Survival Data with Explanatory Processes: A Full Nonparametric Bayesian Analysis

by

Jean-Marie Rolin*

March 1999

Abstract

In this paper, a very general model of survival data with (exclusive or inclusive) right censoring, explanatory processes and a baseline predictable hazard function is considered in the context of nonparametric Bayesian analysis. Particular cases are semiparametric proportional hazards and multiplicative intensity models. If the baseline predictable hazard function is a Levy process, then its distribution conditionally on the censoring times and the explanatory processes is still a Levy process and is completely described. In a semiparametric case, the posterior distribution of the parameter is also obtained. These posterior distributions are computed for Beta processes and Gamma processes in the proportional hazards and multiplicative intensity models. The non informative case provides a new likelihood for the parameters even in case of ties, contrary to the Cox likelihood.

AMS MOS primary: 62N05
secundary: 62F15,62G05

Key words: Bayesian nonparametric, Levy process, Beta process, Gamma process, Right censoring, Multiplicative Intensity Model, Proportional Hazards Model.

---

*Institut de statistique and CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium.
Financial support from the contracts "Projet d'Actions de Recherche Concertées" nos. 93/98-164 and 98/03-217 of the Belgian Government is gratefully acknowledged.

1 Introduction

The aim of this paper is to present a unified approach to nonparametric and semiparametric Bayesian analysis of survival data. This approach encompasses much of the previous work involving a unique baseline distribution (i.e., Aalen's type models will not be considered). A review of the different problems encountered in survival data and of the corresponding literature is provided in Florens, Mouchart and Rolin (1999). The present paper is theoretical in nature but provides, under weak assumptions on the data and on the distribution of the variables of interest, and under general priors, a full description of the posterior distributions. Such a general framework will provide arguments for choosing a sensible model for Bayesian analysis of survival data. Practical issues such as simulation and asymptotic approximations of the distributions of functionals will be presented in subsequent papers.

In survival analysis, the primary interest lies in a positive random variable T representing a lifetime. The distribution of this random variable depends on an explanatory process θ and on a baseline distribution function F that is assumed to be the distribution of T when the explanatory process is irrelevant. The explanatory process θ may be unobserved, partially or totally observed. When unobserved, this corresponds to an heterogeneity model as used by sociometricians and econometricians. The explanatory process may be a function of observable processes and of a (Euclidian) parameter. In that case, the model is semiparametric. A complete observation of the explanatory process is a particular case of Aalen's model (Aalen, 1978). A precise description of the distribution of this lifetime is given in Section 2.

The presence of right censoring is a typical feature of survival data. The distribution of the censoring variable is left unspecified but its relations with the survival time and the explanatory process is presented in Section 3.

Section 4 extends the basic model to a sampling model for the observations. The model is such that the inference may be performed in the model conditional on the censoring variables and on the explanatory processes. The distribution of the censoring variables and the explanatory processes is therefore left unspecified.

Section 5 deals with the choice of a prior distribution for the baseline distribution F. In a stochastic process approach, the most general family of priors used in the literature consists in considering that the distribution function associated with F, i.e., $F([0,t])$, is a neutral to the right process (Docksum, 1974). This is equivalent to say that the hazard function defined by

$$L([0,t]) = -\ln F((t,\infty])$$

is a Levy process, i.e., an increasing process with independent increments. An example of such a specification is given by Kalbfleisch (1978) who considers that the hazard function is a Gamma process. It is also equivalent to say that the "predictable" hazard function defined by

$$L^{-}([0,t]) = \int_{[0,t]} F([s,\infty])^{-1}\, F(ds)$$

is a Levy process (e.g., see Rolin (1998) or Florens, Mouchart and Rolin (1999)). Such a specification has been successfully introduced by Hjort (1990) who defines the predictable hazard function to be a Beta process. The advantage of specifying the distribution of the predictable hazard function is shown in Rolin (1998). This is due to the fact that the predictable hazard function has a σ-finite expectation, and this is not necessary true for the hazard function. Therefore, this specification is used in this paper.

Using the fact that the general model is absolutely continuous with respect to the simple model without explanatory processes, Section 6 provides the posterior distribution of the baseline distribution (hazard measure or predictable hazard measure) conditionally on the explanatory processes and the censoring variables. Integration of the density with respect to the nuisance parameter (the baseline distribution) provides the density of the observations conditionally on the explanatory processes or in the semiparametric case the likelihood of the parameter.

Section 7 applies the general results obtained in the preceding section to two particular models: the proportional hazards model and the multiplicative intensity model.

In Section 8, the likelihood of the parameter of the semiparametric model for noninformative prior distributions is derived by limiting arguments for Beta priors for the predictable hazard function and for extended Gamma priors on the hazard function. This provides new likelihoods for the parameter of the semiparametric model, challenging the famous partial (or conditonal) Cox's likelihood, and moreover providing solutions to the controversial case of ties.

2 The Probability Model

The variable of interest is a lifetime, i.e., a nonnegative random variable T, defined on an abstract probability space $(\Omega\,,\,\mathcal{M})$ with values in $(\overline{\mathbb{R}}^{+}\,,\,\overline{\mathcal{B}}^{+})$ where $\overline{\mathbb{R}}^{+}=[0,\infty]$ and $\overline{\mathcal{B}}^{+}$ are the Borel sets of $\overline{\mathbb{R}}^{+}$. Such an assumption will be denoted as

$$T(\omega)\,:\,(\Omega\,,\,\mathcal{M})\,\rightarrow\,(\overline{\mathbb{R}}^{+}\,,\,\overline{\mathcal{B}}^{+})\,. \tag{1}$$

Note that the lifetime may be zero or infinite. This means that a dead-born or eternal situation is conceivable. Indeed, in the specification of a competing risks model, the introduction of independent risks assumption implies that the distributions of these independent risks may be defective (e.g., see Mouchart and Rolin (1995)).

A general explanatory process θ is associated with the lifetime, i.e., a random measurable function defined on $(\overline{\mathbb{R}}^{+}\times\Omega\,,\,\overline{\mathcal{B}}^{+}\otimes\mathcal{M})$ with values in a general measurable space $(S,\mathcal{S})$ denoted as

$$\theta(t)\,(\omega)\,:\,(\overline{\mathbb{R}}^{+}\times\Omega\,,\,\overline{\mathcal{B}}^{+}\otimes\mathcal{M})\,\rightarrow\,(\,S\,,\,\mathcal{S}\,)\,. \tag{2}$$

Note that the state space of the explanatory process is not necessarily a topological space and the explanatory process is not required to be right continuous with left limits as usually done. Only the joint measurability is imposed.

A baseline random distribution is also associated with the lifetime, i.e., a random function

$$F(B)\,(\omega)\,:\,\overline{\mathcal{B}}^{+}\times\Omega\,\rightarrow\,\mathbb{R}^{+} \tag{3}$$

with the properties:

(i) $\quad F(B)\,(\cdot)$ is a random variable $\forall\, B\in\overline{\mathcal{B}}^{+}$

(ii) $\quad F(\cdot)\,(\omega)$ is a probability measure on $(\overline{\mathbb{R}}^{+},\overline{\mathcal{B}}^{+})$ $\quad\forall\,\omega\in\Omega\,.$

On the abstract probability space, two probabilities are considered : P, the probability of the model; and P_0, a reference probability. The following assumptions define these two probabilities: the explanatory process θ and the baseline distribution F are independent with respect to P and P_0; and the distribution of θ and the distribution of F with respect to P and P_0 are the same. These assumptions will be written as follows:

$$\begin{array}{ll} (i) & F \perp\!\!\!\perp \theta\ ;\ P \text{ and } F \perp\!\!\!\perp \theta\ ;\ P_0 \\ (ii) & P \circ \theta^{-1} = P_0 \circ \theta^{-1} \\ (iii) & P \circ F^{-1} = P_0 \circ F^{-1} . \end{array} \tag{4}$$

Now, F is the distribution of T conditional on (θ, F) with respect to P_0, i.e.,

$$P_0 [\, T \leq t \mid \theta , F \,] = F([0,t]) . \tag{5}$$

This implies that T and θ are independent conditional on F with respect to P_0, i.e.,

$$T \perp\!\!\!\perp \theta \mid F\ ; P_0 . \tag{6}$$

The distribution of θ is left unspecified and F is a neutral to the right process. The description of the model will be completed by specifying the distribution of T conditional on (θ, F) with respect to P, i.e.,

$$F_\theta(\, [0,t]\,) = P [\, T \leq t \mid \theta , F \,] . \tag{7}$$

Now, as recalled in the introduction, F is a neutral to the right process if and only if the associated hazard function, L, is a Levy process (e.g., see Rolin (1998)). The continuous part of such a process is deterministic and therefore has no interest in this analysis. Hence, the prior specification implies that F is a purely discrete distribution. Note however that in order to be a "good" prior distribution in the nonparametric sense, F must have infinitely many jumps in every open interval of the positive real line. Indeed, this is a necessary condition to insure that under a usual topology (defined either by weak or by pointwise convergence) the set of all possible realizations of F is dense in a large family of distributions (e.g., see Ferguson (1973)). Therefore, the specification of the model may be done in describing the jumps of the predictable hazard function associated to F_θ which are merely the instantaneous probabilities of dying. Hence, the description of the model is completed by specifying

$$L_\theta^-(t) \;=\; \frac{P\,[\,T = t \;\mid \theta, F\,]}{P\,[\,T \geq t \;\mid \theta, F\,]} \;=\; k[\,\theta(t)\,,\, L^-(t)\,]\,, \tag{8}$$

where k is a measurable function defined on $S \times [0,1]$ with values in $[0,1]$, i.e.,

$$k(x,u) \;:\; (S \times [0,1]\,,\, \mathcal{S} \otimes \mathcal{B} \cap [0,1]) \;\rightarrow\; ([0,1]\,,\, \mathcal{B} \cap [0,1])\,, \tag{9}$$

such that

$$k(x,0) \;=\; 0 \qquad \forall\, x \,\in\, S\,. \tag{10}$$

This means that the distribution function F_θ has the following form:

$$F_\theta(\,[0,t]\,) \;=\; 1 - \prod_{0 \leq s \leq t} \{1 - k[\,\theta(s)\,,\, L^-(s)\,]\}\,. \tag{11}$$

Two special cases of such a model have been considered in the literature. The first one is the multiplicative intensity model especially used in the sampling theory framework (e.g., see Fleming and Harrington (1991) or Andersen, Bogdan, Gill and Keiding (1993)) stating that

$$k(x,u) \;=\; a(x)\,u\,, \tag{12}$$

where

$$a(x) \;:\; (S\,,\, \mathcal{S}) \;\rightarrow\; ([0,1]\,,\, \mathcal{B} \cap [0,1])\,, \tag{13}$$

is a known function. This means that

$$L_\theta^-(t) \;=\; a[\,\theta(t)\,]\,L^-(t)\,. \tag{14}$$

The second one, the most popular, has been used in particular in Kalbfleisch's (1978) pioneering paper and is the proportional hazards model. In our framework, this amounts to setting

$$k(x,u) \;=\; 1 - (1-u)^{\,b(x)}\,, \tag{15}$$

where

$$b(x) \,:\, (S\,,\,\mathcal{S}) \;\rightarrow\; (\mathrm{I\!R}^+\,,\,\mathcal{B}^+)\,, \tag{16}$$

is a known function. Such a definition implies that

$$L_\theta(t) \;=\; b[\,\theta(t)\,]\,L(t)\,. \tag{17}$$

Indeed, in general, if the continuous parts of the hazard function and of the predictable hazard function are the same, the jumps are in bijection (e.g., see Florens, Mouchart and Rolin (1999)) because

$$L(t) \;=\; -\ln \frac{F((t,\infty])}{F([t,\infty])} \;=\; -\ln[\,1 - L^-(t)\,]\,. \tag{18}$$

Remark: Formula (14) is in fact equivalent to

$$L^-_\theta(\,[0,t]\,) \;=\; \int_{[0,t]} a[\,\theta(s)\,]\,L^-(ds)\,,$$

and Formula (17) is equivalent to

$$L_\theta(\,[0,t]\,) \;=\; \int_{[0,t]} b[\,\theta(s)\,]\,L(ds)\,.$$

In a sampling theory framework, the baseline distribution is generally supposed to be continuous and therefore the two models described here are the same. However, the estimation of the baseline distribution is very often discrete, and consequently these two models provide two different estimators of L_θ as for instance the Nelson and Peterson estimators (e.g., see Miller (1981)).

A semiparametric model is obtained by defining

$$\theta(t) \;=\; c[\,\beta\,,\,X(t)\,]\,, \tag{19}$$

where β is a parameter and X is an observed explanatory process with values in some measurable spaces. More precisely,

$$\beta\,(\,\omega\,) \,:\, (\,\Omega\,,\,\mathcal{M}\,) \;\rightarrow\; (\,A\,,\,\mathcal{A}\,)\,, \tag{20}$$

$$X(t)\,(\omega) \,:\, (\,\overline{\mathrm{I\!R}}^+ \times \Omega\,,\,\overline{\mathcal{B}}^+ \otimes \mathcal{M}\,) \rightarrow (\,U\,,\,\mathcal{U}\,)\,, \tag{21}$$

and

$$c(a,u) : (A \times U, \mathcal{A} \otimes \mathcal{U}) \to (S, \mathcal{S}) \tag{22}$$

is a known function. It is also assumed that, a priori,

$$\begin{array}{ll} (i) & F \perp\!\!\!\perp \beta \perp\!\!\!\perp X \; ; \; P \quad \text{and} \quad F \perp\!\!\!\perp \beta \perp\!\!\!\perp X \; ; \; P_0 \\ (ii) & P \circ (F, \beta, X)^{-1} = P_0 \circ (F, \beta, X)^{-1} . \end{array} \tag{23}$$

The model then implies that

$$T \perp\!\!\!\perp (X, \beta) \mid \theta, F \; ; \; P \quad \text{and} \quad T \perp\!\!\!\perp (X, \beta) \mid F \; ; \; P_0 . \tag{24}$$

Typically β is a Euclidian parameter, i.e., $A = \mathbb{R}^p$, and X is a multivariate explanatory process, $U = \mathbb{R}^d$, and very often $p = d$. In the proportional hazards model, a famous example is provided by the Cox model (adopted by Kalbfleisch (1978)) for which

$$b[c[\beta, X(t)]] = e[\beta, X(t)] = \exp[\beta' X(t)] . \tag{25}$$

In the multiplicative intensity model, the factor has to be smaller than one, and so a logistic model is proposed, i.e.,

$$a[c[\beta, X(t)]] = d[\beta, X(t)] = \frac{\exp[\beta' X(t)]}{1 + \exp[\beta' X(t)]} . \tag{26}$$

Most of our results are consequences of the following very simple lemma:

Lemma 2.1 *On (T, θ, F), P is absolutely continuous with respect to P_0 and the density of T conditional on (θ, F) or equivalently the marginal density (Radon-Nikodym derivative) of (T, θ, F) is given by*

$$p(T \mid \theta, F) = p(T, \theta, F) = \frac{L_\theta^-(T) \exp[-L_\theta([0,T))]}{L^-(T) \exp[-L([0,T))]} .$$

■

Proof: Formula (8) may be rewritten as

$$P[\,T = t \mid \theta, F\,] \;=\; L_\theta^-(t)\,P[\,T \geq t \mid \theta, F\,]\,.$$

But

$$P[\,T \geq t \mid \theta, F\,] \;=\; \exp[\,-L_\theta(\,[0,t)\,)\,]\,,$$

and therefore,

$$P[\,T = t \mid \theta, F\,] \;=\; p\,(\,t \mid \theta\,,\,F\,)\,P_0[\,T = t \mid \theta, F\,]\,.$$

The proof is completed by noticing that Assumptions (4) imply that P is absolutely continuous with respect to P_0 on $(\theta\,,\,F)$ and moreover,

$$p\,(\theta\,,\,F) \;=\; 1\,.$$

This implies that

$$p\,(T\,,\,\theta\,,\,F) \;=\; p\,(T \mid \theta\,,\,F)$$

(for definitions of marginal and conditional densities with respect to a probability measure see, e.g., Florens, Mouchart and Rolin (1990), Section 1.4).

■

3 The Censoring Mechanism and the Observation

As noted earlier, the presence of censoring is a typical feature of survival data. In this paper, only right censoring is considered. This censoring is described through a finite random variable (the only thing observable by a statistician), i.e.,

$$C(\omega)\,:\,(\Omega\,,\,\mathcal{M}) \;\rightarrow\; (\mathbb{R}^+\,,\,\mathcal{B}^+) \tag{27}$$

and the observed lifetime, Y, is given by

$$Y \;=\; \min\,(\,T\,,\,C\,)\,. \tag{28}$$

The observation also includes an indicator function describing the cause of the removal. The censoring mechanism may be *exclusive*, i.e.,

$$A = \mathbb{1}_{\{T \leq C\}}\,, \tag{29}$$

or *inclusive*, i.e.,

$$\tilde{A} = \mathbb{1}_{\{T < C\}} \tag{30}$$

(e.g., see Ferguson and Phadia (1979)). In what follows, the results are stated for both censoring cases but the proofs will only be given for the exclusive censoring.

