Semi- and Non-parametric Bayesian Analysis of Duration Models with Dirichlet Priors: A Survey

J.-P. Florens *
M. Mouchart †
J.-M. Rolin ‡

October 1998

Abstract

The object of this paper is to review the main results obtained in semi- and non-parametric Bayesian analysis of duration models. Standard nonparametric Bayesian models for independent and identically distributed observations are reviewed in line with Ferguson's pioneering papers and recent results on the characterization of Dirichlet processes are also discussed. Next we present recent results on nonparametric treatment of censoring and of heterogeneity in the context of mixtures of Dirichlet processes. The final section considers a Bayesian semiparametric version of the proportional hazards model.

[*] *Acknowledgement:* Supports from GREMAQ and IDEI, Universit des Sciences Sociales (Toulouse), from the Fond National de la Recherche Scientifique and from the "projet d'Actions de Recherche Concertés ARC 93/98 are gratefully acknowledged. Comments by Jean-Francois Richard, anonymous referees associated and joint editors have been useful for improving the quality of the exposition.

Key words: Nonparametric Bayesian Statistics. Dirichlet measures. Neutral to the right processes. Beta processes. Duration data. Censored observations. Heterogeneity. Proportional hazards model.

[*] GREMAQ and IDEI, Université des Sciences Sociales, Toulouse, France.
[†] CORE and Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgique.

Résumé

L'objectif de cet article est de présenter les résultats principaux obtenus dans l'analyse bayésienne semi-paramétrique ou non-paramétrique des modèles de durée. Les résultats fondamentaux du modèle independent identically distributed de base sont rappelés en suivant les travaux initiaux de Ferguson et en utilisant des résultats récents relatifs aux représentations du processus de Dirichlet.

La considération de mélanges de processus de Dirichlet permet d'étudier l'impact de la présence de données censurées et l'introduction de l'hétérogénéité non observée. La dernière section examine le traitement bayésien des modèles semi-paramétriques à risques proportionnels.

1 Introduction.

Duration models (or more generally their extensions to counting processes) are widely used in many areas of applied statistics: life table, actuarial statistics, biometrics, clinical trials, labor market econometrics, reliability theory, and so on.

There exists a vast literature discussing duration models from theoretical as well as empirical viewpoints within a sampling framework. Surveys also are available, see e.g., Andersen and Keiding (1995), Florens, Fougère and Mouchart (1996) and the books of Fleming and Harrington (1991) and Andersen, Borgan, Gill and Keiding (1993).

Duration models are not only important for the statistical analysis of data in many fields, they also display characteristic features of particular interest to the statistician. Among these features, one may mention a prominent interest in nonparametric methods, made possible by the availability of large data sets. Another specific feature of duration models comes from the relevance of alternative characterizations of the sampling distributions, namely their survivor function and their integrated or instantaneous hazard functions. This particular role may be appreciated, for instance, in the specification of conditional models, a usual class of models in econometrics, where the proportional hazards model is most naturally written in terms of a hazard function and displays general properties different from those of the standard regression models. Although not specific to duration models, the necessity of taking into account unobservable components of heterogeneity or censoring problems, actually present in most data sets, emphasizes even more the particularities of duration models.

The object of this paper is to review the main results obtained in this field from a Bayesian point of view, with a particular interest in nonparametric and semiparametric models whose prior specification often involves a Dirichlet process.

Both parametric and nonparametric Bayesian models are frequently used in reliability theory. On one hand, they are used to combine expert opinions that are expressed in the prior distribution and information provided by the data (see e.g. Kalbfleisch and Prentice (1980)). On the other hand, the Bayesian approach provides statistical procedures having some nice properties such as admissibility of the estimators and asymptotic theory based on martingales. In particular, it is known that the Sursala-Van Ryzin estimator smooths the Kaplan-Meier estimator, and is everywhere well defined. This may not be the case for the Kaplan-Meier estimator, and this creates some difficulties in studying its asymptotic properties.

Nonparametric Bayesian methods have rapidly expanded after Ferguson's

(1973, 1974) presentation of the Dirichlet process as a naturel conjugate prior specification to the empirical process. Soon thereafter, the extension to mixtures of Dirichlet processes was considered by Antoniak (1974) and the treatment of censored data was addressed by Suzarla and Van Ryzin (1976). Additional contributions along these lines appeared later, including Lahiri and Dong Ho (1988), Phadia and Susarla (1983) and Tsai (1986). A noticeable extension by Hjort (1990) treats prior specifications based on the Beta process (see also Rolin (1998)). Florens, Mouchart and Rolin (1990, ch. VIII) provide general tools of analysis whereas Florens, Mouchart and Rolin (1992) consider a general model of mixtures for handling problems of heterogeneity. Rolin (1983, 1992a and b) and Sethuraman (1994) further contribute to the analysis of finer properties of the Dirichlet process. This review also gives credit to the original work of Ruggiero (1989,1994) and Hakizamungu (1992). Some unpublished references, in the more general field of non- or semi-parametric Bayesian models, are Erkanli, Müller and West (1993), McEachern (1992), and West (1990, 1992). In this review we only concentrate on non- or semi-parametric models used in the analysis of duration data.

This review is organized in four steps. In the next section, we review the standard nonparametric Bayesian model for independent identically distributed observations, in line with the pioneering papers of Ferguson (1973, 1974). This section also includes more recent results giving more precise characteristics of the Dirichlet process from the point of view of random measures. Section 3 handles the independent identically distributed case with censoring. Section 4 treats the heterogeneity problem in the context of mixtures of Dirichlet processes.The last section considers recent developments of a semiparametric version of the proportional hazards model.

The presentation is made simpler by treating separately these topics even though, in actual applications, heterogeneity, censoring and exogenous variables all arise together.

2 Nonparametric duration models without censoring.

2.1 A nonparametric Bayesian model.

Even though earlier Bayesian papers had discussed nonparametric methods, Ferguson's (1973, 1974) papers have been most influential in motivating new contributions over the last twenty years. In this section we summarize the basic results obtained in that direction, limiting the presentation to the particular case of duration models, *viz* models for non-negative observations, although the basic model is considerably more general.

Whereas a statistical model in sampling theory is a family of *sampling distributions* indexed by a *parameter*, a Bayesian model is characterized by a unique probability measure on the product space parameter $\times$ observation. This probability is obtained by endowing a sampling theory model with a probability measure on the parameter space, to be called the *prior probability*, and by treating the sampling model as a conditional probability on the sample space given the parameter. Bayesian methods aim at analysing *posterior* and *predictive* distributions. The former are distributions for parameters conditionally upon observations, and the latter average sampling distributions using prior probability as weight function.

In the case of duration models, the sample space is $\mathbb{R}_+$, the positive part of the real line, endowed with its Borel sets. In the case of nonparametric models, the parameter space is an appropriate subset of the set of all probability measures on the sample space or of some of their transformations, such as the survivor function or the hazard functions.

Thus, the basic nonparametric Bayesian model for analyzing duration data may be described as follows. The sampling process is I.I.D and the sample space for a sample of size one is $(\mathbb{R}_+, \mathcal{B}_+)$. For a given sample size n, the data are therefore $t_1, \cdots, t_n$ where each t_i, $1 \leq i \leq n$, is independently generated by a common sampling probability, say Φ, an element of the set of all probability measures on $(\mathbb{R}_+, \mathcal{B}_+)$. For the sake of simplicity, one also writes $\Phi(A)$ instead of $P(t_i \in A|\Phi)$ for any $A \in \mathcal{B}_+$.

For the prior specification, note that the prior probability makes Φ a random probability measure on $(\mathbb{R}_+, \mathcal{B}_+)$ and should therefore be viewed as a stochastic process whose trajectories are probability measures on $(\mathbb{R}_+, \mathcal{B}_+)$. Following Ferguson (1973, 1974) a workable choice, which is also natural conjugate to the empirical process, is the Dirichlet process. The basic intuition and the main features of this specification may be approached through the finite case.

So, let us assume as a first step in the presentation that the sampling probability Φ is restricted to give positive probabilities to a fixed finite set say $(a_1, \cdots, a_k)$. Thus Φ is characterized by a point θ of the simplex S_k of $\mathbb{R}^k : \{\theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k_+ \text{ such that } \theta_1 + \ldots + \theta_k = 1\}$ with the interpretation that $\theta_j = \Phi(\{a_j\}), \quad 1 \leq j \leq k$. An observed duration t may be represented by a vector of binary variables $x = (x_1, \cdots, x_k)$ with $x_j = \mathbb{I}_{\{t=a_j\}}$ and the sampling probability may then be written as

$$p(x|\theta) = \prod_{1 \leq j \leq k} \theta_j^{x_j} \qquad (2.1)$$

A natural conjugate prior distribution for the sampling process is provided by the Dirichlet distribution whose density may be written as

$$\begin{aligned} p(\theta) &= f_{Di}(\theta|n_0, p_0) \\ &= \Gamma(n_0) \prod_{1\leq j\leq k} \frac{\theta_j^{n_0 p_{0j}-1}}{\Gamma(n_0 p_{0j})} \mathbb{1}_{\{\theta\in S_k\}} \end{aligned} \tag{2.2}$$

where $(n_0, p_0) \in \mathbb{R}_+ \times S_k$ are the natural parameters in a Dirichlet distribution. If we have n I.I.D observations of x i.e. $x_i = (x_{i1}, \cdots, x_{ij}, \cdots, x_{ik})$ for $i = 1, \cdots, n$, the vector of sampling proportions $p = (p_1, \cdots, p_j, \cdots, p_k)$ with $p_j = n^{-1} \sum\limits_{1\leq i\leq n} x_{ij}$ constitutes a sufficient statistics. The latter multiplied by n has a multinomial distribution. The natural conjugate property implies that the posterior distribution of θ is again Dirichlet with parameters

$$n_* = n_0 + n \tag{2.3}$$

and

$$p_* = \frac{n_0 p_0 + np}{n_0 + n}. \tag{2.4}$$

Suppose now that instead of restricting Φ to have a fixed finite support $\{a_1, \cdots, a_k\}$ we consider a finite fixed partition of $\mathbb{R}_+$, $\{B_1, \cdots, B_k\}$ and restrict the parameter of interest to be $\theta = (\theta_j)_{1\leq j\leq k}$ with $\theta_j = P(t_i \in B_j|\Phi) = \Phi(B_j)$. It is then natural to retain from an I.I.D sample of durations $t_1, \cdots, t_n$ the proportions of observations in each B_j , i.e., $p_j = n^{-1} \sum_{1\leq i\leq n} \mathbb{1}_{\{t_i\in B_j\}}$. If the prior specification of Φ is such that it implies a Dirichlet distribution on θ, the analysis of the case of finite support may be exactly repeated without any change. The transition from the Dirichlet distribution to the Dirichlet process is achieved by switching from a fixed to an arbitrary partition and by replacing the prior parameter $p_0 = (p_{01}, \cdots, p_{0k}) \in S_k$ by a probability measure P_0 on the sample space $(\mathbb{R}_+, \mathcal{B}_+)$. More specifically, the sampling probability Φ on $(\mathbb{R}_+, \mathcal{B}_+)$ is said to be distributed as a Dirichlet process with parameter (n_0, P_0), denoted by $\Phi \sim \mathcal{D}i(n_0, P_0)$ if, for any measurable partition $(B_1, \cdots, B_k)$ of $\mathbb{R}_+$, the random vector $(\Phi(B_1), \cdots, \Phi(B_k))$ is distributed as a Dirichlet distribution with parameters n_0 and $(P_0(B_1), \cdots, P_0(B_k))$. From the properties of the Dirichlet distribution it may be verified that the system $(\Phi(B_1), \cdots, \Phi(B_k))$, defined for any finite partition of $\mathbb{R}_+$, induces a projective system and therefore, by Kolmogorov theorem, uniquely characterizes the law of the process generating Φ.

