The Annals of Statistics
1997, Vol. 25, No. 4, 1762–1780

BETA-STACY PROCESSES AND A GENERALIZATION OF THE PÓLYA-URN SCHEME

BY STEPHEN WALKER AND PIETRO MULIERE

Imperial College, London, and Università di Pavia

A random cumulative distribution function (cdf) F on $[0, \infty)$ from a beta-Stacy process is defined. It is shown to be *neutral to the right* and a generalization of the Dirichlet process. The posterior distribution is also a beta-Stacy process given independent and identically distributed (iid) observations, possibly with right censoring, from F. A generalization of the Pólya-urn scheme is introduced which characterizes the discrete beta-Stacy process.

1. Introduction. Let $\mathscr{F}$ be the space of cumulative distribution functions (cdfs) on $[0, \infty)$. This paper considers placing a probability distribution on $\mathscr{F}$ by defining a stochastic process F on $([0, \infty), \mathscr{A})$, where $\mathscr{A}$ is the Borel σ-field of subsets, such that $F(0) = 0$ a.s., F is a.s. nondecreasing, a.s. right continuous and $\lim_{t\to\infty} F(t) = 1$ a.s. Thus, with probability 1, the sample paths of F are cdf's.

Previous work includes the Dirichlet process [Ferguson (1973, 1974)], *neutral to the right* processes [Doksum (1974)], the extended gamma process [Dykstra and Laud (1981)], the beta process [Hjort (1990)] and Pólya trees [Lavine (1992, 1994), Mauldin, Sudderth and Williams (1992)].

The purpose of this paper is twofold: (1) to introduce a new stochastic process which generalizes the Dirichlet process, in that more flexible prior beliefs are able to be represented, and, unlike the Dirichlet process, is conjugate to right censored observations, and (2) to introduce a generalization of the Pólya-urn scheme in order to characterize the discrete time version of the process.

The property of conjugacy to right censored observations is also a feature of the beta process; however, with the beta process the statistician is required to consider hazard rates and cumulative hazards when constructing the prior. The beta-Stacy process only requires considerations on the distribution of the observations. The process is shown to be *neutral to the right*.

The present paper is restricted to considering the estimation of an unknown cdf on $[0, \infty)$, although it is trivially extended to include $(-\infty, \infty)$. Finally, for ease of notation, F is written to mean either the cdf or the corresponding probability measure.

The organization of the paper is as follows. In Section 2 the process is defined and its connections with other processes given. We also provide an

Received April 1995; revised October 1996.

AMS 1991 *subject classifications*. Primary 62C10; secondary 60G09.

Key words and phrases. Bayesian nonparametrics, beta-Stacy process, Dirichlet process, generalized Dirichlet distribution, generalized Pólya-urn scheme, Lévy process, neutral to the right process.

interpretation for the parameters of the process in terms of the mean and variance of F. Section 3 considers the construction and posterior distributions of the discrete time process, while Section 4 considers the process in continuous time. In Section 4.4 we present a numerical example illustrating the process in continuous time. In Section 5 a generalization of Pólya's urn scheme is introduced which characterizes the discrete time process.

2. Definition of the process. First some preliminary information is required. Let $\mathscr{B}(\alpha, \beta)$ for $\alpha, \beta > 0$ represent the beta distribution. For the purposes of the paper it is convenient to define $\mathscr{G}(\alpha_1, \beta_1, \dots, \alpha_m, \beta_m)$ for $\alpha_j, \beta_j > 0$ to represent the generalized Dirichlet distribution, introduced by Connor and Mosimann (1969). The density function is given, up to a constant of proportionality, by

$$\begin{aligned} & y_1^{\alpha_1-1}(1-y_1)^{\beta_1-1} \times \frac{y_2^{\alpha_2-1}(1-y_1-y_2)^{\beta_2-1}}{(1-y_1)^{\alpha_2+\beta_2-1}} \times \cdots \\ & \times \frac{y_m^{\alpha_m-1}(1-y_1-\cdots-y_{m-1}-y_m)^{\beta_m-1}}{(1-y_1-\cdots-y_{m-1})^{\alpha_m+\beta_m-1}} \\ & \times I\left\{(y_1, \dots, y_m)\colon y_j \ge 0,\ \sum_{j=1}^m y_j \le 1\right\}, \end{aligned} \tag{1}$$

