Simulation of Posterior Distributions in Nonparametric Censored Analysis

Jean-Pierre FLORENS *
Jean-Marie ROLIN **

July 2001

Abstract

We analyse the model in which the latent durations T_i are i.i.d. generated by a distribution F. The statistician observes $Y_i = \min(T_i, C_i)$ and $A_i = 1\!\!I_{\{T_i \leq C_i\}}$ where C_i is a censoring time. The prior probability on F is a Dirichlet process. Hjort (1990) shows that the posterior distribution is a neutral to the right process whose hasard function is a Beta process. Lo (1993) has the same type of results with different assumptions on censoring times. For a large class of specifications on censoring times, we exhibit a representation of the posterior process which has the following form : $F = \sum_j F_j F^j$ where j indexes the intervals between censoring times, the F_j 's are product of independent Beta distributed random variables and the F^j 's are independent Dirichlet processes. Using powerful representations of Dirichlet processes (Rolin (1992) and (1993), Sethuraman (1994), Florens and Rolin (1994)) we deduce from this property a very efficient way to simulate various functionals of F.

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

Key words: Bayesian nonparametric, Dirichlet process, Beta process, right censoring, simulation of trajectories.

---

* GREMAQ and IDEI, Université des Sciences Sociales de Toulouse.
** Institut de statistique, Université catholique de Louvain.
Research support from the contract "Projet d'Actions de Recherche Concertées" (PARC No. 98/03-217) of the Belgian Government is gratefully acknowledged.

1 Introduction

The reference model in Bayesian nonparametric statistics has been introduced by Ferguson (1973). In this model, the observations T_i , $1 \leq i \leq n$, are independent and identically distributed and the commun distribution F is a priori a Dirichlet measure of parameter $n_0 F_0$ to be written as $F \sim \mathcal{D}i(\, n_0 F_0 \,)$ where n_0 is a non negative number and F_0 a given probability measure. This prior specification is "natural conjugate" in the sense that the posterior distribution of F is still a Dirichlet measure; more precisely : $F \mid T_1^n \sim \mathcal{D}i(\, n_* F_{n_*} \,)$ where $n_* = n_0 + n$, $n_* F_{n_*} = n_0 F_0 + n F_n$ and F_n is the empirical measure of the sample.

In case of survival data, a Dirichlet measure $\mathcal{D}i(n_0 F_0)$ verifies several simple and interesting properties. For example, the cumulative hazard function $H(\,[0,t]\,) = -\ln F(\,(t,\infty]\,)$ is a Lévy process having easy finite dimensional distributions (see, e.g., Florens, Mouchart and Rolin (1999)). However, other characteristics of a Dirichlet measure are difficult to derive analytically. For example, real functionals like $F(f) = \int f\, dF$ have, in general, a complex distribution known only in very particular cases (see, e.g., Hannum, Hollander, Landberg (1981), Yamato (1984), Cifarelli, Regazzini (1990), Diaconis, Kemperman (1995) and Florens, Rolin (1994)). Asymptotic approximations have been exhibited (see Lo (1983), (1986) and (1987)) but the Bayesian viewpoint is more focused upon small sample properties and simulation methods are then very useful. In particular, very efficient representations of the trajectories of a Dirichlet measure (a purely discrete probability measure) providing the distributions of the locations and the sizes of the jumps give the foundation of a simulation strategy (Bayesian bootstrap), (see Rolin (1992) (1993), Florens and Rolin (1994) and Sethuraman (1994)).

This representation is more complex in the case of censored observations. Let us assume that T_i is the duration of interest and that C_i is a censoring time. The statistician only observes $Y_i = \min(\,T_i\,,\,C_i\,)$ and $A_i = 1\!\!I_{\{T_i \leq C_i\}}$. An important literature has been developed on this model which appears to be very relevant for survival analysis. The process generating the censoring times may be specified in several ways. The censoring times may be fixed or may be assumed to be generated by a stochastic mechanism, depending or not on the generation process of the durations. Our assumptions on the censoring times will be made precise in the next section.

In case of censored observations, the natural conjugate property of the Dirichlet prior on the probability distribution of the duration disappears. Susarla and Van Ryzin (1976) have computed the posterior expectations of the integer powers of the survival function showing that the posterior distribution is not a Dirichlet measure and they conjectured that the posterior distribution was a mixture of Dirichlet measures (Antoniak (1974)). This conjecture was proved later on by Blum and Susarla (1977). The class of prior distributions has been extended to the wider class of neutral to the right processes which includes the Dirichlet measures (Docksum (1974), Ferguson and Phadia (1979)). This class of distributions is also "natural conjugate". This result implies that the Susarla-Van Ryzin's posterior distributions are neutral to the right. The family of neutral to the right processes is not the smallest family of processes verifying this closure property as shown by Hjort (1990) who introduced the family of Beta processes (see also Lo (1993) who defined Beta neutral processes). The Dirichlet prior measures arise from particular Beta processes and a consequence of Hjort's analysis is that the Susarla-Van Ryzin's posterior distributions are particular Beta processes.Hjort (1990) a-na-lyzes a model conditional on censoring times but Lo (1993) obtains identical results in case of a joint specification of T_i and C_i .

The Beta processes are usually defined through some characteristics of the distribution of the "predictable" form of the hazard function defined as $H^-([0,t]) = \int_{[0,t]} F([s,\infty])^{-1}F(ds)$ and up to the best of our knowledge no representation of their trajectories is available. The Beta processes still induce purely discrete probability measures but the distributions of the locations and the sizes of the jumps have not a tractable expression and we cannot extend the simulation of the functionals of F realized in the case of Dirichlet processes.

Note however that a simulation of the trajectories using Markov Chain Monte Carlo methods for Dirichlet priors has been proposed by Doss (1994). Another different simulation strategy has been proposed by Damien, Laud, Smith (1995) and (1996) for general neutral to the right processes priors. But these simulation techniques must rely on more complicated schemes. The latter strategy is more general but only simulates the probabilities of intervals and the method hardly extends 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.

The objective of this paper is to propose a representation theorem of a sub-class of Beta processes from which efficient simulations of functionals may be derived. This sub-class of Beta processes is the sub-class of posterior distributions obtained from a Dirichlet prior measure after the observation of right censored observations and is then the class of distributions derived by Susarla and Van Ryzin.

Our paper is organized as follows: the model and some general properties are considered in Section 2 and the main representation theorem and its implications are developed in Section 3. Practical considerations on the implementation of the simulation and a revisit of the Kaplan-Meyer data set used by Susarla-Van Ryzin are exposed in Sections 4 and 5. The main proofs are given in an Appendix.

2 Model Specification

Let (T_i, C_i), $1 \leq i \leq n$, be a sequence of positive latent variables where T_i is the duration of interest and C_i is a censoring time. The sampling distribution of this sequence is indexed by a parameter (F, G) where F is a probability measure on $(\overline{\mathbb{R}}^+, \overline{\mathcal{B}}^+)$ and G is a probability measure on $(\mathbb{R}^{+n}, \mathcal{B}^{+n})$. We assume

- H1 : The sequence $T_1^n = \{ T_i : 1 \leq i \leq n \}$ and $C_1^n = \{ C_i : 1 \leq i \leq n \}$ are sampling independent, i.e.,

$$T_1^n \perp\!\!\!\perp C_1^n \mid F, G. \tag{2.1}$$

- H2 : The T_i, $1 \leq i \leq n$, are independently and identically distributed and the distribution of T_i is F, i.e.,

$$\mathop{\perp\!\!\!\perp}_{1 \leq i \leq n} T_i \mid F, G \qquad T_i \perp\!\!\!\perp G \mid F \qquad T_i \mid F \sim F. \tag{2.2}$$

- H3 : The joint distribution of C_1^n is G, i.e.,

$$C_1^n \perp\!\!\!\perp F \mid G \qquad C_1^n \mid G \sim G. \tag{2.3}$$

The Bayesian model requires a prior specification on (F, G). We first only assume the following property on such a prior:

- H4 : F and G are a priori independent, i.e.,

$$F \perp\!\!\!\perp G. \tag{2.4}$$

The variables T_i, $1 \leq i \leq n$, are not all available but the statistician only observes (Y_i, A_i, C_i), $1 \leq i \leq n$, where

$$Y_i = \min(T_i, C_i) \qquad A_i = \mathbb{1}_{\{T_i \leq C_i\}}. \tag{2.5}$$

The independence assumption H1 is motivated by an identification argument. Indeed, it is known that any probability measure on (Y_i, A_i) may be considered as deduced from the product of two independent probabilities on T_i and C_i. Equivalently, in the absence of specific restrictions, any distribution on $(T_i\,,\,C_i)$ is observationally equivalent to a product of probability measures (see, e.g., Mouchart and Rolin (1995)).

Using elementary manipulations of conditional independence (see, e.g., Florens, Mouchart and Rolin (1990)), we deduce from assumptions H1 to H3 the following properties :

i) The observed random variables are independent given C_1^n, i.e.,

$$\perp\!\!\!\perp_{1\leq i\leq n}(Y_i\,,\,A_i)\mid F\,,\,G\,,\,C_1^n\,. \tag{2.6}$$

ii) The distribution of $(Y_i\,,\,A_i)$ conditionally on $(F\,,\,G\,,\,C_1^n)$ only depends on F and C_i, i.e.,

$$(Y_i\,,\,A_i)\perp\!\!\!\perp(G\,,\,C_1^n)\mid F\,,\,C_i\,, \tag{2.7}$$

and that implies

$$(Y_1^n\,,\,A_1^n)\perp\!\!\!\perp G\mid F\,,\,C_1^n\,. \tag{2.8}$$

These two properties are summarized in saying that $(Y_1^n\,,\,A_1^n)$ is C_1^n -conditionally independent.