Let us write P_n for the empirical process, i.e.,

$$P_n(B) = \frac{1}{n} \sum_{1\leq i\leq n} \mathbb{1}_{\{t_i\in B\}} = \frac{1}{n} \sum_{1\leq i\leq n} \delta_{t_i}(B) \qquad B \in \mathcal{B}_+ \tag{2.5}$$

where δ_{t_i} is the unit mass at t_i. The statistic P_n is sufficient for independent identically distributed sampling and the posterior probability of Φ is again a Dirichlet process with parameters

$$n_* = n_0 + n \tag{2.6}$$

$$P_* = \frac{n_0 P_0 + n P_n}{n_0 + n}. \tag{2.7}$$

In most applications, the location parameter of the prior distribution, P_0, is a continuous probability measure whereas the empirical process, P_n, is discrete. Thus, P_*, the location parameter of the posterior distribution, is typically a mixed probability measure on $I\!R_+$ smoothing P_n.

The description of this Bayesian model is now completed by describing the predictive process through the following sequence.

(i) the first observation t_1 is predictively distributed as P_0,

(ii) the observation t_{i+1} conditionally on $(t_1, \cdots, t_i)$ is predictively distributed according to $\frac{1}{n_0+i}(n_0 P_0 + i P_i)$ where P_i is the empirical process of $(t_1, \cdots, t_i)$, i.e.,

$$P_i(B) = \frac{1}{i} \sum_{1 \le j \le i} 1\!\!I_{\{t_j \in B\}} \qquad B \in \mathcal{B}_+. \tag{2.8}$$

An important feature of this predictive process is to generate ties. Specifically, one has, for instance

$$P(t_2 = t_1) = \frac{1}{n_0 + 1} \tag{2.9}$$

when P_0 is continuous. As a consequence, the predictive distribution of $t_1, \cdots, t_n$, may also be characterized globally in an alternative way based on the following remark. The information contained in $(t_1, \cdots, t_n)$ is equivalently described by $(C_n, (t_{(j)})_{1\le j\le p})$ where $(t_{(j)})_{1\le j\le p}$ is the vector of distinct values taken by $(t_1, \cdots, t_n)$ where the values are ordered according to the order of appearance in $(t_1, \cdots, t_n)$ and C_n is a partition of $\{1, 2, \cdots, n\}$ into p non empty elements $(1 \le p \le n)$, namely $C_n = \{I_j : 1 \le j \le p\}$, I_j is a non-empty subset of $\{1, \cdots, n\}$ corresponding to the indices $i \in \{1, \cdots, n\}$ for which the t_i's are equal to $t(j)$. Note that p is a function of C_n, namely $p = |C_n|$ where $|C_n|$ is the cardinality (the number of elements) of C_n.

Therefore the distribution of $(t_1, \cdots, t_n)$ is equivalently described by the distribution of C_n, running over all partitions of $\{1, \cdots, n\}$ into $1, 2, \cdots, n$ elements, and the distribution of $(t_{(j)})_{1\le j\le p}$, conditionally on C_n.

The marginal distribution of C_n is somewhat involved and is given with some details in Blackwell and MacQueen (1973), Antoniak (1974), Yamato (1984) and Rolin (1992b). The distribution of $(t_{(j)})_{1\le j\le p}$, conditionally on C_n may be described more easily. It depends essentially on p, the number of distinct values, and $(t_{(j)})_{1\le j\le p}$ are otherwise independent identically distributed with distribution P_0 when P_0 is continuous.

2.2 Some properties of the Dirichlet process.

The previous section suggests that the structure of the nonparametric Bayesian model under a Dirichlet prior specification is simple and provides a workable approach for the evaluation of the posterior distribution as well as of the predictive distribution. In this section, some properties of the model reinforce this idea but other properties shed light on some subtle aspects of the Dirichlet process and require that it be handled with special care.

2.2.1 Moments.

Let Φ be a random probability on $(I\!R_+, \mathcal{B}_+)$ distributed as a Dirichlet process with parameter (a, M). Thus in the context of section 2.1, a may be taken as n_0 or n_* and M as P_0 or P_*. From the definition of the Dirichlet process, it should be clear that, for any $B \in \mathcal{B}_+$, the random variable $\Phi(B)$ follows a Beta distribution with parameters $(aM(B), aM(B^c))$; in particular one has

$$E[\Phi(B)] = M(B) \tag{2.10}$$

and

$$V[\Phi(B)] = \frac{M(B)M(B^c)}{a+1}. \tag{2.11}$$

Thus, P_0 may be interpreted as a "prior guess" on Φ and n_0 as a measure of prior precision. Similarly, P_* is the posterior expectation of the process and n_* characterizes its posterior precision. Consequently, P_* may be viewed as a natural Bayesian estimator of Φ built as a convex combination of the prior guess P_0 and the empirical process P_n and convergent in as far as it has the same asymptotic behavior as P_n.

2.2.2 The trajectories of the Dirichlet process.

Proper understanding of Bayesian nonparametric models under a Dirichlet prior specification requires a serious analysis of its trajectories, both their structure and their support. Furthermore, knowledge of the trajectories is crucial for designing efficient numerical procedures of simulation and for analytical derivations. Details on these issues are given in Florens and Rolin (1994). In particular, it is shown there how to estimate moments under a Dirichlet prior specification. The relationship between classical bootstrapping and simulation of the posterior distribution under a noninformative prior specification is also discussed.

When a random probability Φ on $(\mathbb{R}_+, \mathcal{B}_+)$ is distributed as a Dirichlet process with parameters (a, M), the structure and the support of its trajectories depend crucially on the location parameter M. To show this, let us decompose M into its purely discrete and continuous parts, i.e., let

$$S = \{x \in \mathbb{R}_+ : M(\{x\}) > 0\} = \{a_j : j \in I\}$$

(where $I \subset \mathbb{N}$ is a countable set)

$$a_d = aM(S) \qquad a_c = aM(S^c)$$

$$M_d(B) = M(B|S) \qquad M_c(B) = M(B|S^c) \quad \forall\ B \in \mathcal{B}_+.$$

Clearly, one has

$$aM = a_c M_c + a_d M_d \tag{2.12}$$

Let us now do the same decomposition on Φ, namely let

$$\alpha = \Phi(S)$$

$$\Phi_d(B) = \Phi(B|S) \qquad \Phi_c(B) = \Phi(B|S^c)$$

so that

$$\Phi = (1-\alpha)\Phi_c + \alpha\Phi_d \tag{2.13}$$

It has been shown in Rolin (1992a) and Sethuraman (1994) that

(i) α, Φ_c, Φ_d are independent

(ii) α has a Beta distribution with parameters (a_d, a_c)

(iii) Φ_c is distributed as a Dirichlet process with parameters (a_c, M_c)

and

(iv) Φ_d is distributed as a Dirichlet process with parameters (a_d, M_d).

Since M_d has a countable support S, Φ_d has the same support with probability 1 and $\{\Phi_d(\{a_j\}) : j \in I\}$ has a Dirichlet distribution with parameters $\{aM(\{a_j\}) : j \in I\}$. In particular, $\Phi_d(\{a_j\})$ has a Beta distribution with parameters $(aM(\{a_j\}), a - aM(\{a_j\}))$. Furthermore, the support of this Dirichlet process is the set of all probabilities with support S. Now, Ferguson (1973) has shown that the trajectories of Φ_c are almost surely discrete. In other words, Φ_c may be represented as

$$\Phi_c = \sum_{1 \leq j < \infty} \gamma_j \delta_{\tau_j} \tag{2.14}$$

almost surely (and not only in distribution). A first description of the distribution of $(\tau_j)_{1\leq j<\infty}$ and $(\gamma_j)_{1\leq j<\infty}$ has been provided by Ferguson (1973) :

(i) $(\tau_j)_{1\leq j<\infty}$ is an infinite independent identically distributed sample from M_c,

(ii) the sequence $(\gamma_j)_{1\leq j<\infty}$ is normalized and decreasing ,i.e., such that $\sum_j \gamma_j = 1$ and $0 < \gamma_{j+1} < \gamma_j$ with probability one, and may be represented as follows:

$$\gamma_j = (\sum_{1\leq l<\infty} J_l)^{-1} J_j \tag{2.15}$$

where the sequence $(J_j)_{1\leq j\leq\infty}$ is markovian and decreasing ($J_{j+1} < J_j$) and its distribution functions are

$$P[J_1 \leq t] = \exp\{-a_c \int_t^\infty u^{-1} e^{-u} du\} \qquad t \geq 0 \tag{2.16}$$

$$P[J_{j+1} \leq t | J_1, \cdots, J_j] = \exp\{-a_c \int_t^{J_j} u^{-1} e^{-u} du\} \quad 0 \leq t \leq J_j, \tag{2.17}$$

and

(iii) furthermore, the sequences $(\tau_j)_{1\leq j<\infty}$ and $(J_j)_{1\leq j<\infty}$ are independent and, eventually, so are also the sequences $(\tau_j)_{1\leq j<\infty}$ and $(\gamma_j)_{1\leq j<\infty}$.

A second description has been obtained in Rolin (1992b) and Sethuraman (1994) where it is shown that there exists an infinite permutation of $\mathbb{N}$ such that, keeping the same representation of Φ_c as in (2.14), (i) and the second part of (iii) remain valid and (ii) becomes (ii*) the sequence $(\gamma_j)_{1\leq j<\infty}$ may be represented as follows:

$$\gamma_j = \beta_j \prod_{1\leq \ell \leq j-1} (1-\beta_\ell) \tag{2.18}$$

where $(\beta_j)_{1\leq j<\infty}$ is an infinite independent identically distributed sample of the Beta distribution with parameters $(1, a_c)$.

