IIASA
International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-97-082 / November

Neutral to the Right Processes from a Predictive Perspective: A Review and New Developments

Pietro Muliere (pmuliere@eco.unipv.it)
Stephen Walker (s.walker@ic.ac.uk)

Approved by
Giovanni Dosi (dosi@iiasa.ac.at)
Leader, *TED Project*

About the Authors

Pietro Muliere is professor at the Dipartimento di Economia Politica e Metodi Quantitativi, Università degli Studi di Pavia in Italy.

Stephen Walker is from the Department of Mathematics, Imperial College, London.

Acknowledgement

The work of the first author is partially supported by Progetto strategico CNR (Decisioni statistiche: teoria e applicazioni). An earlier version of the paper was presented at the conference "Bayesian Nonparametrics", Belgirate, Italy, 1997. The paper was completed while the first author was visiting the International Institute for Applied Systems Analysis, Laxenburg, Austria.

Abstract

This paper presents a Bayesian nonparametric approach to survival analysis based on arbitrarly right censored data. The first aim will be to show that the *neutral to the right* process is the natural prior to use in this context. Secondly, the properties of a particular neutral to the right process, the *beta-Stacy* process are examined. Finally, the connections between some Bayesian bootstraps and the beta-Stacy process are investigated.

KEY WORDS: Bayesian bootstrap, Censoring, Exchangeability, Neutral to the right process, Predicition

AMS 1991 Subject classifications: Primary 62A15 - Secondary 62M20.

Contents

1 Introduction **1**

2 Preliminaries **2**

3 A predictive approach **3**

4 The beta-Stacy process **6**

5 Exchangeable neutral urn scheme **10**

6 Bayesian bootstraps **11**

References **13**

Neutral to the Right Processes from a Predictive Perspective: A Review and New Developments

Pietro Muliere
Stephen Walker

1 Introduction

This paper deals with survival analysis from incomplete observations, in particular right censored data. We are interested in the predictive distribution for a future observation given previous observations. We do this in a Bayesian nonparametric framework in which we assign a prior distribution to the space of survival curves.

The aim of this paper is twofold:

i) to show that the neutral to the right process (Doksum, 1974) is the natural nonparametric prior in the presence of right censored data;

ii) to discuss the properties of a particular neutral to the right process, the beta-Stacy process.

Here we discuss aspects of a Bayesian nonparametric analysis of survival time data, where we assume $X_1, X_2, \cdots$ is an infinite sequence of survival times, and we are witness to $X_1, \cdots, X_n$.

The general predictive approach related to a sequence of random variables $\{X_i\}$, defined on a probability space $(\Omega, \mathcal{B}, P)$, involves the evaluation of the probability of an event, dependent on the future realisations of some variables of the sequence, when the outcome of a finite number of variables of the same sequence have been observed. The main predicitive hypothesis will be the exchangeability of the sequence.

A.1 **Exchangeability**. Our fundamental assumption concerning the sequence is the *exchangeablility* of the sequence of random variables $X_1, X_2, \cdots$, with each X_i defined on $\Omega = (0, \infty)$. From de Finetti's representation theorem (de Finetti, 1937) there exists a random distribution function F, conditionally on which $X_1, X_2, \cdots$ are i.i.d. from F. That is, there exists a unique probability (or de Finetti) measure, defined on the space of probability measures on Ω, such that the joint distribution of $X_1, \cdots, X_n$, for any n, can be written as

$$P(X_1 \in A_1, \cdots, X_n \in A_n) = \int \left\{ \prod_{i=1}^{n} F(A_i) \right\} \mu(dF),$$

where μ is the de Finetti (or prior) measure (Hewitt and Savage, 1955).

From a predictive point of view the problem under consideration, reduces to the computation of the conditional probability

$$P(X_{n+1} \in A | X_1, X_2, \cdots, X_n)$$

for some set $A \in \mathcal{B}$. The assumption of exchangeability implies:

$$P(X_{n+1} \in A | X_1, X_2, \cdots, X_n) = E(F(A) | X_1, X_2, \cdots, X_n).$$

Unfortunately, when de Finetti (1935) suggested the general predictive approach, the nonparametric priors were not known. We had to wait for the papers of Freedman (1963), Fabius (1964), and, in particular, the seminal papers of Ferguson (1973,1974) and Doksum (1974).

A.2 **Nonparametric analysis**. If we want to make as few assumptions about the form of the distribution function, then we can adopt a nonparametric approach. In a Bayesian framework we are required to specify a prior distribution on the space of all distribution functions defined on $(0, \infty)$.

The first Bayesian nonparametric approach to determine $E(F(A)|X_1, X_2, \cdots, X_n)$, with censored data, was made by Susarla and Van Ryzin (1976), who used the Dirichlet process as a prior for F. The standard nonparametric estimator of a survival curve from censored data is the product limit estimator, introduced by Kaplan and Meier (1958). The Susarla-Van Ryzin estimator reduces to this Kaplan-Meier estimator as the weight of the prior information tends to zero. Their result was generalized to prior distributions neutral to the right by Ferguson and Phadia (1979).

Many other classes of prior which yield tractable solutions have been used in inferential problems regarding F. We mention: the extended gamma process (Dykstra and Laud, 1981), the beta process (Hjort, 1990) and the Polya trees (Muliere and Walker,1997a).