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

$$y_1^{\alpha_1-1} \cdots y_m^{\alpha_m-1}(1-y_1-\cdots-y_m)^{\beta_m-1} I\left\{(y_1, \dots, y_m)\colon y_j \ge 0,\ \sum_{j=1}^m y_j \le 1\right\},$$

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

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

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

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

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

Unlike the Dirichlet process which can be defined by the distribution of the joint probabilities of any finite measurable partition of $[0, \infty)$, the beta-Stacy process is defined via the Lévy process and the representation of neutral to the right processes introduced by Doksum (1974).

Consider a Lévy process $Z(t)$ such that the following hold:

1. $Z(t)$ has nonnegative independent increments;
2. $Z(t)$ is nondecreasing a.s.;
3. $Z(t)$ is right continuous a.s.;
4. $Z(t) \to \infty$ a.s. as $t \to \infty$;
5. $Z(0) = 0$ a.s.

For such a process there exist at most countably many fixed points of discontinuity at time points $t_1, t_2, \dots$ with jumps $S_1, S_2, \dots$, independent nonnegative random variables. Then $X(t) = Z(t) - \sum_{t_k \le t} S_k$ is a nondecreasing process with independent increments and with no fixed points of discontinuity and therefore has a Lévy formula [Lévy (1936)] for the moment generating function of the process given by

$$\log E \exp\{-\phi X(t)\} = -\phi b(t) + \int_0^\infty (\exp(-\phi v) - 1)\, dN_t(v),$$

where b is nondecreasing and continuous, with $b(0) = 0$, and N_t is a Lévy measure satisfying:

1. for every Borel set B, $N_t(B)$ is continuous and nondecreasing;
2. for every real $t > 0$, $N_t(\cdot)$ is a measure on the Borel sets of $(0, \infty)$;
3. $\int_0^\infty v(1+v)^{-1}\, dN_t(v) < \infty$;
4. $\int_0^\infty v(1+v)^{-1}\, dN_t(v) \to 0$ as $t \to 0$.

Since b represents a nonrandom component it is not considered in this paper and we assume it to be identically zero.

Let $c(\cdot)$ be a positive function, let $G \in \mathscr{F}$ be right continuous and let $\{t_k\}$ be the countable set of points of discontinuity of G, that is, $G\{t_k\} = G(t_k) - G(t_k-) > 0$ for all k. Now put $G_c(t) = G(t) - \sum_{t_k \le t} G\{t_k\}$ so that $G_c(\cdot)$ is continuous.

DEFINITION 2. F is a beta-Stacy process on $([0,\infty), \mathscr{A})$ with parameters $c(\cdot)$ and G, written $F \sim \mathscr{S}(c(\cdot), G)$, if, for all $t \ge 0$, $F(t) = 1 - \exp(-Z(t))$, where Z is a Lévy process with Lévy measure for $Z(t)$ given, for $v > 0$, by

$$dN_t(v) = \frac{dv}{(1 - \exp(-v))} \int_0^t \exp(-vc(s)G(s,\infty))c(s)\, dG_c(s) \tag{2}$$

and moment generating function given by

$$\log E \exp\{-\phi Z(t)\} = \sum_{t_k \le t} \log E \exp(-\phi S_k) + \int_0^\infty (\exp(-v\phi) - 1)\, dN_t(v), \tag{3}$$

where $1 - \exp(-S_k) \sim \mathscr{B}(c(t_k)G\{t_k\}, c(t_k)G[t_k, \infty))$.

It is easy to show that

$$EF(t) = 1 - \exp\left(-\int_0^t dG_c(s)/G(s,\infty)\right) \prod_{t_k \le t} (1 - G\{t_k\}/G(t_k,\infty)),$$

which can be written as the product integral [Gill and Johansen (1990)], given by

$$1 - \prod_{[0,t]} (1 - dG(s)/G(s,\infty)). \tag{4}$$

Expression (4) gives $EF(t) = G(t)$ [Hjort (1990), page 1260]. Here a straightforward result is given as a lemma.