It should be remarked that (i) the distribution of (τ_j) (of (γ_j)) depends only on M_c (on a_c), (ii) even though M_c is continuous, the trajectories of Φ_c are almost surely discrete but the infinitely many jumps are randomly located and the support of Φ_c is almost surely dense in the support of M_c. In other words, for any set B such that $M_c(B) > 0$, $\Phi_c(B)$ is almost surely a strictly positive random variable. Furthermore, any probability absolutely continuous with respect to M_c is in the pointwise support of the Dirichlet process (recall that the pointwise convergence is defined by $M_n(B) \to M(B) \quad \forall\, B \in \mathcal{B}_+$).

Let us now consider the consequences of a Dirichlet prior specification to the Bayesian model.

When P_0 is a discrete probability measure with jumps at a fixed set $S = \{a_j : j \in I\} \subset I\!R_+$ with $I \subset \mathbb{N}$, the sampling probabilities Φ will be almost surely discrete with the same support S, both a priori and a posteriori for almost any observation $t_1, \cdots, t_n$ because P_* will have the same support as P_0. Furthermore, the support of the Dirichlet process is the set of all probabilities on S, both a priori and a posteriori. Also the predictive distributions generating t_1 and $(t_{i+1}|t_1, \cdots, t_i)$ will all be probabilities supported by S. Note that in such a case, the model is unable to handle an observation falling outside S.

When P_0 is continuous, the prior trajectories are characterized as in (2.14). For the posterior trajectories, let us remark that $P_{*c} = P_0$ and $P_{*d} = P_n$. Therefore, if $(t_{(j)})_{1\leq j\leq p}$ is the set of distinct values taken by $(t_1, \ldots, t_n)$ and if n_j is the number of t_i's that are equal to $t_{(j)}$, $1 \leq j \leq p$, P_n may be written as

$$P_n = \sum_{1\leq j\leq p} \frac{n_j}{n} \delta_{t_{(j)}}$$

Therefore, according to (2.12), Φ may be represented a posteriori as

$$\Phi = (1-\alpha_n)\Phi_c + \alpha_n \Phi_d$$

where

$$\alpha_n = \sum_{1\leq j\leq p} \Phi(\{t_{(j)}\})$$

and

$$\begin{aligned}\Phi_d &= \alpha_n^{-1} \sum_{1\le j\le p} \Phi(\{t_{(j)}\})\delta_{t_{(j)}} \\ &= \sum_{1\le j\le p} \Phi_d(\{t_{(j)}\})\delta_{t_{(j)}}.\end{aligned}$$

In this representation, conditionally on $(t_1, \ldots, t_n)$,

(i) α_n, Φ_c, Φ_d are independent.

(ii) α_n has a Beta distribution with parameters (n, n_0)

(iii) Φ_c is a Dirichlet process with parameters (n_0, P_0)

(iv) Φ_d is a Dirichlet process with parameters (n, P_n)

or equivalently,

(iv*) $\{\Phi_d(\{t_{(j)}\}) : 1 \le j \le p\}$ has a Dirichlet distribution with parameters $\{n_j : 1 \le j \le p\}$.

It should be remarked that: (i) the normalized part of Φ outside the observations, Φ_c , is not revised by the observations, i.e., is distributed as Φ a priori, (ii) the normalized part of Φ at the observations, Φ_d , has a distribution independent of the prior distribution,i.e., does not depend on (n_0, P_0). Finally, Φ is a convex combination of Φ_c and Φ_d the coefficient of which (α_n) has a distribution depending only on the prior precision and the sample size, i.e., (n_0, n).

"Non-informative" prior specification raises problems. A natural suggestion has indeed been to consider that $n_0 \to 0$. In this latter case, the behaviour of the posterior process is rather natural: the posterior process tends to a Dirichlet process with parameter (n, P_n). Note however the discontinuity at $n_0 = 0$ where the prior distribution has the pathological representation of a random jump process: $\Phi = \delta_t$ where t is P_0-distributed.

2.2.3 Uses and Extensions of Dirichlet process in duration models.

When modeling duration data, it is natural to give structure to one of the following transformations of the sampling probability Φ: either the survivor function

$$\Sigma(t) = \Phi((t, \infty)) \tag{2.19}$$

or the cumulative hazard function defined in either one of the following two non equivalent ways:

$$\Lambda(t) = -\ln \Sigma(t) \tag{2.20}$$

$$\hat{\Lambda}(t) = \int_{[0,t]} [\Phi([u,\infty))]^{-1}\Phi(du) \tag{2.21}$$

Historically, Bayesian statistical analysis of duration models has used definition (2.20) to define neutral to the right processes (Doksum (1974)) and to modelize proportional hazards models (see section 5). Definition (2.21), as underlined by a referee, has a better probabilistic meaning in relation with Doob-Meyer decomposition of $1\!\!I_{\{t_i \leq t\}}$ ($\hat{\Lambda}(\min(t_i,t))$ is a previsible process and $1\!\!I_{\{t_i \leq t\}} - \hat{\Lambda}(\min(t_i,t))$ is a martingale). It serves as a cornerstone of martingale estimators (see, e.g., Fleming and Harrington (1991) and Andersen, Borgan, Gill and Keiding (1993)). Its utility in Bayesian analysis has been shown in the definition of Beta processes introduced by Hjort (1990). Some confusions have arisen in the literature despite the fact, as will be seen later on, that they are both useful. It is therefore interesting to present the relations existing between the two definitions.

Because $\Sigma(t)$ is non-increasing and right-continuous such that $\Sigma(0) \leq 1$ and $\Sigma(\infty) = 0$, $\Lambda(t)$ and $\hat{\Lambda}(t)$ are both non-decreasing, right-continuous such that $0 \leq \hat{\Lambda}(0) \leq \Lambda(0)$ and $\hat{\Lambda}(\infty) \leq \Lambda(\infty) = \infty$. Therefore, both $\Lambda(t)$ and $\hat{\Lambda}(t)$ can be viewed as cumulative distribution functions of σ-finite measures on $I\!\!R_+$. Let consider the decomposition of Λ and $\hat{\Lambda}$ into their discrete and continuous parts

$$\Lambda(t) = \Lambda_c(t) + \Lambda_d(t) \tag{2.22}$$

$$\hat{\Lambda}(t) = \hat{\Lambda}_c(t) + \hat{\Lambda}_d(t) \tag{2.23}$$

We note that the continuous parts always coincide

$$\Lambda_c(t) = \hat{\Lambda}_c(t) \quad \forall \ t$$

but the discrete parts are different since

$$\begin{aligned} \Lambda_d(t) &= \sum_{0 \leq s \leq t} \ln \frac{\Sigma(s-)}{\Sigma(s)} \qquad (2.24) \\ &= -\sum_{0 \leq s \leq t} \ln[1 - \{\hat{\Lambda}(s) - \hat{\Lambda}(s-)\}] \end{aligned}$$

$$\begin{aligned} \hat{\Lambda}_d(t) &= \sum_{0 \leq s \leq t} \left[1 - \frac{\Sigma(s)}{\Sigma(s-)}\right] \qquad (2.25) \\ &= \sum_{0 \leq s \leq t} [1 - \exp -\{\Lambda(s) - \Lambda(s-)\}]. \end{aligned}$$

Thus, from the logarithmic inequality $\ln x \geq 1 - x^{-1}$, we conclude that $\Lambda(t) \geq \hat{\Lambda}(t)$ for all t. Definition (2.20) has the advantage to be easily inverted; indeed

$$\Sigma(t) = e^{-\Lambda(t)}. \tag{2.26}$$

But inversion of definition (2.21) is slightly more complicated, since

$$\Sigma(t) = e^{-\hat{\Lambda}_c(t)} \prod_{0 \leq s \leq t} [1 - \{\hat{\Lambda}(s) - \hat{\Lambda}(s-)\}]. \tag{2.27}$$

Formula (2.27) is a simple expression of the so called product limit integral in the case of mixed cumulative hazard function.

The process $\hat{\Lambda}(t)$ has some nicer properties than the process $\Lambda(t)$. One of them, as seen in the definition(2.21), relies on the fact that $\hat{\Lambda}(t)$ has jumps of size smaller than one. Indeed, $\hat{\Lambda}(t) - \hat{\Lambda}(t-) = P[t_i = t \mid t_i \geq t]$.

Once Φ is distributed as a Dirichlet process with parameters (a, M), the law of the survivor process Σ, which is purely discrete in view of (2.13) and (2.14), is easily characterized. Indeed for any k and any ordered k-tuple $s_1 < s_2 < \cdots < s_k$, the joint distribution of the random vector $(\Sigma(s_1), \cdots, \Sigma(s_k))$ is obtained from the joint distribution of $1 - \Sigma(s_1) = \Phi((0, s_1]), \Sigma(s_1) - \Sigma(s_2) = \Phi((s_1, s_2]), \cdots, \Sigma(s_{k-1}) - \Sigma(s_k) = \Phi((s_{k-1}, s_k])$, $\Sigma(s_k) = \Phi((s_k, \infty))$, which is a Dirichlet distribution with parameters $n_0[1 - S_0(s_1)], n_0[S_0(s_1) - S_0(s_2)], \cdots, n_0[S_0(s_{k-1}) - S_0(s_k)], n_0 S_0(s_k)$ where S_0 is the survivor function associated to P_0. For the laws of Λ and $\hat{\Lambda}$ note first that both Λ and $\hat{\Lambda}$ have independent (non negative) increments. For the case of Λ, this is a direct consequence of a basic property of the Dirichlet process. Indeed for any k and any ordered k-tuple $s_1 < s_2 < \cdots < s_k$, the Dirichlet process makes $\Sigma(s_{j+1})(\Sigma(s_j))^{-1}$ independent of $(\Sigma(s_1), \Sigma(s_2), \cdots, \Sigma(s_j))$ and consequently of $(\Sigma(s_{i+1}))(\Sigma(s_i))^{-1} \quad \forall i < j$); on taking logarithm one obtains the independence of the increments of Λ in view of definition (2.20). The same properties hold for $\hat{\Lambda}(t)$ because, from (2.24) and (2.25), $\hat{\Lambda}(t)$ is a bijective transformation of $\Lambda(t)$ (same locations of jumps and bijective transformations of jump heights) and eventually keeps the same independence properties. Furthermore, as $(\Sigma(s_{j+1}))(\Sigma(s_j))^{-1}$ has a Beta distribution with parameters $(aM((s_{j+1}, \infty)), aM((s_j, s_{j+1}]))$ (see e.g. Rolin 1992a), the distribution of each increment of Λ is log-Beta with the same parameters.