Concerning the dependence relations between C and $(T\,,\,\theta\,,\,F\,)$, the distribution of the censoring variable may depend on the explanatory process but is assumed to be independent of T and F with respect to P and P_0 and has the same distribution under P and P_0. To be more precise, the following assumptions hold:

$$\begin{array}{ll} (i) & C \perp\!\!\!\perp (T\,,\,F\,) \mid \theta\,;\ P \text{ and } C \perp\!\!\!\perp (T\,,\,F\,) \mid \theta\,;\ P_0 \\ (ii) & P \circ (C\,,\,\theta\,)^{-1} = P_0 \circ (C\,,\,\theta\,)^{-1}\,. \end{array} \tag{31}$$

Using properties of conditional independence (e.g., see Florens, Mouchart and Rolin (1990)), it may be shown that Assumptions (4) and (31) are equivalent to the following set of assumptions:

$$\begin{array}{ll} (i) & F \perp\!\!\!\perp (C\,,\,\theta\,)\,;\ P \text{ and } F \perp\!\!\!\perp (C\,,\,\theta\,)\,;\ P_0 \\ (ii) & T \perp\!\!\!\perp C \mid \theta\,,\,F\,;\ P \text{ and } T \perp\!\!\!\perp (C\,,\,\theta\,) \mid F\,;\ P_0 \\ (iii) & P \circ (C\,,\,\theta\,,\,F\,)^{-1} = P_0 \circ (C\,,\,\theta\,,\,F\,)^{-1}\,. \end{array} \tag{32}$$

Using the fact that $(Y\,,\,A)$ is a function of $(T\,,\,C\,)$, these independence properties imply, by elementary properties of conditional independence, the following property:

$$(Y\,,\,A\,,\,F\,) \perp\!\!\!\perp \theta \mid C\,;\ P_0\,. \tag{33}$$

In a semiparametric model, Assumptions (31) become

$$\begin{array}{ll} (i) & C \perp\!\!\!\perp (T\,,\,\beta\,,\,F\,) \mid X\,; P \text{ and } C \perp\!\!\!\perp (T\,,\,\beta\,,\,F\,) \mid X\,; P_0 \\ (ii) & P \circ (C\,,\,X\,)^{-1} = P_0 \circ (C\,,\,X\,)^{-1} \end{array} \tag{34}$$

i.e., the distribution of the censoring variable may depend on the observed explanatory process but does not depend on the parameters and the lifetime. Here again, Assumptions (23) and (34) are equivalent to the following set of assumptions:

$$\begin{array}{ll} (i) & \beta \perp\!\!\!\perp F \perp\!\!\!\perp (C, X) \,; P \text{ and } \beta \perp\!\!\!\perp F \perp\!\!\!\perp (C, X) \,; P_0 \\ (ii) & T \perp\!\!\!\perp C \mid X, \beta, F \,; P \text{ and } T \perp\!\!\!\perp (C, X, \beta) \mid F \,; P_0 \\ (iii) & P \circ (C, X, \beta, F)^{-1} = P_0 \circ (C, X, \beta, F)^{-1}. \end{array} \tag{35}$$

The fundamental Lemma 2.1 may be extended to censored observations into a more fundamental lemma as follows:

Lemma 3.1 *For both censoring cases, P is absolutely continuous with respect to P_0 and the densities of the censored observations conditional on (C, θ, F) are given by*

(i) for exclusive censoring,

$$p(Y, A \mid C, \theta, F) = \frac{L_\theta^-(Y)^A [1 - L_\theta^-(Y)]^{1-A} \exp[-L_\theta([0, Y))]}{L^-(Y)^A [1 - L^-(Y)]^{1-A} \exp[-L([0, Y))]}$$

and moreover,

$$p(Y, A, C, \theta, F) = p(Y, A \mid C, \theta, F);$$

(ii) for inclusive censoring,

$$p(Y, \tilde{A} \mid C, \theta, F) = \frac{L_\theta^-(Y)^{\tilde{A}} \exp[-L_\theta([0, Y))]}{L^-(Y)^{\tilde{A}} \exp[-L([0, Y))]}$$

and moreover,

$$p(Y, \tilde{A}, C, \theta, F) = p(Y, \tilde{A} \mid C, \theta, F).$$

■

Proof: The proof is a simple matter of computation. Indeed,

$$P[Y = t, A = 1 \mid C, \theta, F] = P[T = t, T \leq C \mid C, \theta, F]$$

$$= P[\, T \,=\, t \mid C\,,\, \theta\,,\, F\,]\, \mathbb{1}_{\{\, t \leq C\,\}}$$
$$= L^{-}_{\theta}\,(t)\, \exp[\, -\, L_{\theta}(\, [0,t)\,)\,]\, \mathbb{1}_{\{\, t \leq C\,\}}$$

by property (32) (ii) and, by the same property,

$$P[\, Y \,=\, t\,,\, A \,=\, 0 \mid C\,,\, \theta\,,\, F\,] \,=\, P[\, C \,=\, t\,,\, T \,>\, C \mid C\,,\, \theta\,,\, F\,]$$
$$= P[\, T \,>\, t \mid C\,,\, \theta\,,\, F\,]\, \mathbb{1}_{\{\, t = C\,\}} \,=\, \exp[\, -\, L_{\theta}(\, [0,t]\,)\,]\, \mathbb{1}_{\{\, t = C\,\}}$$
$$= [\, 1 \,-\, L^{-}_{\theta}\,(t)\,]\, \exp[\, -\, L_{\theta}(\, [0,t)\,)\,]\, \mathbb{1}_{\{\, t = C\,\}}\,.$$

These two relations may be combined into

$$P[\, Y \,=\, t\,,\, A \,=\, a \mid C\,,\, \theta\,,\, F\,]$$
$$= p\,(t\,,\, a \mid C\,,\, \theta\,,\, F\,)\, P_0[\, Y \,=\, t\,,\, A \,=\, a \mid C\,,\, \theta\,,\, F\,]\,.$$

The proof is completed by noticing that Property (32) (iii) is equivalent to

$$p\,(\, C\,,\, \theta\,,\, F\,) \,=\, 1\,.$$

■

Remark that the two densities of Lemma 3.1 do not depend on C. This is a remarkable feature of these censoring models. This peculiar property will be analyzed in more detail at the end of the next section.

4 The Sampling Model

Let $\{\, (\, T_i\,,\, C_i\,,\, \theta_i\,)\,,\, 1 \,\leq\, i \,\leq\, n\,\}$ be a sample of size n from the model described in Sections 2 and 3, denoted by the abbreviated $(\, T_1^n\,,\, C_1^n\,,\, \theta_1^n\,)$ notation. It is generally assumed that a sample is composed of independent, identically distributed elements, but this is too strong. As in Section 3, the baseline distribution is independent of the explanatory processes and the censoring variables; the distributions of the explanatory processes and of the censoring variables with respect to P and P_0 are the same but this distribution is left unspecified. This means that

$$(i) \qquad F \perp\!\!\!\perp (C_1^n\,,\,\theta_1^n)\ ;\ P \ \text{ and } \ F \perp\!\!\!\perp (C_1^n\,,\,\theta_1^n)\ ;\ P_0 \qquad (36)$$

$$(ii) \qquad P \circ (C_1^n\,,\,\theta_1^n\,,\,F)^{-1} = P_0 \circ (C_1^n\,,\,\theta_1^n\,,\,F)^{-1}\,.$$

The lifetimes T_i's are assumed to be conditionally independent, and the distribution of T_i depends only on the corresponding explanatory process and is the one described in Section 1 and specified by $L^-_{\theta_i}$ with respect to P and by L^- with respect to P_0. More precisely, the following assumptions hold:

$$(i) \qquad \mathop{\perp\!\!\!\perp}_{1\le i\le n} T_i \mid C_1^n,\theta_1^n,F\ ;\ P \ \text{ and } \ \mathop{\perp\!\!\!\perp}_{1\le i\le n} T_i \mid C_1^n,\theta_1^n,F\ ;\ P_0 \quad (37)$$

$$(ii) \qquad T_i \perp\!\!\!\perp (C_1^n,\theta_1^n) \mid \theta_i\,,F\ ;\ P \ \text{ and } \ T_i \perp\!\!\!\perp (C_1^n,\theta_1^n) \mid F\ ;\ P_0\,.$$

As in Section 3, since the distribution of T_i with respect to P_0 does not depend on θ_i , $T_1^n \perp\!\!\!\perp (C_1^n,\theta_1^n) \mid F\ ;\ P_0$ and, along with (36) (i), this is equivalent to

$$(T_1^n,F) \perp\!\!\!\perp (C_1^n,\theta_1^n)\ ;\ P_0\,. \qquad (38)$$

From this model of conditional independences involving latent variables, i.e., $(\,T_1^n\,,\,C_1^n\,)$, the models of observable variables, i.e., either $(\,Y_1^n\,,\,A_1^n\,)$ or $(\,Y_1^n\,,\,\tilde{A}_1^n\,)$, may be deduced by elementary properties of conditional independence (e.g., see Florens, Mouchart and Rolin (1990), Section 2.2). For the exclusive censoring, with respect to P, the model is

$$(i) \qquad \mathop{\perp\!\!\!\perp}_{1\le i\le n} (Y_i\,,\,A_i) \mid C_1^n\,,\,\theta_1^n\,,\,F\ ;\ P \qquad (39)$$

$$(ii) \qquad (Y_i\,,\,A_i) \perp\!\!\!\perp (C_1^n\,,\,\theta_1^n) \mid C_i\,,\,\theta_i\,,\,F\ ;\ P\,,$$

and with respect to P_0,

$$(i) \qquad \mathop{\perp\!\!\!\perp}_{1\le i\le n} (Y_i\,,\,A_i) \mid C_1^n\,,\,\theta_1^n\,,\,F\ ;\ P_0 \qquad (40)$$

$$(ii) \qquad (Y_i\,,\,A_i) \perp\!\!\!\perp (C_1^n\,,\,\theta_1^n) \mid C_i\,,\,F\ ;\ P_0\,.$$

This, along with (36) (i), implies

$$(\, Y_1^n \,,\, A_1^n \,,\, F \,) \;\perp\!\!\!\perp\; \theta_1^n \;\mid\; C_1^n \,;\, P_0 \,. \tag{41}$$

Replacing A by $\tilde{A}$ provides the model for the inclusive censoring.

Under these conditional independences, the densities of the censored observations are simply the product of the individual densities given in Lemma 3.1, i.e.,

$$p\,(\, Y_1^n \,,\, A_1^n \;\mid\; C_1^n \,,\, \theta_1^n \,,\, F \,) \;=\; \prod_{1\le i\le n} p\,(\, Y_i \,,\, A_i \;\mid\; C_i \,,\, \theta_i \,,\, F \,) \tag{42}$$

and

$$p\,(\, Y_1^n \,,\, \tilde{A}_1^n \;\mid\; C_1^n \,,\, \theta_1^n \,,\, F \,) \;=\; \prod_{1\le i\le n} p\,(\, Y_i \,,\, \tilde{A}_i \;\mid\; C_i \,,\, \theta_i \,,\, F \,)\,. \tag{43}$$

As mentioned in the introduction, the baseline distribution is a neutral to the right process, or equivalently, the hazards functions are Levy processes. As a consequence, due to independence properties, it is interesting to express these densities in terms of the order statistics defined as follows:

Definition 4.1 *For the censored observations, the distinct values of the observed lifetimes in increasing order are defined by*

$$\{\, Y_i \;,\; 1 \le i \le n \,\} \;=\; \{\, Z_j \;,\; 1 \le j \le M \,\}$$

with

$$0 \le Z_1 \;<\; Z_2 \;<\; \cdots \;<\; Z_M \;<\; \infty\,.$$

The sets of indices of the dead, censored and at risk persons at time t and the cardinality of these sets are defined by

$$I(t) \;=\; \{i : Y_i = t \,,\, A_i = 1\} \;\text{ and }\; D(t) \;= \mid I(t) \mid,$$

$$J(t) \;=\; \{i : Y_i = t \,,\, A_i = 0\} \;\text{ and }\; R(t) \;= \mid I(t) \mid,$$

and

$$K(t) \;=\; \{i : Y_i > t\} \;\text{ and }\; N(t) \;=\mid K(t) \mid \;.$$

$\blacksquare$

A simple computation (details are provided in Appendix 1) yields the following lemma:

Lemma 4.1 *The conditional densities of the censored observations are given by*

(i) for exclusive censoring

$$p\,(\,Y_1^n\,,\,A_1^n\;\mid\;C_1^n\,,\,\theta_1^n\,,\,F\,)\;=\;\prod_{1\le j\le M}\;\prod_{Z_{j-1}<t<Z_j}\frac{V_\theta(t\,,\,L^-\,(t))}{V(t\,,\,L^-\,(t))}$$

$$\times\;\prod_{1\le j\le M}\frac{W_\theta(Z_j\,,\,L^-\,(Z_j))}{W(Z_j\,,\,L^-\,(Z_j))}$$

(ii) and for inclusive censoring

$$p\,(\,Y_1^n\,,\,\tilde{A}_1^n\;\mid\;C_1^n\,,\,\theta_1^n\,,\,F\,)\;=\;\prod_{1\le j\le M}\;\prod_{Z_{j-1}<t<Z_j}\frac{V_\theta(t\,,\,L^-\,(t))}{V(t\,,\,L^-\,(t))}$$

$$\times\;\prod_{1\le j\le M}\frac{\tilde{W}_\theta(Z_j\,,\,L^-\,(Z_j))}{\tilde{W}(Z_j\,,\,L^-\,(Z_j))}$$

where for both cases,

$$V_\theta\,(t\,,\,u)\;=\;\prod_{i\,\in\,K(t)}\,[1-k[\,\theta_i(t)\,,\,u\,]\,]\quad\text{and}\quad V\,(t\,,\,u)\;=\;(1-u)^{\,N(t)}\,.$$

(i) For exclusive censoring,

$$W_\theta\,(t\,,\,u)\;=\;\prod_{i\,\in\,K(t)\,\cup\,J(t)}\,[1-k[\,\theta_i(t)\,,\,u\,]\,]\;\times\;\prod_{i\,\in\,I(t)}\,k[\,\theta_i(t)\,,\,u\,]$$

and

$$W\,(t\,,\,u) \;=\; u^{\,D(t)}\;\;(1-u)^{\,N(t)+R(t)}\,.$$

(ii) For inclusive censoring,

$$\tilde{W}_\theta\,(t\,,\,u) \;=\; \prod_{i\,\in\,K(t)} [1-k[\,\theta_i(t)\,,\,u\,]\,] \;\times\; \prod_{i\,\in\,I(t)} k[\,\theta_i(t)\,,\,u\,]$$

and

$$\tilde{W}\,(t\,,\,u) \;=\; u^{\,D(t)}\;\;(1-u)^{\,N(t)}\,.$$

■

As noted at the end of Section 3, in these sampling models, the two conditional densities do not depend on the censoring variables, which leads to the following corollary:

Corollary 4.2 *For exclusive censoring,*

$$p\,(\,Y_1^n\,,\,A_1^n\;|\;C_1^n\,,\,\theta_1^n\,,\,F\,)\;=\;p\,(\,Y_1^n\,,\,A_1^n\;|\;\theta_1^n\,,\,F\,)\,,$$

and for inclusive censoring,

$$p\,(\,Y_1^n\,,\,\tilde{A}_1^n\;|\;C_1^n\,,\,\theta_1^n\,,\,F\,)\;=\;p\,(\,Y_1^n\,,\,\tilde{A}_1^n\;|\;\theta_1^n\,,\,F\,)\,.$$

■

However, this does not imply that $(\,Y_1^n\,,\,A_1^n\,)$ is independent of C_1^n conditional on $(\,\theta_1^n\,,\,F\,)$ with respect to P, because such a property is clearly not true with respect to P_0. However, this property has some implications on the inference process. In Florens and Rolin (1998), it is shown that, in censoring models, whatever the prior distribution on F is,

$$F \;\perp\!\!\!\perp\; C_1^n \;\;|\;\; Y_1^n\,,\,A_1^n\;;\;P_0\,. \tag{44}$$

This means that the knowledge of unobserved censoring variables does not improve the inference on F. Combining this result with Property (41), it may be seen, by elementary properties of conditional independences, that

$$F \;\perp\!\!\!\perp\; (\,C_1^n\,,\,\theta_1^n\,) \;\;|\;\; Y_1^n\,,\,A_1^n\;;\;P_0\,. \tag{45}$$