LEMMA 1. *If U_k is a sequence of nonnegative random variables such that $U_k \le 1$ a.s., $\{U_k\}$ is a.s. nondecreasing and $EU_k \to 1$, then $U_k \to 1$ a.s.*

Since $G(t) \to 1$ as $t \to \infty$, $F(t) \le 1$ a.s. and $F(t)$ is a.s. nondecreasing, it follows from this lemma that $F(t) \to 1$ a.s. (This can be seen by considering an increasing sequence of time points $\{s_k\}$ such that $s_k \to \infty$). It follows that a.s. $F \in \mathscr{F}$ and that F is a neutral to the right process [Doksum (1974)]. In addition, as Z has no nonrandom component, Z increases only in jumps a.s. and F is with probability 1 a discrete member of $\mathscr{F}$.

According to Ferguson and Phadia (1979) the fundamental result of Doksum (1974) on the posterior of a process neutral to the right is that $F|\Theta$ is also neutral to the right for any observation of the type $\Theta = \theta$ or $\Theta > \theta$, where Θ is a random sample from F. It is shown later that $F|\Theta$ is also beta-Stacy and hence the conjugacy property of the process.

REMARK 1. The beta-Stacy process generalizes the Dirichlet process, which can be seen more easily if G is taken to be continuous, since, if $c(s) = c > 0$ for all $s \ge 0$, then (2) becomes

$$\begin{aligned} dN_t(v) &= \frac{dv}{(1-e^{-v})}\left[v^{-1}\exp(-vc\{1-G(s)\})\right]_0^t \\ &= \frac{dv}{v(1-e^{-v})}\left(\exp(-vc\{1-G(t)\}) - e^{-vc}\right), \end{aligned}$$

the Lévy measure given in Ferguson (1974) which represents the Lévy process corresponding to the Dirichlet process when viewed as a neutral to the right process.

Here the Lévy process Z which leads to the beta-Stacy process is considered with more general parameters. Let the right continuous measure $\alpha(\cdot)$ [$\alpha(0) = 0$] and the positive function $\beta(\cdot)$ both be defined on $[0,\infty)$, and let $\{t_k\}$ be the countable set of points of discontinuity of $\alpha(\cdot)$. Put $\alpha_c(t) = \alpha(t) - \sum_{t_k \le t} \alpha\{t_k\}$ so that $\alpha_c(\cdot)$ is a continuous measure.

DEFINITION 3. Z is a log-beta process on $([0,\infty), \mathscr{A})$, with parameters $\alpha(\cdot)$ and $\beta(\cdot)$, if Z is a Lévy process with Lévy measure for $Z(t)$ given, for $v > 0$, by

$$dN_t(v) = \frac{dv}{(1-\exp(-v))}\int_0^t \exp(-v(\beta(s)+\alpha\{s\}))\,d\alpha_c(s) \tag{5}$$

and with moment generating function given by

$$\log E\exp\{-\phi Z(t)\} = \sum_{t_k\le t}\log E\exp(-\phi S_k) + \int_0^\infty (\exp(-\phi v)-1)\,dN_t(v), \tag{6}$$

where $1-\exp(-S_k)\sim \mathscr{B}(\alpha\{t_k\},\beta(t_k))$.

Hjort (1990) considered Lévy processes given by $dA = 1-\exp(-dZ)$ and used A (a beta process) as a prior distribution for the space of cumulative hazard functions.

REMARK 2. If A is a beta process and $dZ = -\log(1-dA)$, then $F(t) = 1-\exp\{-Z(t)\}$ is a beta-Stacy process. Therefore if interest is in $F[a,b) = \exp\{-Z[0,a)\}(1-\exp\{-Z[a,b)\})$ [and note that $Z(0,a)$ and $Z[a,b)$ are independent], then one should consider the beta-Stacy process, whereas if interest is in the cumulative hazard $A[a,b)$, then one should consider the beta process (in particular, when it comes to sampling $F[a,b)$ and $A[a,b)$).