The paper is organized as follows: after some preliminaries (Section 2) we discuss in Section 3 a functional predictive approach to the selection of the prior. In Section 4 a particular prior, the beta-Stacy process, is discussed. In Section 5, we present an exchangeable neutral urn scheme and finally in section 6, we consider some Bayesian bootstraps which arise in the limit as the weight of prior information goes to zero.

2 Preliminaries

Let $\mathcal{B}(\alpha, \beta)$ for $\alpha, \beta > 0$ represent the beta distribution. For the purposes of the paper it is convenient to define $\mathcal{G}(\alpha_1, \beta_1, \cdots, \alpha_m, \beta_m)$ for $\alpha_j, \beta_j > 0$ to represent the generalised Dirichlet distribution, introduced by Connor and Mosimann (1969). The density function is given, up to a constant of proportionality, by

$$\begin{aligned} & y_1^{\alpha_1 - 1}(1 - y_1)^{\beta_1 - 1} \\ & \times \frac{y_2^{\alpha_2 - 1}(1 - y_1 - y_2)^{\beta_2 - 1}}{(1 - y_1)^{\alpha_2 + \beta_2 - 1}} \\ & \cdots \\ & \times \frac{y_m^{\alpha_m - 1}(1 - y_1 - \cdots - y_{m-1} - y_m)^{\beta_m - 1}}{(1 - y_1 - \cdots - y_{m-1})^{\alpha_m + \beta_m - 1}} I\{(y_1, \cdots, y_m) : y_j \geq 0, \sum_{j=1}^m y_j \leq 1\}, \end{aligned} \quad (1)$$

where I denotes the indicator function. The usual Dirichlet distribution, $\mathcal{D}(\alpha_1, \cdots, \alpha_m, \beta_m)$, with density proportional to

$$y_1^{\alpha_1 - 1} \cdots y_m^{\alpha_m - 1}(1 - y_1 - \cdots - y_m)^{\beta_m - 1} I\{(y_1, \cdots, y_m) : y_j \geq 0, \sum_{j=1}^m y_j \leq 1\}$$

follows if $\beta_{j-1} = \beta_j + \alpha_j$ for all $j = 2, \cdots, m$,

DEFINITION 2.1 $\mathcal{C}(\alpha, \beta, \xi)$ with $\alpha, \beta > 0$ and $0 < \xi \leq 1$ is said to be the beta-Stacy distribution if the density function is given by

$$\frac{1}{B(\alpha,\beta)} y^{\alpha-1} \frac{(\xi - y)^{\beta-1}}{\xi^{\alpha+\beta-1}} I_{(0,\xi)}(y),$$

where $B(\alpha, \beta)$ is the usual beta function.

Note that if $Y \sim \mathcal{C}(\alpha, \beta, \xi)$ then $Y/\xi \sim \mathcal{B}(\alpha, \beta)$ and the usual beta distribution arises if $\xi = 1$. The name beta-Stacy is taken from the paper of Mihram and Hultquist (1967).

The definition of a *neutral to the right process* (NTR) is given in the following:

DEFINITION 2.2 (Doksum, 1974) The random distribution function F is said *neutral to the right* if for each $k > 1$ and $t_1 < t_2 < \cdots < t_k$ there exists nonnegative independent random variables $V_1, \cdots, V_k$ such that

$$(F(t_1), F(t_2), \cdots, F(t_k)) =_{\mathcal{L}} \left(V_1, 1 - (1 - V_1)(1 - V_2), \cdots, 1 - \prod_{i=1}^{k} (1 - V_i) \right).$$

The equations

$$F(t_i) = 1 - \prod_{j \leq i} (1 - V_j), \quad i = 1, \cdots, k$$

yield

$$F(t_i) - F(t_{i-1}) = V_i \prod_{j=1}^{i-1} (1 - V_j)$$

and

$$\frac{F(t_i) - F(t_{i-1})}{1 - F(t_{i-1})} = V_i.$$

F is NTR essentially means that the normalized increments

$$F(t_1), [F(t_2) - F(t_1)]/[1 - F(t_1)], \cdots, [F(t_k) - F(t_{k-1}]/[1 - F(t_{k-1}]$$

are independent for all $t_1 < \cdots < t_k$.

The fundamental result for process neutral to the right is :

THEOREM 2.1 (Ferguson, 1974) *If F is NTR and, given F, $X_1, X_2, \cdots, X_n$ is a sample from F, then the posterior distribution for F is also neutral to the right.*

Ferguson and Phadia (1979) extended this theorem to cover the case of censored data.

3 A predictive approach

Our problem is to make inference about the unknown cumulative distribution function. In the nonparametric framework the function F is itself the parameter and so we need to define a prior on the space of all distribution functions on $(0, \infty)$. It is a task to adequately express our prior knowledge on such a large space. Our approach is to suggest the form for the predictive bearing in mind the type of observation available.

A.3 **Censored data**. A number of individuals are observed from an entry time until a particular event (such as death) occurs. Often, the exact time of death is not known for all individuals; for some, it is only known that the event had not yet happened at some specified time and in this case the observation is *right censored*. See Andersen et al. (1993) for several examples. Formally the model considered is the following: consider n individuals with survival times $X_1, X_2, \cdots, X_n$. Each X_i corresponds either to the time of death or it is only known that the time of death is greater than X_i. We represent the data as

$$(X_1, \delta_1), \cdots, (X_n, \delta_n)$$

where $\delta_i = 1$ if death occurred and $\delta_0 = 1$ if censoring occurred. Whenever we now write X_i, we mean (X_i, δ_i).