Now,

$$
\begin{aligned}
& p\,(\,Y_1^n\,,\,A_1^n\,,\,C_1^n\,,\,\theta_1^n\,) \\
& \qquad = E_0\,[\,p\,(\,Y_1^n\,,\,A_1^n\,,\,C_1^n\,,\,\theta_1^n\,,\,F\,)\ \mid\ Y_1^n\,,\,A_1^n\,,\,C_1^n\,,\,\theta_1^n\,] \\
& \qquad = E_0\,[\,p\,(\,Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,,\,F\,)\ \mid\ Y_1^n\,,\,A_1^n\,,\,C_1^n\,,\,\theta_1^n\,] \\
& \qquad = E_0\,[\,p\,(\,Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,,\,F\,)\ \mid\ Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,] \\
& \qquad = p\,(\,Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,)\,.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
p\,(\,F\ \mid\ Y_1^n\,,\,A_1^n\,,\,C_1^n\,,\,\theta_1^n\,) &= \frac{p\,(\,Y_1^n\,,\,A_1^n\,,\,C_1^n\,,\,\theta_1^n\,,\,F\,)}{p\,(\,Y_1^n\,,\,A_1^n\,,\,C_1^n\,,\,\theta_1^n\,)} \\
&= \frac{p\,(\,Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,,\,F\,)}{p\,(\,Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,)} \\
&= p\,(\,F\ \mid\ Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,)\,.
\end{aligned}
$$

This implies the following proposition stating that, whatever the prior distribution on the baseline distribution is, the knowledge of the censoring variables does not improve the inference on F in a model with explanatory processes:

Proposition 4.3 *In the sampling model defined by Assumptions* (36) *and* (37)*, the following properties hold:*

$$F \perp\!\!\!\perp C_1^n \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,;\ P$$

and

$$F \perp\!\!\!\perp C_1^n \mid Y_1^n\,,\,\tilde{A}_1^n\,,\,\theta_1^n\,;\ P\,.$$

■

In the semiparametric model, all assumptions and properties of (conditional) independence still hold in replacing θ_i by $(\,X_i\,,\,\beta\,)$ and θ_1^n by $(\,X_1^n\,,\,\beta\,)$. But Assumption (36) (i) is replaced by

$$F \perp\!\!\!\perp \beta \perp\!\!\!\perp (\,C_1^n\,,\,X_1^n\,)\,;P \ \text{ and }\ F \perp\!\!\!\perp \beta \perp\!\!\!\perp (\,C_1^n\,,\,X_1^n\,)\ ;P_0\,. \qquad (46)$$

This Assumption, along with Property (41), implies

$$\beta \perp\!\!\!\perp (Y_1^n , A_1^n , C_1^n , X_1^n , F) \; ; \; P_0 , \tag{47}$$

and this property along with Property (45) implies

$$F \perp\!\!\!\perp \beta \perp\!\!\!\perp (C_1^n , X_1^n) \mid Y_1^n , A_1^n \; ; P_0 , \tag{48}$$

which implies

$$(\beta , F) \perp\!\!\!\perp C_1^n \mid Y_1^n , A_1^n , X_1^n \;\; P_0 . \tag{49}$$

By the same argument as above,

$$p (Y_1^n , A_1^n , C_1^n , X_1^n) = p (Y_1^n , A_1^n , X_1^n) , \tag{50}$$

and this means as before that, whatever the prior distribution of the baseline distribution can be, the inference on the parameters is not improved by the knowledge of the censoring variables.

Proposition 4.4 *In the semiparametric model defined by Assumptions* (46) *,* (36) (*ii*) *and* (37) *, the following properties hold:*

$$(\beta , F) \perp\!\!\!\perp C_1^n \mid Y_1^n , A_1^n , X_1^n \; ; \; P$$

and

$$(\beta , F) \perp\!\!\!\perp C_1^n \mid Y_1^n , \tilde{A}_1^n , X_1^n \; ; \; P .$$

■

Remark: In a sampling theory framework, the generalized maximum likelihood principle introduced by Kiefer and Wolfowitz (1956) imposes the predictable hazard measure to be of the following form:

$$L^- = \sum_{1 \le j \le M} \lambda_j \, \varepsilon_{Z_j} . \tag{51}$$

Therefore, in a semiparametric model, the likelihood becomes parametric and is given, under exclusive censoring, by

$$\hat{p} (Y_1^n , A_1^n \mid X_1^n , \beta , \lambda_1^M) = \prod_{1 \le j \le M} W_\theta(Z_j , \lambda_j) . \tag{52}$$

For inclusive censoring, i.e., $\tilde{A}_1^n$ instead of A_1^n, W_θ has to be replaced by $\tilde{W}_\theta$.

5 The Prior Distribution and Preliminary Results

5.1 The Specification of the Prior Distribution

To elicitate the prior distribution of the baseline distribution F, the Bayesian statistician must first provide an initial guess of this distribution through its expectation. So, let

$$F_0(B) = E[F(B)] \qquad \forall\, B \in \overline{\mathcal{B}}^{+}. \tag{53}$$

Also, let

$$\begin{aligned} t_0 &= \inf\{t : F_0([0,t]) > 0\} \\ t_1 &= \inf\{t : F_0([0,t]) = 1\} \\ I &= \{t : F_0(t) > 0\}. \end{aligned} \tag{54}$$

Typically, the statistician chooses $t_0 = 0$, $t_1 = \infty$ and $I = \emptyset$, but here we examine the general case. This means that death occurs in the interval $[t_0, t_1]$ and there is a positive probability of death at fixed points in I. Note that the statistician may equivalently specify either the hazard function of F_0, denoted by L_0, or the predictable hazard function of F_0, denoted by L_0^-.

The elicitation of the prior is completed by specifying F in such a way that the baseline distribution function $F([0,t])$ is a neutral to the right process (Doksum (1974)). This is equivalent to specifying the hazard function defined by

$$L([0,t]) = -\ln F((t,\infty]) \tag{55}$$

as a Levy process, i.e., an increasing right-continuous process with independent increments. This is also equivalent to specifying the predictable hazard function defined by

$$L^-([0,t]) = \int_{[0,t]} \frac{F(ds)}{F([s,\infty])} \tag{56}$$

as a Levy process with jumps in $(0,1]$. The case of jumps of size 1 is considered to include Poisson processes, but as will be seen later on, this corresponds to a degenerate case.

Such processes have been widely studied in the literature (e.g., see Ferguson and Klass (1972), Rolin (1983), Karr (1986)). These processes may also be viewed as cumulative functions of "*purely*" random σ-finite measures on $\overline{\mathbb{R}}^+$ defined by the Lebesgue-Stieltjes integrals

$$L\,(B) \;=\; \int_B dL\,(\,[0,t]\,) \qquad\qquad \forall\; B\,\in\,\overline{\mathcal{B}}^+ \tag{57}$$

and

$$L^-\,(B) \;=\; \int_B dL^-\,(\,[0,t]\,) \qquad\qquad \forall\; B\,\in\,\overline{\mathcal{B}}^+\,. \tag{58}$$

Such measures are called purely random because the following property holds:

$$\underset{1\leq\ell\leq k}{\perp\!\!\!\perp}\; L(B_\ell)\;, \quad \forall\,\{B_\ell\,,1\leq\ell\leq k\}\,\subset\,\overline{\mathcal{B}}^+ \;\text{ partition of }\; \overline{\mathbb{R}}^+\,. \tag{59}$$

Note that in general an increasing right continuous function may be decomposed into a sum of a continuous function and a pure jump or discrete increasing function. In such a decomposition, the continuous parts of L and L^- are the same and the jumps are in bijection since, as seen before,

$$L(t) \;=\; -\ln[\,1 - L^-(t)\,]\,. \tag{60}$$

For a Levy process, in the decomposition into its continuous part and its discrete part, the continuous part is deterministic. It is difficult to imagine a statistical experiment in which it will be interesting to introduce such a deterministic component. Therefore, the baseline distribution F is a discrete distribution.

For the representation and path properties of the purely random measure L^-, the approach of Rolin (1998) will be used.
First of all, L^- may be written as

$$L^- \;=\; \tilde{L}^- \,+\, \sum_{t\in I} L^-(t)\;\varepsilon_t\,, \tag{61}$$

(where ε_t is the point mass at t) with the following property:

$$\underset{t\in I}{\perp\!\!\!\perp}\; L^-(t)\;\perp\!\!\!\perp\;\tilde{L}^-\,. \tag{62}$$

The distributions of $L^-(t)\,,\ t\,\in\,I$, are arbitrary and $\tilde{L}^-$ is a Levy process without fixed point of discontinuity.

The counting measure on $(\mathbb{R}^+ \times \mathbb{R}_0^+, \mathcal{B}^+ \otimes \mathcal{B}_0^+)$ defined by

$$N^-(D) = |\{t : (t, \tilde{L}^-(t)) \in D\}| \tag{63}$$

is a Poisson process with expectation

$$N_0^-(D) \doteq E[N^-(D)], \tag{64}$$

which is called the Levy measure of the Levy process $\tilde{L}^-$ and which characterizes entirely its distribution. This measure is supported by $[t_0, t_1) \times (0,1]$ and is σ-finite on $[t_0, t_1) \times (0,1]$ (because, for $t_0 \leq t < t_1$, $\tilde{L}^-([0,t]) < \infty$ a.s. since $\tilde{L}_0^-([0,t]) < \infty$).

Therefore, $\forall f : (\mathbb{R}^{+2}, \mathcal{B}^{+2}) \to (\mathbb{R}^+, \mathcal{B}^+)$ with the property that $f(t, 0) = 0 \quad \forall t \in \mathbb{R}^+$,

$$\int_{\mathbb{R}^+ \times \mathbb{R}_0^+} f(t,u)\, N^-(dt, du) = \sum_{t \in \mathbb{R}^+} f[t, \tilde{L}^-(t)], \tag{65}$$

$$E[\sum_{t \in \mathbb{R}^+} f[t, \tilde{L}^-(t)]] = \int_{\mathbb{R}^+ \times \mathbb{R}_0^+} f(t,u)\, N_0^-(dt, du), \tag{66}$$

and

$$E[\exp\{-\sum_{t \in \mathbb{R}^+} f[t, \tilde{L}^-(t)]\}] \tag{67}$$

$$= \exp\{-\int_{\mathbb{R}^+ \times \mathbb{R}_0^+} [1 - e^{-f(t,u)}]\, N_0^-(dt, du)\}.$$

A similar decomposition and representation may be done for the hazard function L with counting measure and Levy measure denoted respectively by N and N_0.
Now, particular cases of Formula (67) provide

$$E[e^{-p\,\tilde{L}^-([0,t])}] \tag{68}$$

$$= \exp\{-\int_{[0,t] \times (0,1]} [1 - e^{-pu}]\, N_0^-(ds, du)\},$$

and using Formula (60),

$$E\,[\,e^{-p\,\tilde{L}(\,[0,t]\,)}\,] \tag{69}$$

$$= \exp\{-\int_{[0,t]\,\times\,(0,1]} [\,1-(1-u)^p\,]\; N_0^-\,(ds,du)\,\}\,.$$

The expectations of the hazards functions are obtained from Formula (66) under both specifications (L and L^-) and are

$$\begin{aligned} E\,[\,\tilde{L}^-(\,[0,t]\,)\,] &= \int_{[0,t]\,\times\,(0,1]} u\; N_0^-\,(ds,du) \\ &= \int_{[0,t]\,\times\,(0,\infty]} (1-e^{-v})\; N_0\,(ds,dv)\;, \end{aligned} \tag{70}$$

$$\begin{aligned} E\,[\,\tilde{L}(\,[0,t]\,)\,] &= \int_{[0,t]\,\times\,(0,1]} \ln(\,\frac{1}{1-u}\,)\; N_0^-\,(ds,du) \\ &= \int_{[0,t]\,\times\,(0,\infty]} v\; N_0\,(ds,dv)\,. \end{aligned} \tag{71}$$

The second expression of (70) shows that the expectation of the predictable hazard function is finite for $t < t_1$, while the expectation of the hazard function (Formula (71)) may be infinite.

A most remarkable fact in specifying the predictable hazard function to be a Levy process is the following: the expectation of the predictable hazard function is the predictable hazard function of the expectation of the baseline distribution, i.e.,

$$L_0^-\,(\,[0,t]\,) \;=\; E\,[\,L^-\,(\,[0,t]\,)\,]\,. \tag{72}$$

This property may be rewritten as

$$E\,[\int_{[0,t]} \frac{F(ds)}{F(\,[s,\infty]\,)}\,] = \int_{[0,t]} \frac{F_0(ds)}{F_0(\,[s,\infty]\,)}\,. \tag{73}$$

Indeed, Formulae (69) and (70) give, for $p = 1$,

$$E\,[\,e^{-\,\tilde{L}(\,[0,t]\,)}\,] \;=\; e^{-E\,[\,\tilde{L}^{-}(\,[0,t]\,)\,]}\,. \tag{74}$$

Therefore, using (62),

$$\begin{aligned} L_0\,(\,[0,t]\,) \;&=\; -\ln E[\,e^{-L\,(\,[0,t]\,)}\,] \qquad\qquad (75)\\ &=\; -\ln E[\,e^{-\tilde{L}\,(\,[0,t]\,)}\,] - \sum_{0\le s\le t} \ln E[\,e^{-L\,(s)}\,]\\ &=\; E[\,\tilde{L}^{-}\,(\,[0,t]\,)\,] - \sum_{0\le s\le t} \ln[\,1 \;-\; E[\,L^{-}\,(s)\,]\,]\,. \end{aligned}$$

In particular, this implies that the expectation of $\tilde{L}^{-}$ is the continuous part of the predictable hazard function (and of the hazard function) of the expectation of the baseline distribution denoted by L_{0c}.

By Formula (70), since $L_{0\,c}$ is a σ-finite measure, N_0^{-} may then be decomposed as follows:

$$N_0^{-}\,(dt\,,\,du) \;=\; \frac{1}{u}\,L_{0c}\,(dt)\,R_0^{-}\,(t\,,\,du) \tag{76}$$

where R_0^{-} is a probability transition on $(t_0, t_1) \times (0,1]$.
This decomposition allows to rewrite Formula (69) in the form

$$E\,[\,e^{-\,p\,\tilde{L}(\,[0,t]\,)}\,] \;=\; \exp\{\,-\int_0^t r_0\,(p,s)\,L_{0\,c}\,(ds)\,\}\,, \tag{77}$$

where

$$r_0\,(p\,,\,t) \;=\; \int_{(0,1]} \frac{1-(1-u)^p}{u}\,R_0^{-}\,(t\,,\,du)\,. \tag{78}$$

Remark that L^{-} is an homogeneous Levy process if this probability transition does not depend on t, i.e., if

$$R_0^{-}\,(t\,,\,du) \;\equiv\; R_0^{-}\,(\,du\,)\,.$$

All the preceding considerations lead to the following definition of a suitable prior distribution for the baseline distribution:

Definition 5.1 *The prior distribution of the baseline distribution F with prior expectation F_0 is such that L^- is a Levy process defined by*

(i) *a probability transition, R_0^-, on $(t_0, t_1) \times (0, 1]$,*

(ii) *arbitrary prior distributions of $L^-(t)$ for $t \in I$ subject to $E[L^-(t)] = L_0^-(t)$.*

■

Two particular cases of Levy processes will be considered in this paper. The specification of the probability transition will depend only on a strictly positive measurable function defined on $\mathbb{R}^+$, i.e.,

$$c_0(t) : (\overline{\mathbb{R}}^+, \overline{\mathcal{B}}^+) \to (\mathbb{R}_0^+, \mathcal{B}_0^+). \tag{79}$$

The first one, introduced by Hjort (1990), defines L^- as a Beta process.

Definition 5.2 *L^- is a Beta process of parameter c_0, a positive Borel function defined on $\overline{\mathbb{R}}^+$, and $L_0^- = E(L^-)$, to be written as $L^- \sim \mathcal{B}e(c_0, L_0^-)$ if*

(i) $\quad L^-(t) \sim Be[c_0(t) L_0^-(t), c_0(t)[1 - L_0^-(t)]] \qquad \forall\, t \in I$

(ii) $\quad R_0^-(t, du) = c_0(t)(1-u)^{c_0(t)-1} du$

i.e., is a $Be(1, c_0(t))$ distribution.

■

It is interesting to note that, for such a Beta process,

$$r_0(p, t) = c_0(t)\{\psi[c_0(t) + p] - \psi[c_0(t)]\} \tag{80}$$

where ψ is the digamma function (i.e., $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$ and $\Gamma(x)$ is the gamma function).

This family of prior distributions has well-known particular cases. The most famous is the Dirichlet process. Indeed, $F \sim \mathcal{D}i\,(n_0\,F_0)$ if and only if $L^- \sim \mathcal{B}e\,(c_0\,,\,L_0^-)$ with $c_0(t) = n_0\,F_0\,([t\,,\,\infty])$ (e.g., see Hjort (1990) or Rolin (1998)). The simple homogeneous process introduced by Ferguson and Phadia (1979) for L is equivalent to setting $L^- \sim \mathcal{B}e\,(c_0\,,\,L_0^-)$ with $c_0(t) \equiv c_0$ (e.g., see Rolin (1998)). Moreover, the specification of F to be a Beta-neutral process introduced by Lo (1993) or to be a Beta-Stacy process introduced by Walker and Muliere (1997) is equivalent to specifying L^- as a Beta process.