A generalization of the Dirichlet process may be obtained by relaxing the log-Beta distribution property of the increments of Λ. More specifically, a positive right-continuous stochastic process indexed by $t \geq 0$ with independent non negative increments is called a *Levy process* and the associated random σ-finite measure is called a *purely random* measure. From (2.20), assuming Λ to be a Levy process is equivalent to assuming that Σ is neutral to the right, i.e., for any k and any ordered k-tuple $0 = s_0 < s_1 < s_2 < \ldots < s_k$, the random variables $\Sigma(s_{j+1})(\Sigma(s_j))^{-1}$ are mutually independent (for more details see, e.g., Doksum (1974)). It may be shown (see, e.g., Ferguson and Klass (1972)) that the continuous part of any Levy process is deterministic and that its discrete part is also a Levy process. Furthermore, Λ is a Levy process if and only if $\hat{\Lambda}$ is a Levy process, in which case they share the same deterministic continuous part but have different discrete parts. An example of such a Levy process, as a tool for modelling neutral to the right processes, is given by the Beta processes introduced by Hjort (1990).

To describe this extension, let us first assume that M is supported by a fixed finite ordered set $(a_1, ..., a_k)$. Then, if $a_0 = 0$, let us define

$$\hat{\Lambda}(\{a_j\}) = \hat{\Lambda}(a_j) - \hat{\Lambda}(a_{j-1}) = 1 - \frac{\Sigma(a_j)}{\Sigma(a_{j-1})} = P(t_i = a_j \mid t_i \geq a_j). \quad (2.28)$$

From the above discussion, $\hat{\Lambda}(\{a_j\})_{1\leq j\leq k}$, are independent and have Beta distributions with parameters $(aM((a_{j-1}, a_j]), aM((a_j, \infty)))$. It follows that

$$E[\hat{\Lambda}(\{a_j\})] = \hat{L}(\{a_j\}) = 1 - \frac{M((a_j, \infty))}{M([a_{j-1}, \infty))}. \quad (2.29)$$

Therefore,we may write that $\hat{\Lambda}(\{a_j\})$ has a Beta distribution with parameters $(c(a_j)\hat{L}(\{a_j\}), c(a_j)[1 - \hat{L}(\{a_j\})])$ where $c(a_j) = aM((a_{j-1}, \infty))$.

Now, the extension is obtained by saying that $\hat{\Lambda}$ is a discrete Beta process with parameters c and $\hat{L}$ if $\hat{\Lambda}(\{a_j\})$ is distributed as above with c an arbitrary function.

In the general case, if Φ is distributed as a Dirichlet process with parameter (a, M), $\hat{\Lambda}$ will be said to be distributed as a Beta process with parameter $(c, \hat{L})$ denoted by $\hat{\Lambda} \sim \mathcal{B}e(c, \hat{L})$ where

$$c(t) = aM([t, \infty)) \quad (2.30)$$

and

$$\hat{L}(t) = \int_{[0,t]} M([u, \infty))^{-1} M(du), \quad (2.31)$$

i.e., $\hat{L}$ is a cumulative hazard function of the probability M (see (2.21)).

Therefore, if a priori $\Phi \sim \mathcal{D}i(n_0, P_0)$ or equivalently $\hat{\Lambda} \sim \mathcal{B}e(c_0, \hat{L}_0)$, then a posteriori, $\Phi | t_1, t_2, ..., t_n \sim \mathcal{D}i(n_* P_*)$ or equivalently, $\hat{\Lambda} | t_1, t_2, ..., t_n \sim \mathcal{B}e(c_*, \hat{L}_*)$.

Now, clearly

$$c_*(t) = n_* P_*([t,\infty)) = c_0(t) + nP_n([t,\infty)) \tag{2.32}$$

and since (2.31) entails that $n_0 P_0(du) = c_0(u)\hat{L}_0(du)$,

$$\begin{aligned}\hat{L}_*(t) &= \int_{[0,t]} P_*([u,\infty))^{-1} P_*(du) \\ &= \int_{[0,t]} \frac{c_0(u)\hat{L}_0(du) + nP_n(du)}{c_0(u) + nP_n([u,\infty))}.\end{aligned} \tag{2.33}$$

Hjort (1990) has shown that taking c_0 to be an arbitrary positive measurable function on $I\!R_+$, then $\hat{\Lambda} \sim \mathcal{B}e(c_0, \hat{L}_0)$ is precisely defined as a prior specification and that a posteriori, $\hat{\Lambda}|t_1, t_2, ...t_n \sim \mathcal{B}e(c_*, \hat{L}_*)$ where c_* and $\hat{\Lambda}_*$ are defined by the last members of (2.32) and (2.33). Therefore, Beta processes form a natural conjugate family larger than Dirichlet processes (n_0, a real number, is replaced by a measurable function c_0) and smaller than the neutral to the right processes or equivalently the Levy processes.

3 Nonparametric duration models with censored observations.

One of the main features of duration data sets is the presence of censored observations. In this section, we want to show the extensions of the results presented in Section 2 to nonstochastically right-censored durations. Let us first recall that, in this case, the sample is generated as follows: for any $i = 1, \cdots, n$, c_i is a fixed duration, τ_i is a latent duration independently and identically distributed from an unknown probability Φ and we observe

$$\begin{aligned} t_i &= \min(\tau_i, c_i) \\ d_i &= I\!I_{\{\tau_i \le c_i\}} \end{aligned} \tag{3.1}$$

The unknown functional parameter is endowed with a prior probability (in this paper, a Dirichlet process) and we want to analyse the posterior probability of Φ given $(t_i, d_i)_{1 \le i \le n}$.

The main results of this analysis are the following:

i) The family of Dirichlet processes is not closed under such a sampling scheme. In other words, the posterior probability deduced from a Dirichlet prior and a censored sample is not a Dirichlet process. However, the class of neutral to the right processes is closed for the Bayesian inference as shown by Fergusson and Phadia (1979).

ii) Elements of this class are not as simple as Dirichlet processes but some of their characteristics can be obtained analytically. This is in particular the case for the expectation of the survivor function. An application of this computation is the posterior expected survivor function constructed through a Dirichlet prior (see Susarla and Van Ryzin (1976) and (1978), Blum and Susarla (1977)).

iii) Beta processes that constitute a strict subclass of neutral to the right processes are still closed for sampling with right-censoring as shown by Hjort (1990). This implies in particular that the posterior of a Dirichlet process prior is a Beta process.

In order to point out the essential elements of this inference procedure, we will start by the treatment of the finite case in which the support of Φ is finite. If we assume that the values c_i are also elements of this set, we may describe the sample using different counting statistics.

Let $\{a_1, \cdots, a_k\}$ be the support of Φ (with $a_1 < a_2 < \cdots < a_k$) and let n be the sample size. For any $j = 1, \cdots, k$ we define

$$e_j = \sum_{1 \leq i \leq n} 1\!\!I_{\{t_i = a_j, d_i = 1\}}, \tag{3.2}$$

the number of non censored durations equal to a_j

$$h_j = \sum_{1 \leq i \leq n} 1\!\!I_{\{t_i = a_j\}}, \tag{3.3}$$

the number of censored or non censored durations equal to a_j

$$n_j = \sum_{j \leq \ell \leq k} h_\ell = \sum_{1 \leq i \leq n} 1\!\!I_{\{t_i \geq a_j\}},$$

the number of individuals at risk in a_j. Note that, in particular, $n_1 = n = \sum_{1 \leq j \leq k} h_j$ and $h_j - e_j$ is the number of censored durations at a_j. Furthermore, the statistic $(h_j, e_j)_{1 \leq j \leq k}$ is obviously sufficient.

The probability Φ on $\{a_1, \cdots, a_k\}$ may be described in different ways :

i) the sequence $(\theta_j)_{1 \leq j \leq k}$, $\theta_j > 0$, $\sum_{1 \leq j \leq k} \theta_j = 1$, are the probabilities of the a_j's, i.e, $\theta_j = \Phi(\{a_j\})$, from which we can construct the survivor function

$$\Sigma(t-) = \sum_{a_j \geq t} \theta_j = \sum_{1 \leq j \leq k} \theta_j 1\!\!I_{\{a_j \geq t\}}. \tag{3.4}$$

ii) Φ is also characterized by the hazard rates sequence $(\lambda_j)_{1\leq j\leq k}$ defined by

$$\lambda_j = \hat{\Lambda}(\{a_j\}) = \frac{\theta_j}{\Sigma_j} \quad \text{where } \Sigma_j = \Sigma(a_j-). \tag{3.5}$$

In particular, the well known product formula connects θ_j and Σ_j to λ_j

$$\theta_j = \lambda_j \prod_{1\leq \ell<j} (1-\lambda_\ell) \quad \Sigma_j = \prod_{1\leq \ell<j} (1-\lambda_\ell). \tag{3.6}$$

The likelihood of the sufficient statistics may be written in terms of the probabilities or in terms of the hazard rates

$$\ell((h_j, e_j)_{1\leq j\leq k} \mid (\theta_j)_{1\leq j\leq k}) = \prod_{1\leq j\leq k} \theta_j^{e_j} \Sigma_j^{h_j - e_j} \tag{3.7}$$

or

$$\ell((h_j, e_j)_{1\leq j\leq k} \mid (\lambda_j)_{1\leq j\leq k}) = \prod_{1\leq j\leq k-1} \lambda_j^{e_j} (1-\lambda_j)^{n_j - e_j}. \tag{3.8}$$

Let us now consider the prior probability. With the same notation as in Section 2.1, a Dirichlet prior distribution on (θ_j) has density

$$m((\theta_j)_{1\leq j\leq k}) \propto \prod_{1\leq j\leq k} \theta_j^{n_0 p_{0j} - 1} \tag{3.9}$$

with respect to the uniform density restricted to the simplex in $\mathbb{R}^k$. Under the change of variables underlying (3.5) and (3.6), the prior density defined in (3.9) may be rewritten in terms of the (λ_j)'s as

$$m((\lambda_j)_{1\leq j\leq k}) \propto \prod_{1\leq j\leq k-1} \lambda_j^{n_0 p_{0j} - 1} (1-\lambda_j)^{n_0 S_{0\ j+1} - 1} \tag{3.10}$$

with respect to the uniform measure restricted to the set of hazard rates in $[0,1]^{k-1}$. In (3.10) S_0 is the expected prior survivor function

$$S_0(t-) = \sum_{1\leq j\leq k} p_{0j} \mathbb{1}_{\{a_j \geq t\}} \text{ and } S_{0j} = S_0(a_j-) \tag{3.11}$$

Thus the parameters $(\lambda_1, \cdots, \lambda_{k-1})$ are independently Beta distributed with parameter $(n_0 p_{0j}, n_0 S_{0\ j+1})$, $1 \leq j \leq k-1$. Note the restrictions among the parameters of the independent distributions of λ_j implied by (3.11). According to (2.29), an interesting property is the following:

$$E(\lambda_j) = \frac{n_0 p_{0j}}{n_0 p_{0j} + n_0 S_{0\ j+1}} = \frac{p_{0j}}{S_{0j}} = \lambda_{0j}. \tag{3.12}$$

In other words, the prior expected hazard rates are the hazard rates of the prior expected distribution.