First, we will consider the discrete case when each $X_i \in \Omega = \{1, 2, \cdots\}$. To develop the theory we study the consequences of the following assumption concerning the predictive:

$$P(X_{n+1} = k | X_1, \cdots, X_n) = f_k(n_1, \cdots, n_k, m_k), \tag{2}$$

for some suitable f_k, where $n_k = \sum_{1 \le i \le n} I(X_i = k)$ and $m_k = \sum_{1 \le i \le n} I(X_i > k)$. This condition turns out to be an extension of Johnson's sufficientness postulate (Zabell, 1982).

REMARK 3.1 In the 1920's the English philosopher W.E. Johnson discovered a characterisation of the Dirichlet distribution and process (Zabell, 1982). An appropriate extension of Johnson's sufficientness postulate to the case of recurrent Markov exchangeable sequence is introduced by Zabell (1995). In the present note Johnson's result is extended to the case of a neutral to the right exchangeable sequence.

It is possible to show that the assumption of exchangeability, combined with (2), implies a neutral to the right process prior for the sequence.

THEOREM 3.1 (Walker and Muliere, 1977b) *An exchangeable sequence $X_1, X_2, \cdots$ with each X_i defined on $\Omega = \{1, 2, \cdots\}$ has a neutral to the right prior if, and only if, for each $n = 1, 2, \cdots$ and $k \in \Omega$,*

$$P(X_{n+1} = k | X_1, \cdots, X_n) = f_k(n_1, \cdots, n_k, m_k), \tag{3}$$

where $n_k = \sum_{1 \le i \le n} I(X_i = k)$ and $m_k = \sum_{1 \le i \le n} I(X_i > k)$.

REMARK 3.2. The expression (2) has an intuitive justification for censored or truncated data. Nevertheless in practical applications the condition (2) on the predictive may or may not be an adequate description of our state of knowledge. Consequently it is argued that the neutral to the right prior seems inappropriate in that the fundamental assumption concerning the sequence (aside from that of exchangeability), when (2) is hard to justify. Why should an observation $> k$ not matter where it occurs (as far as $P(X_{n+1} = k | X_1, \cdots, X_n)$ is concerned) but which is not the case for observation $< k$.

In order to make the theorem useful in applications, we need to specify the form of the function f_k. In this respect, we present the general construction of the NTR prior, assuming exchangeability and the form of the prediction.

If F is from a NTR process prior on $\Omega = \{1, 2, \cdots\}$ then, by construction, if F_k denotes the random mass assigned to $\{1, 2, \cdots, k\}$, then:

$$F_k = 1 - \prod_{j=1}^{k} (1 - V_j) \tag{4}$$

where the V_j are mutually independent random variables defined on $(0, 1)$. Let $E(1 - V_j) = q_j$. It is easy to verify that the random measure F defined by (4) is almost surely a random probability distribution on Ω if, and only if, $\prod_{i=1}^{\infty} q_j = 0$.

COROLLARY 3.1 *An exchangeable sequence $X_1, X_2, \cdots$, with each X_i defined on the space $\Omega = \{1, 2, \cdots\}$, has a neutral to the right prior if, and only if,*

$$P(X_{n+1} = k | X_1, \cdots, X_n) = \frac{E\left\{V_k^{n_k+1}(1 - V_k)^{m_k} \prod_{j<k} V_j^{n_j}(1 - V_j)^{m_j+1}\right\}}{E\left\{V_k^{n_k}(1 - V_k)^{m_k} \prod_{j<k} V_j^{n_j}(1 - V_j)^{m_j}\right\}}, \tag{5}$$

where $n_k = \sum_{1 \le i \le n} I(X_i = k)$ and $m_k = \sum_{1 \le i \le n} I(X_i > k)$.

Proof. If the exchangeable sequence has neutral to the right prior, then the condition (5) is surely satisfied. In order to prove the sufficiency of the condition (5), define $T_1 = V_1$ and for $k = 2, 3, \cdots$ define $T_k = V_k(1 - V_{k-1}) \cdots (1 - V_1)$ so that (5) can be written as

$$\frac{E\{T_k^{n_k+1} \prod_{j \ne k} T_j^{n_j}\}}{E\{T_k^{n_k} \prod_{j \ne k} T_j^{n_j}\}},$$

using $m_k + n_k = m_{k-1}$ with $m_0 = n$, leading to

$$P(X_1 = k_1, \cdots, X_n = k_n) = E\left\{\prod_k T_k^{n_k}\right\}. \tag{6}$$

Clearly $T = (T_1, T_2, \cdots)$ represents a random draw from a neutral to the right process prior provided we have $\sum_k T_k = 1$ a.s., which is satisfied if $\prod_k \{1 - EV_k\} = 0$. We have shown, (6), that given T, the X_is are i.i.d. and $P(X_1 = k | T) = T_k$ where T is from a neutral to the right process: by construction,

$$T_k = F_k - F_{k-1} \tag{7}$$

$F_0 = 0$, completing the proof.

Corollary 3.1 suggests the form for f_k, given by

$$f_k(n_1, \cdots, n_k, m_k) = g_k(n_k, m_k) \prod_{j<k} (1 - g_j(n_j, m_j)),$$

where

$$g_k(n_k, m_k) = \frac{E\left(V_k^{n_k+1}(1 - V_k)^{m_k}\right)}{E\left(V_k^{n_k}(1 - V_k)^{m_k}\right)}.$$

For tractability, it is obvious that we will need the distribution of V_ks. The most convenient distribution for V_ks is the beta distributions, say $V_k \sim \mathcal{B}(\alpha_k, \beta_k)$. This is the subject of Section 4.

