(t, a) = \int_t^\infty S_T(s) dH_a(s) \tag{3.7}$$

where $H_a(s)$ is a first form of an integrated hazard function:

$$H_a(s) = -\ ln\ S_a(s) \tag{3.8}$$

Thus H_a is a non-decreasing function such that $H_a(0) = 0$ and may accordingly be viewed as the cumulative function of a σ-finite measure on $I\!R_+$; H_a may therefore be used for integration purposes exactly as F_T in (1.7). Note also that the same form applied to the observed duration may be written as:

$$H_T(s) = -ln\ S_T(s) = \sum_{1 \leq a \leq p} H_a(s) \tag{3.9}$$

3.3 The absolutely continuous case

In order to obtain explicitly the data density we now assume that the marginal distributions of the latent durations admit a density $f_a(t)$ say, and that $S_*(t, t, \ldots, t)$ is continuously differentiable or, at least, everywhere differentiable. In such a case, the marginal hazard functions

$$h_a(t) = -\frac{d}{dt}\ ln\ S_a(t) = \frac{1}{S_a(t)} f_a(t) \tag{3.10}$$

are equal to the conditional ones:

$$h_a(t) = h_a^c(t) \tag{3.11}$$

and the data density becomes:

$$\begin{aligned} L_{T,A}(t,a) &= h_a(t) S_T(t) \\ &= f_a(t) P(A = a | \tau_a = t) \end{aligned} \tag{3.12}$$

3.4 The case of no common discontinuities

As mentionned in the introduction, it may be interesting to introduce (mixed or) discrete distributions for the latent durations relative to one or to all causes. In the framework of independent risks models, the general formulae may become untractable if ties among latent durations are allowed with strictly positive probabilities but remain tractable if no such ties may happen. This is the topic of this subsection.

Thus, let us assume:

$$P(\tau_i = \tau_j) = 0 \quad \forall i \neq j \tag{3.13}$$

This means, in case of independence, that no pair of marginal distributions have common values with strictly positive probability. Using (3.4) we write the sub-distribution (2.6) as:

$$S(t,a) = \int_{(t,\infty)} \prod_{i<a} S_i(s) \prod_{i>a} S_i(s-) dF_a(s) \tag{3.14}$$

Note that when $F_a(s) - F_a(s-) = P(\tau_a = s) > 0$, one has that $S_i(s) = S_i(s-) \quad \forall i \neq a$ because these are no common discontinuit and when $P(\tau_a = s) = 0$, the jumps of some S_i at s are inconsequential for the integral. The integral in (3.14) is therefore unaffected when replacing $S_i(s)$ by $S_i(s-)$ (and vice-versa); by so-doing we obtain:

$$S(t,a) = \int_{(t,\infty)} \prod_{i\neq a} S_i(s-) dF_a(s) \tag{3.15}$$

This formula may also be written as:

$$S(t,a) = \int_{(t,\infty)} S_T(s-) dL_a(s) \tag{3.16}$$

where $L_a(t)$ is another form of an integrated hazard function, the so-called "predictable form" or the "compensator form" in reference to the Doob-Meyer decomposition of the underlying process, defined as:

$$L_a(t) = \int_{[0,t]} [S_a(s-)]^{-1} dF_a(s) \tag{3.17}$$

Similarly to $H_a(t), L_a(t)$ is non-decreasing such that $L_a(0) = 0$ and may therefore also be used for integration purposes as in (1.7). More specifically, L_a and F_a have the same support and their respective differentials are characterized by:

$$dF_a(s) = S_a(s-)dL_a(s) \tag{3.18}$$

which justifies (3.16). Some care has to be taken when this support is bounded.

In general (i.e. without conditions of continuity or of no common discontinuities), one has:

$$L_a(t) \leq H_a(t) \qquad \forall t \in I\!R_+ \tag{3.19}$$

with equality in the continuous case (i.e. when $P(\tau_a = t) = 0 \quad \forall t$). Furthermore they have the same continuous part and have their jumps at the same location as those of S_a (or of F_a). The heights of the jumps are different but S_a, H_a and L_a are all bijective transformations of each other.