The second one is the (extended) Gamma process introduced by Dykstra and Laud (1981) (see also Walker, Damien and Laud (1998)). Let G be a simple Gamma measure with parameter G_0, i.e., a purely random measure such that, for any $B \in \overline{\mathcal{B}}^+$, $G\,(B) \sim Ga\,[\,1\,,\,G_0\,(B)\,]$, then the (extended) Gamma measure is defined as

$$L\,(B) \;=\; \int_B [\,c_0\,(t)\,]^{-1}\,G\,(dt)\,. \tag{81}$$

This implies that

$$L_0\,(B) \;=\; \int_B \ln\,(\,1 \,+\, \frac{1}{c_0\,(t)}\,)\,G_0\,(dt)\,. \tag{82}$$

Some elementary computations gives, in our framework, the following definition:

Definition 5.3 *L is a Gamma process of parameter c_0, a positive Borel function defined on $\mathbb{R}^+$, and $L_0^- = E\,(L^-)$, to be written as $L \sim \mathcal{G}a\,(c_0\,,\,L_0^-)$ if*

$$(i) \qquad L\,(t) \sim Ga\,[\,c_0(t)\,,\,[\,\ln\,(\,1 \,+\, \frac{1}{c_0\,(t)}\,)\,]^{-1}\,L_0\,(t)\,] \qquad \forall\, t \in I$$

$$(ii) \qquad R_0^-\,(t\,,\,du) \;=\; [\,\ln\,(\,1 \,+\, \frac{1}{c_0\,(t)}\,)\,]^{-1}\;\frac{u\,(1-u)^{\,c_0(t)\,-\,1}}{\ln\,(\dfrac{1}{1-u})}\,du\,.$$

■

Note that, for such a Gamma process,

$$r_0\,(p\,,\,t) \;=\; \frac{\ln\,[\,c_0\,(t) \;+\; p\,] \;-\; \ln\,[\,c_0\,(t)\,]}{\ln\,[\,c_0\,(t) \;+\; 1\,] \;-\; \ln\,[\,c_0\,(t)\,]}\,. \tag{83}$$

The homogeneous particular case, i.e., $c_0\,(t) \;\equiv\; c_0$, has been used in the pioneering paper of Kalbfleisch (1978). Note that in such a case, $L(B) \;\sim\; Ga\,(\,c_0\,,\,G_0\,(B)\,) \quad \forall\, B \,\in\, \overline{\mathcal{B}}^{+}$.

The non informative case amounts to specifying L^- as a (extended) Poisson process defined as follows (note that $T \;\sim\; I\,(p)$ means that T is a Bernoulli or indicator random variable, i.e., with values in $\{0\,,\,1\}$ and p is the probability that $T \;=\; 1$):

Definition 5.4 *L^- is a Poisson process of parameter $L^-_0 \;=\; E\,(\,L^-\,)$, to be written as $L^- \;\sim\; \mathcal{P}o\,(\,L^-_0\,)$ if*

$$(i) \qquad L^-\,(t) \;\sim\; I\,(\,L^-_0\,(t)\,) \qquad \forall\, t \,\in\, I$$

$$(ii) \qquad R^-_0\,(t\,,\,du) \;=\; \varepsilon_1\,(du)$$

i.e., is a point mass at 1.

■

Now, let

$$S \;=\; \inf\,\{\,t \,:\, L^-\,(\,[0,t]\,) \;=\; 1\,\}\,. \tag{84}$$

Then, $L\,(\,[0,t]\,) \;=\; 0$ for $t \,<\, S$ and $L\,(\,[0,t]\,) \;=\; \infty$ for $t \,\geq\, S$ and therefore,

$$F \;=\; \varepsilon_S\,. \tag{85}$$

This means that, in this case, L is no longer a Levy process and F is no longer a neutral to the right process but this corresponds to the idea of a non informative prior distribution. At last, it can be seen that, when $c_0\,(t) \;=\; n_0\; d_0\,(t)$, then, for both Beta and Gamma processes, $R^-_0\,(t\,,\,du)$ converges weakly to $\varepsilon_1\,(du)$ as n_0 tends to 0.

5.2 The Reference Posterior Distribution

In case of no censoring, Docksum (1974) derives, up to a Radon-Nikodym derivative, the posterior distribution of the observations with respect to P_0. Ferguson (1974) gives an alternative description of this posterior distribution (see also Rolin (1983) for another description). For censored observations, the Bayesian estimator and some properties of the posterior distribution are obtained by Susarla and Van Ryzin (1976) and Blum and Susarla (1977). Ferguson and Phadia (1979) generalizes these results as follows: if F is a priori a neutral to the right process, then F is a posteriori a neutral to the right process for exclusive or inclusive censoring. Moreover, this paper describes the posterior distribution with a little bit of trouble for the posterior distribution of the jump at an uncensored observation outside the fixed points of discontinuity, but shows that this trouble disappears for an homogeneous Levy process and for the Dirichlet process. Hjort (1990) proves that Beta processes constitute a natural conjugate family of prior distributions for right censored observations and describes completely the posterior distribution. Walker and Damien (1998) describes the posterior distribution for one censored observation under a neutral to the right process and this description is extended in Walker, Damien and Laud (1998). According to Rolin (1998) and with the notations of Definition 4.1 and Lemma 4.1, the posterior distribution of F under P_0 may be described as follows:

Proposition 5.1 *Let* $J = \{ Y_i : 1 \leq i \leq n \} = \{ Z_j : 1 \leq j \leq M \}$, *then, a posteriori, the hazard measure of the baseline distribution is a Levy process that can be written as*

$$L = \tilde{L} + \sum_{t \in I \cap J^c} L(t)\, \varepsilon_t + \sum_{1 \leq j \leq M} L(Z_j)\, \varepsilon_{Z_j} ,$$

and similarly,

$$L^- = \tilde{L}^- + \sum_{t \in I \cap J^c} L^-(t)\, \varepsilon_t + \sum_{1 \leq j \leq M} L^-(Z_j)\, \varepsilon_{Z_j} .$$

Now,

(i) $\quad E_0 [e^{-p \tilde{L}([0,t])} \mid Y_1^n , A_1^n]$

$$= \exp\{ -\int_0^t L_{0c}(ds) \int_{(0,1]} \frac{1-(1-u)^p}{u} \, V(s,u) \, R_0^-(s,du) \} .$$

(ii) $\quad \forall\, t \in I \cap J^c$,

$$P_0 [\, L^-(t) \in du \mid Y_1^n , A_1^n \,] \propto V(t,u) \, P_0 [\, L^-(t) \in du \,] .$$

(iii) $\quad$ If $Z_j \in I$,

$$P_0 [\, L^-(Z_j) \in du \mid Y_1^n , A_1^n \,] \propto W(Z_j , u) \, P_0 [\, L^-(Z_j) \in du \,] .$$

(iv) $\quad$ If $Z_j \in I^c$ and $D_j > 0$,

$$P_0 [\, L^-(Z_j) \in du \mid Y_1^n , A_1^n \,] \propto W(Z_j , u) \, \frac{1}{u} \, R_0^-(Z_j , du) .$$

(v) $\quad$ If $Z_j \in I^c$ and $D_j = 0$,

$$L^-(Z_j) = 0 .$$

If exclusive censoring, i.e., A_1^n, is replaced by inclusive censoring, i.e., $\tilde{A}_1^n$, it suffices to replace W by $\tilde{W}$ in (iii) and (iv).

■

In the next section, one of the main results of this paper, i.e., the posterior distribution of F with respect to P will have exactly the same form as in the above proposition. Indeed, it suffices to replace V, W and $\tilde{W}$ by V_θ, W_θ and $\tilde{W}_\theta$.

Note that if $N_{n*}^-(dt, du)$ stands for the posterior Levy measure of L^- with respect to P_0, then (i) means that

$$N_{n*}^-(dt, du) = V(t,u) \, N_0^-(dt, du) . \tag{86}$$

Note also that Proposition 5.1 (iv) provides an interpretation of the transition probability R_0^-. Indeed, for a single non censored observation falling outside the fixed points of discontinuity, this is translated into

$$P_0 [L^- (T) \in du \mid T] = R_0^- (T , du) . \tag{87}$$

From this proposition, it is also possible to compute the posterior expectation of the baseline distribution, under exclusive or inclusive censoring, for Beta, Gamma or Non informative priors. But such results appeared (sometimes in special cases) in the papers cited at the beginning of this subsection and some of these computations will be provided in the general model, i.e., under P.

6 Bayesian Analysis of the General Model

This section presents the main results of this paper. It extends considerably known results in several directions: the general model can be particularized to proportional hazards models and multiplicative intensity models, under general neutral to the right priors, having Beta and Gamma processes as particular cases, and for data with exclusive or inclusive censoring and presenting ties.
The Kalbfleisch's pioneering paper considers a semiparametric proportional hazards model with homogeneous Gamma prior whose parameter has a density and deals more or less successfully with the presence of ties. Extensions of these results for a neutral to the right prior but for constant covariates has been obtained in Wild and Kalbfleisch (1981). For Dirichlet prior distributions, Ruggiero (1994) and Gouget and Raoult (1998) obtain some results on the posterior density of the parameter of a semiparametric proportional hazards model in the presence of ties. Walker, Damien and Laud (1998) extends the paper of Wild and Kalbfleisch (1981) to time dependent covariates with no fixed point of discontinuity in the prior and in the absence of ties in the semiparametric proportional hazards model. They obtain the posterior distribution of the neutral to the right process conditional on the parameter and the posterior distribution of the parameter conditional on the baseline distribution. They complete their Bayesian analysis by

MCMC methods to obtain estimations of the parameter. To the best of the author's knowledge, these papers are the major contributions on that subject.

6.1 Predictive Densities of the Censored Observations

The main problem encountered by researchers to compute the predictive densities was the consequence of the methodology they used: computing at first the predictive (i.e., integrated on the baseline distribution) survival distribution of the sample and obtaining by differentiation the densities with respect to Lebesgue measure (under the assumption of absolute continuity of the expectation of the baseline survival function). To do so, every configuration of ties had to be considered (e.g., see Kalbfleisch (1978), Ruggiero (1994) and Gouget and Raoult (1998)).

In this paper, the computation of the predictive density with respect to P_0, is straightforward. Indeed, it suffices to compute

$$p\,(\,Y_1^n\,,\,A_1^n\,\mid\,\theta_1^n\,)\;=\;E_0\,[\,p\,(\,Y_1^n\,,\,A_1^n\,\mid\,\theta_1^n\,,\,F\,)\,\mid\,Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]\,. \quad (88)$$

Now, Property (45), i.e.,

$$F \perp\!\!\!\perp (\,C_1^n\,,\,\theta_1^n\,)\;\mid\;Y_1^n\,,\,A_1^n\,;\;P_0$$

means that the distribution of F conditional on $(\,Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,)$ with respect to P_0, does not depend on θ_1^n and is therefore the distribution presented in Proposition 5.1. The densities exhibited in Lemma 4.1 are then product of independent elements and this yields the following theorem whose proof is given in details in Appendix 2 :

Theorem 6.1 *The predictive densities of the censored observations are given by*

(i) for exclusive censoring

$$p\,(\,Y_1^n\,,\,A_1^n\,\mid\,\theta_1^n\,)\;=\;\overline{V}_\theta\;\times\;\prod_{t\,\in\,I\,\cap\,J^c}\overline{V}_\theta\,(t)\;\times\;\prod_{1\leq j\leq M}\overline{W}_\theta\,(\,Z_j\,)\,,$$

(ii) and for inclusive censoring

$$p\,(\,Y_1^n\,,\,\tilde{A}_1^n\;\mid\;\theta_1^n\,)\;=\;\overline{V}_\theta\;\times\prod_{t\,\in\,I\,\cap\,J^c}\overline{V}_\theta\,(t)\;\times\prod_{1\le j\le M}\overline{\tilde{W}_\theta}\,(\,Z_j\,)\,,$$

where, for exclusive censoring,

$$(i)\quad \overline{V}_\theta = \exp\,[\,-\int_0^\infty L_{0c}\,(dt)\int_{(0,1]}\frac{V\,(t,u)-V_\theta\,(t,u)}{u}\,R_0^-\,(t,du)\,]\,.$$

$$(ii)\quad \overline{V}_\theta\,(t)\;=\;\frac{\int_{(0,1]}\;V_\theta\,(t\,,\,u)\;P\,[\,L^-(t)\;\in\;du\,]}{\int_{(0,1]}\;V\,(t\,,\,u)\;P\,[\,L^-(t)\;\in\;du\,]}\,.$$

(iii) If $Z_j \in I$,

$$\overline{W}_\theta\,(Z_j)\;=\;\frac{\int_{(0,1]}\;W_\theta\,(Z_j\,,\,u)\;P\,[\,L^-(Z_j)\;\in\;du\,]}{\int_{(0,1]}\;W\,(Z_j\,,\,u)\;P\,[\,L^-(Z_j)\;\in\;du\,]}\,.$$

(iv) If $Z_j \in I^c$ and $D_j > 0$,

$$\overline{W}_\theta\,(Z_j)\;=\;\frac{\int_{(0,1]}\;W_\theta\,(Z_j\,,\,u)\;u^{-1}\;R_0^-\,(Z_j\,,\,du)}{\int_{(0,1]}\;W\,(Z_j\,,\,u)\;u^{-1}\;R_0^-\,(Z_j\,,\,du)}\,.$$

(v) If $Z_j \in I^c$ and $D_j = 0$,

$$\overline{W}_\theta\,(Z_j)\;=\;1\,.$$

For inclusive censoring, $\overline{\tilde{W}_\theta}\,(Z_j)$ *is defined as in* (iii) *and* (iv) *by replacing* W_θ *and* W *by* $\tilde{W}_\theta$ *and* $\tilde{W}$.

■

6.2 Semiparametric Models

This section uses the preceding general result for the analysis of semiparametric models.

Note first that the predictive density given in Theorem 6.1 for exclusive censoring (mutatis mutandis, for inclusive censoring) is equivalently written as

$$p\,(\,Y_1^n\,,\,A_1^n\;|\;\theta_1^n\,)\;=\;p\,(\,Y_1^n\,,\,A_1^n\;|\;\beta\,,\,X_1^n\,)\,. \tag{89}$$

This density, being integrated on the nuisance parameter, (i.e., F) is a genuine sampling density (i.e., of expectation 1 with respect to the dominating probability measure) contrary to likelihoods maximized on the nuisance parameter. In the Bayesian framework, the posterior density of β with respect to P_0 is given by

$$p\,(\,\beta\;|\;Y_1^n\,,\,A_1^n\,,\,X_1^n\,)\;=\;\frac{p\,(\,Y_1^n\,,\,A_1^n\;|\;\beta\,,\,X_1^n\,)}{p\,(\,Y_1^n\,,\,A_1^n\;|\;X_1^n\,)}\,, \tag{90}$$

where

$$p\,(\,Y_1^n\,,\,A_1^n\,|\,X_1^n\,)=E_0\,[\,p\,(\,Y_1^n\,,\,A_1^n\;|\;\beta\,,\,X_1^n\,)\,|\,Y_1^n\,,\,A_1^n\,,\,X_1^n\,]\,. \tag{91}$$

Now, Property (47) implies that

$$\beta \perp\!\!\!\perp (\,Y_1^n\,,\,A_1^n\,,\,X_1^n\,)\,;\,P_0\,, \tag{92}$$

and therefore, if

$$P\circ\beta^{-1}\;=\;P_0\circ\beta^{-1}\;=\;m_0\,, \tag{93}$$

$$p\,(\,Y_1^n\,,\,A_1^n\,|\,X_1^n\,)\;=\;\int_A p\,(\,Y_1^n\,,\,A_1^n\;|\;\beta\,,\,X_1^n\,)\,m_0(d\beta)\,. \tag{94}$$

Hence, this gives the following corollary:

Corollary 6.2 *In a semiparametric model, the posterior density of the parameter is given by*

(i) for exclusive censoring

$$p\,(\beta\,|\,Y_1^n\,,\,A_1^n\,,\,X_1^n)\;\propto\;e^{-\,V_c(\beta)}\;\times\prod_{t\in I\cap J^c}V_t\,(\beta)\;\times\prod_{1\le j\le M}W_j(\beta)\,,$$

(ii) and for inclusive censoring

$$p\,(\beta \mid Y_1^n\,,\, A_1^n\,,\, X_1^n) \,\propto\, e^{-\,V_c(\beta)} \,\times \prod_{t \in I \cap J^c} V_t\,(\beta) \,\times \prod_{1 \le j \le M} \tilde{W}_j(\beta)\,,$$

where, for exclusive censoring,

$$(i) \quad V_c(\beta) \,=\, \int_0^\infty L_{0c}\,(dt) \int_{(0,1]} \frac{1 \,-\, V_\theta\,(t,u)}{u}\, R_0^-\,(t,du)\,.$$

$$(ii) \quad V_t\,(\beta) \,=\, \int_{(0,1]} V_\theta\,(t,u)\; P\,[\,L^-(t) \,\in\, du\,]\,.$$

(iii) If $Z_j \,\in\, I$,

$$W_j(\beta) \,=\, \int_{(0,1]} W_\theta\,(Z_j\,,\, u)\; P\,[\,L^-(Z_j) \,\in\, du\,]\,.$$

(iv) If $Z_j \,\in\, I^c$ and $D_j \,>\, 0$,

$$W_j(\beta) \,=\, \int_{(0,1]} W_\theta\,(Z_j\,,\, u)\, u^{-1} R_0\,(Z_j\,,\, du\,)\,.$$

(v) If $Z_j \,\in\, I^c$ and $D_j \,=\, 0$,

$$W_j(\beta) \,=\, 1\,.$$

For inclusive censoring, $\tilde{W}_j(\beta)$ *is defined as in* (iii) *and* (iv) *by replacing* W_θ *by* $\tilde{W}_\theta$ *.*

■

In the next section, the posterior distribution of the baseline distribution in this very general model is completely described.