The posterior probability on the θ_j's has the density

$$m((\theta_j)_{1\le j\le k}|(h_j,e_j)_{1\le j\le k}) \propto \prod_j \theta_j^{n_0p_{0j}+e_j-1}(\sum_{a_{j'}\ge a_{j+1}} \theta_j)^{h_j-e_j} \tag{3.13}$$

$$\propto \prod_j \theta_j^{n_0p_{0j}+e_j-1}\Sigma_{j+1}^{h_j-e_j}, \tag{3.14}$$

which is clearly not a Dirichlet distribution and the lack of closedness property of the class of the Dirichlet distributions is thus demonstrated.

The posterior density on the λ_j's is equal to

$$m((\lambda_j)_{1\le j\le k-1} \mid (h_j,e_j)_{1\le j\le k}) \propto \tag{3.15}$$

$$\prod_{1\le j\le k-1} \lambda_j^{n_0p_{0j}+e_j-1}(1-\lambda_j)^{n_0S_{0\ j+1}+n_j-e_j-1}$$

It follows from (3.15) that, a posteriori, λ_j's , $1\le j\le k-1$, are still independent. Any hazard rate λ_j follows a posteriori a Beta distribution whose parameters are equal to $n_0p_{0j}+e_j$ and $n_0S_{0\ j+1}+n_j-e_j$, but the previously noted restriction on the parameters of the independent distribution of λ_j no longer holds. Thus, the sequence of $k-1$ independent Beta probabilities looks like the prior specification but, apart from the case of non censored data, these distributions cannot be derived from a Dirichlet posterior on $(\theta_j)_{1\le j\le k}$. If, however, we compute the survivor function from the usual product formula (3.6) and exploit the posterior mutual independence of the λ_j's we find that

$$\begin{aligned} S_{nj}=E(\Sigma_j|(h_j,e_j),1\le j\le k) &= E(\textstyle\prod_{1\le \ell<j}(1-\lambda_\ell)|(h_j,e_j),1\le j\le k) \\ &= \textstyle\prod_{1\le \ell<j}(1-E(\lambda_\ell|(h_j,e_j),1\le j\le k)) \\ &= \textstyle\prod_{1\le \ell<j}(1-\dfrac{n_0p_{0j}+e_j}{n_0S_{0j}+n_j}) \end{aligned} \tag{3.16}$$

The Bayesian estimation $S_n(t)$ is deduced from the S_{nj} using the property that the survivor function is constant between the jumps and is right continuous at the jumps (and therefore everywhere).

The same computations show that this analysis goes through if, instead of a Dirichlet prior, we specify $\hat{\Lambda}$ to be a discrete Beta process. Indeed, in this

case, a priori the λ_j's are independently and Beta distributed with parameter $(c_{0j}\lambda_{0j}, c_{0j}(1-\lambda_{0j}))$ where c_{0j}, $1 \leq j \leq k-1$, are arbitrary constants. Hence a posteriori the λ_j's are independently and Beta distributed with parameter $(c_{*j}\lambda_{*j}, c_{*j}(1-\lambda_{*j}))$ where $c_{*j} = c_{0j} + n_j$ and $\lambda_{*j} = \dfrac{c_{0j}\lambda_{0j} + e_j}{c_{0j} + n_j}$.
These relations, in the finite case, extend formulas (2.32) and (2.33) in the case of censoring.

Let us consider the general case in which the functional parameter Φ is not constrained by a finite support condition and is distributed as a neutral to the right prior process. Its associated integrated hazard function $\Lambda(t) = -\ln \Sigma(t)$ (where $\Sigma(t) = \Phi((t,\infty))$) is a Levy process. Such a process has been defined in section 2.2.3. Let us recall that a Levy process is increasing with independent increments. For any sequence $s_1 < s_2 < \cdots < s_k$, the random variables $r_j = \Lambda(s_j) - \Lambda(s_{j-1})$ are independently distributed with densities $m_j(r_j)$, $1 \leq j \leq k$. In the case of a Dirichlet prior distribution, we have seen that the r_j are log-Beta distributed. It is sufficient to prove that for any observation t_i (censored or not) the posterior measure given t_i is still a Levy process. In what follows, we give a heuristic argument giving a clue to a more formal proof.

First let us assume that t_i is noncensored. In order to compute the posterior distribution of Λ restricted to any finite family of increments, we just have to consider the marginalized likelihood given this family. This likelihood reduces to the probability of the interval $(s_\ell, s_{\ell+1}]$ to which the observed t_i belongs. Then the posterior probability of $(r_j)_{1\leq j\leq k}$ is proportional to

$$
\begin{aligned}
& m((r_j):\ 1 \leq j \leq k \mid s_\ell < t_i \leq s_{\ell+1}) \propto \\
& [\prod_{1\leq j\leq k} m_j(r_j)](\Sigma(s_\ell) - \Sigma(s_{\ell+1})) \\
& = [\prod_{1\leq j\leq k} m_j(r_j)][\exp - \sum_{1\leq j\leq \ell} r_j][1 - \exp - r_{\ell+1}] \\
& = [\prod_{1\leq j\leq \ell} m_j(r_j)\exp - r_j][m_{\ell+1}(r_{\ell+1})(1 - \exp - r_{\ell+1})][\prod_{\ell+2\leq j\leq k} m_j(r_j)]
\end{aligned}
\tag{3.17}
$$

Thus, the r_j's are a posteriori independent. The posterior density of the first ℓ are proportional to $m_j(r_j)\exp(-r_j)$, the posterior density of $r_{\ell+1}$ is proportional to $m_{\ell+1}(r_{\ell+1})(1 - \exp - r_{\ell+1})$ and the distribution of the last increments is identical a priori and a posteriori.

An identical computation can be done in the case of a censored observation $t_i \in (s_\ell, s_{\ell+1}]$. The marginalized likelihood on the finite family of increments is equal to $\Sigma(s_\ell) = \exp(-\sum_{1\le j\le \ell} r_j)$ and the posterior distribution of the r_j's is identical to the previous one excepted for $r_{\ell+1}$ which is not revised by the observation.

Given a prior distribution on $\Lambda(t)$, one may derive the prior distribution of any sequence of increments and compute the posterior distribution using previous arguments sequentially on the sample.

In the special case of a Beta prior specification, i.e., $\hat{L} \sim \mathcal{B}e(c_0, \hat{L}_0)$, Hjort (1990) has shown that a posteriori $\hat{L}|(t_i, d_i)_{1\le i\le n} \sim \mathcal{B}e(c_*, \hat{L}_*)$ as in the case of no censoring where

$$c_*(t) = c_0(t) + \sum_{1\le i\le n} 1\!\!I_{\{t_i \ge t\}} \tag{3.18}$$

and

$$\hat{L}_*(t) = \int_{[0,t]} \frac{c_0(s)\hat{L}_0(ds) + \sum_{1\le i\le n} d_i \delta_{t_i}(ds)}{c_0(s) + \sum_{1\le i\le n} 1\!\!I_{\{t_i\ge s\}}} \tag{3.19}$$

For a Dirichlet process prior, the posterior is therefore characterized by (3.18) and (3.19) where $c_0(t) = n_0 S_0(t-)$.

In particular (Sursala, Van Ryzin (1976)), the posterior expectation of the survival function is given by

$$\begin{aligned} E(\Sigma(t)|(t_i, d_i)_{1\le i\le n}) \quad &= \frac{n_0 S_0(t) + n(t)}{n_0 + n} \\ &\prod_{1\le j\le \ell} \frac{n_0 S_0(u_j) + n(u_j) + f_j}{n_0 S_0(u_j) + n(u_j)} \qquad t \in [u_\ell, u_{\ell+1}) \end{aligned} \tag{3.20}$$

where $u_1, u_2, \cdots, u_k$ are the distinct observed values of the sample, $n(t)$ is the number of individuals at risk after t i.e. $n(t) = \sum_{1\le i\le n} 1\!\!I_{\{t_i > t\}}$, f_j is the number of censored observations at u_j and S_0 is the prior survivor function. Let us remark that, in absence of censoring, (3.20) reduces to the usual result given in section 2.

A more detailed characterization of the posterior distribution for a Dirichlet process prior with censored observations will be given in a future paper providing easy simulation to analyze posterior distributions of functionals of the survival function.

For general Beta processes and more generally Levy processes, no easy to simulate descriptions of the trajectories are available. Simulation techniques must rely on more complicated schemes simulating probabilities of intervals as in Damien, Laud and Smith (1995) and (1996). However this method does not lead to simulations of the distributions of functionals. In the case of non informative priors, Lo (1987) and (1993) provides a Bayesian Bootstrap for censored durations.

4 Heterogeneity and Mixture of Dirichlet Processes.

4.1 Introduction

The last two sections of this paper are devoted to the study of models conditional on a variable or a function of variables describing individual heterogeneity. In other words, the first step in the specification consists in describing the law of the duration t_i conditionally on θ_i, a variable representing individual characteristics of the ith individual, i.e., one needs to specify

$$\Sigma(t|\theta_i) = P(t_i > t|\theta_i, \Phi)$$

Two characteristics have to be taken into account: on one hand, whether θ_i observable or not, and on the other hand, the class of conditional models to be considered (usually a proportional hazards or an accelerated lifetime model).

As a first step, θ_i may be considered as a function of observed explanatory variables z_i and of an unknown structural parameter β (the same for each individual). This is known as observed heterogeneity and the z_i's are also known as treatment variables or covariates. Econometric literature has also emphasized the interest in considering θ_i as an unobserved realization of a random variable (see e.g., Heckman (1981), Heckman and Singer (1982), (1984a) and (1984b) and Lancaster (1990)). The main reference for identification problems in such models is Elbers and Ridder (1982). A motivation for this model is the following. Suppose that each individual has a duration generated by the exponential law of parameter $\lambda\theta_i$. Given θ_i, the hazard rate of the ith individual is time independent but if θ_i is random, the marginal distribution of t_i has a decreasing hazard rate. This is known as "spurious time dependence". The same argument of heterogeneity is also used to explain the u-shaped observed hazard rates in reliability theory.

At last, two types of conditional models for duration data are generally considered: accelerated lifetime models and proportional hazards models. In the first type of models, the observed lifetime t_i is written as $\theta_i\tau_i$ where τ_i is a basic lifetime and θ_i appears as an acceleration factor and therefore,

$$\Sigma(t|\theta_i) = \Sigma(\theta_i^{-1}t)$$

where Σ is the survival function of τ_i's.