In terms of the distribution of the observed duration, T, one also defines:

$$L_T(t) = \int_{[0,t]} [S_T(s-)]^{-1} dF_T(s) \tag{3.20}$$

From (3.16) it may be checked that when there are no common discontinuities:

$$L_T(t) = \sum_{a \in A} L_a(t) \tag{3.21}$$

4 Identification Problems

4.1 General results

The competing risks model may be viewed as follows. Considering a random vector of latent durations $(\tau_1 \dots \tau_p)$, only the minimum element, $T = \min\ \tau_j$, and the first coordinate where the minimum is attained, $A = \min\{j|\tau_j = T\}$, are observed. It is therefore to be expected that i.i.d. observations of such a process will not tell us whether the coordinates τ_j are independent in probability or not, even asymptotically. In this section, we show that this intuition is correct by giving a precise theorem and we draw some important conclusions for model specification.

We first notice that observing only the minimum of the latent durations (and its coordinate) implies that if one of the marginal distribution is truncated at, say, m then modifying the S_* at the right of m for the other coordinates will not modify the associated sub-distributions, and consequently its data density. More precisely we have the following theorem:

Theorem 4.1.

Let $S_*(t_1, \dots, t_p)$ and $S_{**}(t_1, \dots, t_p)$ be two survivor functions and $S_*(t, a)$ and $S_{**}(t, a)$ be their associated sub-distributions of the corresponding competing risks models. Let also $m_* = \inf\{t|S_{*T}(t) = 0\}$ where $S_{*T}(t) = \sum_a S_*(t, a)$. If :

$$m_* < \infty$$

$$S_*(t_1, \dots, t_p) = S_{**}(t_1, \dots t_p) \qquad \forall t_i \leq m_* \quad \forall i = 1, \dots p$$

then:

$$S_*(t, a) = S_{**}(t, a) \qquad \blacksquare$$

The proof is straightforward once it is noticed that $S_{*T}(m_*) = 0$ implies $S_*(t, a) = 0 \quad \forall t > m_*$ and $\forall a$ and therefore the interval of integration, in (2.6) becomes (t, m_*) instead of (t, ∞).

Note also that this theorem is also valid in the independent case. This theorem is a theorem of model identification rather than parameter identification: two models with different tails at the right of m_* produce a same data density and therefore, even asymptotically, no data will distinguish them.

The main identification problem concerns however the independence of the latent durations. More specifically, we have the following result.

Theorem 4.2.

If the sub-distributions $\{S(t,a) : 1 \leq a \leq p\}$ *have no common discontinuities, then there exists a distribution of independent latent variables* $S^I_*(t_1 \dots t_p) = \prod_a S^I_a(t_a)$ *such that:*

$$S(t,a) = S^I(t,a) \quad \forall a, \forall t \tag{4.1}$$

where $S^I(t,a)$ *is the sub-distribution associated to* $S^I_*(t_1, \dots, t_p)$.

Furthermore, if $m = \inf\{t | S_T(t) = 0\}$ *is infinite,* S^I_* *is unique.* ∎

Proof

We prove the theorem by constructing an independent survivor function S^I_* which satifies (4.1). Note first that (4.1) implies:

$$S_T(t) = S^I_T(t) = \sum_a S(t,a) \tag{4.2}$$

Therefore, S^I_* must imply an integrated hazard function L^I_a such that:

$$S(t,a) = \int_{(t,\infty)} S_T(s-) \quad dL^I_a(s) = S^I(t,a) \tag{4.3}$$

which implies:

$$dL_a^I(t) = -[S_T(t-)]^{-1} dS(t,a) \qquad \forall t \in [0,m] \tag{4.4}$$

or, by (2.6):

$$\begin{aligned} L_a^I(t) &= -\int_{[0,t]} [S_T(s-)]^{-1} dS(s,a) \qquad \forall t \in [0,m] \qquad (4.5) \\ &= \int_{[0,t]} [S_T(s-)]^{-1} P[A=a|\tau_a = s] dF_a(s) \qquad \forall t \in [0,m] \end{aligned}$$

Note that the absence of common discontinuities in the sub-distributions is a weaker condition than the absence of common discontinuities in the marginal distributions. But even this weaker condition is not necessary as is illustrated in an example displayed in section 4.3.

The practical importance of this theorem is as follows. To each competing risks model with dependent risks one may associate another model with independent risks which is observationally equivalent: data will never be able to distinguish the two models. In other words, the choice among those two models cannot be based on the data. Furthermore, the meaning of the parameters will crucially depend on which model is chosen. In particular, parameters measuring the association among latent durations in the first models will have a different meaning in the second model.