Under assumptions H1 to H4 the observed model may be separated through a Bayesian cut defined by the three independence conditions (see Florens, Mouchart and Rolin (1990))

$$F\perp\!\!\!\perp G \qquad C_1^n\perp\!\!\!\perp F\mid G \qquad (Y_1^n\,,\,A_1^n)\perp\!\!\!\perp G\mid F\,,\,C_1^n\,. \tag{2.9}$$

Therefore, without loosing information, the inference may be totally separated into the inference on G through the marginal model generating C_1^n and the inference on F through the conditional model generating $(Y_1^n\,,\,A_1^n)$ given C_1^n. Equivalently C_1^n may be considered as "known" or "fixed" for the estimation of F. Moreover F and G are a posteriori independent, i.e.,

$$F\perp\!\!\!\perp G\mid Y_1^n\,,\,A_1^n\,,\,C_1^n\,. \tag{2.10}$$

We have in fact a stronger result : the inference on F does not require the knowledge of the censoring times if $(Y_i\,,\,A_i)$, $1\leq i\leq n$, are observed. This means that the knowledge of "unactive" censoring times (C_i greater than T_i) is unnecessary.

Lemma 2.1 *Under assumptions H1 to H4 the following conditional independence relation is verified:*

$$F \perp\!\!\!\perp C_1^n \mid Y_1^n\,,\, A_1^n\,.$$

■

Proof: see Appendix A.1

■

This, along with (2.10), is equivalent to

$$F \perp\!\!\!\perp (\, C_1^n\,,\, G\,) \mid Y_1^n\,,\, A_1^n\,. \tag{2.11}$$

In conclusion, under H.1 to H.4 , whatever is the distribution of C_1^n , i.e., G , the posterior distribution of F , i.e., the distribution of F conditionally on $(Y_1^n\,,\, A_1^n)$, will be the same as the posterior distribution of F obtained in the model conditional on C_1^n , i.e., in the model considering that the censoring times are fixed and known. This result extends slightly Lemma 7.1 of Lo (1993) that shows that, under the supplementary assumption of independence of the censoring times, $F \perp\!\!\!\perp G \mid Y_1^n\,,\, A_1^n$. Such results show that comparisons between marginal and conditional models are irrelevant.

3 Posterior Distribution under a Dirichlet Prior Specification

Let us now specify the prior distribution on F . We assume

- H5 : F is a Dirichlet measure with parameters $n_0 \in \mathbb{R}^+$ and F_0 , a probability measure on $(\,\overline{\mathbb{R}}^+\,,\, \overline{\mathcal{B}}^+\,)$, i.e.,

$$F \,\sim\, \mathcal{D}i(n_0\, F_0)\,. \tag{3.1}$$

Let us recall (see Ferguson (1973)) that the Dirichlet measure is entirely characterized by the finite dimensional distributions of $\{\, F(B_\ell) : \, 1 \le \ell \le k \,\}$ where $\{\, B_\ell \,:\, 1 \le \ell \le k \,\}$ is a (non trivial) measurable partition of $\overline{\mathbb{R}}^+$. Namely, the distribution of this vector on the $k-1$ dimensional simplex of $[0,1]^k$ is a Dirichlet distribution of parameter $\{n_0\, F_0(B_\ell) \,:\, 1 \le \ell \le k\}$, i.e.,

$$\{\, F(B_\ell) \,:\, 1 \le \ell \le k \,\} \,\sim\, Di[\,\{\, n_0\, F_0(B_\ell) \,:\, 1 \le \ell \le k \,\}\,]\,. \tag{3.2}$$

This distribution may be characterized as follows:

$$E\left[\prod_{1\le\ell\le k} F(B_\ell)^{a_\ell}\right] = \frac{\Gamma(n_0)}{\Gamma(n_0+a)} \prod_{1\le\ell\le k} \frac{\Gamma[n_0F_0(B_\ell)+a_\ell]}{\Gamma[n_0F_0(B_\ell)]} 1\!\!1_{\{F_0(B_\ell)>0\}} \tag{3.3}$$

$$\forall \quad a_\ell > -n_0F_0(B_\ell)\ , \quad 1\le\ell\le k\ , \quad \text{and} \quad a = \sum_{1\le\ell\le k} a_\ell\ .$$

This prior specification entails three remarks:

(i) The prior distribution is defined on the structural parameter (or on the parameter of interest) F and not on the reduced parameter (or sufficient parameter), i.e. the probability measure generating the actual data $(Y_1^n\,,\,A_1^n)$ conditionally on C_1^n. Recall that we assume $F \perp\!\!\!\perp C_1^n$.

(ii) We only consider as a prior a Dirichlet measure and not a member of a larger class of natural conjugate priors such as neutral to the right processes or Beta-neutral processes.

(iii) Thanks to the result of section 2, the prior distribution on G, the nuisance parameter, may be left unspecified.

The description of the posterior distribution requires some definitions and notations.

Let $\{a_j : 1\le j\le m\}$ be the distinct censoring times in increasing order, i.e., $\{C_i : 1\le i\le n\} = \{a_j : 1\le j\le m\}$ and $0\le a_1 < a_2 < \ldots < a_m < \infty$. They generate the following (measurable) partition of $\overline{I\!\!R}^+$:

$$\begin{aligned} &B_1 = [0, a_1] \\ &B_j = (a_{j-1}, a_j] \qquad 2\le j\le m \\ &B_{m+1} = (a_m, \infty]\ . \end{aligned} \tag{3.4}$$

Under this partition, F (and F_0) may be decomposed into marginal and conditional probability measures as follows:

$$F_j = F(B_j) \qquad 1\le j\le m+1 \tag{3.5}$$

and

$$F^j(\,\cdot\,) = F(\,\cdot \mid B_j\,) = \frac{F(\,\cdot\cap B_j\,)}{F(\,B_j\,)} \qquad 1\le j\le m+1\ . \tag{3.6}$$

Clearly

$$F = \sum_{1\le j\le m+1} F_j\, F^j\ . \tag{3.7}$$

We use the same notations in adding the subscript 0 for the marginal and conditional probabilities defined from F_0 .

The next proposition is derived from a basic general property of Dirichlet measures (i.e., not related to the structure of $(\overline{I\!R}^+ , \overline{\mathcal{B}}^+)$ or to the particular partition we consider (see, e.g., Rolin (1992) or Florens, Rolin (1994)) .

Proposition 3.1 If $F \sim \mathcal{D}i(n_0 F_0)$, then

(i) $\underset{1\leq j\leq m+1}{\perp\!\!\!\perp} F^j \perp\!\!\!\perp \{F_j : 1 \leq j \leq m+1\}$,

(ii) $F^j \sim \mathcal{D}i(n_0 F_{0j} F_0^j) \qquad 1 \leq j \leq m+1$,

(iii) $\{F_j : 1 \leq j \leq m+1\} \sim Di[\{n_0 F_{0j} : 1 \leq j \leq m+1\}]$.

■

Now, the marginal probabilities $\{F_j : 1 \leq j \leq m+1\}$ may be reparametrized in terms of survival probabilities and in terms of hazard probabilities that will be particularly useful in this context.

Let us define the survival probabilities by

$$\begin{aligned} S_0 &= 1 \\ S_j &= F(\,(a_j, \infty]\,) = \sum_{j+1\leq \ell\leq m+1} F_j \qquad 1 \leq j \leq m \qquad (3.8) \\ S_{m+1} &= 0 \end{aligned}$$

so that

$$F_j = S_{j-1} - S_j \quad 1 \leq j \leq m+1 \ , \qquad (3.9)$$

and the hazard probabilities by

$$\begin{aligned} H_1 &= F_1 = 1 - S_1 \\ H_j &= \frac{F(\,(a_{j-1}, a_j]\,)}{F(\,(a_{j-1}, \infty]\,)} = \frac{F_j}{S_{j-1}} = 1 - \frac{S_j}{S_{j-1}} \qquad 2 \leq j \leq m \qquad (3.10) \\ H_{m+1} &= 1 \ . \end{aligned}$$

This gives, in terms of hazard probabilities, the following product representations of the marginal probabilities :

$$F_j = H_j \prod_{1 \le \ell < j} (1 - H_\ell) \qquad 1 \le j \le m+1 \tag{3.11}$$

and of the survival probabilities:

$$S_j = \prod_{1 \le \ell \le j} (1 - H_\ell) \qquad 1 \le j \le m+1 \; . \tag{3.12}$$

We use the same notations in adding the subscript 0 for the survival and hazard probabilities defined from F_0.

Now by a known property of the Dirichlet distribution (see, e.g., Rolin (1983)), we have the following proposition :

Proposition 3.2 If $F \sim \mathcal{D}i(n_0\, F_0)$, then

(i) $\perp\!\!\!\perp_{1 \le j \le m} H_j$

(ii) $H_j \sim Be[\, n_0 F_{0\,j}, n_0 S_{0\,j}\,] \qquad 1 \le j \le m\,,$

i.e., H_j has a beta distribution of parameters $n_0 F_{0\,j}$ and $n_0 S_{0\,j}$.