In the proportional hazards model, the observed lifetime has, conditionally on θ_i, a survival function given by

$$\Sigma(t|\theta_i) = \Sigma(t)^{\theta_i}$$

or equivalently has a hazard function given by

$$\Lambda(t|\theta_i) = \theta_i \Lambda(t).$$

In this paper, we shall only consider two alternative combinations of conditional models and heterogeneity. In this section, we analyse an accelerated lifetime model with un-observable heterogeneity. In the next section we discuss proportional hazards model with observable explanatory variables.

The interest in the accelerated lifetime model with unobserved heterogeneity lies in the fact that integration of the unobservable variable produces a "smoothing" of the trajectories of the Dirichlet process and provides an estimation of the density and of the hazard rate. This is a Bayesian version of kernel estimators (see, e.g., Lo and Weng (1989)). In the proportional hazards model, however, the discrete character of the Dirichlet process is preserved after integration since the locations of the jumps are unchanged. Moreover, the neutral to the right property is lost and this makes the computation of the posterior much more difficult. However such a model may sometimes be useful and may be considered as a byproduct of the computations sketched in the next section.

In the proportional hazards model with $\theta_i = a(\beta, z_i)$ the marginal likelihood on β is easily obtained and provides an alternative to the well-known Cox's partial likelihood model. On the contrary, the semiparametric analysis of the accelerated lifetime model with $\theta_i = a(\beta, z_i)$ with a Dirichlet process prior for Φ, i.e. $\Phi \sim \mathcal{D}i(n_0 P_0)$ has little interest. Indeed, as far as all the observations are distinct, the posterior distribution of β is identical to the posterior distribution obtained in the parametric model where the τ_i's are independent identically distributed. P_0 (see e.g. Bunke (1981)).

4.2 A simple model.

Let us consider n observed durations $(t_1, \cdots, t_i, \cdots, t_n)$ along with the multiplicative decomposition

$$t_i = \theta_i \tau_i \qquad 1 \leq i \leq n. \tag{4.1}$$

Both θ_i's and τ_i's are unobservable and independent identically distributed For identifiability, we assume that the distribution of the θ_i's is known and is characterized by a density function. In contrast,the distribution of the τ_i's is unknown and assumed to be a priori distributed as a Dirichlet process.

In the context of the heterogeneity problem, model (4.1) may be interpreted as follows. The duration τ_i of each individual i is assumed to be generated by a same unknown distribution but accelerated by a factor θ_i reflecting unobservable individual characteristics. A simple solution to the standard problem of identification in accelerated time models is provided by the assumption that the distribution of the factor θ_i is known.

It is interesting to notice that, because the t_i's can be represented as the product of two quantities their sampling distribution is *a.s.* smooth, i.e. admits a density *a.s.*, in spite of the Dirichlet specification. In other words, the introduction of θ_i acts similarly to a smoothing kernel, a main difference being a multiplicative convolution rather than an additive one. The multiplicative model is indeed more natural than the additive one when dealing with non negative random variables. One of the results of this model is to produce a Bayesian analogue to kernel estimators.

Let us now be more specific about the basic assumptions underlying (4.1).

(A.1) $(\theta_i)_{1\leq i\leq n}$ are independent identically distributed; their common distribution, denoted by Q, is known, supported by $I\!R_+$ and admits a known density q.

(A.2) $(\tau_i)_{1\leq i\leq n}$ are independent identically distributed; the common distribution denoted by Γ, is unknown and distributed a priori as a Dirichlet process with parameters (n_0, G_0) where G_0 is a probability measure supported by $I\!R_+$.

(A.3) The θ_i's and the τ_i's are jointly independent in the sampling, i.e.,

$$(\theta_i)_{1\leq i\leq n} \perp\!\!\!\perp (\tau_i)_{1\leq i\leq n} | \Gamma \tag{4.2}$$

These assumptions imply that the t_i's are independent identically distributed with common distribution denoted by Φ, being a multiplicative convolution between Q and Γ and will be denoted by $Q.\Gamma$. More precisely

$$\begin{aligned} \Phi([0,t]) = Q.\Gamma([0,t]) &= \int_{I\!R_+} \Gamma([0,\frac{t}{\theta}])Q(d\theta) \\ &= \int_{I\!R_+} Q([0,\frac{t}{\tau}])\Gamma(d\tau). \end{aligned} \tag{4.3}$$

Because Q admits a density q, (4.3) may be rewritten as

$$\Phi([0,t]) = \int_0^t \int_{I\!R_+} \frac{1}{\tau} \; q(\frac{u}{\tau}) \; \Gamma(d\tau) du. \tag{4.4}$$

Thus Φ is dominated by Lebesgue measure and admits a density φ defined as

$$\varphi(t) = \int_{\mathbb{R}_+} \frac{1}{\tau} \; q(\frac{t}{\tau}) \; \Gamma(d\tau). \tag{4.5}$$

When $\Gamma \sim \mathcal{D}i(n_0, G_0)$ the distribution of Φ, derived from (4.3), is an example (a particular case) of a mixture of Dirichlet processes (see, e.g., Antoniak (1974)). One may easily check that

$$E[\Phi] = Q.E(\Gamma) = Q.G_0 \tag{4.6}$$

$$\frac{d}{dt} E[\Phi([0,t])] = \int_{\mathbb{R}_+} \frac{1}{\tau} \; q(\frac{t}{\tau}) \; G_0(d\tau) \tag{4.7}$$

When G_0 is a continuous probability measure, results given in Section 2.2.2 imply that $\varphi(t)$ may be represented as

$$\varphi(t) = \sum_{1 \leq k < \infty} \gamma_k \frac{1}{\sigma_k} q(\frac{t}{\sigma_k}) \tag{4.8}$$

and may accordingly be easily simulated using (2.18) and the fact that $(\sigma_k)_{1 \leq k < \infty}$ is an infinite independent identically distributed sample from G_0.

The posterior distribution of Φ is somewhat more involved. Let us have a look its expectation and first evaluate it conditionally on $(\tau_1, \cdots, \tau_n)$.

$$E[\Phi \mid t_1, \cdots, t_n, \tau_1, \cdots, \tau_n] = \frac{n_0}{n_0 + n}(Q.G_0) + \frac{n}{n_0 + n}(Q.G_n) \tag{4.9}$$

where G_n is the empirical distribution of the τ_i's, i.e.,

$$G_n = n^{-1} \sum_{1 \leq i \leq n} \delta_{\tau_i}. \tag{4.10}$$

In order to integrate the τ_i's out, it is convenient to represent $(\tau_1, \cdots, \tau_n)$ through three components namely,(1) p, the number of the different values of the τ_i's,(2) a partition C_n of $\{1, \cdots, n\}$ into p elements $(I_j)_{1 \leq j \leq p}$ where each I_j is a set of indices corresponding to identical values of τ_i (thus C_n represents the partition of the configuration of ties in the τ_i's), and finally (3) the p-vector of distinct values of the τ_i's: $(\tau_{(j)})_{1 \leq j \leq p}$. Therefore G_n may also be written as

$$G_n = \frac{1}{n} \sum_{1 \leq j \leq p} n_j \delta_{\tau_{(j)}} \tag{4.11}$$

where $n_j = |I_j|$ and consequently $\sum_{1\le j\le p} n_j = n$. Note also that

$$t_i = \tau_{(j)}\theta_i \qquad \forall i \in I_j \tag{4.12}$$

where this expression is based on the configuration C_n. Let us now denote by G_n^j the conditional probability of $\tau_{(j)}$ conditionally on $t_1, \cdots, t_n$ and C_n. Thus

$$G_n^j(A) = E[\mathbb{1}_A(\tau_{(j)}) \mid t_1, \cdots, t_n, C_n] \tag{4.13}$$

It may be shown, and it is intuitively rather natural, that G_n^j depends only on those t_i's for which $i \in I_j$ and may be evaluated easily using standard techniques for evaluating posterior distributions; this is so, in particular, because the θ_i's are independent identically distributed with known distribution Q and the $\tau_{(j)}$'s are a priori indepently distributed according to G_0. Later on we give an example for evaluating G_n^j. If we first integrate $(\tau_{(j)})_{1\le j\le p}$ out from (4.9), conditionally on the configuration C_n, we obtain

$$E[\Phi \mid t_1, t_2, \ldots, t_n, C_n] = \frac{n_0}{n_0+n}(Q \cdot G_0) + \frac{1}{n_0+n}\sum_{1\le j\le p} n_j\ (Q \cdot G_n^j) \tag{4.14}$$

Just as above, the properties of the trajectories of the Dirichlet process may be used to generate the posterior distribution of the density $\varphi(t)$ given $(t_1, t_2, \ldots, t_n)$ and the configuration of ties, C_n, using the representation

$$\varphi(t) = (1-\alpha_n)\sum_{1\le k<\infty} \gamma_k \frac{1}{\sigma_k} q(\frac{t}{\sigma_k}) + \alpha_n \sum_{1\le j\le p} \beta_{(j)} \frac{1}{\sigma_{(j)}} q(\frac{t}{\sigma_{(j)}}), \tag{4.15}$$

where α_n has a Beta distribution with parameter (n, n_0), $\{\beta_{(j)} : 1 \le j \le p\}$ has a Dirichlet distribution with parameter $\{n_j : 1 \le j \le p\}$ and $\sigma_{(j)}$ are independently generated from G_n^j, $1 \le j \le p$.

Since the configuration C_n is unknown, we may, at least formally, integrate C_n out from (4.14) conditionally on $(t_1, t_2, \ldots, t_n)$, to obtain

$$\begin{aligned} E[\Phi \mid t_1, t_2, \ldots, t_n] \;=\; & \frac{n_0}{n_0+n}(Q \cdot G_0) \qquad (4.16) \\ & + \frac{1}{n_0+n}\sum_{C_n\in\mathcal{C}_n} P(C_n|t_1, \ldots, t_n) \sum_{1\le j\le p} n_j\ (Q \cdot G_n^j) \end{aligned}$$

where $\mathcal{C}_n$ is the set of all partitions of $\{1, 2, \ldots, n\}$. In order to simulate $\varphi(t)$ conditionally on $(t_1, t_2, \ldots, t_n)$ we may use formula (4.15), if we first generate C_n conditionally on $(t_1, t_2, \ldots, t_n)$.

Unfortunately, an exact evaluation of (4.16) is close to impossible for reasonably large sample size n. Indeed, the evaluation of $P(C_n|t_1, \cdots, t_n)$ is rather involved in view of the analysis given in section 2.3 and of the fact that $|\mathcal{C}_n|$ increases dramatically with n.

Different strategies can be envisaged to address this difficulty. Two of these are the following. The first consists in selecting arbitrarily a configuration C_n of ties. This is leaving the strict Bayesian framework as far as one is conditioning on a non-available information. Although general theorems may ensure the convergence of $E(\Phi \mid t_1, \cdots, t_n)$, the convergence of $E(\Phi \mid t_1, \cdots, t_n, C_n)$ requires specific hypotheses for arbitrary choices of C_n.