4.2 The absolutely continuous case

When S_*, the joint survivor function of the latent durations, is continuously partially differentiable, formulae (2.12) and (3.10) show that the associated independent risks model is constructed as

$$h_a^I(t) = h_a^c(t) \tag{4.6}$$

This result may indeed be obtained from (4.5) as follows. From (2.5), (2.9), we obtain

$$h_a^c(t) = \frac{P(A = a|\tau_a = t)f_a(t)}{S_T(t)}; \tag{4.7}$$

furthermore $dF_a(s) = f_a(s)ds$ in the absolutely continuous case. Therefore (4.5) may be written as:

$$L_a^I(t) = \int_{[0,t]} h_a^c(s)ds \tag{4.8}$$

As $L_a^I(t) = H_a^I(t)$, we conclude (4.6) which means that the hazard functions of the marginal distributions of the equivalent model are equal to the conditional hazard functions of the non independent model.

4.3 Some examples

In this section, some examples illustrate the construction of an associated model with independent latent durations but with same subdistributions, under various degrees of complexity. The first one is a case where the distribution of the latent duration, S_*, is continuously differentiable and is treated by the (instantaneous) hazard functions whereas the other examples involves partially differentiable S_* with discountinuities in its density; the second example illustrates a structure rather general for being distribution-free whereas the last example is specific to a particular distribution of latent durations.

These examples illustrate not only some techniques of construction of associated models but also the change in the meaning of the parameters according to whether they appear in the original or in the associated model.

4.3.1 A continuously differentiable example

Let us consider the following joint survivor function of the latent durations ($p = 2$)

$$S_*(t_1, t_2) = exp[1 - \alpha_1 t_1 - \alpha_2 t_2 - exp[\alpha_{12}(\alpha_1 t_1 + \alpha_2 t_2)]]$$

where $\alpha_1 > 0, \alpha_2 > 0$ and $\alpha_{12} > -1$. Let us remark that $S_*(t_1, t_2)$ is continuously differentiable and that τ_1 and τ_2 are not independent unless $\alpha_{12} = 0$. From (2.5) (2.13) and (2.15) we obtain for $a = 1, 2$.

$$\begin{array}{ll} S_T(t) & = exp[1 - (\alpha_1 + \alpha_2)t - exp[\alpha_{12}(\alpha_1 + \alpha_2)t]] \\ h_a^c(t) & = \alpha_a\{1 + \alpha_{12} exp[\alpha_{12}(\alpha_1 + \alpha_2)t]\} \\ h_T(t) & = (\alpha_1 + \alpha_2)\{1 + \alpha_{12} exp[\alpha_{12}(\alpha_1 + \alpha_2)t]\} \end{array}$$

Note also that:

$$S_a(t) = exp[1 - \alpha_a t - exp[\alpha_{12}\alpha_a t]$$

The equivalent independent risks model is defined by

$$h_a^I(t) = h_a^c(t) = \alpha_a\{1 + \alpha_{12} exp[\alpha_{12}(\alpha_1 + \alpha_2)t]\}$$

By formula (3.10) we obtain:

$$S_a^I(t) = exp[\frac{\alpha_a}{\alpha_1 + \alpha_2} - \alpha_a t - \frac{\alpha_a}{\alpha_1 + \alpha_2} exp[\alpha_{12}(\alpha_1 + \alpha_2)t]]$$

and therefore,

$$\begin{array}{ll} S_*^I(t_1, t_2) & = S_1^I(t_1) S_2^I(t_2) \\ & = exp[1 - \alpha_1 t_1 - \alpha_2 t_2 - \frac{\alpha_1}{\alpha_1 + \alpha_2} exp[\alpha_{12}(\alpha_1 + \alpha_2)t_1] \\ & \quad - \frac{\alpha_2}{\alpha_1 + \alpha_2} exp[\alpha_{12}(\alpha_1 + \alpha_2)t_2]] \end{array}$$

■

Note that in $S_*(t_1, t_2)$, the parameter α_{12} measures how closely the latent durations τ_1 and τ_2 are associated whereas in S_*^I, the parameter α_{12} is common to the two marginal distributions of the independent latent durations τ_1 and τ_2.