■

Note that

$$E(H_j) = H_{0\,j} \qquad 1 \le j \le m \; . \tag{3.13}$$

Now, the main result of the paper states that the independence relations (i) in Proposition 3.1 and in Proposition 3.2 still hold a posteriori and provide the posterior distributions of F^j, $1 \le j \le m+1$, and of H_j , $1 \le j \le m$. We first introduce some notations. Let

$$F_n = \frac{1}{n} \sum_{1 \le i \le n} \varepsilon_{Y_i} \tag{3.14}$$

be the empirical probability measure of the observations and

$$F_{un} = \frac{1}{n} \sum_{1 \le i \le n} A_i \, \varepsilon_{Y_i} \tag{3.15}$$

be the empirical subprobability measure of the observed deaths.

Just as before, we define

$$F^j_{un}(\,\cdot\,) = F_{un}(\,\cdot \mid B_j\,) \qquad 1 \le j \le m+1\,, \tag{3.16}$$

$$F_{un\,j} = F_{un}(B_j) = \frac{1}{n} D_j \qquad 1 \le j \le m+1\,, \tag{3.17}$$

$$S_{n\,j} = F_n(\,(a_j, \infty]\,) = \frac{1}{n} N_j \quad 1 \le j \le m \tag{3.18}$$

where D_j is the number of observed deaths in B_j and N_j is the number of individuals at risk just after time a_j. Let L_j be the number of censoring times at a_j, i.e.,

$$L_j = \sum_{1 \le i \le n} (1 - A_i)\, \mathbb{I}_{\{\,Y_i = a_j\,\}} \qquad 1 \le j \le m\,, \tag{3.19}$$

then we have

$$N_{j-1} = N_j + D_j + L_j \qquad 1 \le j \le m\,. \tag{3.20}$$

We can now state the main result of this paper.

Theorem 3.3 Under the assumptions H1 to H5 ,

(i) $\displaystyle\perp\!\!\!\perp_{1 \le j \le m+1} F^j \perp\!\!\!\perp \{F_j : 1 \le j \le m+1\} \mid Y_1^n\,,\, A_1^n\,,$

(ii) $F^j \mid Y_1^n\,,\, A_1^n \sim \mathcal{D}i(\,n_0\, F_{0\,j}\, F^j_0 + n F_{un\,j}\, F^j_{un}\,) \qquad 1 \le j \le m+1\,,$

(iii) $\displaystyle F_j \;=\; H_j \prod_{1 \le \ell < j} (1 - H_\ell) \qquad 1 \le j \le m+1\,,$

(iv) $\displaystyle\perp\!\!\!\perp_{1 \le j \le m} H_j \mid Y_1^n\,,\, A_1^n\,,$

(v) $H_j \mid Y_1^n\,,\, A_1^n \sim Be[\,n_0\, F_{0\,j} + D_j\,,\, n_0\, S_{0\,j} + N_j + L_j\,] \qquad 1 \le j \le m\,.$

■

The proof is given in Appendix A.2.

Let us remark that, in view of the posterior distribution of F^j given in (ii), by Propositions 3.1 and 3.2, Theorem 3.3 is also true for any finer partition that the one we used.

Note that (v) implies

$$E[\,H_j \mid Y_1^n\,,\, A_1^n\,] = \frac{n_0 F_{0\,j} + D_j}{n_0 S_{0\;j-1} + N_{j-1}} \qquad 1 \le j \le m\,. \tag{3.21}$$

The main object of inference is the survival function S_t , i.e.,

$$S_t \;=\; F(\,(t,\infty]\,)\,. \tag{3.22}$$

Now for $a_{j-1} \le t < a_j$, we note that

$$S_t \;=\; F_j\, F^j(\,(t,a_j]\,) + S_j \tag{3.23}$$

or in terms of hazard probabilities,

$$S_t = \prod_{1\le \ell<j} (1-H_\ell)\,\cdot\,\{H_j\, F^j(\,(t,a_j]\,) + 1 - H_j\}\,. \tag{3.24}$$

By a known property of the Beta distribution (namely : $X \perp\!\!\!\perp Y$, $X \sim Be(a,b)$, $Y \sim Be(a+b,d)$ imply $XY \sim Be(a,b+d)$), we obtain that

$$\begin{aligned} & H_j\, F^j(\,(t,a_j]\,) + 1 - H_j \mid Y_1^n\,,\, A_1^n \\ & \sim Be[\,n_0F_0(\,(t,\infty]\,) + nF_n(\,(t,\infty]\,)\,,\, n_0F_0(\,(a_{j-1},t]\,) + nF_{un}(\,(a_{j-1},t]\,)\,]\,. \end{aligned} \tag{3.25}$$

Therefore, as noticed by Susarla-Van Ryzin, S_t is expressed as a product of independent beta-distributed random variables and may then be simulated without approximations (see Devroye (1986) section 9.4). By taking expectation, we recover the Susarla-Van Ryzin estimator. Indeed, using (3.21) and (3.25), we obtain

$$\begin{aligned} & E[\,S_t \mid Y_1^n\,,\, A_1^n\,] \\ & \quad = \frac{n_0S_{0\,t} + nS_{n\,t}}{n_0S_{0\,j-1} + N_{j-1}} \cdot \prod_{1\le\ell<j} \frac{n_0S_{0\,\ell} + N_\ell + L_\ell}{n_0S_{0\,\ell-1} + N_{\ell-1}} \\ & \quad = S_{n_*\,t} \cdot \prod_{1\le\ell<j} \{1 + \frac{L_\ell}{n_0S_{0\,\ell} + N_\ell}\} \end{aligned} \tag{3.26}$$

where

$$S_{n_*\,t} = \frac{n_0S_{0\,t} + nS_{n\,t}}{n_0+n} \;=\; \frac{n_0F_0(\,(t,\infty]\,) + nF_n(\,(t,\infty]\,)}{n_0+n}\,. \tag{3.27}$$

This is indeed the correct expression of the Susarla-Van Ryzin estimator.

More generally, if f is a positive Borel function defined on $(\,\overline{I\!R}^+\,,\, \overline{\mathcal{B}}^+\,)$, we obtain as a corollary of Theorem 3.3 ,

$$E[\,F(f) \mid Y_1^n\,,\, A_1^n\,] \;=\; \sum_{1\le j\le m+1} E[\,F_j \mid Y_1^n\,,\, A_1^n\,]\; E[\,F^j(f) \mid Y_1^n\,,\, A_1^n\,]\,, \tag{3.28}$$

$$E[\, F_j \mid Y_1^n\,,\, A_1^n\,] \;=\; \frac{n_0 F_{0\,j} + D_j}{n_0 S_{0\,j-1} + N_{j-1}} \;\cdot \prod_{1 \le \ell < j} \{1 - \frac{n_0 F_{0\,\ell} + D_\ell}{n_0 S_{0\,\ell-1} + N_{\ell-1}}\} \qquad (3.29)$$

and

$$E[\, F^j(f) \mid Y_1^n\,,\, A_1^n\,] \;=\; \frac{n_0 F_{0\,j}\, F_0^j(f) + n F_{un\,j}\, F_{un}^j(f)}{n_0 F_{0\,j} + D_j} \qquad (3.30)$$

where

$$F_{0\,j}\, F_0^j(f) \;=\; \int_{B_j} f\, dF_0 \qquad (3.31)$$

and

$$n F_{un\,j} F_{un}^j(f) \;=\; \sum_{1 \le i \le n} A_i\, f(Y_i)\, \mathbb{1}_{B_j}(\,Y_i\,)\,. \qquad (3.32)$$

Finally, in Section 2, we proved that the posterior distribution depends on the active censoring times only while in this section we consider all the values of the censoring times. Note that a censoring time a_j is not active if and only if $L_j = 0$. In view of (3.26), it is clear that such a censoring time has no effect on the posterior expectation of the survival distribution and , from Lemma 2.1, the same is true for the posterior distribution of F.

Proposition 3.4 Under assumptions H1 to H5 , Theorem 3.3 is still valid if we take into account only the a_j 's for which $L_j > 0$.