Up to now, we have assumed the X_is are uncensored. Here we discuss the predictive in the presence of right censoring, assumed to be noninformative. The data can be summarized as $\{n_k, m_k\}_{k=1}^{\infty}$, where the n_k and m_k have been defined, as in Theorem 3.1, for example. Here we note that the predictive (5) is a function of $\{n_j, m_j\}$ for all $j \leq k$ and can therefore 'cope' with censored observations — no more work is required on our part.

4 The beta-Stacy process

In this section we discuss the NTR process when the V_ks are independent $\mathcal{B}(\alpha_k, \beta_k)$ random variables, and we call this the beta-Stacy process. Our motivation for working with beta-Stacy process stems from the fact that this process encapsulates virtually all the NTR processes mentioned in the literature.

The discrete case. First, we notice the conjugacy property of such a choice. From (5) we see that $P(X_1 = k) = E(V_k)$ and it is easy to show that $P(X_{n+1} = k|X_1, \cdots, X_n) = E(V_k^*)$, where $V_k^* \sim \mathcal{B}(\alpha_k + n_k, \beta_k + m_k)$, giving

$$P(X_{n+1} = k|X_1, \cdots, X_n) = \frac{\alpha_k + n_k}{\beta_k + \alpha_k + n_k + m_k} \prod_{j=1}^{k-1} \frac{\beta_j + m_j}{\beta_j + \alpha_j + n_j + m_j}. \tag{8}$$

It is also of interest to point out that if

$$\begin{aligned} Y_1 &\sim \mathcal{C}(\alpha_1, \beta_1, 1), \\ Y_2|Y_1 &\sim \mathcal{C}(\alpha_2, \beta_2, 1 - Y_1), \\ &\cdots \\ Y_k|Y_{k-1}, \cdots, Y_1 &\sim \mathcal{C}(\alpha_k, \beta_k, 1 - F_{k-1}), \end{aligned} \tag{9}$$

where $F_k = \sum_{j=1}^{k} Y_j$, then F defined by the countable sequence of r.v. Y_k is from a beta-Stacy process and, for any $m > 1$,

$$\mathcal{L}(Y_1, \cdots, Y_m) = \mathcal{G}(\alpha_1, \beta_1, \cdots, \alpha_m, \beta_m).$$

If we put some constraints on the parameters of the beta distribution it is possible to obtain different processes belonging the class of NTR. Observe that the Dirichlet process arises when we constrain

$$\beta_j = \sum_{k>j} \alpha_k \qquad \left(\sum_{j=1}^{\infty} \alpha_j < \infty \right).$$

Under this condition, and with no censored observations, (5) becomes:

$$P(X_{n+1} = k|X_1, \cdots, X_n) = \frac{\alpha_k + n_k}{\sum_{j=1}^{\infty} \alpha_j + n} \tag{10}$$

since $\alpha_k + \beta_k = \beta_{k-1}$, $n_k + m_k = m_{k-1}$ and $n_1 + m_1 = n$. Expression (10) is identified as a sequence of predictive probabilities obtained from a Polya-urn scheme, and as such, characterizes the Dirichlet process (Blackwell and MacQueen, 1973). We do not understand why the constraint $\beta_j = \sum_{k>j} \alpha_k$ is appropriate, other than providing a simple form for the predictive. It is obvious that the simplification of (8) into (10) fails if censored observations are present, and in this respect the Dirichlet process is not a natural prior in the presence of censoring.

To see this we note that if $X_1, \cdots, X_n$, with each $X_i \in \{t_k : k \geq 1\}$, is an i.i.d. sample, possibly with right censoring (with X_i being the censoring time if applicable), from an unknown F, defined by the countable sequence of random variables $\{Y_k\}$ in (9), then the likelihood function, assuming that there are no censoring times or exact observations for $t > t_L$, is given by

$$l(y_1, y_2, \cdots, y_L|\text{data}) \propto y_1^{n_1} \cdots y_L^{n_L}(1 - y_1)^{r_1} \cdots (1 - y_1 - \cdots - y_L)^{r_L} \times I,$$

where n_k is the number of exact observations at t_k, r_k is the number of censoring times at t_k $(X > t_k)$, I is the indicator function given in (1) and $n = n_1 + \cdots + n_L + r_1 + \cdots + r_L$.

The generalized Dirichlet distribution is clearly seen to be a conjugate prior, the Dirichlet distribution is not. We can repeat our result from the predictive approach using the likelihood and prior:

THEOREM 4.1 (Walker and Muliere, 1977a) *Let $X_1, \cdots, X_n$, with each $X_i \in \{t_k : k \geq 1\}$, be an i.i.d. sample, possibly with right censoring, with an unknown F. If F is from a discrete time beta-Stacy process with parameters $\{\alpha_k, \beta_k\}$ and jumps at $\{t_k\}$, then, given $X_1, \cdots, X_n$, the posterior distribution for F is also a discrete time beta-Stacy process with jumps at $\{t_k\}$ and parameters $\{\alpha_k^*, \beta_k^*\}$ where*

$$\alpha_k^* = \alpha_k + n_k \ \text{ and } \ \beta_k^* = \beta_k + m_k \tag{11}$$

and m_k is the sum of the number of exact observations in $\{t_j : j > k\}$ and censored observations in $\{t_j : j \geq k\}$; that is, $m_k = \sum_{j>k} n_j + \sum_{j \geq k} r_j$.