6.3 Posterior Distribution of the Baseline Distribution

The posterior density of the baseline distribution conditional on the explanatory processes is given by

$$\begin{aligned} p\,(\,F \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,) \quad &= \quad \frac{p\,(Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,,\,F\,)}{p\,(\,Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,)} \\ &= \quad \frac{p\,(Y_1^n\,,\,A_1^n \;\mid\; \theta_1^n\,,\,F\,)}{p\,(\,Y_1^n\,,\,A_1^n \;\mid\; \theta_1^n\,)}\,, \end{aligned} \tag{95}$$

where the numerator is given in Lemma 4.1 and the denominator in Theorem 6.1. This expression shows that this posterior density is again a product of independent factors. However, this nonparametric density of the baseline distribution (a measure derived from a neutral to the right process) is not very useful. The distribution of a Levy process, the hazard function L in this context, is generally characterized by its Laplace transform. In view of Proposition 5.1, the hazard function is a sum (and then the survival function is a product) of independent factors with respect to P_0. This shows that the representation of L as a sum of independent terms provided in Proposition 5.1 under P_0 is still valid under P, and the distribution of each term may be computed separately. This yields the following theorem whose a detailed proof is given in Appendix 3:

Theorem 6.3 *A posteriori, the hazard measure of the baseline distribution is a Levy process that can be written as*

$$L \;=\; \tilde{L} \;+\; \sum_{t \,\in\, I \,\cap\, J^c} L\,(t)\;\varepsilon_t \;+\; \sum_{1 \le j \le M} L\,(Z_j)\;\varepsilon_{Z_j}\,,$$

and similarly,

$$L^- \;=\; \tilde{L}^- \;+\; \sum_{t \,\in\, I \,\cap\, J^c} L^-\,(t)\;\varepsilon_t \;+\; \sum_{1 \le j \le M} L^-\,(Z_j)\;\varepsilon_{Z_j}\,.$$

Now, for exclusive censoring,

$$(i) \qquad E\,[\,e^{-p\,\tilde{L}\,(\,[0,t]\,)} \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$= \exp\{ - \int_0^t L_{0c}(ds) \int_{(0,1]} \frac{1-(1-u)^p}{u} \, V_\theta\,(s,u)\; R_0^-(s,du) \,\} \,.$$

(ii) $\quad \forall\, t \,\in\, I \,\cap\, J^c \,,$

$$P\,[\,L^-(t) \in du \;\mid\; Y_1^n \,,\, A_1^n \,,\, \theta_1^n\,] \;\propto\; V_\theta\,(t,u)\; P\,[\,L^-(t) \,\in\, du\,] \,.$$

(iii) $\quad$ If $\;Z_j \,\in\, I\,,$

$$P\,[\,L^-(Z_j) \in du \;\mid\; Y_1^n \,,\, A_1^n \,,\, \theta_1^n\,]$$

$$\propto\; W_\theta\,(\,Z_j\,,\,u\,)\; P\,[\,L^-(\,Z_j\,) \in du\,] \,.$$

(iv) $\quad$ If $\;Z_j \,\in\, I^c\;$ and $\;D_j \,>\, 0\,,$

$$P\,[\,L^-\,(\,Z_j\,) \,\in\, du \;\mid\; Y_1^n \,,\, A_1^n \,,\, \theta_1^n\,]$$

$$\propto\; W_\theta\,(\,Z_j\,,\,u\,)\; u^{-1}\; R_0^-\,(\,Z_j\,,\,du\,) \,.$$

(v) $\quad$ If $\;Z_j \,\in\, I^c\;$ and $\;D_j \,=\, 0\,,$

$$L^-\,(\,Z_j\,) \;=\; 0 \,.$$

For inclusive censoring, it suffices to replace W_θ by $\tilde{W}_\theta$ in (iii) and (iv).

∎

Note again that if $N_{n_*}^-\,(\,dt,du)$ stands for the posterior Levy measure of L^- under P, i.e., in the general model, then (i) means that

$$N_{n_*}^-\,(\,dt,du) \;=\; V_\theta\,(t,u)\,N_0^-\,(\,dt,du) \,. \tag{96}$$

The Bayesian analysis of the general model is therefore (theoretically) complete. The rest is a matter of computation.

To simplify the presentation (without much loss of generality), it is assumed, from now on, that the hazard function has no prior fixed point of discontinuity, i.e., $I = \emptyset$ and therefore, $L_0 = L_0^- = L_{0c}$.

6.4 Bayesian Estimators of the Survival Function

The aim of this section is to provide Bayesian estimators (under a square error loss function) of the survival distribution of the baseline distribution conditional on the explanatory processes, i.e.,

$$F_{n*}\,(\,(t,\infty]\,) \;=\; E\,[\,F(\,(t,\infty]\,)\;|\;Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]\,. \tag{97}$$

Since F is a posteriori a neutral to the right process, thanks to Property (72),

$$F_{n*}\,(\,(t,\infty]\,) \;=\; e^{-L_{n*\,c}\,(\,[0,t]\,)} \;\times \prod_{\{j\,:\,Z_j\,\leq\,t\}} [\,1 - L^-_{n*}\,(Z_j)\,]\,, \tag{98}$$

where,

$$L^-_{n*}\,(\,[0,t]\,) \;=\; E\,[\,L^-(\,[0,t]\,)\;|\;Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]\,. \tag{99}$$

Consequently,

$$L_{n*\,c}\,(\,[0,t]\,) \;=\; E\,[\,\tilde{L}^-(\,[0,t]\,)\;|\;Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,] \tag{100}$$

$$=\; \int_0^t L_0\,(ds) \int_{(0,1]} V_\theta\,(s,u)\,R_0^-\,(s,du)\,,$$

and

$$L^-_{n*}\,(Z_j) \;=\; E\,[\,L^-(\,Z_j\,)\;|\;Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,] \tag{101}$$

$$=\; \frac{\int_{(0,1]} W_\theta\,(Z_j\,,\,u)\;R_0^-\,(Z_j\,,\,du)}{\int_{(0,1]} W_\theta\,(Z_j\,,\,u)\;u^{-1}\;R_0^-\,(Z_j\,,\,du)}\,.$$

Note that there is no special additional difficulty for computing $E\,[\,\{F(\,(t,\infty]\,)\}^p\;|\;Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$. In particular, the variance of the baseline survival distribution can easily be obtained.

6.5 Prediction of the Survival Function

Let T_0 be a future lifetime of an individual with explanatory process θ_0 satisfying the assumptions of Section 4 (with $C_0 = \infty$). Then, its survival function is given by

$$F_{\theta_0}\,(\,(t,\infty]\,) \;=\; P\,[T_0 > t \mid \theta_0\,,\,F\,] \,=\, e^{-\,L_{\theta_0}\,(\,[0,t]\,)} \qquad (102)$$

$$=\; \prod_{0\,\leq\, s\,\leq\, t} [\,1 - L^{-}_{\theta_0}\,(s)\,]\,.$$

The posterior Bayesian estimator of this survival function is given by

$$F_{\theta_0\,n_*}\,(\,(t,\infty]\,) \qquad (103)$$

$$=\, E\,[\,F_{\theta_0}(\,(t,\infty]\,) \mid Y_1^n\,,\,A_1^n\,,\,\theta_0^n\,]$$

$$=\, e^{-L_{\theta_0\,n_*\,c}\,(\,[0,t]\,)} \;\times \prod_{\{j\,:\,Z_j\,\leq\, t\}} [\,1 - L^{-}_{\theta_0\,n_*}\,(Z_j)\,]\,,$$

where

$$L_{\theta_0\,n_*\,c}\,(\,[0,t]\,) \,=\, E\,[\,\tilde{L}^{-}_{\theta_0}(\,[0,t]\,) \mid Y_1^n\,,\,A_1^n\,,\,\theta_0^n\,] \qquad (104)$$

$$=\, \int_0^t L_0\,(ds) \int_{(0,1]} \frac{k\,[\,\theta_0\,(s),u\,]}{u}\, V_\theta\,(s,u)\, R_0^{-}\,(s,du)\,,$$

and

$$L^{-}_{\theta_0\,n_*}\,(Z_j) \,=\, E\,[\,L^{-}_{\theta_0}(\,Z_j\,) \mid Y_1^n\,,\,A_1^n\,,\,\theta_0^n\,] \qquad (105)$$

$$=\, \frac{\int_{(0,1]}\, k\,[\,\theta_0\,(Z_j)\,,u\,]\, W_\theta\,(Z_j\,,\,u)\,u^{-1}\, R_0^{-}\,(Z_j\,,\,du)}{\int_{(0,1]}\, W_\theta\,(Z_j\,,\,u)\,u^{-1}\, R_0^{-}\,(Z_j\,,\,du)}\,.$$

Here again, there is no special additional difficulty for computing $E\,[\,\{F_{\theta_0}(\,(t,\infty]\,)\}^p \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$. In particular, the variance of the survival distribution can easily be obtained.

It is worthwhile to consider the prior Bayesian estimator of this survival function given by

$$F_{\theta_0\,0}\,(\,(t,\infty]\,)\;=\;E\,[\,F_{\theta_0}(\,(t,\infty]\,)\mid\theta_0\,]\;=\;e^{\,-L_{\theta_0\,0}\,(\,[0,t]\,)}\;, \qquad (106)$$

where

$$L_{\theta_0\,0}\,(\,[0,t]\,)\;=\;E\,[\,L^-_{\theta_0}(\,[0,t]\,)\;\mid\;\theta_0\,] \qquad (107)$$

$$=\int_0^t L_0\,(ds)\int_{(0,1]}\frac{k\,[\,\theta_0\,(s),u\,]}{u}\,R^-_0\,(s,du)\,.$$

This shows that the prior expectation of the survival function is a proportional hazards model.

7 Two Particular Models

In this section, explicit computations are made in the two particular models presented in Section 2, viz., the proportional hazards model and the multiplicative intensity model.

7.1 The Proportional Hazards Model

To the best of the author's knowledge, only this model has been considered in Bayesian survival analysis. Recall that the proportional hazards model is defined by

$$L_\theta\,(t)\;=\;b\,[\,\theta(t)\,]\,L\,(t)\;\doteq\;\varepsilon(t)\,L(t)\,, \qquad (108)$$

and this amounts to choose

$$k[\,\theta(t)\,,\,u\,]\;=\;1\;-\;(1-u)^{\,\varepsilon(t)}\,. \qquad (109)$$

/noindent Note first that, a priori, by Formula (107),

$$L_{\theta\,0}\,(\,[0,t]\,)\;=\;E\,[\,L^-_\theta\,(\,[0,t]\,)\;\mid\;\theta\,]\;=\;\int_0^t r_0\,[\,\varepsilon\,(s),s\,]\,L_0\,(ds)\,, \qquad (110)$$

where the function r_0 is defined in Formula (78) and, for Beta or Gamma prior specification, takes the following form:

$$r_0(p,t) = \frac{\varphi[\,c_0(t) + p\,] - \varphi[\,c_0(t)\,]}{\varphi[\,c_0(t) + 1\,] - \varphi[\,c_0(t)\,]}\,, \tag{111}$$

where $\varphi(x) = \psi(x)$ for a Beta prior specification (Formula (80)) and $\varphi(x) = \ln(x)$ for a Gamma prior specification (Formula (83)).

In this proportional hazards model, the functions V_θ , W_θ and $\tilde{W}_\theta$ are relatively simple and easy to integrate with a Beta prior specification. This may explain the popularity of this model. Indeed,

$$V_\theta(t\,,\,u) = (1-u)^{\,\hat{N}(t)}\,, \tag{112}$$

where

$$\hat{N}(t) = \sum_{i\,\in\,K(t)} \varepsilon_i(t)\,, \tag{113}$$

$$W_\theta(Z_j\,,\,u) = (1-u)^{\,\hat{N}_j+\hat{R}_j} \prod_{i\,\in\,I_j} [\,1-(1-u)^{\,\varepsilon_{ij}}\,]\,, \tag{114}$$

and

$$\tilde{W}_\theta(Z_j\,,\,u) = (1-u)^{\,\hat{N}_j} \prod_{i\,\in\,I_j} [\,1-(1-u)^{\,\varepsilon_{ij}}\,]\,, \tag{115}$$

where

$$\hat{R}(t) = \sum_{i\,\in\,J(t)} \varepsilon_i(t)\,, \tag{116}$$

and, by the following general convention: for a function or a set $f(t)$, f_j is a short notation for $f(Z_j)$.
This simplicity is clear from the fact that I_j is generally a small set since it represents the indices of the uncensored observations at time Z_j and is a singleton in case of no ties.

The posterior distribution of the baseline distribution is entirely characterized by the following theorem whose proof is given in Appendix 4:

Theorem 7.1 *In the proportional hazards model with exclusive censoring,*

$$(i) \qquad E\,[\,e^{-p\,\tilde{L}\,(\,[0,t]\,)} \mid Y_1^n, A_1^n, \theta_1^n\,] \;=\; \exp\{-\int_0^t r_{n_*}(p,s)\,L_0(ds)\,\}\,,$$

where

$$r_{n_*}(p\,,\,t) \;=\; r_0\,(\,\hat{N}\,(t)+p\,,\,t) \;-\; r_0\,(\,\hat{N}\,(t)\,,\,t)\,.$$

$$(ii) \qquad E\,[\,\{1-L^-\,(Z_j)\}^{\,p} \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$= \frac{\sum_{0\le k\le D_j}(-1)^{\,k-1}\sum_{\{L\subset I_j:|L|=k\}} r_0(\hat{N}_j+\hat{R}_j+\varepsilon_{L\,j}+p\,,\,Z_j)}{\sum_{0\le k\le D_j}(-1)^{\,k-1}\sum_{\{L\subset I_j:|L|=k\}} r_0(\hat{N}_j+\hat{R}_j+\varepsilon_{L\,j}\,,\,Z_j)}\,,$$

where

$$\varepsilon_{L\,j} \;=\; \sum_{i\in L}\varepsilon_{i\,j}\,.$$

For inclusive censoring, it suffices to replace $\hat{R}_j$ *by* 0 *.*

■

Note that, in expression (ii), the function $r_0\,(\,p\,,\,t\,)$ may be replaced by the function $\psi\,[\,c_0\,(t)+p\,]$ for a Beta prior specification and by $\ln\,[\,c_0\,(t)+p\,]$ for a Gamma prior specification.

The Bayesian estimator of the baseline survival function is obtained by setting $p=1$ and the posterior variance of the baseline survival function can be deduced from the above relations with $p=2$. However no real simplification occurs in these expressions. The posterior distribution of the survival function of an individual with explanatory process θ_0, is obtained from the above Theorem by replacing p by $p\,\varepsilon_0\,(t)$. Indeed,

$$1-L^-_{\theta_0}\,(t) \;=\; [\,1-L^-(t)\,]^{\,\varepsilon_0\,(t)}\,. \tag{117}$$

Note also that, when $D_j = 1$,

$$E\,[\,\{1 - L^-\,(Z_j)\}^{\,p}\ \mid\ Y_1^n\,,\ A_1^n\,,\ \theta_1^n\,] \tag{118}$$

$$= \frac{r_0\,(\,\hat{N}_j + \hat{R}_j + \varepsilon_j + p\,,\ Z_j\,) - r_0\,(\,\hat{N}_j + \hat{R}_j + p\,,\ Z_j\,)}{r_0\,(\,\hat{N}_j + \hat{R}_j + \varepsilon_j\,,\ Z_j\,) - r_0\,(\,\hat{N}_j + \hat{R}_j\,,\ Z_j\,)}\,.$$

In a semiparametric proportional hazards model, as presented in Section 2, $\varepsilon_i\,(t)\ =\ e\,[\,\beta\,,\ X_i(t)\,]$ and, in particular, the Cox's model specifies $\varepsilon_i(t)\ =\ \exp\,[\,\beta' X_i(t)\,]$. The posterior density of the parameter β is given in the following theorem:

Theorem 7.2 *In a semiparametric proportional hazards model with exclusive censoring,*

$$p\,(\beta \mid Y_1^n\,,\ A_1^n\,,\ X_1^n)\ \propto\ e^{-\,V(\beta)}\ \times \prod_{1 \le j \le M} W_j(\beta)\,,$$

where

$$V(\beta)\ =\ \int_0^{Z_M} r_0\,(\hat{N}(t)\,,\ t)\,L_0\,(dt)\,,$$

and

$$W_j(\beta)\ =\ \sum_{0 \le k \le D_j} (-1)^{\,k-1} \sum_{\{L \subset I_j : |L| = k\}} r_0\,(\,\hat{N}_j + \hat{R}_j + \varepsilon_{L\,j}\,,\ Z_j\,)\,.$$

Results for inclusive censoring are obtained by setting $\hat{R}_j\ =\ 0$.