■

Thanks to the comment following Theorem 3.3, we may consider the finer partition generated by the observed durations. If Z_j , $1 \le j \le M$, are the distinct observed durations in increasing order, i.e., $\{Y_i \,:\, 1 \le i \le n\} = \{Z_j \,:\, 1 \le j \le M\}$ and $0 \le Z_1 < Z_2 < \ldots < Z_M < \infty$, we consider the following partition of $\overline{\mathbb{R}}^+$:

$$\begin{aligned} &B_1 = [0, Z_1]\,, \\ &B_j = (Z_{j-1}, Z_j] \qquad 2 \le j \le M\,, \\ &B_{M+1} = (Z_M, \infty]\,. \end{aligned} \qquad (3.33)$$

We use the same notations as before and note that, for this partition,

$$\begin{aligned} D_j &= \sum_{1 \le i \le n} A_i\, \mathbb{1}_{\{\,Y_i = Z_j\,\}}\,, \\ L_j &= \sum_{1 \le i \le n} (1 - A_i)\, \mathbb{1}_{\{\,Y_i = Z_j\,\}}\,, \\ N_j &= \sum_{1 \le i \le n} \mathbb{1}_{\{\,Y_i > Z_j\,\}}\,, \end{aligned} \qquad (3.34)$$

i.e., D_j is the number of deaths at Z_j, L_j the number of losses at Z_j and N_j the number of individuals at risk just after Z_j. Then Theorem 3.3 holds for this finer partition but (ii) becomes

$$F^j \mid Y_1^n \, , \, A_1^n \; \sim \; \mathcal{D}i(\, n_0 F_{0\,j} F_0^j + D_j \, \varepsilon_{Z_j} \,) \qquad 1 \leq j \leq M+1 \tag{3.35}$$

and formula (3.30) is simplified into

$$E[\, F^j(f) \mid Y_1^n \, , \, A_1^n \,] \; = \; \frac{n_0 F_{0\,j} \, F_0^j(f) + D_j \, f(Z_j)}{n_0 F_{0\,j} + D_j} \qquad 1 \leq j \leq M+1 \, . \tag{3.36}$$

If we analyze the jumps of the distribution function at the observations, i.e.,

$$F(Z_j) \; = \; S_{Z_j -} \; - \; S_{Z_j} \, , \tag{3.37}$$

we obtain, using (3.24),

$$F(Z_j) \; = \; F^j(Z_j) \, H_j \prod_{1 \leq \ell < j} (1 - H_\ell) \tag{3.38}$$

and

$$F^j(Z_j) \mid Y_1^n \, , \, A_1^n \; \sim \; Be[\, n_0 F_0(Z_j) + D_j \, , \, n_0 F_{0\,j} - n_0 F_0(Z_j) \,] \, . \tag{3.39}$$

Therefore

$$E[\, F(Z_j) \mid Y_1^n \, , \, A_1^n \,] \; = \; \frac{n_0 F_0(Z_j) + D_j}{n_0 + n} \prod_{1 \leq \ell < j} \{ \, 1 \, + \, \frac{L_\ell}{n_0 S_{0\,\ell} + N_\ell} \, \} \, . \tag{3.40}$$

This shows in particular that, if F_0 is atomeless, the expectation of the posterior distribution has jumps only at the observed deaths, i.e., at points Z_j for which $D_j \, > \, 0$.

4 Implementation of the simulation

Theorem 3.3 (ii) shows that the F^j's , $1 \leq j \leq m+1$, are the conditional probability measures of a random probability measure F^p, satisfying

$$F^p \mid Y_1^n \, , \, A_1^n \;\; \sim \;\; \mathcal{D}i \, n_0 F_0 + n F_{un} \,) \, . \tag{4.1}$$

According to the representation of trajectories of Dirichlet measures when F_0 is atomeless (see Rolin (1992) and (1993), Florens, Rolin (1994) and Sethuraman (1994)), we may write

$$F^p \; = \; (1 - \pi) F^a + \pi F^s \tag{4.2}$$

where

$$\begin{array}{ll} (i) & \pi \perp\!\!\!\perp F^a \perp\!\!\!\perp F^s \mid Y_1^n \,,\, A_1^n \,, \\ (ii) & \pi \mid Y_1^n \,,\, A_1^n \sim Be[\, n_u \,,\, n_0 \,] \,, \\ (iii) & F^a \mid Y_1^n \,,\, A_1^n \sim \mathcal{D}i(\, n_0 \, F_0 \,) \,, \\ (iv) & F^s \mid Y_1^n \,,\, A_1^n \sim \mathcal{D}i(\, n \, F_{un} \,) \,, \end{array} \tag{4.3}$$

where n_u is the total number of observed deaths, i.e.,

$$n_u = \sum_{1 \leq j \leq M} D_j \,. \tag{4.4}$$

F^a is then the prior Dirichlet measure and F^a has the following representation:

$$F^a = \sum_{1 \leq k < \infty} \alpha_k \, \varepsilon_{\tau_k} \tag{4.5}$$

where

$$\begin{array}{ll} (i) & \alpha_k = \delta_k \prod_{1 \leq \ell < k} (1 - \delta_\ell) \\ (ii) & \delta_1^\infty \perp\!\!\!\perp \tau_1^\infty \\ (iii) & \tau_k \quad , 1 \leq k < \infty, \text{ are i.i.d. } F_0 \\ (iv) & \delta_k \quad , 1 \leq k < \infty, \text{ are i.i.d. } Be(\, 1 \,,\, n_0 \,) \,. \end{array} \tag{4.6}$$

Since F^a has an (infinite) countable number of jumps, in practice, it must be truncated and replaced by

$$\tilde{F}^a = (\sum_{1 \leq k \leq K} \alpha_k)^{-1} \sum_{1 \leq k \leq K} \alpha_k \, \varepsilon_{\tau_k} = \sum_{1 \leq k \leq K} \overline{\alpha}_k \, \varepsilon_{\tau_k} \,. \tag{4.7}$$

The approximation error is given by (see, e.g., Rolin (1993) or Florens, Rolin (1994))

$$\sup_{B \in \overline{\mathcal{B}}^+} |F^a(B) - \tilde{F}^a(B)| = \prod_{1 \leq k \leq K} (1 - \delta_k) = e_K \tag{4.8}$$

Its value may be easily computed for each simulation. Moreover, $-\ln e_K \sim \Gamma(n_0 \,,\, K)$ and so, as K is large, $\ln e_K$ is a.s. of order $-\frac{K}{n_0}$. The choice of K is therefore related to the value of n_0.

F^s may be considered as the posterior Dirichlet measure for a non informative prior, i.e., corresponding to $n_0 = 0$ and is often called the sampling Dirichlet measure. It may be represented almost surely as

$$F^s = \sum_{1 \leq j \leq M} \beta_j \, \varepsilon_{Z_j} \tag{4.9}$$

where

$$\{ \beta_j \,:\, 1 \le j \le M \} \mid Y_1^n \,,\, A_1^n \;\sim\; Di[\{ D_j \,:\, 1 \le j \le M \}] \,. \tag{4.10}$$

Using a known relation between Beta and Gamma distributions (namely: $X \sim \Gamma(1,a)\,,\, Y \sim \Gamma(1,b)\,,\, X \perp\!\!\!\perp Y$ is equivalent to $X+Y \perp\!\!\!\perp X/(X+Y)\,,\, X+Y \sim \Gamma(1,a+b)$ and $X/(X+Y) \sim Be(a,b)$), we obtain an equivalent representation of F^p, namely

$$F^p = \frac{\gamma\, F^a + \sum\limits_{1\le i\le n} \gamma_i\, A_i\, \varepsilon_{Y_i}}{\gamma + \sum\limits_{1\le i\le n} \gamma_i\, A_i} \tag{4.11}$$

where

$$\begin{array}{ll} (i) & \gamma \perp\!\!\!\perp \gamma_1^n \,, \\ (ii) & \gamma \sim \Gamma(1, n_0)\,, \\ (iii) & \gamma_i \quad , 1 \le i \le n, \;\; \text{are i.i.d.} \;\; Exp(1)\,. \end{array} \tag{4.12}$$

We can now describe the simulation of the posterior distribution of $F(f) = \int f\, dF$ where f is a Borel measurable function defined on $(\overline{I\!\!R}^+ \,,\, \overline{\mathcal{B}}^+)$ such that $\int |f|\, dF_0 < \infty$.

1. Generation of

- $\delta_k \quad , 1 \le k \le K\,, \qquad$ i.i.d. $Be(1, n_0)\,,$
- $\tau_k \quad , 1 \le k \le K\,, \qquad$ i.i.d. $F_0\,,$
- $\gamma_i \quad , 1 \le i \le n\,, \qquad$ i.i.d. $Exp(1)\,,$
- $\gamma \sim \Gamma(1, n_0)\,.$

2. Computation of

- $\alpha_k = \delta_k \prod\limits_{1\le \ell < k} (1-\delta_\ell) \qquad 1 \le k \le K\,,$
- $\overline{\alpha}_k = (\sum\limits_{1\le \ell \le K} \alpha_\ell)^{-1}\, \alpha_k \qquad 1 \le k \le K\,.$

3. Computation of the prior terms

- $\tilde{D}_j^a = \gamma \sum\limits_{1\le k\le K} \overline{\alpha}_k\, 1\!\!1_{B_j}(\tau_k) \qquad 1 \le j \le m+1\,,$

- $\tilde{D}^a_j(f) = \gamma \sum_{1 \le k \le K} \overline{\alpha}_k \, f(\tau_k) \, \mathbb{1}_{B_j}(\tau_k) \qquad 1 \le j \le m+1 \,.$

4. Computation of the sampling terms

   - $\tilde{D}^s_j = \sum_{1 \le i \le n} \gamma_i \, A_i \, \mathbb{1}_{B_j}(Y_i) \qquad 1 \le j \le m+1 \,,$
   - $\tilde{L}^s_j = \sum_{1 \le i \le n} \gamma_i \, (1 - A_i) \, \mathbb{1}_{B_j}(Y_i) \qquad 1 \le j \le m+1 \,,$
   - $\tilde{D}^s_j(f) = \sum_{1 \le i \le n} \gamma_i \, A_i \, f(Y_i) \, \mathbb{1}_{B_j}(Y_i) \qquad 1 \le j \le m+1 \,.$