REMARK 3. Let $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy the condition given by

$$\prod_k\left(1-\frac{\alpha\{t_k\}}{\beta(t_k)+\alpha\{t_k\}}\right)\exp\left(-\int_0^\infty\frac{d\alpha_c(s)}{\beta(s)+\alpha\{s\}}\right)=0. \tag{7}$$

Condition (7) can be written in the form of a product integral given by

$$\prod_{[0,\infty]}\left(1-\frac{d\alpha(s)}{\beta(s)+\alpha\{s\}}\right)=0.$$

This condition becomes $\int_0^\infty d\alpha(s)/\beta(s)=\infty$ if $\alpha(\cdot)$ is continuous. If now $F(t) = 1-\exp(-Z(t))$ and $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy condition (7), then $F(t)$ is beta-Stacy with parameters G and $c(\cdot)$ given by

$$G(t)=1-\prod_{t_k\le t}\left(1-\frac{\alpha\{t_k\}}{\beta(t_k)+\alpha\{t_k\}}\right)\exp\left(-\int_0^t\frac{d\alpha_c(s)}{\beta(s)+\alpha\{s\}}\right)$$

or

$$G(t)=1-\prod_{[0,t]}\left(1-\frac{d\alpha(s)}{\beta(s)+\alpha\{s\}}\right) \tag{8}$$

and

$$c(t)=\frac{\beta(t)}{G[t,\infty)}. \tag{9}$$

Similarly $\alpha(\cdot)$ and $\beta(\cdot)$ can be obtained from $c(\cdot)$ and G via

$$\alpha(t)=\int_0^t c(s)\,dG_c(s)+\sum_{t_k\le t}c(t_k)G\{t_k\}$$

and

$$\beta(t)=c(t)G[t,\infty).$$

If $\alpha(\cdot)$ is continuous, then G is continuous and given by

$$G(t) = 1 - \exp\biggl(-\int_0^t d\alpha(s)/\beta(s)\biggr).$$

REMARK 4. If $1 - \exp(-v)$ is replaced by v in the Lévy measure for the log-beta process, then it can be shown that, for continuous $\alpha(\cdot)$,

$$\log E \exp\{-\phi Z(t)\} = -\int_0^t \log(1 + \phi/\beta(s))\, d\alpha(s),$$

which characterizes the extended gamma process [Dykstra and Laud (1981)]. Also the Lévy measure for the beta process, $dL_t(s)$ with support on $(0, 1)$ [Hjort (1990)], can be obtained via a simple transformation of the Lévy measure from the log-beta process:

$$dL_t(s) = \frac{1}{1-s}\, dN_t(-\log(1-s)).$$

2.1. *Prior specifications.* It is important to be able to use any available prior information to center F and express uncertainty in F about this centering, that is, to assign arbitrarily $EF(t)$ and $\operatorname{var} F(t)$. We do this by considering the first two moments of $S(t) = 1 - F(t)$. Using the Lévy representation of a beta-Stacy process (without fixed points of discontinuity), it follows that

$$\mu(t) = -\log\{ES(t)\} = \int_0^\infty \int_0^t \exp(-v\beta(s))\, d\alpha(s)\, dv$$

and

$$\begin{aligned}\lambda(t) &= -\log\{E[S^2(t)]\}\\ &= \int_0^\infty \int_0^t \left\{\frac{1-\exp(-2v)}{1-\exp(-v)}\right\} \exp(-v\beta(s))\, d\alpha(s)\, dv.\end{aligned}$$

Note it is necessary that $0 < \mu(t) < \lambda(t) < 2\mu(t)$, which corresponds to $\{ES(t)\}^2 < E[S^2(t)] < ES(t)$. The first of these conditions is satisfied when

$$\int_0^t d\alpha(s)/\beta(s) = \mu(t),$$

that is, when

$$d\alpha(t) = \beta(t)\, d\mu(t).$$

The second condition becomes, using the transformation $u = 1 - \exp(-v)$,

$$\lambda(t) = \int_0^t \{2 - 1/(1 + \beta(s))\}\, d\mu(s),$$

leading to

$$d\lambda(t)/d\mu(t) = 2 - 1/(1 + \beta(t))$$

and hence the solution.