∎

Here again, in the last expression, the function $r_0\,(\,p\,,\ t\,)$ may be replaced by the function $\psi\,[\,c_0\,(t) + p\,]$ for a Beta prior specification and by $\ln\,[\,c_0\,(t) + p\,]$ for a Gamma prior specification . /vskip 10pt It is interesting to note that, in all these results, the explanatory process is closely linked to the prior specification through the function r_0.

7.2 The Multiplicative Intensity Model

Recall that the multiplicative intensity model is defined by

$$L_\theta^-(t) \;=\; a\,[\,\theta(t)\,]\,L^-(t) \;\doteq\; \delta(t)\,L^-(t) \tag{119}$$

and this means that

$$k[\,\theta(t)\,,\,u\,] \;=\; \delta(t)\,u\,. \tag{120}$$

Note that a priori, by Formula (107),

$$L_{\theta\,0}\,(\,[0,t]\,) \;=\; E\,[\,L_\theta^-(\,[0,t]\,)\;\mid\;\theta\,] \;=\; \int_0^t \delta\,(s)\,L_0\,(ds)\,. \tag{121}$$

This is a remarkable property of this model. Indeed, whatever the transition probability of the neutral to the right process is, the prior expectation of the survival function is a proportional hazards model with the same accelerating function.

In the multiplicative intensity model, the functions V_θ, W_θ and $\tilde{W}_\theta$ are less simple and less easy to integrate with the Beta or Gamma prior specification. Indeed,

$$V_\theta\,(t,u) \;=\; \prod_{i\,\in\,K(t)} [\,1-\delta_i\,(t)\,u\,]\,, \tag{122}$$

$$W_\theta\,(Z_j,u) \;=\; \prod_{i\,\in\,I_j} \delta_{i\,j} \;\times\; u^{D_j} \;\times \prod_{i\,\in\,K_j\,\cup\,J_j} [\,1-\delta_{i\,j}\,u\,]\,, \tag{123}$$

and

$$\tilde{W}_\theta\,(Z_j,u) \;=\; \prod_{i\,\in\,I_j} \delta_{i\,j} \;\times\; u^{D_j} \;\times\; \prod_{i\,\in\,K_j} [\,1-\delta_{i\,j}\,u\,]\,. \tag{124}$$

The posterior distribution is given in the following theorem whose proof is provided in Appendix 5:

Theorem 7.3 *In the multiplicative intensity model with exclusive censoring,*

(i) $$E\,[\, e^{-p\,\tilde{L}\,(\,[0,t]\,)} \mid Y_1^n, A_1^n, \theta_1^n\,] \;=\; \exp\{\, -\int_0^t r_{n_*}(p,s)\, L_0(ds)\,\}\,,$$

where

$$r_{n_*}(p\,,\,t) \;=\; \sum_{0 \le k \le N(t)} \mu_k\,(t)\,[\, r_0\,(\,k+p\,,\,t) \;-\; r_0\,(\,k\,,\,t)\,]\,,$$

and

$$\mu_k\,(t) \;=\; \prod_{i\,\in\,K(t)} [\,1-\delta_i(t)\,] \;\times \sum_{\{L\subset K(t)\,:\,|L|\,=\,k\}} \;\prod_{i\,\in\,L} \frac{\delta_i(t)}{1-\delta_i(t)}\,.$$

(ii) $$E\,[\,\{1-L^-\,(Z_j)\}^p \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$=\; \frac{\sum_{0\le k\le N_j+R_j} \gamma_{j\,k}\; s_0\,[\,D_j\,,\,p+k\,,\,Z_j\,]}{\sum_{0\le k\le N_j+R_j} \gamma_{j\,k}\; s_0\,[\,D_j\,,\,k\,,\,Z_j\,]}\,,$$

where

$$\gamma_{j\,k} \;=\; \sum_{\{L\subset K_j\cup J_j\,:\,|L|\,=\,k\}} \;\prod_{i\,\in\,L} \frac{\delta_{i\,j}}{1-\delta_{i\,j}}\,,$$

and

$$s_0\,(m,p,t) \;=\; \int_{(0,1]} u^{\,m-1}\,(1-u)^p\, R_0\,(t,du)\,.$$

For inclusive censoring, it suffices to replace J_j *by* $\emptyset$ *and* R_j *by* 0*.*

■

Note first that the function s_0 may be expressed in terms of the function r_0. Indeed,

$$s_0(m,p,t) \;=\; \sum_{0\le k\le m} C_m^k\,(-1)^{k-1}\,r_0(p+k,t)\,. \tag{125}$$

For a Beta prior specification, the function s_0 has a known analytic form and in (ii), the function $s_0(m,p,t)$ may replaced by

$$\frac{\Gamma[\,c_0(t)+p\,]}{\Gamma[\,c_0(t)+p+m\,]}\,.$$

For Gamma prior specification, such a closed form is not available and $s_0(m,p,t)$ may be replaced by

$$\sum_{0\le k\le m} C_m^k\,(-1)^{k-1}\,\ln[\,c_0(t)+p+k\,]\,.$$

Note also that, $s_0(1,p,t)$ may be replaced by $[\,c_0(t)+p\,]^{-1}$ for a Beta prior specification and by $\ln[\,1+[\,c_0(t)+p\,]^{-1}\,]$ for a Gamma prior specification. Here again, no real simplification occurs in the computation of the Bayesian estimator or the posterior variance of the baseline survival function. For the prediction of the survival function of an individual with explanatory process θ_0 it suffices to note that

$$\begin{aligned} L^-_{\theta_0}(\,[0,t]\,) \;&=\; \int_{[0,t]} \delta_0(s)\,L^-(ds) \\ &=\; \int_0^t \delta_0(s)\tilde{L}^-(ds) \;+\; \sum_{\{\,j\,:\,Z_j\le t\}} \delta_0(\,Z_j\,)\,L^-(\,Z_j\,)\,. \end{aligned} \tag{126}$$

In a semiparametric multiplicative intensity model, as presented in Section 2, $\delta_i(t) \;=\; d(\,\beta, X_i(t)\,)$ and it was suggested to use a logistic model for this accelerating factor. The posterior density of the parameter is given in the following theorem:

Theorem 7.4 *In a multiplicative intensity model with exclusive censoring,*

$$p(\beta \mid Y_1^n\,,\,A_1^n\,,\,X_1^n) \;\propto\; e^{-\,V(\beta)} \;\times \prod_{1\le j\le M} W_j(\beta)\,,$$

where

$$V(\beta) \,=\, \int_0^{Z_M} r_{n_*}(t)\, L_0\,(dt)$$

with

$$r_{n_*}(t) \,=\, \sum_{0 \le k \le N(t)} \mu_k\,(t)\, r_0\,(\,k\,,\,t\,)\,,$$

and

$$W_j(\beta) \,=\, \nu_j \sum_{0 \le k \le N_j + R_j} \gamma_{j\,k}\; s_0\,[\,D_j\,,\,k\,,\,Z_j\,]$$

with

$$\nu_j \,=\, \prod_{i \in I_j} \delta_{i\,j} \,\times \prod_{i \in K_j \,\cup\, J_j} [\,1 - \delta_{i\,j}\,]\,.$$

Results for inclusive censoring are obtained by setting $J_j \,=\, \emptyset$ *and* $R_j \,=\, 0\,.$

■

The results obtained in Theorem 7.3 and Theorem 7.4 may seem more complicated than those obtained in Theorem 7.1 and Theorem 7.2 because they are expressed by summations over a possibly large number of terms (the sample size at most). However, these terms are product of two coefficients; the first one only depends on the explanatory process and the observations while the second one only depends on the prior distribution and possibly on the number of deaths in case of ties. Therefore, the multiplicative intensity model enjoys a reasonable robustness with respect to the choice of the prior distribution. Moreover these last summations of positive terms are easier to simulate than the alternate summations given in Theorem 7.1 and 7.2.

8 Noninformative Prior Specification

According to Definition 5.1, the expectation of the square of the baseline survival function is given by

$$E\,[\,F\,(\,(t,\infty]\,)^2\,] \,=\, e^{-2L_{0c}(\,[0,t]\,)} \tag{127}$$

$$\times \; e^{\int_0^t L_{0c}\,(ds) \int_{(0,1]} u\, R_0^-\,(s, du)} \times \prod_{0 \le s \le t} E\,[\,\{1 - L^-\,(s)\}^2\,]\,.$$

It implies that the variance of the baseline survival function is maximized by choosing L^- to be a Poisson process as in Definition 5.4. Note that in that case,

$$\begin{aligned} r_0\,(p\,,\,t) &= 1 \qquad \text{for} \qquad p \,>\, 0 \\ &= 0 \qquad \text{for} \qquad p \,=\, 0\,. \end{aligned} \tag{128}$$

This implies that $F \;=\; \varepsilon_S$ where S has distribution F_0. This may be still quite informative if F_0 is very concentrated. So, a second step in searching for noninformative priors is to choose F_0 with a large variance and this amounts to make L_0^- tend to 0.
It is easily seen that a Beta or a Gamma prior specification with $c_0\,(t) \;=\; n_0\, d_0\,(t)$ converges to the Poisson process presented in Definition 5.4 as $n_0 \;\to\; 0$.

The next section analyses the effect of noninformative priors on the inference in proportional hazards model.

8.1 Proportional Hazards Model

From Theorem 7.1 (i), it is readily seen that for a Poisson process,

$$E\,[\,e^{-p\,\tilde{L}\,(\,[0,t]\,)} \mid Y_1^n, A_1^n, \theta_1^n\,] \;=\; \exp\{\,-\int_0^t \mathbb{1}_{\{Z_M \le s\}}\, L_0(ds)\,\}\,. \tag{129}$$

Therefore, a posteriori, the predictable hazard measure can be written as

$$L^- \;=\; \sum_{\{\,j\,:\,D_j\,>\,0\,\}} L^-\,(Z_j)\,\varepsilon_{Z_j} \,+\, \varepsilon_S \tag{130}$$

where, a priori, S is distributed as F_0 and, a posteriori,

$$P\,[\,S \,>\, t \;\mid\; Y_1^n, A_1^n, \theta_1^n\,] \;=\; P\,[\,S \,>\, t \;\mid\; S \,>\, Z_M\,]\,. \tag{131}$$

It is also readily seen that, when $D_M > 0$, then, for inclusive censoring, and, for exclusive censoring in the case $R_M = 0$,

$$L^-(Z_M) = 1 . \tag{132}$$

Hence, S is therefore redundant because $F[(t, \infty]] = 0 \quad \forall\, t \geq Z_M$.

For the other cases, a straightforward application of Theorem 7.1 (ii) leads to an indetermination (i.e., $0/0$). However, for Beta or Gamma prior specification, it has been seen from Formula (111) that, in this expression, the function $r_0(p, t)$ can be replaced by the function $\varphi[c_0(t) + p]$ where $\varphi(x) = \psi(x)$ for a Beta prior specification (Formula (80)) and $\varphi(x) = \ln(x)$ for a Gamma prior specification (Formula (83)). This solves the indetermination problem and gives by letting n_0 tend to 0,

$$E[\{1 - L^-(Z_j)\}^p \mid Y_1^n, A_1^n, \theta_1^n] \tag{133}$$

$$= \frac{\sum_{0 \leq k \leq D_j} (-1)^{k-1} \sum_{\{L \subset I_j : |L| = k\}} \varphi(\hat{N}_j + \hat{R}_j + \varepsilon_{Lj} + p)}{\sum_{0 \leq k \leq D_j} (-1)^{k-1} \sum_{\{L \subset I_j : |L| = k\}} \varphi(\hat{N}_j + \hat{R}_j + \varepsilon_{Lj})} .$$

It may be shown that these results (Formulae (130), (131) and (133)) are still valid if the set of fixed points of discontinuity (I) is not empty.

It may be realized that noninformative prior specification does not simplify the inference on the baseline survival distribution.

Concerning the posterior density of the parameter in a semiparametric proportional hazards model, it is readily seen from Theorem 7.2, that

$$V(\beta) = L_0([0, Z_M)) \tag{134}$$

does not depend on β and this gives the following proposition:

Proposition 8.1 *In the semiparametric proportional hazards model with noninformative priors deduced from a Beta or a Gamma prior distribution,*

$$p(\beta \mid Y_1^n, A_1^n, X_1^n) \propto \prod_{\{j : D_j > 0\}} W_j(\beta) ,$$

and

$$W_j(\beta) \;=\; \sum_{0\le k\le D_j} (-1)^{k-1} \sum_{\{L\subset I_j : |L|=k\}} \varphi\,[\,A_{jL}\,(\,\beta\,)\,]\,.$$

For a Beta prior specification,

$$\varphi\,(x) \;=\; \psi\,(x)\,.$$

For a Gamma prior specification,

$$\varphi\,(x) \;=\; \ln\,(x)\,.$$

For exclusive censoring,

$$A_{jL}\,(\,\beta\,) \;=\; \sum_{i\,\in K_j\cup J_j\cup L} e\,[\,\beta\,,\,X_i\,(Z_j)\,]\,.$$

For inclusive censoring,

$$A_{jL}\,(\,\beta\,) \;=\; \sum_{i\,\in K_j\cup L} e\,[\,\beta\,,\,X_i\,(Z_j)\,]\,.$$

■

Note that, in case of no ties,

$$W_j(\beta) \;=\; \varphi\,[\,A_j\,(\,\beta\,)\,] \;-\; \varphi\,[\,A_{j+1}\,(\,\beta\,)\,]\,, \tag{135}$$

where

$$A_j\,(\,\beta\,) \;=\; \sum_{\{\,i\,:\,Y_i\,\ge\,Z_j\,\}} e\,[\,\beta\,,\,X_i\,(Z_j)\,]\,. \tag{136}$$

This proposition is relatively straightforward and deserves to be compared to the very painful deductions made by Kalbfleisch (1978) in the Gamma prior distribution case and by Ruggiero (1994) and Gouget and Raoult (1998) in the Beta prior distribution case when facing ties in the observations.

These Bayesian posterior densities are obtained through integration of the nuisance parameter under a noninformative limiting case of a family of prior distributions on this nuisance parameter. In case of constant explanatory processes, these densities only depend on the rank statistics of the sample. Cox's likelihood is obtained by marginalization on the rank statistics of the sample and so, some connections between the two likelihoods may be expected. However, these likelihoods are not only quite different from the Cox's likelihood but depend strongly on the considered family of prior distributions.

8.2 Multiplicative Intensity Model

From Theorem 7.3 (i), it is readily seen that for a Poisson process,

$$E\,[\, e^{-p\,\tilde{L}\,(\,[0,t]\,)} \mid Y_1^n, A_1^n, \theta_1^n\,] \;=\; \exp\{\,-\int_0^t \mu_0\,(s)\,L_0(ds)\,\}\,, \tag{137}$$

and, from Theorem 7.3 (ii) or Theorem 6.3 (iv), that

$$E\,[\,\{1 - L^-\,(Z_j)\}^p \;\mid\; Y_1^n\,,\, A_1^n\,,\, \theta_1^n\,] \;=\; 0\,. \tag{138}$$

Therefore, a posteriori, the predictable hazard measure can be written as

$$L^- \;=\; \sum_{\{\,j\,:\,D_j\,>\,0\,\}} \varepsilon_{\,Z_j} \;+\; \varepsilon_{\,S}\,, \tag{139}$$

where S is a random variable such that

$$P\,[\,S \;>\; t \;\mid\; Y_1^n\,,\, A_1^n\,,\, \theta_1^n\,] \;=\; \exp\{\,-\int_0^t \mu_0\,(s)\,L_0(ds)\}\,. \tag{140}$$

Therefore, if

$$I_u \;=\; \inf\,\{\,j\;:\;D_j\;>\;0\}\,, \tag{141}$$

then

$$F \;=\; \varepsilon_{\,S\,\wedge\,Z_{I_u}}\,. \tag{142}$$

It may be shown that, when the set of fixed points of discontinuity is not empty, the representation is still valid but S is F_1-distributed where

$$L_1^-\,(\,[0\,,\,t]\,)\;=\;\int_{[0\,,\,t]} \frac{\nu\,(s)}{1-[\,1-\nu\,(s)\,]\,L_0^-\,(s)}\,L_0^-(ds)\,.$$

A second step into the direction of noninformative prior, i.e., $L_0^- = 0$, leads to $F = \varepsilon_{Z_{I_u}}$.