Another strategy (proposed and largely used by Escobar (1994) and Escobar and West (1994) relies on simulation methods (post-data sampling). Here Gibbs sampling is a natural candidate for generating $(\tau_i)_{1\leq i\leq n}$ conditionally on $(t_i)_{1\leq i\leq n}$. This distribution is rather complicated and almost unmanageable if n is large but the conditional distributions are however rather simple. Indeed, conditionally on $(\tau_{i'})_{i'\neq i}$ and $(t_{i'})_{i'\neq i}$, (see formula (2.24)) the distribution of τ_i is given by

$$G_0^i = \frac{n_0}{n_0+n-1} G_0 + \frac{1}{n_0+n-1} \sum_{i'\neq i} \delta_{\tau_{i'}} \tag{4.17}$$

and the conditional distribution of t_i given τ_i has density

$$h(t_i \mid \tau_i) = \frac{1}{\tau_i} q\left(\frac{t_i}{\tau_i}\right) \tag{4.18}$$

Then by Bayes theorem the conditional distribution of τ_i given t_i (and $(\tau_{i'})_{i'\neq i}$ and $(t_{i'})_{i'\neq i}$) is deduced from (4.18) and (4.19), i.e.,

$$P[\tau_i \in A | \tau_{i'}, i' \neq i, ; t_i, 1 \leq i \leq n] = \frac{n_0 h(t_i) G_0^{t_i}(A) + \sum_{i'\neq i} \frac{1}{\tau_{i'}} q\left(\frac{t_i}{\tau_{i'}}\right) \delta_{\tau_{i'}}(A)}{n_0 h(t_i) + \sum_{i'\neq i} \frac{1}{\tau_{i'}} q\left(\frac{t_i}{\tau_{i'}}\right)} \tag{4.19}$$

where $G_0^{t_i}$ is the conditional distribution of τ_i given t_i computed with G_0 as the prior distribution for τ_i. The predictive density $h(t_i)$ is the density of $Q.G_0$ (formula (4.7)).

Starting from an initial value of $(\tau_i)_{1\leq i\leq n}$, the Gibbs sampling procedure will generate sequentially τ_i from (4.20) to finally obtain a draw of $(\tau_i)_{1\leq i\leq n}$ conditional on $(t_i)_{1\leq i\leq n}$. From this, a realization of the posterior distribution of Γ or of Φ may be computed.

A review of a lot of techniques for estimating smoothly densities in the Bayesian framework is provided by Hjort (1996).

4.3 A particular case.

In this section, we give a simple example of the computations suggested above.

Let Q be the inverse gamma distribution with parameter $(\theta_0, \nu_0) \in \mathbb{R}_+^2$ the density of which is given by

$$q(\theta) = \frac{\theta_0^{\nu_0}}{\Gamma(\nu_0)} \theta^{-(\nu_0+1)} e^{-\theta_0/\theta}, \tag{4.20}$$

and let G_0 be the gamma distribution with parameter $(\tau_0, \mu_0) \in \mathbb{R}_+^2$ whose density is given by

$$g_0(\tau) = \frac{1}{\Gamma(\mu_0)} \tau_0^{-\mu_0} \tau^{\mu_0 - 1} e^{-\tau/\tau_0} \tag{4.21}$$

Thus the distribution of t_i conditionally on τ_i is the an inverse-gamma distribution with parameter $(\theta_0\tau_i, \nu_0)$ since its density is given by $\dfrac{1}{\tau_i} q\left(\dfrac{t}{\tau_i}\right)$

Simple computations show that the distribution of τ_i conditionally on t_i, denoted above as $G_0^{t_i}$, is the gamma distribution with parameter $\left[\left(\dfrac{1}{\tau_0} + \dfrac{\theta_0}{t_i}\right)^{-1}, \mu_0 + \nu_0\right]$. Furthermore the distribution of t_i, denoted above as $Q \cdot G_0$ is the Fisher distribution with parameter $(\theta_0\tau_0, \mu_0, \nu_0)$ the density of which is given by

$$g(t) = \frac{\Gamma(\mu_0 + \nu_0)}{\Gamma(\mu_0)\Gamma(\nu_0)} \frac{(\theta_0\tau_0)^{-\mu_0-\nu_0-1}}{\left(1 + \frac{t}{\theta_0\tau_0}\right)^{\mu_0+\nu_0}}. \tag{4.22}$$

Now conditionally on $\tau_{(j)}$ and C_n, the t_i's for $i \in I_j$ are distributed independently following the inverse gamma distribution with parameter $(\theta_0\tau_{(j)}, \nu_0)$. Therefore, the likelihood is proportional to

$$\prod_{i \in I_j} \tau_{(j)}^{\nu_0} e^{-\theta_0 \tau_{(j)}/t_i} = \tau_{(j)}^{n_j\nu_0} e^{-\theta_0\tau_{(j)} \sum_{i\in I_j} t_i^{-1}}.$$

By Bayes' theorem, the distribution of $\tau_{(j)}$ conditionally on $\{t_i; 1 \leq i \leq n\}$ and C_n denoted before as G_n^j is the gamma distribution with parameter

$$\left[\left(\frac{1}{\tau_0} + \theta_0 \sum_{i\in I_j} \frac{1}{t_i}\right)^{-1}, \mu_0 + n_j\nu_0\right].$$

Finally, by the same computation as before, $Q \cdot G_n^j$ is the Fisher distribution with parameter

$$\left[\left(\frac{1}{\theta_0\tau_0} + \sum_{i\in I_j} \frac{1}{t_i}\right)^{-1}, \mu_0 + n_j\nu_0, \nu_0\right].$$

This implies that the posterior expectation of Γ conditionally on C_n has a density given by a convex combination of gamma densities and that the posterior expectation of Φ conditionally on C_n has a density given by a convex combination of Fisher densities. Notice that, in such a case, $(t_1, t_2, \ldots, t_n, C_n)$ has been reduced, by sufficiency, into $y_n = ((n_j, \sum_{i \in I_j} t_i^{-1}),\ 1 \le j \le p)$. This reduction has been made possible because the mixed distribution, i.e., the distribution $(t_i|\tau_i)$, is a member of the exponential family.

5 Semiparametric model with proportional hazards

The last section of this article is devoted to a Bayesian treatment of the semiparametric analysis of duration models conditional on observed explanatory variables. We restrict the attention to fixed explanatory variables (i.e., only dependent on the individuals but not on time) acting through a proportional hazards model. Moreover we assume for the sake of exposition that the data are observed without censoring and we do not introduce an unobserved heterogeneity component.

The sample is now defined by the sequence (t_i, z_i) , $i = 1, \cdots, n$, where t_i is a duration and z_i a vector of explanatory variables. The observations are assumed to be independent and the sampling process is characterized by the distribution of t_i conditionally on z_i. This conditional probability Φ_i may be characterized by its survivor function Σ_i which is assumed to satisfies

$$\Sigma_i(t) = \Sigma(t)^{a(\beta, z_i)}. \tag{5.1}$$

Here Σ is an unknown survivor function of a baseline probability Φ and $a(\beta, z_i)$ is a known positive function of an unknown vector of parameters β and of the explanatory variables z_i. A common choice for a is $a(\beta, z_i) = \exp \beta' z_i$.

The specification (5.1) is equivalent to

$$\Lambda_i(t) = a(\beta, z_i)\Lambda(t) \tag{5.2}$$

where $\Lambda_i(t)$ is the conditional integrated hazard associated to $\sum_i$ (i.e., $\Lambda_i(t) = -\ln \sum_i(t))$ and $\Lambda(t)$ the integrated hazard associated to the baseline survivor function $\sum(t)$. Relation (5.2) justifies the name " proportional hazards model".

The sampling model is then indexed by a functional parameter (the baseline probability Φ or equivalently $\sum$ or Λ) and by a vector of parameters β.

The Bayesian specification is completed by the choice of a prior distribution. Given β, Φ is endowed with a Dirichlet process with parameters (n_0, Φ_0) that are possibly dependent on β. The vector β has a prior density $m(\beta)$.

Another model has been used by Kalbfleisch (1978), specifying a Gamma process on Λ, i.e., a Levy process with gamma distributed increments.

This section is essentially devoted to the computation of the posterior distribution of β. This means that the functional parameter is treated as a nuisance parameter and integrated out analytically. However the posterior distribution of β must be treated numerically (in order to compute, for example, its moments or its marginal densities). There is no suitable choice of $m(\beta)$ that would simplify the posterior computations and this density will be left unspecified.

In order to simplify the computations we assume that the observed sample contains no ties and, therefore, has n distinct durations. This assumption may be easily verified and the following derivations may be extended to samples with ties (see Ruggiero (1989)). Under the independence property one may assume without loss of generality that $t_1 < t_2 < \cdots < t_n$.

Let us start with the joint sampling survivor function

$$\begin{aligned}\prod_{1\leq i\leq n} \Sigma_i(t_i) &= \prod_{1\leq i\leq n} \Sigma(t_i)^{a(\beta,z_i)} \\ &= \exp\{-\sum_{1\leq i\leq n} a(\beta, z_i)\Lambda(t_i)\} \qquad (5.3) \\ &= \exp\{-\sum_{1\leq i\leq n} a(\beta, z_i) \sum_{1\leq j\leq i} \gamma_j\}\end{aligned}$$

where $\gamma_j = \Lambda(t_j) - \Lambda(t_{j-1})$ for $j > 1$ and $\gamma_1 = \Lambda(t_1)$. Therefore,

$$\prod_{1\leq i\leq n} \Sigma_i(t_i) = \prod_{1\leq i\leq n} \exp\{-\gamma_i A_i(\beta)\} \qquad (5.4)$$

where $A_i(\beta) = \sum_{i\leq j\leq n} a(\beta, z_j)$. Note that the joint sampling survivor function depends on the functional parameter Λ through the sequence $(\gamma_i)_{1\leq i\leq n}$ only. As a consequence, integrating out the functional parameter Λ may be performed by integrating out the γ_i's from(5.4). Furthermore, the Dirichlet prior on Φ implies that the γ_i are a priori independent with log-Beta prior density, i.e.,

$$m(\gamma_i) = \frac{\Gamma(n_0\Sigma_0(t_{i-1}))}{\Gamma(n_0\Sigma_0(t_i))\Gamma(n_0\Sigma_0(t_{i-1}) - n_0\Sigma_0(t_i))} \times (e^{-\gamma_i})^{n_0\Sigma_0(t_i)}(1-e^{-\gamma_i})^{n_0(\Sigma_0(t_{i-1})-\Sigma_0(t_i))-1}. \tag{5.5}$$