REMARK 4.1. We have seen that the Dirichlet process is not conjugate with respect right censored data whereas the beta-Stacy is. If the prior is a Dirichlet process then the posterior, given right censored data, is a beta-Stacy process.

Bernoulli Trips. We can also understand the beta-Stacy process in terms of an exchangeable process. We introduce a simple concept and method for modeling multiple state processes based on an exchangeable sampling scheme (Bernoulli trip), suggested by Walker (1996). A Bernoulli trip is a reinforced random walk (Coppersmith and Diaconis, 1987,1988; Pemantle, 1988) on a tree which characterizes the space for which a prior is required. An observation in this space corresponds to an unique path or branch of the tree. The path corresponding to this observation is reinforced; that is, the probability of a future observation following this path is increased; thus, after n observations, a maximum of n paths have been reinforced.

To construct the Bernoulli trip, we define a sequence $Z_1, Z_2, \cdots$ of independet random variables defined on $(0, 1)$ such that, for all $j = 1, 2, \cdots$, and $r, s > 0$

$$E(Z_j^r(1 - Z_j)^s)$$

exists. Let $Y_1, Y_2, \cdots$ be independent Bernoulli random variables such that:

$$P(Y_j = 1) = \frac{E(Z_j^{r_j+1}(1 - Z_j)^{s_j})}{E(Z_j^{r_j}(1 - Z_j)^{s_j})}.$$

where $s_j=\sum_{k>j}^{\infty} r_k$ and $\sum_{j=1}^{\infty} r_j<\infty$. Note that

$$P(Y_j=0)=1-P(Y_j=1)=\frac{E(Z_j^{r_j}(1-Z_j)^{s_j+1})}{E(Z_j^{r_j}(1-Z_j)^{s_j})}.$$

A Bernoulli trip along the positive integers involves sampling the Y_j in turn, starting at $j=1$. The trip is completed, at the integer X, whenever the event

$$\mathcal{E}_X=\{Y_1=0,\cdots,Y_{X-1}=0,Y_X=1\}$$

occurs. Along the trip whenever $Y_j=0$ the current s_j is replaced by s_j+1 and whenever $Y_j=1$ the current r_j is replaced by r_j+1. A second trip on completion of the first trip, and so on, involves returning to $j=1$ and repeating the scheme (always keeping the updated $\{r_j,s_j\}$). After the nth trip let the updated parameters be $\{r_j(n),s_j(n)\}$ so that, in particular,

$$n_j(n)=r_j(n)-r_j$$

is the number of trips completed at j and

$$m_j(n)=s_j(n)-s_j$$

is the number of trips completed at integers greater than j; that is,

$$m_j(n)=\sum_{k>j} n_k(n).$$

A Bernoulli trip is censored at X if only $\{Y_1=0,\cdots,Y_{X-1}\}$ are sampled and $\{Y_X,Y_{X+1},\cdots\}$ are not sampled. Therefore, it is only known that the particular trip in question is completed somewhere $>X$. The censoring occurs at random and independently of the Y_js, so that the updating mechanism for such a trip is given by

$$s_j\to s_j+1 \quad \text{for } j=1,\cdots,X.$$

Then X_1 characterises the first walk, X_2 the second walk, and so on. We can write the joint probability of the first n walks following particular paths. From this it is possible to show (Walker, 1996) that $X_1,X_2,\cdots,X_n$ are exchangeable random variables for all n, and the exchangeable process, $X_1,X_2,\cdots$, has a neutral to the right process prior. In particular, taking $r_j=0$ and $Z_j\sim\mathcal{B}(\alpha_j,\beta_j)$ for $\alpha_j,\beta_j>0$:

$$P(Y_j=1)=\frac{\alpha_j}{\alpha_j+\beta_j}$$

and

$$P(Y_j=1|X_1,X_2,\cdots,X_n)=\frac{\alpha_j+n_j}{\alpha_j+\beta_j+n_j+m_j}.$$

Therefore, for any n,

$$P(X_{n+1}=k|X_1,\cdots,X_n)=\frac{\alpha_k+n_k}{\beta_k+\alpha_k+n_k+m_k}\prod_{j=1}^{k-1}\frac{\beta_j+m_j}{\beta_j+\alpha_j+n_j+m_j}. \tag{12}$$

which characterizes the discrete time version of the beta-Stacy process. These trips can be extended to modeling multiple state processes in an obvious way (Walker, 1996).

The continuous case. It is possible to define the continuous time beta-Stacy process using

Lévy theory (Lévy , 1936). It is well known (Doksum, 1974) that a random distribution function F on the real line is NTR if it can be expressed as $F(t) = 1 - \exp[-Z(t)]$, where Z is a Lévy process satisfying $Z(0) = 0$ a.s. and $\lim_{t\to\infty} Z(t) = \infty$ a.s. Let α be a continuous measure and β a positive function: F is a beta-Stacy process, with parameters α and β, if the Lévy measure for Z is given by

$$dN_t(v) = \frac{dv}{1 - e^{-v}} \int_0^t \exp\left[-v\beta(s)\right] d\alpha(s).$$

It can be shown that F is almost surely a random probability measure under the condition $\int d\alpha(s)/\beta(s) = +\infty$. The beta-Stacy process generalizes the Dirichlet process, which is obtained when α is a finite measure and $\beta(s) = \alpha(s, \infty)$. Compare this constraint with the discrete constraint. The simple homogeneous process (Ferguson and Phadia, 1979) arises when β is constant.