So, in a multiplicative intensity model, the posterior distribution of the baseline distribution is degenerate. This is a little bit disappointing but not so strange, because, in this model, the baseline distribution is in some sense an extremal case never observed since $\delta_i(t) < 1$.

However, prediction is still feasible since,

$$L_{\theta_0}^- = \sum_{\{j : D_j > 0\}} \delta_0(Z_j)\, \varepsilon_{Z_j} + \delta_0(S)\, \varepsilon_S . \tag{143}$$

Concerning the posterior distribution of the parameter in a semi-parametric multiplicative intensity model, it is readily seen from Corollary 6.2 that, for a Poisson process,

(i) $\quad V_c(\beta) = \int_0^{Z_M} [1 - \mu_0(t)] L_{0c}(dt) .$

(ii) $\quad V_t(\beta) = 1 - [1 - \mu_0(t)]\, L_0^-(t) .$

(iii) $\quad$ If $Z_j \in I$ and $D_j = 0$,

$$W_j(\beta) = [1 - L_0^-(Z_j)] + \nu_j\, L_0^-(Z_j) .$$

(iv) $\quad$ If $Z_j \in I$ and $D_j > 0$,

$$W_j(\beta) = \nu_j\, L_0^-(Z_j) .$$

(v) $\quad$ If $Z_j \in I^c$ and $D_j > 0$,

$$W_j(\beta) = \nu_j .$$

(vi) $\quad$ If $Z_j \in I^c$ and $D_j = 0$,

$$W_j(\beta) \;=\; 1\,.$$

Therefore, a second step toward noninformative prior distribution yields the following proposition:

Proposition 8.2 *In a totally noninformative semiparametric multiplicative intensity model, the posterior distribution of the parameter is given by*

$$p\,(\beta \mid Y_1^n\,,\, A_1^n\,,\, X_1^n) \;\propto\; \prod_{\{\,j\,:\,D_j\,>\,0\}} W_j(\beta)\,,$$

where

(i) for exclusive censoring,

$$W_j(\beta) \;=\; \prod_{i \in I_j} d\,[\,\beta\,,\, X_i\,(Z_j)\,] \;\times \prod_{i \in K_j\,\cup\, J_j} [\,1 - d\,[\,\beta\,,\, X_i\,(Z_j)\,]\,]$$

and

(ii) for inclusive censoring,

$$W_j(\beta) \;=\; \prod_{i \in I_j} d\,[\,\beta\,,\, X_i\,(Z_j)\,] \;\times\; \prod_{i \in K_j} [\,1 - d\,[\,\beta\,,\, X_i\,(Z_j)\,]\,]\,.$$

■

In case of no ties, the two densities are the same and with obvious notations,

$$W_j(\beta) \;=\; d\,[\,\beta\,,\, X_{(j)}\,(\,Y_{(j)}\,)\,] \prod_{j\,<\,k\,\leq\,M} [\,1 - d\,[\,\beta\,,\, X_{(k)}\,(\,Y_{(j)}\,)\,]\,]\,. \quad (144)$$

This proposition is particularly interesting. Firstly, these posterior densities arise from a genuine Bayesian model specifying a Poisson process for the predictable hazard function. As noted at the beginning of this section, they may be seen as a noninformative limit case of any family of prior distributions (such as the Beta family or the Gamma family). Secondly, they are obtained under general assumptions including explanatory processes, censoring, possibly depending on these explanatory processes, and possible equalities in the observed lifetimes.

Comparison between these posterior densities and the Cox's likelihood, which is only justified under the assumption of no equality in the observed lifetimes, is easy since for the Cox's likelihood,

$$W_j(\beta) \;=\; \frac{d\,[\,\beta\,,\,X_{(j)}\,(\,Y_{(j)}\,)\,]}{\sum\limits_{j\,<\,k\,\leq\,M} d\,[\,\beta\,,\,X_{(k)}\,(\,Y_{(j)}\,)\,]}\,. \tag{145}$$

Therefore, it is of a great interest to study, in a sampling theory framework, the properties of this new likelihood since under the assumption of a continuous baseline distribution the multiplicative intensity model and the proportional hazards model are the same.

9 Some Conclusions

The general sampling model presented in Section 4 shows the irrelevance of the sometimes lengthy discussions comparing model with fixed (non random) censoring times versus model with independent and identically distributed censoring times, in the point of view of inference on the baseline distribution or on the parameter of a semiparametric model (e.g., see Susarla and Van Ryzin (1976), Ferguson and Phadia (1979) and especially Lo (1993)).

A key fact of this paper is the absolute continuity of the model with explanatory process with respect to the simple model (without explanatory variables). Such a property provides elegant and straightforward proofs in the most general case of the posterior distribution of the baseline distribution, and of the posterior distribution of the parameter of a semiparametric model. The complexity of the early descriptions of this last posterior distribution is shown, e.g., in Kalbfleisch (1978), Ruggiero (1994) and Gouget and Raoult (1999).

Concerning the choice of the prior distribution, the advantage of specifying the predictable hazard measure as in Definition 5.1 is shown in Section 5. Indeed, its expectation is σ-finite and is the predictable hazard measure of the expectation of the baseline distribution. Moreover, in the subsequent decomposition of the Levy measure, the transition probability receives a nice interpretation in the posterior distribution for a sample of size one.

The family of Beta processes for the predictable hazard function seems more adequate than the family of Gamma processes for the hazard function. Indeed, in the model without explanatory variables, only the first one is natural conjugate. For a Gamma process specification, the prior distribution of the fixed points of discontinuity are Gamma distributed but this is no longer true for the posterior distribution.

In general, it is rather natural to suppose that the true model for the observations is a proportional hazard model with a continuous baseline distribution function. A Bayesian statistician may therefore either use a proportional hazards model or a multiplicative intensity model since the generated baseline distribution functions are discrete. If the computations are easier in the proportional hazards model, the multiplicative intensity model presents several advantages. The prior expectation is a proportional hazards model with the same accelerating factor. There exists some kind of separation between the role played by the explanatory process and the impact of the prior specification. Moreover, the noninformative case provides a more sensible posterior distribution for the parameter of a semiparametric model in the sense that this distribution is independent of the chosen family of prior distributions.

Appendix

A.1 Proof of Lemma 4.1

The density of the censored observations conditional on the explanatory processes and on the baseline distribution function is given by

$$\begin{aligned} p\,(\,Y_1^n\,,\,A_1^n\;|\;\theta_1^n\,,\,F\,) &= \prod_{1\,\leq\,i\,\leq\,n} p\,(\,Y_i\,,\,A_i\;|\;\theta_i\,,\,F\,) \\ &= \prod_{1\,\leq\,j\,\leq\,M} \;\; \prod_{i\,\in\,I_j\,\cup\,J_j} p\,(\,Z_j\,,\,A_i\;|\;\theta_i\,,\,F\,)\,. \end{aligned}$$

where $I_j = I(Z_j)$ and $J_j = J(Z_j)$. The numerator of this density is the product of two terms. The first one is

$$\begin{aligned} &\prod_{1\,\leq\,j\,\leq\,M} \;\; \prod_{i\,\in\,I_j\,\cup\,J_j} L^-_{\theta_i}(Z_j)^{\,A_i}\;[1 - L^-_{\theta_i}(Z_j)]^{\,1-A_i} \\ &= \prod_{1\,\leq\,j\,\leq\,M} \left[\prod_{i\,\in\,I_j} L^-_{\theta_i}(Z_j) \;\times\; \prod_{i\,\in\,J_j} [1 - L^-_{\theta_i}(Z_j)]\right]\,, \end{aligned}$$

and the second one is

$$\begin{aligned} &\prod_{1\,\leq\,j\,\leq\,M} \;\; \prod_{i\,\in\,I_j\,\cup\,J_j} \exp[\,-\,L_{\theta_i}(\,[0, Z_j)\,)\,] \\ &= \prod_{1\,\leq\,j\,\leq\,M} \;\; \prod_{i\,\in\,I_j\,\cup\,J_j} \;\; \prod_{1\,\leq\,k\,\leq\,j} \exp[\,-\,L_{\theta_i}(\,(Z_{k-1}, Z_k)\,)\,] \\ &\quad\times \prod_{1\,\leq\,j\,\leq\,M} \;\; \prod_{i\,\in\,I_j\,\cup\,J_j} \;\; \prod_{1\,\leq\,k\,\leq\,j-1} \exp[\,-\,L_{\theta_i}(\,Z_k\,)\,]\,. \end{aligned}$$

The first term of the right hand side of the last equality may be written as

$$\begin{aligned} &\prod_{1\,\leq\,j\,\leq\,M} \;\; \prod_{j\,\leq\,k\,\leq\,M} \;\; \prod_{i\,\in\,I_k\,\cup\,J_k} \exp[\,-\,L_{\theta_i}(\,(Z_{j-1}, Z_j)\,)\,] \\ &= \prod_{1\,\leq\,j\,\leq\,M} \;\; \prod_{i\,\in\,K_{j-1}} \exp[\,-\,L_{\theta_i}(\,(Z_{j-1}, Z_j)\,)\,] \end{aligned}$$

$$= \prod_{1 \le j \le M} \prod_{i \in K_{j-1}} \prod_{Z_{j-1} < t < Z_j} \exp[\, - L_{\theta_i}(t)\,]$$
$$= \prod_{1 \le j \le M} \prod_{Z_{j-1} < t < Z_j} \prod_{i \in K_{j-1}} [\,1 - L_{\theta_i}^{-}(t)\,]\,,$$

where $K_{j-1} = K(Z_{j-1})$. Noticing that, for $Z_{j-1} < t < Z_j$, $K(t) = K_{j-1}$, this first term is then equal to

$$\prod_{1 \le j \le M} \prod_{Z_{j-1} < t < Z_j} V_\theta(t\,,\, L^{-}\,(t))\,.$$

Similarly, the second term of the right hand side may be written as

$$\prod_{1 \le j \le M} \prod_{i \in I_j \cup J_j} \prod_{1 \le k \le j-1} [\,1 - \, L_{\theta_i}^{-}(\,Z_k\,)\,]$$
$$= \prod_{1 \le j \le M-1} \prod_{j+1 \le k \le M} \prod_{i \in I_k \cup J_k} [\,1 - \, L_{\theta_i}^{-}(\,Z_j\,)\,]$$
$$= \prod_{1 \le j \le M-1} \prod_{i \in K_j} [\,1 - \, L_{\theta_i}^{-}(\,Z_j\,)\,]\,.$$

Combining this last term with the first term of the numerator of the density provides

$$\prod_{1 \le j \le M} \left[\prod_{i \in I_j} L_{\theta_i}^{-}(Z_j) \;\times \prod_{i \in K_j \cup J_j} [1 - L_{\theta_i}^{-}(Z_j)] \right]$$
$$= \prod_{1 \le j \le M} W_\theta(\,Z_j\,,\, L^{-}\,(\,Z_j\,))\,.$$

The denominator is obtained by replacing $L_{\theta_i}^{-}$ by L^{-}. This gives

$$V(t\,,\, L^{-}\,(t)) \;=\; [\,1 - L^{-}(t)\,]^{\,N(t)}\,,$$

and

$$W(t\,,\, L^{-}\,(t)) \;=\; L^{-}(t)^{\,D(t)} \;\; [\,1 - L^{-}(t)\,]^{\,N(t)+R(t)}\,.$$

■

A.2 Proof of Theorem 6.1

The first factor of the conditional density of the exclusively censored observations may be written as

$$\prod_{1 \le j \le M} \; \prod_{Z_{j-1} < t < Z_j} \frac{V_\theta(t\,,\, L^-\,(t))}{V(t\,,\, L^-\,(t))} \;=\; \prod_{0 \le t < \infty} \frac{V_\theta(t\,,\, \tilde{L}^-\,(t))}{V(t\,,\, \tilde{L}^-\,(t))} \times \prod_{t \,\in\, I \,\cap\, J^c} \frac{V_\theta(t\,,\, L^-\,(t))}{V(t\,,\, L^-\,(t))}\,.$$

It is therefore, under P_0, a product of independent terms.
Now, by Proposition 5.1 (i) (or Formula (86)) and Formula (67),

$$\begin{aligned} & E_0\,[\prod_{0 \le t < \infty} \frac{V_\theta(t\,,\, \tilde{L}^-\,(t))}{V(t\,,\, \tilde{L}^-\,(t))} \;\mid\; Y_1^n\,,\, A_1^n\,,\, \theta_1^n\,] \\ &= \exp\{-\int_0^\infty L_{0c}\,(dt) \int_0^1 [\,1 - \frac{V_\theta(t,u)}{V(t,u)}\,]\, V(t,u)\, u^{-1} R_0^-\,(t, du)\,\} \\ &= \exp\{-\int_0^\infty L_{0c}\,(dt) \int_0^1 [\,V(t,u) - V_\theta(t,u)\,]\, u^{-1} R_0^-\,(t, du)\,\}\,. \end{aligned}$$

By Proposition 5.2 (ii), for $t \,\in\, I \,\cap\, J^c$,

$$\begin{aligned} & E_0\,[\frac{V_\theta(t\,,\, L^-\,(t))}{V(t\,,\, L^-\,(t))} \;\mid\; Y_1^n\,,\, A_1^n\,,\, \theta_1^n\,] \\ &= \frac{\int_0^1 \frac{V_\theta(t,u)}{V(t,u)}\, V(t,u)\, P_0\,[\,L^-(t) \in du\,]}{\int_0^1 V(t,u)\, P_0\,[\,L^-(t) \in du\,]} \\ &= \frac{\int_0^1 V_\theta(t,u)\, P_0\,[\,L^-(t) \in du\,]}{\int_0^1 V(t,u)\, P_0\,[\,L^-(t) \in du\,]}\,. \end{aligned}$$

Similar computations show that for any Z_j,

$$E_0\,[\frac{W_\theta(Zj\,,\, L^-\,(Z_j))}{W(Z_j\,,\, L^-\,(Z_j))} \;\mid\; Y_1^n\,,\, A_1^n\,,\, \theta_1^n\,] \;=\; \overline{W_\theta}\,(Z_j)\,.$$

■

A.3 Proof of Theorem 6.2

Let f be any nonnegative Borel measurable function, i.e.,

$$f(t) \, : \, (\,\overline{\mathbb{R}}^{+}\,,\,\overline{\mathcal{B}}^{+}\,) \;\to\; (\,\mathbb{R}^{+}\,,\,\mathcal{B}^{+}\,)\,.$$

The distribution of L is entirely characterized by the Laplace transform of

$$L\,(\,f\,) \;=\; \int_{[\,0,\infty\,]} f(t)\,L\,(dt)\,.$$

Therefore, the problem is to compute

$$\begin{aligned}
& E\,[\,e^{-\,L\,(f)} \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,] \\
&= \; E_0\,[\,e^{-\,L\,(f)}\,p\,(\,F \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,) \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,] \\
&= \; \frac{E_0\,[\,e^{-\,L\,(f)}\,p\,(\,Y_1^n\,,\,A_1^n \;\mid\; \theta_1^n\,,\,F\,) \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]}{E_0\,[\,p\,(\,Y_1^n\,,\,A_1^n \;\mid\; \theta_1^n\,,\,F\,) \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]}\,.
\end{aligned}$$

The denominator has been computed in Theorem 6.1. For the numerator, note that, by Proposition 5.1,

$$L\,(f) \;=\; \tilde{L}\,(f) \,+\, \sum_{t\,\in\, I\,\cap\, J^c} f\,(t)\,L\,(t) \,+\, \sum_{1\,\le\, j\,\le\, M} f\,(Z_j)\,L\,(Z_j)\,.$$

Using the expression of the density provided in Lemma 4.1, we obtain

$$\begin{aligned}
& E_0\,[\,e^{-\,L\,(f)}\;p\,(\,Y_1^n\,,\,A_1^n \;\mid\; \theta_1^n\,,\,F\,) \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,] \\
&= \; E_0\,[\,e^{-\,\tilde{L}\,(f)} \prod_{0\le t<\infty} \frac{V_\theta(t\,,\,\tilde{L}^-\,(t))}{V(t\,,\,\tilde{L}^-\,(t))} \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,] \\
& \times \prod_{t\,\in\, I\,\cap\, J^c} E_0\,[\,e^{-\,f(t)\,L\,(t)}\,\frac{V_\theta(t\,,\,L^-\,(t))}{V(t\,,\,L^-\,(t))} \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,] \\
& \times \prod_{1\,\le j\,\le M} E_0\,[\,e^{-\,f(Z_j)\,L\,(Z_j)}\,\frac{W_\theta(Z_j\,,\,L^-\,(Z_j)\,)}{W(Z_j\,,\,L^-\,(Z_j))} \;\mid\; Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]\,.
\end{aligned}$$