After integration of the γ_i 's the joint survivor function of the t_i 's is therefore equal to

$$\begin{aligned} S(t_1,\cdots,t_n|\beta) &= \prod_{1\le i\le n}\int_0^\infty m(\gamma_i)\exp\{-\gamma_i A_i(\beta)\}d\gamma_i \\ &= \frac{\Gamma(n_0)}{\Gamma(n_0\Sigma_0(t_n))}\prod_{1\le i\le n}\frac{\Gamma(n_0\Sigma_0(t_i)+A_i(\beta))}{\Gamma(n_0\Sigma_0(t_{i-1})+A_i(\beta))} \qquad (5.6) \\ &= \frac{\Gamma(n_0)}{\Gamma(n_0+A_1(\beta))}\prod_{1\le i\le n}\frac{\Gamma(n_0\Sigma_0(t_i)+A_i(\beta))}{\Gamma(n_0\Sigma_0(t_i)+A_{i+1}(\beta))}. \end{aligned}$$

The corresponding density is given by

$$\begin{aligned} &\ell(t_1,\cdots,t_n|\beta) \\ &= (-1)^n\frac{\partial^n}{\partial t_1\cdots\partial t_n}S(t_1,\cdots,t_n|\beta) \\ &= (-1)^n\frac{\Gamma(n_0)}{\Gamma(n_0+A_1)}\prod_{1\le i\le n}n_0\Sigma'_{0i} \qquad (5.7) \\ &\quad [\Gamma'(n_0\Sigma_{0i}+A_i)\Gamma(n_0\Sigma_{0i}+A_{i+1}) - \Gamma'(n_0\Sigma_{0i}+A_{i+1})\Gamma(n_0\Sigma_{0i}+A_i)] \\ &\quad [\Gamma(n_0\Sigma_{0i}+A_{i+1})]^{-2} \end{aligned}$$

where Γ' is the derivative of the gamma function and $\Sigma_{0i} = \Sigma_0(t_i)$. The posterior density of β follows from Bayes rule

$$m(\beta|t_i,\cdots,t_n) \propto m(\beta)\ell(t_1,\cdots,t_n|\beta) \tag{5.8}$$

A complete analysis of the marginalized density in this case (computation of score and second derivative) can be found in Hakizamungu (1992).

In (5.6), we have computed the joint marginalized survivor function in the (open) subset of $(I\!R_+)^n$ in which all the durations are different and this function is differentiable on this subset. However the marginalized survivor function is not differentiable on the subsets defined by configurations of ties and the application of Bayes theorem becomes more complicated if the sample has ties (see Ruggiero (1989) for a general description of this computation).

A final, natural question concerns the relation between the Bayesian marginalized likelihood (5.7) and the Cox marginalized likelihood. Note first that the word "marginalized" has different signification in the two approaches; in the Bayesian analysis the marginalization is obtained through an integration of the nuisance parameter Φ using a prior probability and in Cox analysis the marginalization is realized on the rank statistic in the sampling distribution. However some connection between the two results might be expected in the case of "non informative" prior measure on Φ (or on Γ). A natural way is to compute the posterior distribution of β with a uniform prior measure on β, i.e., $m(\beta) = 1$ in (5.8) and to take its limit when $n_0 \to 0$. The result becomes

$$m(\beta | t_1, \cdots, t_n) \propto \frac{1}{\Gamma(A_1)} \prod_{i=1}^{n} \frac{\Gamma'(A_i)\Gamma(A_{i+1}) - \Gamma'(A_{i+1})\Gamma(A_i)}{[\Gamma(A_{i+1})]^2} \tag{5.9}$$

Note that in (5.9) individual durations t_i disappear and the rank statistics become sufficient but this posterior density on β is rather different from the Cox marginalized likelihood.

References

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

Andersen, P.K., N. Keiding (1995), Survival Analysis, Research Report, Department of Biostatistics, University of Copenhagen.

Antoniak, C.E. (1974), Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems, *Annals of Statistics*, **2**, 1152-1174.

Blackwell D., J.B. MacQueen (1973), Ferguson distributions via olya urn schemes, *Annals of Statistics* 1, 353-355.

Blum K., V. Susarla (1977), On the posterior distribution of a Dirichlet process given randomly right censored data, *Stochastic Process. Appl.* 5, 207-211.

Bunke, O. (1981), Bayesian Estimators in Semiparametric Problems, Preprint 102, Humboldt University, Berlin.

Damien P., P.W. Laud, A.F.M. Smith (1995), Approximate random variate generation from infinitely divisible distributions with applications to Bayesian inference, *Journal of the Royal Statistical Society*, Series B, 57(3), 547-563.

Damien P., P.W. Laud, A.F.M. Smith (1996), Implementation of Bayesian non-parametric inference based on Beta processes, *Scandinavian Journal of Statistics*, 23, 27-36.

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

Elbers, C. and G. Ridder (1982), True and Spurious Dependence: The identifiability of the proportional hazards model, *Review of Economic Studies*, **49**, 403-409.

Erkanli, A., Muller, P. and M. West (1993), Curve Fitting using Dirichlet Process Mixtures. Tech. report, ISDS, Duke University.

Escobar, M.D. (1994), Estimating Normal means with a Dirichlet Process prior, *Journal of the American Statistical Association*, **89**, 268-277.

Escobar, A. and M. West (1994), Bayesian Prediction and Density Estimation, Mimeo July 1994.

Ferguson, T. (1973), A Bayesian Analysis of Some Nonparametric Problems, *Annals of Statistics*, **1**, 209-230.

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

Ferguson, T.S. and M.J. Klass (1972), A representation of Independent Processes without Gaussian components, *Annals of Math. Stat.*, **43**, 1634-1643.

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

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

Florens, J.P., D. Fougère and M. Mouchart (1995), Duration Models. In Matyas and Sevestre (eds) *Econometrics of Panel Data*, 2nd Ed. Kluwer Academic Publishers, Dordrecht, 491-536.

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

Florens, J.P., M. Mouchart and J.M. Rolin (1992), Bayesian Analysis of Mixtures: Some Results on Exact Estimability and

Identification. In Bernado, J.M. and al., *Bayesian Statistics* , **4**, Oxford Science Publications, 127-145.

Florens, J.P. and J.M. Rolin (1994), Bayes, Bootstrap, Moments, IS Discussion Paper n∘ 9413, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Hakizamungu, J. (1992), Modèles de Durée: Approche Bayesienne Semiparamétrique, Unpublished Ph.D. Dissertation, Faculté des Sciences, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Heckman, J.J. (1981), Heterogeneity and State Dependence, in: *Studies in Labor Markets*, ed. S. Rosen, Chigago, University of Chigago Press, 91-139.

Heckman, J.J. and B. Singer (1982), The Idendification Problem in Econometric Models for Duration Data, in *Advances in Econometrics*, ed. by W. Hildenbrand, Cambridge University Press, 39-77.

Heckman, J.J., and B. Singer (1984a), Econometric Duration Analysis, *Journal of Econometrics*, 24, 63-132.

Heckman, J.J. and B. Singer (1984b), The Identifiability of the Proportional Hazards Model, *Review of Economic Studies*, **51**, 231-243.

Hjort, N. (1990), Nonparametric Bayes Estimators Based on Beta Processes in Models For Life History Data, *Annals of Statistics*, **18**, 1259-1294.

Hjort,N. (1996), Bayesian approaches to non-and semiparametric density estimation in *Bayesian Statistics 5*, Bernardo, J.M., Berger, J.O., Dawid,A.P. and Smith, A.F.M.(Eds.), Oxford University Press, 223-253.

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

Kalbfleisch, J.D. and R.L. Prentice (1980), *The Statistical Analysis of Failure Time Data*, New York, Wiley.

Korwar, R.M. and M. Hollander (1973), Contributions to the Theory of Dirichlet Processes, *The Annals of Probability*, **1**, 705-711.

Lahiri, P. and P. Dong Ho (1988), Nonparametric Bayes and Empirical Bayes Estimation of the Residual Survival Function at Age t, *Commun. Statist. Theory Meth.*, 4085-4098.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Econometric Society Monograph, New York, Cambridge

University Press.
Lo A.Y.(1987), A large sample study for the Bayesian bootstrap, *Annals of Statistics* 15, 360-375.
Lo A.Y.(1993), A Bayesian bootstrap for censored data, *Annals of Statistics* 21, 100-123.
Lo, A.Y and C.S. Weng (1989), On a Class of Bayesian Nonparametric Estimates: II. Hazard Rates Estimates, *Annals of the Institute of Statistical Mathematics*, 227-245.
McEachern, S. (1992), Estimating normal means with a conjugate style Dirichlet process prior, Tech. report, Dept. of Statistics, Ohio State.
Phadia, E. and V. Susarla (1983), Nonparametric Bayesian Estimation of a Survival with Dependent Censoring Mechanism, *Annals of the Institute of Statistical Mathematics*, 389-400.
Rolin, J.M. (1983), Nonparametric Bayesian Statistic: A Stochastic Process Approach. In: *Specifying Statistical Models, From Parametric to Nonparametric Using Bayesian Approaches.* Lecture Notes in Statistics, Springer, Berlin, 108-133.
Rolin, J.M. (1992a), Some Useful Properties of the Dirichlet Process. Core Discussion Paper, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
Rolin, J.M. (1992b), On the Distribution of Jumps of the Dirichlet Process, Core Discussion Paper, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
Rolin, J.M. (1998), Bayesian Survival Analysis, in: *Encyclopedia of Biostatistics*, New-York, Wiley, 1998.
Ruggiero, M. (1989), Analyse semi-paramétrique des modèles de durées: L'apport des méthodes bayésiennes, Thèse de doctorat, Université d'Aix-Marseille II, France.
Ruggiero, M. (1994), Bayesian Semiparametric estimation of proportional hazard models, *Journal of Econometrics*, **62**, 277-300.
Sethuraman, J. (1994), A Constructive Definition of the Dirichlet Prior, *Statistica Sinica*, **2**, 639-650.
Sursarla, V. and J. Van Ryzin (1976), Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations, *Journal of the American Statistical Association*, **71**, 897-902.
Susarla, V. and J. Van Ryzin (1978a), Empirical Bayes Estimation of a Distribution (Survival) Function from Right-Censored Observations, *Annals of Statistics*, **6**, 740-754.
Susarla V., J. Van Ryzin (1978b), Large sample theory for a

Bayesian nonparametric survival curve estimator based on censored samples, *Annals of Statistics* 6,755-768.

Tsai, W.Y. (1986), Estimation of Survival Curves from Dependent Censorship Models via a Generalized Self-Consistent Property with Nonparametric Bayesian Estimation Application, *Annals of Statistics*, **14**, 238-249.

West, M. (1990), Bayesian Kernel Density Estimation, Tech. report, Department of Statistics, Ohio State.

West, M. (1992), Hyperparameter Estimation in Dirichlet Process Mixture Models, Tech. report, ISDS, Duke University.

Yamato H. (1984), Properties of samples from distributions chosen from a Dirichlet process, *Bulletin of Informatics and Cybernetics* 21, 77-83.