5. Computation of the posterior terms

   - $H_j \; = \; \dfrac{\tilde{D}^a_j + \tilde{D}^s_j}{\sum_{j \le \ell \le m+1} (\tilde{D}^a_\ell + \tilde{D}^s_\ell + \tilde{L}^s_\ell)} \qquad 1 \le j \le m \,,$
   - $F_j \; = \; H_j \prod_{1 \le \ell < j} (1 - H_\ell) \qquad 1 \le j \le m+1 \,,$
   - $F^j(f) \; = \; \dfrac{\tilde{D}^a_j(f) + \tilde{D}^s_j(f)}{\tilde{D}^a_j + \tilde{D}^s_j} \qquad 1 \le j \le m+1 \,,$
   - $F(f) \; = \sum_{1 \le j \le m+1} F_j \, F^j(f) \,.$

It can be verified that in absence of censored data $(A_i = 1 \quad \forall 1 \le i \le n)$, F_j reduces to

$$F_j \; = \; \frac{\tilde{D}^a_j + \tilde{D}^s_j}{\gamma + \sum_{1 \le i \le n} \gamma_i} \qquad 1 \le j \le m+1 \tag{4.13}$$

so that

$$F(f) \; = \; (\gamma + \sum_{1 \le i \le n} \gamma_i)^{-1} \, \{ \, \gamma F^a(f) + (\sum_{1 \le i \le n} \gamma_i) \, F^s(f) \, \} \tag{4.14}$$

and is therefore identical to F^p given in (4.11).

This implementation may be particularized in the case of non informative prior specification $(\, n_0 \; = \; 0 \,)$ with the partition generated by the observations. In such a situation, the posterior distribution is called the "sampling posterior distribution". The hazard probabilities are represented as

$$H_j \;=\; \frac{\sum\limits_{1\le i\le n} \gamma_i\, A_i\; 1\!\!1_{\{Y_i = Z_j\}}}{\sum\limits_{1\le i\le n} \gamma_i\; 1\!\!1_{\{Y_i \ge Z_j\}}} \qquad 1\le j\le M\,. \tag{4.15}$$

From (3.24) and (3.26), we deduce that, for $Z_j \le t < Z_{j+1}$ and $1\le j\le M$,

$$S_t \;=\; S_{Z_j} \;=\; \prod_{1\le \ell\le j} (1-H_\ell) \tag{4.16}$$

and

$$E[\,S_t \mid Y_1^n\,,\,A_1^n\,] \;=\; \prod_{1\le \ell\le j} (1-\frac{D_\ell}{N_{\ell-1}}) \tag{4.17}$$

The second member of (4.17) is the Kaplan-Meyer estimator of the survival function, and (4.15) and (4.16) have been proposed by Lo (1993) as a Bayesian Bootstrap for censored data.

In the sampling theory framework, several authors consider that the Kaplan-Meyer estimator is undefined for $Z_M \le t < \infty$, but according to the generalized maximum likelihood principle, (4.17) must also hold in that case. This implies that if $L_M > 0$, S_t is defective (there is a positive probability not to die) since $H_M \sim Be(D_M, L_M)$.

In the Bayesian framework, some care is needed when we let n_0 tend to zero.
If $D_j > 0$ and $1\le j\le M$, we obtain from (3.35) that $F^j \mid Y_1^n\,,\,A_1^n \sim \mathcal{D}i(\,D_j\,\epsilon_{Z_j}\,)$ but this is equivalent to $F^j = \epsilon_{Z_j}$ a.s. and therefore

$$F^j(f) = f(Z_j) \qquad 1\le j\le M\,. \tag{4.18}$$

If $D_j = 0$ and $1\le j\le M$, from (3.35) we obtain $F^j = \epsilon_{\zeta_j}$ where ζ_j is a F_0^j-distributed random variable (see, e.g., Sethuraman, Tiwari(1982) or Florens, Rolin (1994)). But in that case, since $H_j \mid Y_1^n\,,\,A_1^n \sim Be(\,D_j\,,\,N_j + L_j\,)$, H_j (and therefore F_j) is equal to zero a.s. and we may use (4.18) for all $1\le j\le M$. Similarly, for $j = M+1$, when $n_0 = 0$, $F^{M+1} = \epsilon_{\zeta_{M+1}}$ where ζ_{M+1} is a F_0^{M+1}-distributed random variable. Therefore,

$$F^{M+1}(f) \;=\; f(\zeta_{M+1}) \tag{4.19}$$

and

$$E[\,F^{M+1}(f) \mid Y_1^n\,,\,A_1^n\,] \;=\; \frac{1}{F_0(\,(Z_M,\infty]\,)} \int_{(Z_M,\infty]} f\, dF_0 \tag{4.20}$$

In particular, for $Z_M \leq t < \infty$,

$$S_t \; = \; S_{Z_M} \; 1\!\!I_{\{\zeta_{M+1} > t\}} \tag{4.21}$$

and

$$E[\, S_t \mid Y_1^n \, , \, A_1^n \,] \; = \prod_{1 \leq \ell \leq M} (1 - \frac{D_\ell}{N_{\ell-1}}) \cdot \frac{F_0(\, (t, \infty] \,)}{F_0(\, (Z_M, \infty] \,)} \tag{4.22}$$

As a consequence, if $L_M > 0$ (this implies $S_{Z_M} > 0$ a.s.), the prior specification is still playing a role in the so-called " non informative" situation. However, the Bayesian estimator is no longer defective.

5 Example: Susarla-Van Ryzin revisited

We illustrate the proposed simulation by reworking the example of Kaplan, Meier (1958). The data are presented in Table 1.

Z_j	0.8	1.0	2.7	3.1	5.4	7.0	9.2	12.1
D_j	1	0	0	1	1	0	1	0
L_j	0	1	1	0	0	1	0	1
N_{j-1}	8	7	6	5	4	3	2	1

Table 1: Kaplan-Meier data

This example has been used for Bayesian estimations by Susarla, Van Ryzin (1976) and Ferguson, Phadia (1979).

Three methods of simulation of the densities of the survival probabilities have been proposed by Damien, Laud, Smith (1995) and (1996) (to be referred to as DLS later on). Thanks to an approximate simulation of infinitely divisible distribution, they are able to simulate the hazard function or the "predictable" hazard function of any neutral to the right process.
In the first paper, the method is based on approximate simulations of $H(B_j)$, $1 \leq j \leq M$ (where B_j is given in (3.33)) using the fact that, for a Dirichlet prior specification, the hazard function $H(\, [0,t] \,) = -\ln S_t$ is a Levy process. In the second paper, they perform approximate simulations of the "predictable" hazard function $H^-(\, [0,t] \,) = \int_{[0,t]} F(\, [s, \infty] \,)^{-1} F(ds)$ which is a Beta process when F is a Dirichlet measure. The first method uses Hjort's approximation which considers that on small intervals H^- is approximately Beta-distributed. The second method is based on approximate simulations

of $H^-(B_j), 1 \leq j \leq M$. But in these two last methods, another approximation is necessary to recover S_t from $H^-(\,[0,t]\,)$. Indeed, since H^- is purely discrete,

$$S_t = \prod_{0 \leq s \leq t} (1 - H^-(s)) \tag{5.1}$$

and so they have to approximate $\prod_{Z_{j-1} < s < Z_j} (1 - H^-(s))$ by $1 - H^-(\,(Z_{j-1}, Z_j)\,)$.

Other methods of simulation of the trajectories of the posterior distribution function using Markov Chain Monte Carlo methods have been used by Doss (1994) and Arjas and Gasbara (1994) (see also Doss, Huffer and Lawson (1997)). But these methods require much more extensive computations and the error of approximation is difficult to control.

In the case of Dirichlet prior specification, as mentioned in the comments of Theorem 3.3, no approximate simulation is required since the survival probabilities are product of independent Beta-distributed random variables. However, thanks to this result, we will show that our proposed simulation is much faster and more accurate than the methods described above because it provides complete trajectories of the posterior process. The only approximation is the truncation of the prior simulation.

Note however, that the method proposed by DLS can be used more generally for neutral to the right prior specification.

To compare the results, we used the same prior specification, i.e., $n_0 = 1$ and $S_{0\,t} = e^{-\theta_0 t}$ with $\theta_0 = 0.1$ (the maximum likelihood estimate of θ_0 in the exponential model is 0.0969). We perform the same number of simulations, i.e., $B = 2000$. Computations were made on a Mcintosh Performa 5200 with Matlab 4.2c.1. This clearly shows the great simplicity of our method.

The Kaplan-Meier, prior, sampling and posterior estimations of the survival function appear in Figure 1.

For the posterior survival probabilities at the observations, S_{Z_j} and S_{Z_j-} , $1 \leq j \leq M$, the expectations (St) , the standard deviations $(Ssdt)$ along with the means (Ss) and standard deviations $(Ssds)$ of the simulated samples are reported in Table 2 (t stands for theoretical and s for simulated). It includes (when available) the values obtained by DLS (1995) and (1996). It also gives the intervals with 95 percent posterior probabilities (CI), i.e., estimates of the 2.5 and 97.5 percentiles. Note that estimation of any quantile is easily obtained.