REMARK 4.2. The beta-Stacy is closely related with the beta process (Hjort,1990). With the beta process the statistician is required to consider hazard rates and cumulative hazards when constructing the prior. The beta-Stacy only requires considerations on the distribution of the observations.

The predictive version of (12) in the continuous framework simply involves replacing the product with a *product integral* (Gill and Johansen, 1990):

$$P(X_{n+1} > t | X_1, \cdots, X_n) = \prod_{[0,t]} \left(1 - \frac{d\alpha(s) + dN(s)}{\beta(s) + Y(s)}\right),$$

where $N(s) = \sum_i I(X_i \leq s, \delta_i = 1)$ and $Y(s) = \sum_i I(X_i \geq s)$. The Kaplan-Meier estimator is obtained when $\alpha(.), \beta(.) = 0$.

Here we provide the theory for using a general Z Lévy process for modeling a cumulative distribution function; that is, taking $F(t) = 1 - \exp[-Z(t)]$. We assume the Lévy measure to be of the type

$$dN_t(v) = dv \int_0^t K(v, s) ds$$

and $\int v dN_t(v) < \infty$ and $\int v^2 dN_t(v) < \infty$ which ensure that $E[F(t)]$ and $\mathrm{var}[F(t)]$ both exist. We also assume that there are no fixed points of discontinuity in the prior process.

THEOREM 4.2 (Walker and Muliere, 1997a). *Let Z be a Lévy process with $Z(0) = 0$ and $\lim_{t\to\infty} Z(t) = \infty$. The posterior Lévy process is given by*

$$Z^*(t) = \sum_{X_i \leq t, \delta_i = 1} S_{X_i} + Z_c^*(t),$$

where the S_X are independent jump variables with density function

$$f_x(v) \propto [1 - \exp(-v)]^{N\{x\}} \exp[-vY(x)] K(v, x)$$

and Z_c^* is a Lévy process with Lévy measure

$$dN_t^*(v) = dv \int_0^t \exp[-vY(s)] K(v, s) ds.$$

Here (X_i, δ_i) is the observed data ($\delta_i = 1$ indicating an exact observation) and $Y(s) = \sum_i I(X_i > s)$.

Consequently, the Bayes estimator for the cumulative distribution function, with respect to a quadratic loss function, which coincides with the predictive distribution $P(X_{n+1} > t|X_1, \cdots, X_n)$, is given by

$$\sum_{X_i \le t, \delta_i = 1} E[\exp(-S_{X_i})] + \exp\left(-\int_0^\infty (1-\exp(-v))dN_t^*(v)\right).$$

5 Exchangeable neutral urn scheme

Various schemes have been introduced in the literature for constructing prior distributions on the spaces of probability measures. The Polya urn scheme is, perhaps, the simplest and most concrete way (see: Blackwell and MacQueen ,1973; Mauldin,Sudderth and Williams,1992). In this section *exchangeable neutral* urn schemes are introduced. Let $\{\theta_1, \theta_2, \cdots, \theta_N\}$ represent a finite sample space and consider a sequence of random variables $X = \{X_1, X_2, \cdots\}$ with each $X_j \in \{\theta_1, \theta_2, \cdots, \theta_N\}$. Also introduce the *dummy* space $\{\phi_1, \cdots, \phi_{N-1}\}$, which, again, represents a finite sample space.

Let $V_1, V_2, \cdots,$ be mutually independent random variables, defined on $(0, 1)$, such that, for all $\alpha, \beta, \ge 0$, $E(V_k^\alpha(1-V_k)^\beta)$ exists. Then define, for all $\alpha, \beta, \ge 0$,

$$\lambda_k(\alpha, \beta) = E(V_k^\alpha(1-V_k)^\beta).$$

Take N urns, and in urn $k = 1, 2, \cdots, N-1$, put the elements θ_k and ϕ_k in the ratio $\lambda_k(\alpha_k + 1, \beta_k)$ to $\lambda_k(\alpha_k, \beta_k + 1)$, where the constraint on $\{\alpha_k, \beta_k\}$ is that $\beta_k = \sum_{l>k} \alpha_l$. The urn N has only the element θ_N.

Generate a sequence, $X = (X_1, X_2, \cdots,)$, from $\{\theta_1, \theta_2, \cdots, \theta_N\}$ by starting at urn $k = 1$. Sample the urn; if θ_1 is taken then this is the required first sample from the scheme, that is, $X_1 = \theta_1$. Replace α_1 by $\alpha_1 + 1$. If ϕ_1 is taken then replace β_1 by $\beta_1 + 1$ and go to the next urn. Repeat the procedure until θ_k, which is then the required first sample; that is, $X_1 = \theta_k$, is taken from urn k. Note this happens with probability 1 at the Nth urn (if it is reached).

To obtain the next sample, X_2, and so on, return to the first urn, keeping the new urns, and repeat the procedure, noting that whenever θ_k is sampled then replace α_k by $\alpha_k + 1$, and take θ_k as the required sample, and whenever ϕ_k is sampled then replace β_k by $\beta_k + 1$, and move on the next urn.