4.3.2 An absolutely continuous but not differentiable example

Johnson and Kotz (1972, p265) attribute the following example to Freund. Let the latent duration distribution be

$$\begin{aligned} S_*(t_1, t_2) &= \{1 + \alpha_1(t_1 - t_2)\}e^{-(\alpha_1+\alpha_2)t_1}\mathbb{1}\{t_1 > t_2\} \\ &\quad +\{1 + \alpha_2(t_2 - t_1)\}e^{-(\alpha_1+\alpha_2)t_2}\mathbb{1}\{t_1 < t_2\} \end{aligned}$$

This survivor function is differentiable outside the main diagonal, i.e.:

$$\begin{aligned} D_1D_2S_*(t_1, t_2) &= f(t_1, t_2) \\ &= \alpha_1(\alpha_1 + \alpha_2)e^{-(\alpha_1+\alpha_2)t_1}\mathbb{1}\{t_1 > t_2\} \\ &\quad +\alpha_2(\alpha_1 + \alpha_2)e^{-(\alpha_1+\alpha_2)t_2}\mathbb{1}\{t_1 < t_2\} \end{aligned}$$

and it may be shown that $f(t_1, t_2)$ is a density, even if it is not continuous on the main diagonal. Therefore in this model;

$$P(\tau_1 = \tau_2) = 0$$

We deduce from this that for $a = 1, 2$

$$\begin{aligned} S_T(t) &= e^{-(\alpha_1+\alpha_2)t} \\ S_a(t) &= (1 + \alpha_a t)e^{-(\alpha_1+\alpha_2)t} \end{aligned}$$

By integration of the density, we obtain:

$$S(t,a) = (1 - \frac{\alpha_a}{\alpha_1 + \alpha_2})e^{-(\alpha_1+\alpha_2)t}$$

Using formulae (4.4) and (3.19), we obtain the equivalent independent risks model as follows:

$$\begin{aligned} dH_a^I(t) &= dL_a^I(t) = -S_T(t)^{-1}dS(t,a) \\ &= (\alpha_1 + \alpha_2 - \alpha_a)dt \end{aligned}$$

Therefore:

$$\begin{aligned} S_1^I(t) &= e^{-\alpha_2 t} \\ S_2^I(t) &= e^{-\alpha_1 t} \end{aligned}$$

■

4.3.3 A distribution-free example with positive probability of ties

The following example illustrates the construction of an associated model with independent latent durations but same sub-distributions and shows that the conditions of no common discontinuities is sufficient but not neccessary for obtaining reasonably tractable models.

Let (X_1, X_2, X_3) be three independent non negative random variables with arbitrary distributions. Consider the following two competing risks models:

model 1

$$\begin{aligned} \tau_1 &= X_1 \wedge X_3 \\ \tau_2 &= X_2 \wedge X_3 \\ T &= \tau_1 \wedge \tau_2 \qquad A = \min\{j | T = \tau_j\} \end{aligned}$$

model 2

$$\begin{array}{lll} \tau_1^* & = X_1 \wedge X_3 & \\ \tau_2^* = & = X_2 & \\ T^* & = \tau_1^* \wedge \tau_2^* \qquad & A^* = \min\{j | T^* = \tau_j^*\} \end{array}$$

where, for instance, $X_1 \wedge X_3$ denotes $\min\{X_1, X_3\}$. Clearly $\tau_1 \not\parallel \tau_2$ whereas $\tau_1^* \parallel \tau_2^*$. Let us show that the associated sub-distributions are identical.

$$\begin{array}{lll} \text{(i)} & \{T > t, A = 1\} & = \{X_1 \wedge X_3 > t, X_1 \wedge X_3 \leq X_2 \wedge X_3\} \\ & & = \{X_1 \wedge X_3 > t, X_1 \wedge X_3 \leq X_2\} \\ & & = \{\tau_1^* > t, \tau_1^* \leq \tau_2^*\} \\ & & = \{T^* > t, A^* = 1\} \end{array}$$

$$\begin{array}{lll} \text{(ii)} & \{T > t, A = 2\} & = \{X_2 \wedge X_3 > t, X_1 \wedge X_3 > X_2 \wedge X_3\} \\ & & = \{X_2 \wedge X_3 > t, X_2 < X_1 \wedge X_3\} \\ & & = \{X_2 > t, X_2 < X_1 \wedge X_3\} \\ & & = \{\tau_2^* > t, \tau_2^* < \tau_1^*\} \\ & & = \{T^* > t, A^* = 2\} \end{array}$$