For example, if $ES(t) = \exp\{-at\}$, $a > 0$, and $E[S^2(t)] = \exp\{-b(t)\}$, $b \geq 0$, $b(0) = 0$, then

$$\beta(t) = 1/(2 - a^{-1}\, db/dt) - 1$$

and

$$d\alpha(t) = a\beta(t)\, dt.$$

Note that here we need the condition $a < db/dt < 2a$, which corresponds to the necessary condition $\mu(t) < \lambda(t) < 2\mu(t)$. This also provides an interpretation for $c(\cdot)$ in the alternative parameterization of Definition 2.

3. Discrete time process. Here the discrete time version of the process is studied. Section 3.1 gives the construction and Section 3.2 derives the posterior distributions having observed an iid sequence possibly with right censoring.

3.1. *The construction.* Let $0 < t_1 < t_2 < \cdots$ be a countable sequence of time points in $[0, \infty)$ indexed by $k = 1, 2, \ldots$. With each k there is associated a positive random variable Y_k defined as follows:

$$\begin{aligned} Y_1 &\sim \mathscr{C}(\alpha_1, \beta_1, 1),\\ Y_2 | Y_1 &\sim \mathscr{C}(\alpha_2, \beta_2, 1 - Y_1),\\ &\vdots\\ Y_k | Y_{k-1}, \ldots, Y_1 &\sim \mathscr{C}(\alpha_k, \beta_k, 1 - F_{k-1}), \end{aligned} \tag{10}$$

where

$$F_k = \sum_{j=1}^{k} Y_j$$

and each α_k and β_k is positive. It follows that a.s. $Y_k < 1 - \sum_{j=1}^{k-1} Y_j$ so that a.s. $F_k < 1$. Also, for any $m > 1$,

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

LEMMA 2. *For all* $k \geq 1$,

$$EF_k = \gamma_k + (1 - \gamma_k) EF_{k-1}, \tag{11}$$

where $F_0 = 0$ *a.s. and* $\gamma_k = \alpha_k / (\alpha_k + \beta_k)$.

PROOF. The proof follows immediately from (10) and the construction of F_k. □

LEMMA 3. *If* $\prod_{k=1}^{\infty} (1 - \gamma_k) = 0$, *then* $F_k \to 1$ *a.s.*

PROOF. From Lemma 2 it follows that

$$\prod_{k=1}^{\infty}\left(\frac{1-EF_k}{1-EF_{k-1}}\right)=0,$$

which implies that $\lim_{k\to\infty}(1-EF_k)=0$. Using Lemma 1 combined with $\{F_k\}$ a.s. nondecreasing and $F_k \le 1$ a.s., it is seen that $F_k \to 1$ a.s., completing the proof. □

PROPOSITION 1. *If* $F(0)=0$, $F(t)=\sum_{t_k\le t} Y_k$ *and the condition of Lemma* 2 *is satisfied*, *then*, *with probability* 1, $F\in\mathscr{F}$.

The proof of this follows from the construction of F and Lemmas 2 and 3.

Define the sets $\{B_k\}$ in $[0,\infty)$ by $B_k=(t_{k-1},t_k]$ and the sets $\{A_k\}$ by $A_k=[0,t_k]$. It is possible to center the process on a particular $G\in\mathscr{F}$ with $G(B_k)>0$ for all k: let $\gamma_1=G(B_1)$ and, for $k>1$,

$$\gamma_k=\frac{G(B_k)}{1-G(A_{k-1})};$$

it is then straightforward to show that $EY_k=G(B_k)$. A particular choice of α_k and β_k is given by $\alpha_k=c_kG(B_k)$ and $\beta_k=c_k(1-G(A_k))$ with $c_k>0$.

REMARK 5. If $c_k=c$ for all k, then $\alpha_k+\beta_k=\beta_{k-1}$ with the consequence that F is a discrete time Dirichlet process.