Prior, sampling and posterior densities estimations of the survival probabilities are presented in Figure 2. We use a Beta histogram estimator, i.e.,

$$\tilde{f}(t) = \sum_{1 \leq j \leq k} f_{B\,j} Be(j, k - j + 1)(t) \tag{5.2}$$

where f_{Bj} is the proportion of the B simulated values falling in the interval $(\,(j-1)/k, j/k\,]$ with $k\,=\,100$.

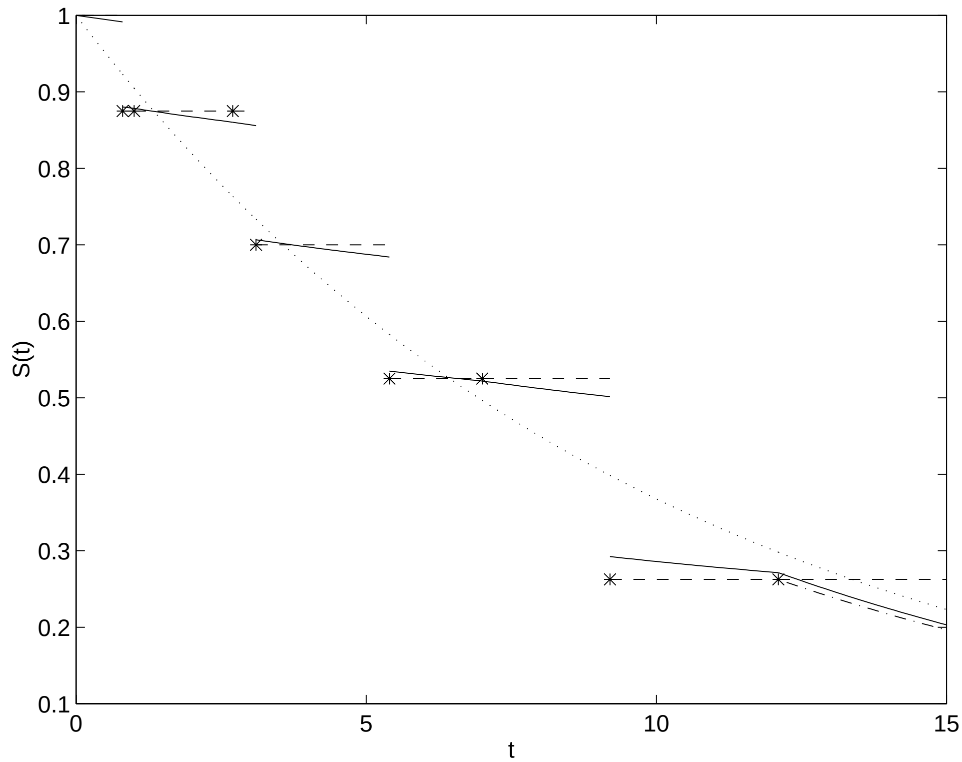

Figure 1 : Kaplan-Meier and Bayesian Estimations.
Dashed line is the Kaplan-Meier estimation, dot the prior, dashdot the sampling and solid the posterior.

	St	Ss	DLS (1995)	DLS (1996)	Ssdt	Ssds	CI	
S_{Z_1-}	0.9915	0.9910	0.990	0.9922	0.0291	0.0306	0.9075	1.0000
S_{Z_1}	0.8803	0.8797			0.1026	0.1052	0.6143	0.9952
S_{Z_2}	0.8783	0.8778	0.921	0.8802	0.1034	0.1055	0.6109	0.9948
S_{Z_3}	0.8603	0.8593	0.898	0.8629	0.1106	0.1127	0.5692	0.9929
S_{Z_4-}	0.8559	0.8552	0.897	0.8587	0.1126	0.1135	0.5683	0.9923
S_{Z_4}	0.7066	0.7097			0.1568	0.1552	0.3526	0.9479
S_{Z_5-}	0.6841	0.6871	0.701	0.6910	0.1608	0.1596	0.3326	0.9408
S_{Z_5}	0.5348	0.5289			0.1758	0.1730	0.2116	0.8523
S_{Z_6}	0.5219	0.5265	0.502	0.5243	0.1762	0.1748	0.1952	0.8487
S_{Z_7-}	0.5014	0.5044	0.416	0.5036	0.1787	0.1775	0.1707	0.8368
S_{Z_7}	0.2924	0.2938			0.1764	0.1759	0.0280	0.6772
S_{Z_8}	0.2714	0.2713	0.275	0.2716	0.1734	0.1732	0.0184	0.6564

Table 2: Posterior Survival Probabilities Estimations

In the special case of sampling estimations ($n_0 = 0$), the densities and the distribution functions of the survival probabilities are analytically computable. The expectations, standard deviations and intervals with 95 percent posterior probabilities along with their simulated values appear in Table 3. The densities and the Beta histogram estimations are presented in Figure 3.

	S	Ssd	CI	
S_{Z_1}	0.8750	0.1102	0.5904	0.9964
	0.8745	0.1142	0.5800	0.9959
S_{Z_4}	0.7000	0.1689	0.3232	0.9551
	0.7035	0.1691	0.3302	0.9580
S_{Z_5}	0.5250	0.1884	0.1621	0.8671
	0.5299	0.1880	0.1766	0.8633
S_{Z_7}	0.2626	0.1865	0.0107	0.6839
	0.2648	0.1859	0.0102	0.6770

Table 3: Sampling Survival Probabilities Estimations
First line: real values
Second line: estimated values

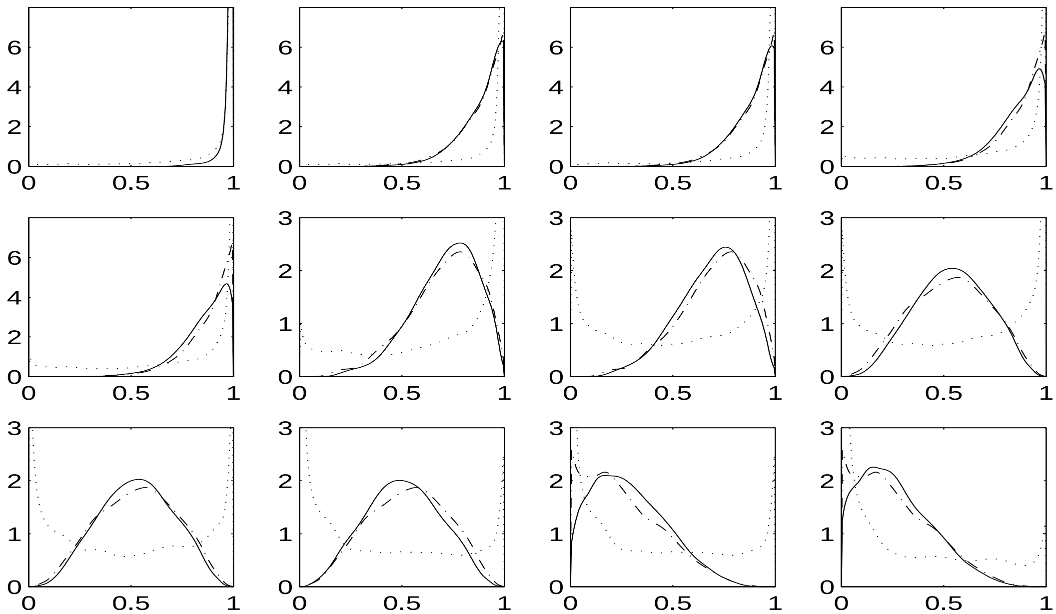

Figure 2: Bayesian Densities of the Survival Probabilities
Dot lines are the prior estimations, dashdot the sampling and solid the posterior

	Top line	S_{Z_1-}	S_{Z_1}	S_{Z_2}	S_{Z_3}
From the left to the right	*Middle line*	S_{Z_4-}	S_{Z_4}	S_{Z_5-}	S_{Z_5}
	Bottom line	S_{Z_6}	S_{Z_7-}	S_{Z_7}	S_{Z_8}

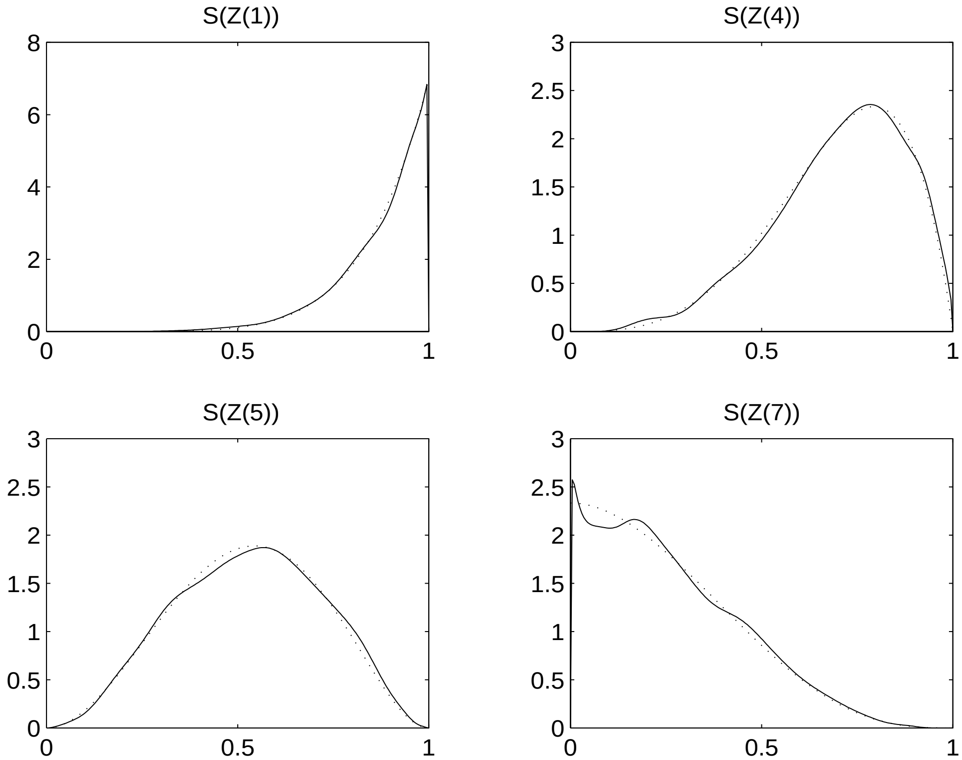

Figure 3 : Sampling Densities of the Survival Probabilities
Dot lines : true densities
Solid lines : simulated densities