This example has been introduced by Marshall Olkin (1967) for the particular case where the X_i are independently distributed as an exponential variate with parameter λ_j. In that particular case,

$$S_*(t_1, t_2) = exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_3 \max(t_1, t_2)\}$$

and therefore:

$$S_T(t) = exp(-\lambda_+ t)$$

where $\lambda_+ = \lambda_1 + \lambda_2 + \lambda_3$

Note that S_* is continuous but not continuously differentiable (on the main diagonal); in particular:

$$P(\tau_1 = \tau_2) = \lambda_3 \lambda_+^{-1}$$

which is strictly positive unless $\lambda_3 = 0$; also τ_1 and τ_2 are not independent unless $\lambda_3 = 0$. The sub-distributions are obtained as follows:

$$\begin{aligned} S(t,a) &= (\lambda_1 + \lambda_3)\lambda_+^{-1} e^{-\lambda_+ t} \quad a = 1 \\ &= \lambda_2 \lambda_+^{-1} e^{-\lambda_+ t} \qquad\qquad a = 2 \end{aligned}$$

We obtain the compensator from (4.5):

$$\begin{aligned} L_a^I(t^*) &= (\lambda_1 + \lambda_3)t^* \quad a = 1 \\ &= \lambda_2 t^* \qquad\qquad a = 2 \end{aligned}$$

and derive the survivor functions of the independent equivalent model:

$$\begin{aligned} S_a^I(t^*) &= e^{-(\lambda_1+\lambda_3)t^*} \quad a = 1 \\ &= e^{-\lambda_2 t^*} \qquad\qquad a = 2 \end{aligned}$$

Thus:

$$S_*^I(t_1^*, t_2^*) = e^{-(\lambda_1+\lambda_2)t_1^* - \lambda_2 t_2^*}$$

Clearly $S_T^I(t) = S_*^I(t,t) = S_T(t) = e^{-\lambda_+ t}$

Note also that:

$$P(A=a|\tau_a=t) = e^{-\lambda_2 t} \qquad a=1$$
$$= \frac{\lambda_2}{\lambda_2+\lambda_3} e^{-\lambda_1 t} \quad a=2$$

In the particular case of exponential distributions, T and A are independent. Indeed:

$$P(A=a|T>t) = \frac{\lambda_1+\lambda_3}{\lambda_+} \quad a=1$$
$$= \frac{\lambda_2}{\lambda_+} \qquad a=2$$

5 Implications for modeling strategies

Let us conclude this survey by some remarks on research strategies.

Firstly, modeling single-transition-multiple exits (point) processes is a delicate matter. The general (contemporary) tool is provided by the sub-distributions and competing risks models is one possible device for specifying sub-distributions. Such a strategy is a flexible one allowing, in particular, to handle cases where data are given in the form of several causes (because of simultaneity of actions or because of incomplete observability of the actual cause). This strategy involves however model identification problems, the model builder should be aware of.

A first, but probably minor, problem is the unidentifiability of the tail of the latent distributions once one of its marginals is bounded in probability. This identification problem is relevant in a non-parametric framework only.

The second, much more substantial, problem is due to the fact of observing only the minimum of the latent durations. It implies that no data will allow one to distinguish between a model with independent latent durations and a model with dependent latent durations. In

other words, the choice between two such models cannot be monitored from the data, but only from a prior knowledge provided by existing theories and experience of the field. As the examples show, however, the meaning of the parameter crucially depends on the model. This is a clear example of the difference between model identification and parameter identification. This is also an example of a situation where a parameter has a meaning relative that a particular model rather than a meaning independent of the experimental context.

REFERENCES

CROWDER, M., 1994, Identifiability crises in Competing Risks, *International Statistical Review*,**62**, 379-391.

DAVID, H.A. and MOESCHBERGER, M.L., 1978, *The Theory of Competing Risks*; London : Griffin.

HECKMAN, J.J. and HONORE, B.E., 1989, The identifiability of the Competing risks models, *Biometrika*,**76**, 325-330.

JOHNSON, N.L. and KOTZ, S., 1972, *Continuous Multivariate Distributions*, New York : Wiley.

KALBFLEISCH, J.D. and PRENTICE, R.L., 1980, *The Statistical Analysis of Failure Time Data*, New York : Wiley.

MARSHALL, A.W. and OLKIN, I., 1967, A multivariate exponential distribution, *Journal of the American Statistical Association*, **62**, 30-44.

MILLER, D.R., 1977, A note on independence of multivariate lifetimes in Competing risks models, *The Annals of Statistics*, **5**, 576-579.