LEMMA 4. *If* $V_k=Y_k/(1-F_{k-1})$, *then*, *for any* $m>1$, *the random variables* $V_1,V_2,\dots,V_m$ *are independent and marginally each* $V_k\sim\mathscr{B}(\alpha_k,\beta_k)$.

PROOF. Consider $p_{Y_1,\dots,Y_m}(y_1,\dots,y_m)$ given, up to a constant of proportionality, by (1). By considering the transformation $V_1=Y_1$, $V_2=Y_2/(1-Y_1),\dots,V_m=Y_m/(1-Y_1-\cdots-Y_{m-1})$, so that $Y_1=V_1$, $Y_2=V_2(1-V_1),\dots,$ $Y_m=V_m\prod_{j=1}^{m-1}(1-V_j)$, the Jacobian of which is $\prod_{j=1}^{m-1}(1-V_j)^{m-j}$, the joint distribution of $V_1,\dots,V_m$ can be obtained, from which the result follows, completing the proof. □

Note the equality in law of $1-\sum_{j=1}^{k}Y_j$ and $\prod_{j=1}^{k}(1-V_j)$, which implies that $-\log(1-F_k)$ can be expressed as the sum of k independent random variables.

3.2. *Posterior distributions.* Throughout this section let $0=t_0<t_1<t_2<\cdots$ be any fixed partition of $[0,\infty)$. Let $\Theta_1,\dots,\Theta_n$, with each $\Theta_i\in\{t_k\colon k\ge 1\}$, be an iid sample, possibly with right censoring (with Θ_i being the censoring time if applicable), from an unknown cdf F on $[0,\infty)$ which is defined by the countable sequence of random variables $\{Y_k\}$ given by (10). Then F is referred to as a discrete time beta-Stacy process with parameters $\{\alpha_k,\beta_k\}$ and jumps at $\{t_k\}$. The likelihood function, assuming that there are no censoring times or exact observations for $t>t_L$, is given by

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

where n_k is the number of exact observations at t_k, r_k is the number of censoring times at t_k, I is the indicator function given in (1) and $n_1 + \cdots + n_L + r_1 + \cdots + r_L = n$. The generalized Dirichlet distribution is clearly seen to be a conjugate prior.

THEOREM 1. *Let $\Theta_1, \ldots, \Theta_n$, with each $\Theta_i \in \{t_k: k \geq 1\}$, be an iid sample, possibly with right censoring, with an unknown cdf F on $[0, \infty)$. If F is a discrete time beta-Stacy process with parameters $\{\alpha_k, \beta_k\}$ and jumps at $\{t_k\}$, then, given $\Theta_1, \ldots, \Theta_n$, the posterior distribution for F is also a discrete time beta-Stacy process with jumps at $\{t_k\}$ but with parameters $\{\alpha_k^*, \beta_k^*\}$, where*

$$\alpha_k^* = \alpha_k + n_k \quad \textit{and} \quad \beta_k^* = \beta_k + m_k, \tag{12}$$

n_k is the number of exact observations at t_k 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, $m_k = \sum_{j>k} n_j + \sum_{j \geq k} r_j$.

PROOF. The proof is immediate on combining the likelihood function with the generalized Dirichlet prior. □

The posterior expectation of $[F(t_k) - F(t_{k-1})]$, which is the predictive probability $p(\Theta_{n+1} = t_k | \Theta_1, \ldots, \Theta_n)$, is given by

$$E(Y_k | \Theta_1, \ldots, \Theta_n) = \frac{\alpha_k^*}{\alpha_k^* + \beta_k^*} \prod_{j=1}^{k-1} \frac{\beta_j^*}{\alpha_j^* + \beta_j^*}. \tag{13}$$

Expression (13) is fundamental to the results obtained in Section 5.

4. Continuous time process. In this section we consider the Lévy process from which the beta-Stacy process is derived and constructed. The posterior processes are also obtained.