It is easy to show (Muliere and Walker, 1997a) that the sequence $X = (X_1, X_2, \cdots)$ is exchangeable, and the predictive probabilities at the $(m+1)$th iteration of the scheme are given by:

$$P(X_{m+1} = k|X_1, X_2, \cdots, X_m) = \frac{\lambda_k(\alpha_k + n_k + 1, \beta_k + m_k)}{\lambda_k(\alpha_k + n_k, \beta_k + m_k)} \prod_{j=1}^{k-1} \frac{\lambda_k(\alpha_j + n_j, \beta_j + m_j + 1)}{\lambda_k(\alpha_j + n_j, \beta_j + m_j)} \tag{13}$$

where $n_k = \sum_{l=1}^m I(X_l = \theta_k)$ and $m_k = \sum_{l>k} n_l$.

REMARK 5.1 Specifying the distribution of the $V_1, V_2, \cdots$ we obtain different predictive distributions and different schemes. If $V_k \sim \mathcal{B}(r_k, s_k)$, where $s_k = \sum_{l>k} r_l$ and $\alpha_k = \beta_k = 0$, the exchangeable neutral scheme is the Polya-urn scheme, with the prior urn containing r_k amounts of θ_k. The Polya-urn scheme on $\{1, 2, \cdots, N\}$ can now be thought of an exchangeable neutral scheme with a constraint on the composition of the prior urns. Without the condition $s_k = \sum_{l>k} r_l$, we obtain the generalised Polya-urn scheme.

6 Bayesian bootstraps

The bootstrap resampling plan introduced by Efron (1979), has a Bayesian counterpart; the Bayesian bootstrap, BB (Rubin, 1981). Both resampling plans are asymptotically equivalent (Lo, 1987; Weng, 1989) and first order equivalent from a predictive point of view (Muliere and Secchi, 1996).

A Bayesian bootstrap for a finite population, FPBB, was introduced by Lo (1988), followed by a censored data Bayesian bootstrap, CDBB, in Lo (1993). Other bootstraps for censored data, include those of Reid (1981), Efron (1981) and Wells and Tiwari (1994). Efron's bootstrap and the CDBB are first order asymptotically equivalent (Lo, 1993), but Akritas (1986) showed that Reid's bootstrap is not asymptotically equivalent to that of Efron.

A *finite population censored data* Bayesian bootstrap, FCBB, was introduced by Muliere and Walker (1977b). The FCBB method is defined in terms of the generalised Polya-urn scheme (Walker and Muliere, 1977a).

The BB simulates (discrete) random probability distributions based on the observed data. If $X_1, X_2, \cdots, X_n$ are (uncensored) observations, then a Bayesian bootstrap simulation is given by:

$$F_{BB} = \sum_{i=1}^{n} W_i \delta_{X_i}, \tag{14}$$

where

$$\mathcal{L}(W_1, W_2, \cdots, W_n) = \mathcal{D}(1, 1, \cdots, 1).$$

The CDBB is obtained via simulation from a discrete time beta-Stacy process with parameters (n_j, m_j); that is, with α and β taken to be identically zero. The FPBB simulates a randon probability distribution F_{FPBB}. If the population size is N and the sample size is n, $(n < N)$, then the FPBB samples the missing $m = N - n$ observations using a Polya-urn scheme: that is,

$$F_{FPBB} = N^{-1}(nF_n + mF_m) \tag{15}$$

where F_n is the empirical distribution of the observations and F_m is the random probability distribution

$$F_m = m^{-1} \sum_{i=n+1}^{n+m} \delta_{E_i}$$

and $E_{n+1}, \cdots, E_{n+m}$ are taken from a Polya-urn scheme where $X_1, X_2, \cdots, X_n$ are the observed data. Explicitly, this involves taking

$$P(E_{n+1} = X_i) = n_i/n$$

and

$$P(E_{n+m} = X_i | E_{n+1}, \cdots, E_{n+m-1}) = \frac{n_i + \sum_{i=1}^{m-1} I(E_{n+1} = X_i)}{n + m - 1}.$$

REMARK 6.1 $\mathcal{L}(F_{FPBB}) \to \mathcal{L}(F_{BB})$ as $m \to \infty$, with n fixed.

A finite population censored data Bayesian bootstrap involves taking

$$F_{FCBB} = N^{-1}(nF_n + mF_m)$$

where now F_n is the Kaplan-Meier nonparametric estimator of the survival distribution and $E_{n+1}, \cdots, E_{n+m}$ are taken from a generalized Polya-urn scheme based on uncensored observations $X_1 < X_2 < \cdots < X_k$ for some $k \leq n$. Explicitly, this involves taking

$$P(E_{n+1} = X_i) = \frac{n_i}{n_i + m_i} \prod_{l=1}^{i-1} \frac{m_l}{n_l + m_l}$$

and $P(E_{n+m} = X_i | E_{n+1}, \cdots, E_{n+m+1}) =$

$$\frac{n_i + \sum_{j=1}^{m-1} I(E_{n+j} = X_i)}{n_i + m_i + \sum_{j=1}^{m-1} I(E_{n+j} \geq X_i)} \prod_{l=1}^{i-1} \frac{m_l + \sum_{j=1}^{m-1} I(E_{n+j} > X_l)}{n_l + m_l + \sum_{j=1}^{m-1} I(E_{n+j} \geq X_l)}.$$

REMARK 6.2

$$\mathcal{L}(F_{FCBB}) \rightarrow \mathcal{L}(F_{CDBB}) \quad \text{as } m \rightarrow \infty \ \ (n \text{ fixed})$$

and

$$\mathcal{L}(F_{FCBB}) = \mathcal{L}(F_{FPBB}) \ \ \text{no censoring.}$$

REMARK 6.3 The FPBB is defined by a Multinomial Dirichlet (MD) point process (Lo,1988). The FCBB is defined by a Multinomial Generalized Dirichlet (MGD) process. This MGD process in a limit is to be a beta-Stacy process.