A more interesting application of our proposed simulation is the estimation of functionals. Indeed, no analytical form of the distribution is available. Moreover, the simulation methods of DLS do not seem to be applicable in such a case. We choose to give the estimation of the expected lifetime, i.e.,

$$E[\,T \mid F\,] = \int_{[0,\infty]} t\,F(dt)\,. \tag{5.3}$$

The prior, sampling and posterior expectations (Ft) and the standard deviations ($Fsdt$) along with the means (Fs),the standard deviations ($Fsds$) and the intervals with 95 percent posterior probabilities (CI) computed from the simulated sample are reported in Table 4.

	Ft	Fs	Fsdt	Fsds	CI	
Prior	10.0000	9.7371	7.0711	6.7759	1.5951	27.9205
Sampling	9.8038	9.7658	4.4814	4.2378	4.7219	21.3353
Posterior	9.8915	9.9048	4.0708	4.0250	5.0066	20.8981

Table 4: Expected Lifetime Estimations

Densities estimations using a normal kernel and a bandwidth equal to $B^{-1/5}$ multiplied by the standard deviations of the sample ($Fsds$) appear in Figure 4.

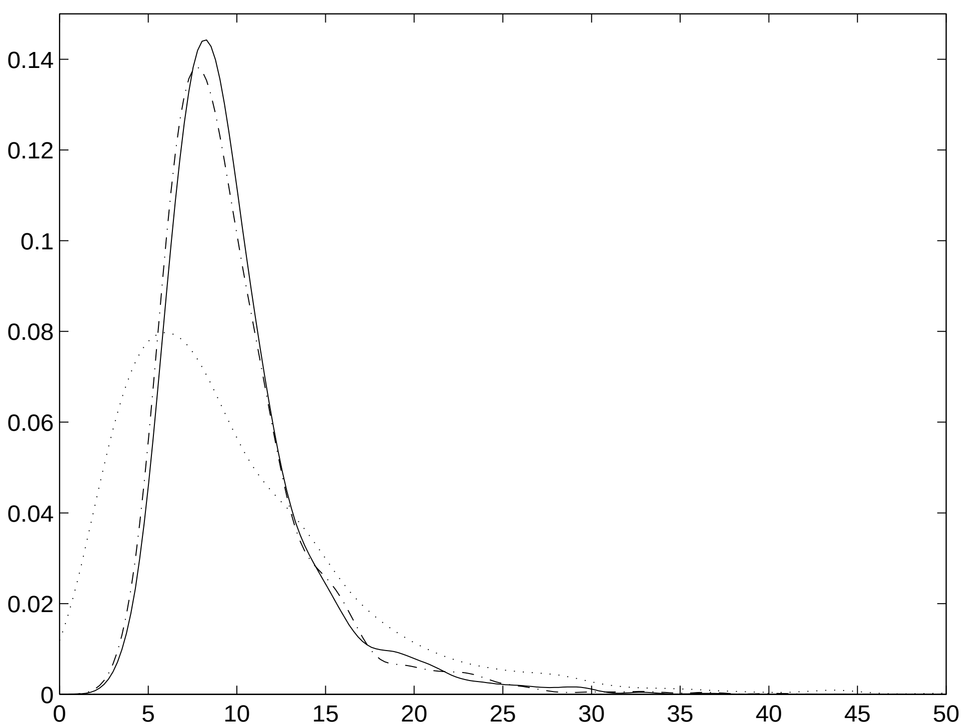

Figure 4 : Bayesian Densities of the Expectation
Dot line is the prior estimation, dashdot the sampling and solid the posterior

Appendix

A.1 Proof of Lemma 2.1

We have to prove that under H1 to H4

$$F \perp\!\!\!\perp C_1^n \mid Y_1^n\,,\, A_1^n\,.$$

First note that under H1, H2, H3 and H4,

$$(F\,,\, T_1^n) \perp\!\!\!\perp (G\,,\, C_1^n)\,.$$

This implies that for any $J \subset \{\,1, 2, \ldots, n\,\}$, if $T_J = \{\,T_i : i \in J\,\}$,

$$F \perp\!\!\!\perp C_1^n \mid T_J\,.$$

Now, let g be a positive Borel measurable function defined on $[0,1]^{\overline{\mathcal{B}}^+}$, i.e., the set of functions defined on $\overline{\mathcal{B}}^+$, the Borel subsets of $\overline{\mathbb{R}}^+$, with values in $[0,1]$, equipped with the product σ-field, and let h_i, $1 \le i \le n$, be positive Borel measurable functions defined on $\mathbb{R}^+ \times \{0\,,\,1\}$. Then

$$E[\, g(F) \prod_{1\le i\le n} h_i(Y_i, A_i) \mid C_1^n\,] = \sum_{J\subset\{1,2,\ldots,n\}} E[\, g(F) \prod_{1\le i\le n} h_i(Y_i, A_i)\, 1\!\!1_{A_J} \mid C_1^n\,]$$

where

$$A_J = \bigcap_{i\in J}\{A_i = 1\} \cap \bigcap_{i\in J^c}\{A_i = 0\} = \bigcap_{i\in J}\{T_i \le C_i\} \cap \bigcap_{i\in J^c}\{T_i > C_i\}\,.$$

Now

$$\begin{aligned} &E[\, g(F) \prod_{1\le i\le n} h_i(Y_i, A_i)\, 1\!\!1_{A_J} \mid C_1^n\,] \\ &= E[\, g(F) \prod_{i\in J} h_i(T_i, 1)\, 1\!\!1_{\{T_i \le C_i\}} \cdot \prod_{i\in J^c} h_i(C_i, 0)\, 1\!\!1_{\{T_i > C_i\}} \mid C_1^n\,]\,. \end{aligned}$$

But, since $\perp\!\!\!\perp_{1\le i\le n} T_i \mid F\,,\, C_1^n$, we have

$$\begin{aligned} &E[\prod_{i\in J} h_i(T_i, 1)\, 1\!\!1_{\{T_i \le C_i\}} \cdot \prod_{i\in J^c} 1\!\!1_{\{T_i > C_i\}} \mid F\,,\, C_1^n\,] \\ &= \prod_{i\in J} E[\, h_i(T_i, 1)\, 1\!\!1_{\{T_i \le C_i\}} \mid F\,,\, C_1^n\,] \cdot \prod_{i\in J^c} F(\,(C_i, \infty]\,) \\ &= E[\prod_{i\in J} h_i(T_i, 1)\, 1\!\!1_{\{T_i \le C_i\}} \cdot \prod_{i\in J^c} F(\,(C_i, \infty]\,) \mid F\,,\, C_1^n\,]\,. \end{aligned}$$

Therefore

$$
\begin{aligned}
&E[\, g(F) \prod_{1\le i\le n} h_i(Y_i, A_i)\, 1\!\!I_{A_J} \mid C_1^n \,] \\
&= E[\, g(F) \prod_{i\in J^c} F(\,(C_i,\infty]\,) \prod_{i\in J} h_i(T_i,1)\, 1\!\!I_{\{T_i \le C_i\}} \mid C_1^n \,] \cdot \prod_{i\in J^c} h_i(C_i, 0)\,.
\end{aligned}
$$