Now, by Proposition 5.1 (i) (or Formula (86)) and Formula (67) ,

$$E_0\,[\,e^{-\,\tilde{L}\,(f)} \prod_{0\le t<\infty} \frac{V_\theta(t\,,\,\tilde{L}^-\,(t))}{V(t\,,\,\tilde{L}^-\,(t))} \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$=\;E_0\,[\prod_{0\le t<\infty} e^{-\,f(t)\,\tilde{L}\,(t)}\,\frac{V_\theta(t\,,\,\tilde{L}^-\,(t))}{V(t\,,\,\tilde{L}^-\,(t))} \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$=\;E_0\,[\prod_{0\le t<\infty} [\,1-\tilde{L}^-\,(t)\,]^{f(t)}\,\frac{V_\theta(t\,,\,\tilde{L}^-\,(t))}{V(t\,,\,\tilde{L}^-\,(t))} \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$=\exp\{-\int_0^\infty L_{0c}(dt)\int_0^1 [1-(1-u)^{f(t)}\frac{V_\theta(t,u)}{V(t,u)}]V(t,u)u^{-1}R_0^-(t,du)\}$$

$$=\exp\{\,-\int_0^\infty L_{0c}(dt)\int_0^1 [V(t,u)-(1-u)^{f(t)}V_\theta(t,u)]u^{-1}R_0^-(t,du)\,\}\,.$$

Dividing this last expression by the same term with $f(t)\,=\,0$ gives

$$E\,[\,e^{-\,\tilde{L}\,(f)} \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$=\;\exp\{\,-\int_0^\infty L_{0c}\,(dt)\int_0^1 \frac{1-(1-u)^{f(t)}}{u}\;V_\theta(t,u)\;R_0^-\,(t,du)\,\}\,.$$

Now, by Proposition 5.2 (ii), for $Z_j\,\in I^c$ and $D_j\,>\,0$,

$$E_0\,[\,e^{-\,f(Z_j)\,L\,(Z_j)}\,\frac{W_\theta(Z_j\,,\,L^-\,(Z_j)\,)}{W(Z_j\,,\,L^-\,(Z_j))} \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$=\;E_0\,[\,[\,1-L^-\,(Z_j)\,]^{f(Z_j)}\,\frac{W_\theta(Z_j\,,\,L^-\,(Z_j)\,)}{W(Z_j\,,\,L^-\,(Z_j))} \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$=\;\frac{\int_0^1(\,1-u\,)^{f(Z_j)}\,\frac{W_\theta(Z_j,u)}{W(Z_j,u)}\;W(Z_j,u)\,u^{-1}\,R_0^-(Z_j,du)}{\int_0^1 W(Z_j,u)\;u^{-1}\,R_0^-(Z_j,du)}$$

$$=\;\frac{\int_0^1(\,1-u\,)^{f(Z_j)}\;W_\theta(Z_j,u)\,u^{-1}\,R_0^-(Z_j,du)}{\int_0^1 W(Z_j,u)\;u^{-1}\,R_0^-(Z_j,du)}\,.$$

Dividing this last expression by the same term with $f(Z_j) = 0$ gives

$$E\,[\,[\,1 - L^-\,(Z_j)\,]^{f(Z_j)} \mid Y_1^n\,,\,A_1^n\,,\,\theta_1^n\,]$$

$$= \frac{\int_0^1 (\,1-u\,)^{f(Z_j)}\,W_\theta(Z_j,u)\,u^{-1}\,R_0^-(Z_j,du)}{\int_0^1 W_\theta(Z_j,u)\,u^{-1}\,R_0^-(Z_j,du)}\,.$$

The proof is similar for the other cases.

■

A.4 Proofs of Theorems 7.1 and 7.2

In the proportional hazards model, $k\,[\,\theta\,(t)\,,\,u\,] = 1 - (1-u)^{\varepsilon\,(t)}$. Therefore,

$$V_\theta\,(t,u) = \prod_{i\,\in\,K(t)} (1-u)^{\varepsilon_i\,(t)} = (1-u)^{\hat{N}\,(t)}\,,$$

where

$$\hat{N}\,(t) = \sum_{i\,\in\,K\,(t)} \varepsilon_i\,(t)\,,$$

and

$$\begin{aligned} W_\theta\,(Z_j\,,\,u) &= \prod_{i\in K_j\cup J_j} (1-u)^{\varepsilon_{ij}} \times \prod_{i\in I_j} [\,1-(1-u)^{\varepsilon_{ij}}\,] \\ &= (1-u)^{\hat{N}_j+\hat{R}_j} \times \sum_{0\le k\le D_j} (-1)^k \sum_{\{\,L\subset I_j\,:\,|L|\,=\,k\}} (1-u)^{\varepsilon\,_{Lj}}\,, \end{aligned}$$

where

$$\hat{R}_j = \sum_{i\,\in\,J_j} \varepsilon_{ij}\,,$$

and

$$\varepsilon\,_{Lj} = \sum_{i\,\in\,L} \varepsilon_{ij}\,.$$

The double summation is translation invariant. Hence,

$$W_\theta\,(Z_j\,,\,u) \;=\; (1-u)^{\hat N_j+\hat R_j} \times \sum_{0\le k\le D_j} (-1)^{\,k-1} \sum_{\{\,L\subset I_j\,:\,|L|\,=\,k\}} [\,1-(1-u)^{\,\varepsilon_{\,Lj}}\,]\,.$$

Now, using Theorem 6.3 (ii) and Formula (78) ,

$$\int_{(0,1]} \frac{1-(1-u)^{\,p}}{u}\, V_\theta\,(t,u)\, R_0^-\,(t,du)$$

$$= \int_{(0,1]} \frac{1-(1-u)^{\,\hat N\,(t)+p}}{u} - \frac{1-(1-u)^{\,\hat N\,(t)}}{u}\, R_0^-\,(t,du)$$

$$= r_0\,(\hat N\,(t)+p\,,\,t) \;-\; r_0\,(\hat N\,(t)\,,\,t)\,.$$

This proves part (i) of Theorem 7.1. Similarly,

$$(1-u)^{\,p}\,(1-u)^{\hat N_j+\hat R_j}\,[\,1-(1-u)^{\,\varepsilon_{\,Lj}}\,]\,u^{\,-1}$$

$$= \frac{1-(1-u)^{\,\hat N_j+\hat R_j+\varepsilon_{\,Lj}+p}}{u} - \frac{1-(1-u)^{\,\hat N_j+\hat R_j+p}}{u}\,.$$

The double summation being translation invariant, this implies

$$\int_{(0,1]} (1-u)^{\,p}\, W_\theta\,(Z_j,u)\, u^{\,-1}\, R_0^-\,(t,du)$$

$$= \sum_{0\le k\le D_j} (-1)^{\,k-1} \sum_{\{\,L\subset I_j\,:\,|L|\,=\,k\}} r_0\,(\,\hat N_j+\hat R_j+\varepsilon_{\,Lj}+p\,,\,Z_j)\,.$$

By Theorem 6.3 (iv) , part (ii) of Theorem 7.1 is obtained by dividing this last term by the same expression with $p\,=\,0$. By Corollary 5.2, the expression of $W_j(\beta)$ in Theorem 7.2 is obtained by taking $p\,=\,0$ in this last equality. Finally,

$$\int_{(0,1]} \frac{1-V_\theta\,(t,u)}{u}\, R_0^-\,(t,du) \;=\; r_0\,(\hat N\,(t)\,,\,t)\,.$$

Since $r_0\,(0\,,\,t) \;=\; 0$, this gives the expression of $V\,(\beta)$ in Theorem 7.2.

■

A.5 Proofs of Theorems 7.3 and 7.4

In the proportional hazards model, $k\,[\,\theta\,(t)\,,\,u\,]\;=\;\delta\,(t)\,u\,.$ Therefore,

$$1 - k\,[\,\theta\,(t)\,,\,u\,]\;=\;[\,1-\delta\,(t)\,]\,[\,1+\eta\,(t)\,(1-u)\,]\,,$$

where

$$\eta\,(t)\;=\;\frac{\delta\,(t)}{1-\delta\,(t)}\,.$$

Therefore,

$$\begin{aligned} V_\theta\,(t,u)\;&=\;\prod_{i\,\in\,K(t)}[\,1-\delta_i\,(t)\,]\;\times\;\prod_{i\,\in\,K(t)}[\,1+\eta_i\,(t)\,(1-u)\,] \\ &=\;\sum_{0\le k\le N(t)}\mu_k\,(t)\,(1-u)^{\,k}\,, \end{aligned}$$

where

$$\mu_k\,(t)\;=\;\prod_{i\,\in\,K(t)}[\,1-\delta_i\,(t)\,]\;\times\;\sum_{\{L\,\subset\,K(t)\,:\,|L|=k\}}\;\prod_{i\,\in\,L}\eta_i\,(t)\,,$$

and

$$\begin{aligned} W_\theta\,(Z_j,u)\;&=\;\prod_{i\,\in\,I_j}\delta_{ij}\;\times\;u^{\,D_j}\;\times\;\prod_{i\,\in\,K_j\,\cup\,J_j}[\,1-\delta_{ij}\,u\,] \\ &=\;\nu_j\sum_{0\le k\le N_j+R_j}\gamma_{jk}\,u^{\,D_j}\,(1-u)^{\,k}\,, \end{aligned}$$

where

$$\nu_j\;=\;\prod_{i\,\in\,I_j}\delta_{ij}\;\times\;\prod_{i\,\in\,K_j\,\cup\,J_j}[\,1-\delta_{ij})\,]\,,$$

and

$$\gamma_{jk}\;=\;\sum_{\{L\,\subset\,K_j\,\cup\,J_j\,:\,|L|=k\}}\;\prod_{i\,\in\,L}\eta_{ij}\,.$$

Now, using Theorem 6.3 (ii) and Formula (78),

$$\begin{aligned} &\int_{(0,1]}\frac{1-(1-u)^p}{u}\,V_\theta\,(t,u)\,R_0^-\,(t,du) \\ =\;&\sum_{0\le k\le N(t)}\mu_k\,(t)\int_{(0,1]}\frac{1-(1-u)^p}{u}\,(1-u)^{\,k}\,R_0^-\,(t,du) \\ =\;&\sum_{0\le k\le N(t)}\mu_k\,(t)\,[\,r_0\,(k+p\,,\,t)\;-\;r_0\,(k\,,\,t)\,]\,. \end{aligned}$$

This proves part (i) of Theorem 7.1. Now,

$$\begin{aligned}
&\int_{(0,1]} (1-u)^{p}\, W_\theta\,(Z_j,u)\, u^{-1}\, R_0^-\,(Z_j,du) \\
=\;& \nu_j \sum_{0\le k\le N_j+R_j} \gamma_{jk} \int_{(0,1]} (1-u)^{p+k}\, u^{D_j-1}\, R_0^-\,(Z_j,du) \\
=\;& \nu_j \sum_{0\le k\le N_j+R_j} \gamma_{jk}\; s_0\,(\,D_j\,,\,k+p\,,\,Z_j\,)\,.
\end{aligned}$$

By Theorem 6.3 (iv), part (ii) of Theorem 7.3 is obtained by dividing this last term by the same expression with $p\,=\,0$. By Corollary 5.2, the expression of $W_j(\beta)$ in Theorem 7.4 is obtained by taking $p\,=\,0$ in this last equality. Now, since $V_\theta\,(t,0)\,=\,1$,

$$1\,-\,V_\theta\,(t,u)\;=\;\sum_{0\le k\le N(t)} \mu_k\,(t)\,[\,1\,-\,(1-u)^{\,k}\,]\,.$$

Therefore,

$$\int_{(0,1]} \frac{1-V_\theta\,(t,u)}{u}\, R_0^-\,(t,du)\;=\;\sum_{0\le k\le N(t)} \mu_k\,(t)\; r_0\,(k\,,\,t)\,.$$

This proves the first part of Theorem 7.4. Finally, since

$$\begin{aligned}
u^{\,m} \;&=\; \sum_{0\le k\le m} C_m^k\,(-1)^{\,k}\,(1-u)^{\,k} \\
&=\; \sum_{0\le k\le m} C_m^k\,(-1)^{\,k-1}\,[\,1-(1-u)^{\,k}\,]\,,
\end{aligned}$$

$$\begin{aligned}
s_0\,(m,p,t)\;&=\;\sum_{0\le k\le m} C_m^k\,(-1)^{\,k-1}\,[\,r_0\,(k+p\,,\,t)\;-\;r_0\,(p\,,\,t)\,] \\
&=\;\sum_{0\le k\le m} C_m^k\,(-1)^{\,k-1}\,r_0\,(k+p\,,\,t)\,.
\end{aligned}$$

■

References

[1] Aalen,O. (1978), Nonparametric Inference for a Family of Counting Processes, *Annals of Statistics* **6**, 701-726.

[2] Andersen P.K., O. Borgan, R.D. Gill, N. Keiding (1993), *Statistical Models Based on Counting Processes*, Springer, New York.

[3] Blum K., V. Susarla (1977), On the Posterior Distribution of a Dirichlet Process Given Randomly Right Censored Data, *Stochastic Processes and their Applications* **5**, 207-211.

[4] Doksum K. (1974), Tailfree and Neutral Random Probabilities and their Posterior Distributions, *Annals of Probability* **2**, 183-201.

[5] Dykstra R.L., P.W. Laud (1981), A Bayesian Nonparametric Approach to Reliability, *Annals of Statistics* **9**, 356-367.

[6] Ferguson T.S. (1973), A Bayesian Analysis of some Nonparametric Problems, *Annals of Statistics* **1**, 209-230.

[7] Ferguson T.S. (1974), Prior Distributions on Spaces of Probability Measures, *Annals of Statistics*, **2**, 615-629.

[8] Ferguson T.S., M.J. Klass (1972), A Representation of Independent Processes without Gaussian Components, *Annals of Mathematical Statistics*, **43**, 1634-1643.

[9] Ferguson T.S., E. Phadia (1979), Bayesian Nonparametric Estimation Based on Censored Data, *Annals of Statistics* **7**, 163-186.

[10] Fleming, T.R. , D.P. Harrington (1991), *Counting Processes and Survival Analysis*, Wiley, New York.

[11] Florens J.P., M. Mouchart, J.M. Rolin (1990), *Elements of Bayesian Statistics*. Marcel Dekker, New York.

[12] Florens J.P., M. Mouchart, J.M. Rolin (1999), Semi- and Nonparametric Bayesian Analysis of Duration Models: A Dirichlet Process Approach, to appear in *International Statistical Review*.

[13] Florens J.P. , J.M. Rolin (1998), Simulation of Posterior Distributions in Nonparametric Censored Analysis, Discussion Paper 9818 Institut de Statistique, Université catholique de Louvain, Belgium.

[14] Gouget N., J.P. Raoult (1999), Computation of Predictive Densities in the Semiparametric Bayesian Cox-Dirichlet Model, to appear in *Journal of Nonparametric Statistics*.

[15] Hjort N.L. (1990), Nonparametric Bayes Estimators Based on Beta Process in Models for Life History Data, *Annals of Statistics* **18**, 1259-1294.

[16] Kalbfleisch J.D. (1978), Nonparametric Bayesian Analysis of Survival Time Data, *Journal of the Royal Statistical Society* B, **40**, 214-221.

[17] Karr A.F. (1986), *Point Processes and their Statistical Inference*, Marcel Dekker, New York.

[18] Kiefer J. , J. Wolfowitz (1956), Consistency of the Maximum Likelihood Estimator in the Presence of infinitely many Incidental Parameters. *Annals of Mathematical Statistics* **27**, 887-906.

[19] Lo A.Y. (1993), A Bayesian Bootstrap for Censored Data, *Annals of Statistics* **21**, 100-123.

[20] Miller R.G. Jr. (1981) , Survival Analysis, Wiley, New York.

[21] Mouchart M., J.M. Rolin (1995), Competing Risks Models: Problems of Modelling and of Identification, Discussion Paper 9507, Institut de Statistique, Université catholique de Louvain, Belgium.

[22] Rolin J.M. (1983), Nonparametric Bayesian Statistic: a Stochastic Process Approach, in *Specifying Statistical Models, from Parametric to Nonparametric using Bayesian or Non-Bayesian Approaches*. Lecture Notes in Statistics, vol 16, Springer, Berlin, 108-133.

[23] Rolin J.M. (1998), Bayesian Survival Analysis, in *Encyclopedia of Biostatistics* Wiley, New-York,**1**, 271-286.

[24] Ruggiero M. (1994), Bayesian Semiparametric Estimation of Proportional Hazard Models, *Journal of Econometrics*, **62**, 277-300.

[25] Susarla V., J. Van Ryzin (1976), Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations, *Journal of the American Statistical Association* **71**, 897-902.

[26] Walker S.G., P. Damien (1998), A Full Bayesian Analysis Involving a Neutral to the Right Process, *Scandinavian Journal of Statistics* **25**, 669-680.

[27] Walker S.G., P. Muliere (1997), Beta-stacy Processes and a Generalisation of the Polya-urn Scheme, *Annals of Statistics* **25**, 1762-1780.

[28] Walker S.G., P. Damien, P.W. Laud (1998), Bayesian Nonparametric Inference for a Continuous Cumulative Hazard Function, First Draft.

[29] Wild C.J., J.D. Kalbfleisch (1981), A Note on a Paper by Ferguson and Phadia, *Annals of Statistics* **9**, 1061-1065.