References

Akritas, M.G. (1986). Bootstrapping the Kaplan-Meier estimate. *J. Amer. Statist. Assoc.* 81, 1032-1038.

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models based on Counting Processes*, Springer-Verlag.

Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Polya-urn schemes. *Ann. Statist.* 1, 353-355.

Connor, R.J. and Mosimann, J.E. (1969). Concepts of independence for proportions with a generalisation of the Dirichlet distribution. *J. Amer. Statist. Assoc.* 64, 194-206.

Coppersmith, D. and Diaconis, P. (1987). Random walk with reinforcement. *unpublished manuscript*

de Finetti, B. (1935), Il problema della perequazione. *Atti della societa' Italiana per il Progresso delle Scienze, (XII Riunione)*, Napoli.

de Finetti, B. (1937). La prevision: ses lois logiques, ses sources subjectives. *Ann. Instit. H. Poincare* 7, 1-68.

Diaconis,P. (1988). Recent progress on de Finetti's notions of exchangeability. *Bayesian Statistics* 3, 111- 125.

Doksum, K.A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* 2, 183-201.

Dykstra, R.L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* 9, 356-367.

Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *Ann. Statist.* 7, 1-26.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* 76, 312-319.

Fabius, J. (1964). Asymtotic Behavior of Bayes estimate. *Ann. Math. Statist.* 35, 846-856.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* 1, 209-230.

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* 2, 615-629.

Ferguson, T.S. and Phadia, E.G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* 7, 163-186.

Freedman, D.A. (1963). On the asymptotic behavior of Bayes' estimate in the discrete case. *Ann. Math. Statist.* 34, 1386-1403.

Gill, R.D. and Johansen, S. (1990). A survey of product integration with a view toward application in survival analysis. *Ann. Statist.* 18, 1501-1555.

Hewitt, E. and Savage, L.J. (1955). Symmetric measures on cartesian products. *Trans. Amer. Math. Soc.* 80, 1501-1555.

Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* 18, 1259-1294.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53, 457-481.

Lévy, P. (1936). *Theorie de l'Addition des Variables Aleatoire.* Gauthiers-Villars, Paris.

Lo, A.Y. (1987). A large sample study of the Bayesian bootstrap. *Ann. Statist.* 15, 360-375.

Lo, A.Y. (1988). A Bayesian bootstrap for a finite population. *Ann. Statist.* 16, 1684-1695.

Lo, A.Y. (1993). A Bayesian bootstrap for censored data. *Ann. Statist.* 21, 100 -123.

Mauldin,R.D.,Sudderth,W.D. and Williams,S.C. (1992). Polya trees and random distributions. *Ann.Statist.* 20, 1203-1221.

Mihram, G.A. and Hultquist, R.A. (1967). A bivariate warning-time/failure-time distribution. *J. Amer. Statist. Assoc.* 62, 589-599.

Muliere, P. and Secchi, P. (1996). Bayesian nonparametric predictive inference and bootstrap techniques. *Ann. Inst. Statist. Math.* 48, 663-673.

Muliere, P. and Walker, S.G. (1997a). A Bayesian nonparametric approach to survival analysis using Polya trees. *Scand. J. Statist.* 24, 331-340.

Muliere, P. and Walker, S.G. (1997b). Extending the family of Bayesian bootstraps and exchangeable urn schemes. *J. R. Statist. Soc. Ser. B*, to appear.

Pemantle, R. (1988). Phase transition in reinforced random walk and RWRE on trees. *Ann. Probab.* 16, 1229-1241.

Reid, N. (1981). Estimating the median survival time. *Biometrika* 68, 601-608.

Rubin, D.B. (1981). The Bayesian bootstrap. *Ann. Statist* 9, 130-134.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete data. *J. Amer. Statist. Assoc.* 71, 897-902.

Walker, S.G. (1996). Bayesian nonparametrics from Bernoulli trips. *Technical Report, Imperial College.* London.

Walker, S.G. and Muliere, P. (1997a). Beta-Stacy Processes and a generalisation of the Polya-urn Scheme. *Ann. Statist.* 25, 1762-1780.

Walker, S.G. and Muliere, P. (1997b). A characterization of a neutral to the right prior via an extension of Johnson's sufficientness postulate. *Technical report, University of Pavia* 55, (1-97).

Wells, M.T. and Tiwari, R.C. (1994). Bootstrapping a Bayes estimator of a survival function with censored data. *Ann. Inst. Statist. Math.* 46, 487-495.

Weng, C.S. (1989). On a second order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.* 17, 705-710.

Zabell, S.L. (1982). W.E.Johnson's "Sufficientness" postulate. *Ann. Statist.* 10, 1091-1099.

Zabell, S.L. (1995). Characterizing Markov exchangeable sequences. *J. Theor. Probab.* 8, 175-178.