Now , since $F \perp\!\!\!\perp C_1^n \mid T_J$,

$$
E[\, g(F) \prod_{i\in J^c} F(\,(C_i,\infty]\,) \mid T_J \,,\, C_1^n \,] \;=\; E[\, g(F) \prod_{i\in J^c} F(\,(C_i,\infty]\,) \mid T_J \,,\, C_{J^c} \,]
$$

where $C_{J^c} \;=\; \{C_i : i \in J^c\}$, and if we define

$$
k(T_J \,,\, C_{J^c}) \;=\; \frac{E[\, g(F) \prod_{i\in J^c} F(\,(C_i,\infty]\,) \mid T_J \,,\, C_{J^c} \,]}{E[\, \prod_{i\in J^c} F(\,(C_i,\infty]\,) \mid T_J \,,\, C_{J^c} \,]}\,,
$$

clearly on A_J , $\quad k(T_J \,,\, C_{J^c}) \;=\; k(Y_J \,,\, Y_{J^c})$ and

$$
\begin{aligned}
&E[\, g(F) \prod_{i\in J^c} F(\,(C_i,\infty]\,) \mid T_J \,,\, C_{J^c} \,] \\
&=\; E[\, k(T_J \,,\, C_{J^c}) \prod_{i\in J^c} F(\,(C_i,\infty]\,) \mid T_J \,,\, C_{J^c}]\,.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&E[\, g(F) \prod_{1\le i\le n} h_i(Y_i, A_i)\, 1\!\!I_{A_J} \mid C_1^n \,] \\
&=\; E[\, k(T_J \,,\, C_{J^c}) \prod_{i\in J} h_i(T_i,1)\, 1\!\!I_{\{T_i \le C_i\}} \prod_{i\in J^c} h_i(C_i,0) F(\,(C_i,\infty]\,) \mid C_1^n \,] \\
&=\; E[\, k(T_J \,,\, C_{J^c}) \prod_{i\in J} h_i(T_i,1)\, 1\!\!I_{\{T_i \le C_i\}} \prod_{i\in J^c} h_i(C_i,0)\, 1\!\!I_{\{T_i > C_i\}} \mid C_1^n \,] \\
&=\; E[\, k(Y_J \,,\, Y_{J^c}) \prod_{1\le i\le n} h_i(Y_i, A_i)\, 1\!\!I_{A_J} \mid C_1^n \,]
\end{aligned}
$$

since $F(\,(C_i,\infty]\,) \;=\; P(\,T_i \;>\; C_i \mid F \,,\, C_1^n \,,\, T_J\,) \quad \forall \quad i \in J^c$.

Therefore,

$$
E[\, g(F) \mid Y_1^n \,,\, A_1^n \,,\, C_1^n \,] \;=\; k(Y_J \,,\, Y_{J^c}) \quad \text{on} \quad A_J
$$

and this implies

$$
E[\, g(F) \mid Y_1^n \,,\, A_1^n \,,\, C_1^n \,] \;=\; E[\, g(F) \mid Y_1^n \,,\, A_1^n \,]\,.
$$

■

A.2 Proof of Theorem 3.3

As shown in Appendix 1, on $A_J = \bigcap_{i \in J} \{A_i = 1\} \cap \bigcap_{i \in J^c} \{A_i = 0\}$,

$$E[\, g(F) \mid Y_1^n \,, A_1^n \,, C_1^n \,] \;=\; \frac{E[\, g(F) \prod_{i \in J^c} F(\,(C_i, \infty]\,) \mid T_J \,, C_{J^c} \,]}{E[\, \prod_{i \in J^c} F(\,(C_i, \infty]\,) \mid T_J \,, C_{J^c} \,]} \,.$$

But, by assumptions H1 to H4,

$$F \mid T_J \,, C_{J^c} \;=\; F \mid T_J \;\sim\; \mathcal{D}i(n_0 F_0 + n F_{un}) \,,$$

since $\sum_{i \in J} \varepsilon_{T_i} = n F_{un}$.

Therefore, by Proposition 3.1 and 3.2, we have

(i) $\perp\!\!\!\perp_{1 \le j \le m+1} F^j \perp\!\!\!\perp \perp\!\!\!\perp_{1 \le j \le m} H_j \mid T_J \,, C_{J^c} \,,$

(ii) $F^j \mid T_J \,, C_{J^c} \;\sim\; \mathcal{D}i(n_0 F_{0\,j}\, F_0^j + n F_{un\,j}\, F_{un}^j) \,,$

(iii) $H_j \mid T_J, C_{J^c} \;\sim\; Be[\, n_0 F_{0\,j} + D_j \,,\, n_0 S_{0\,j} + N_{u\,j} \,]$

where $N_{u\,j} \;=\; \sum_{1 \le i \le n} A_i \, 1\!\!I_{\{Y_i > a_j\}}$.

Now

$$\prod_{i \in J^c} F(\,(C_i, \infty]\,) \;=\; \prod_{1 \le j \le m} F(\,(a_j, \infty]\,)^{L_j}$$

$$= \prod_{1 \le j \le m} \prod_{1 \le \ell \le j} (1 - H_\ell)^{L_j} \;=\; \prod_{1 \le j \le m} (1 - H_j)^{N_{c\,j} + L_j}$$

where $N_{c\,j} = \sum_{1 \le i \le n} (1 - A_i) \, 1\!\!I_{\{Y_i > a_j\}}$.

This clearly implies that

(i) $\perp\!\!\!\perp_{1 \le j \le m+1} F^j \perp\!\!\!\perp \perp\!\!\!\perp_{1 \le j \le m+1} H_j \mid Y_1^n \,, A_1^n \,, C_1^n \,,$

(ii) $F^j \mid Y_1^n \,, A_1^n \,, C_1^n = F^j \mid T_J \,, C_{J^c} \,,$

(iii) $H_j \mid Y_1^n \,, A_1^n \,, C_1^n \;\sim\; Be[\, n_0 F_{0\,j} + D_j \,,\, n_0 S_{0\,j} + N_{u\,j} + N_{c\,j} + L_j \,]$,
i.e.,

(iii') $H_j \mid Y_1^n \,, A_1^n \,, C_1^n \;\sim\; Be[\, n_0 F_{0\,j} + D_j \,,\, n_0 S_{0\,j} + N_j + L_j \,]$.

■

References

[1] Antoniak C. (1974), Mixtures of Dirichlet process with applications to Bayesian nonparametric problems, *Annals of Statistics* 2, 1152-1174.

[2] Arjas E. and Gasbarra D. (1994), Nonparametric Bayesian inference from right censored survival data, using Gibbs sampler, *Statistica Sinica*, 4, 505-524.

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

[4] Cifarelli D.M., E. Regazzini (1990), Distributions functions of means of a Dirichlet process, *Annals of Statistics* 18, 429-442.

[5] 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.

[6] 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.

[7] Devroye L. (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag, New-York.

[8] Diaconis P., J.B.K. Kemperman (1995), Some new tools for Dirichlet priors, in *Bayesian Statistics V: Fifth Valencia International Meeting on Bayesian Statistics*, Oxford Sciences Publications, 589-597.

[9] Doksum K. (1974), Tailfree and neutral random probabilities and their posterior distributions, *Annals of Probability* 2, 183-201.

[10] Doss H. (1994), Bayesian nonparametric estimation for incomplete data via successive substitution sampling, *Annals of Statistics* 22, 1763-1786.

[11] Doss H., F.W. Huffer and K.L. Lawson (1997), Bayesian nonparametric estimation via Gibbs sampling for coherent systems with redundancy, *Annals of Statistics* 25, 1109-1139.

[12] Ferguson T.S. (1973), A Bayesian analysis of some nonparametric problems, *Annals of Statistics* 1, 209-230.

[13] Ferguson T.S., E. Phadia (1979), Bayesian nonparametric estimation based on censored data, *Annals of Statistics* 7, 163-186.

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

[15] Florens J.P., M. Mouchart, J.M. Rolin (1999), Semi- and non-parametric Bayesian analysis of Duration models, *International Statistical Review* 67 (2), 187-210.

[16] Florens J.P., J.M. Rolin (1994), Bayes, Bootstrap, Moments , Discussion Paper 9413, Institut de Statistique, Université catholique de Louvain, Belgium.

[17] Hannum R.C., M. Hollander, N.A. Langberg (1981), Distributional results for random functionals of a Dirichlet process, *Annals of Probability* 7, 665-670.

[18] Hjort N.L. (1990), Nonparametric Bayes estimators based on beta process in models for life history data, *Annals of Statistics* 18, 1259-1294.

[19] Kaplan E. L. , P. Meier (1958), Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* 53, 457-481.

[20] Lo A.Y. (1983), Weak convergence for Dirichlet processes, *Sankhya*, Ser. A, 45, 105-111.

[21] Lo A.Y. (1986), A remark on the limiting posterior distribution of the multiparameter Dirichlet process, *Sankhya*, Ser. A, 48, 247-249.

[22] Lo A.Y. (1987), A large sample study for the Bayesian bootstrap, *Annals of Statistics* 15, 360-375.

[23] Lo A.Y. (1993), A Bayesian bootstrap for censored data, *Annals of Statistics* 21, 100-123.

[24] 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.

[25] 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.

[26] Rolin J.M. (1992), Some useful properties of the Dirichlet process, Discussion Paper 9202, Institut de Statistique, Université catholique de Louvain, Belgium.

[27] Rolin J.M. (1993), On the distribution of jumps of the Dirichlet process, Discussion Paper 9302, Institut de Statistique, Université catholique de Louvain, Belgium.

[28] Rolin J.M. (1998), Survival Analysis, *Encyclopedia of Biostatistics*, 1, 271-286, Wiley, 1998.

[29] Sethuraman J. (1994), A constructive definition of the Dirichlet prior, *Statistica Sinica* 2, 639-650.

[30] Sethuraman J. and R.C. Tiwari (1982), Convergence of Dirichlet measures and the interpretation of their parameters, *Statistical Decision Theory and Related Topics* III 2, 305-315.

[31] Susarla V., J. Van Ryzin (1976), Nonparametric Bayesian estimation of survival curves from incomplete observations, *Journal of the American Statistical Association* 71, 897-902.

[32] Wilks S.S (1962), *Mathematical Statistics*, Wiley, New York.

[33] Yamato H. (1984), Characteristic functions of means of distributions chosen from a Dirichlet process, *Annals of Probability* 12, 262-267.