4.1. *The construction.* For the continuous time process the search is on for a stochastic process $F(t)$, defined on $([0, \infty), \mathscr{A})$, such that, infinitesimally speaking,

$$dF(t) | F(t) \sim \mathscr{C}(d\alpha(t), \beta(t), 1 - F(t)), \tag{14}$$

with $F(0) = 0$, and, with probability 1, $F \in \mathscr{F}$. Note that if $\alpha(\cdot)$ has a fixed point of discontinuity at t, then

$$F\{t\} | F(t-) \sim \mathscr{C}(\alpha\{t\}, \beta(t), 1 - F(t-)).$$

It is shown in this section that such a process exists and is obtained from a log-beta process. First it is shown that the (Lévy) log-beta process exists, with representation given by (5) and (6) and with parameters given by $d\alpha(t) = c(t)\,dG(t)$ and $\beta(t) = c(t)G[t, \infty)$.

THEOREM 2. *Let $G \in \mathscr{F}$ be continuous and let $c(\cdot)$ be a piecewise continuous positive function. There exists a Lévy process Z with Lévy representation given by*

$$\log E \exp\{-\phi Z(t)\} = \int_0^\infty (\exp(-v\phi) - 1)\, dN_t(v),$$

where

$$dN_t(v) = \frac{dv}{(1-\exp(-v))} \int_0^t \exp(-vc(s)G[s,\infty))c(s)\, dG(s).$$

PROOF. The proof is along lines similar to the proof of Theorem 3.1 in Hjort (1990). Let, for $k = 1, 2, \dots,$ $\alpha_{n,k} = c_{n,k}G[(k-1)/n, k/n)$ and $\beta_{n,k} = c_{n,k}G[k/n, \infty)$, where $c_{n,k} = c((k-1/2)/n)$. Then define the independent random variables $V_{n,k} \sim \mathscr{B}(\alpha_{n,k}, \beta_{n,k})$ and the corresponding $W_{n,k} = -\log(1 - V_{n,k})$. Let

$$Z_n(0) = 0 \quad \text{and} \quad Z_n(t) = \sum_{k/n \le t} W_{n,k} \quad \text{for } t \ge 0.$$

The first aim is to show that $\{Z_n\}$ converges in distribution for each Borel set D in $(0, \infty)$ to a Lévy process Z with the required representation. Now $Z_n(t) = -\sum_{k/n\le t} \log(1 - V_{n,k})$ so

$$\log E \exp\{-\phi Z_n(t)\} = \sum_{k/n\le t} \log \frac{\Gamma(\alpha_{n,k} + \beta_{n,k})\Gamma(\beta_{n,k} + \phi)}{\Gamma(\beta_{n,k})\Gamma(\alpha_{n,k} + \beta_{n,k} + \phi)},$$

where $\Gamma(\cdot)$ denotes the gamma function. Using a result contained in the proof of Lemma 1 from Ferguson (1974),

$$\begin{aligned}
&\log E \exp\{-\phi Z_n(t)\} \\
&\quad = \sum_{k/n\le t} \sum_{l=0}^{\infty} \log \frac{(\alpha_{n,k} + \beta_{n,k} + \phi + l)(\beta_{n,k} + l)}{(\alpha_{n,k} + \beta_{n,k} + l)(\beta_{n,k} + \phi + l)} \\
&\quad = \sum_{k/n\le t} \int_0^\infty (\exp(-v\phi) - 1) \frac{\exp(-\beta_{n,k}v)(1 - \exp(-\alpha_{n,k}v))}{v(1-\exp(-v))}\, dv \\
&\quad = \int_0^\infty \frac{\exp(-v\phi) - 1}{v(1-\exp(-v))} \sum_{k/n\le t} \exp(-\beta_{n,k}v)(1 - \exp(-\alpha_{n,k}v))\, dv.
\end{aligned}$$

It can now be seen that

$$\sum_{k/n\le t} \exp(-\beta_{n,k}v)(1-\exp(-\alpha_{n,k}v)) \to v \int_0^t \exp(-vc(s)G[s,\infty))c(s)\, dG(s),$$

which implies that

$$\begin{aligned}
&\log E \exp\{-\phi Z_n(t)\} \\
&\quad \to \int_0^\infty \frac{\exp(-v\phi) - 1}{1-\exp(-v)} \int_0^t \exp(-vc(s)G[s,\infty))c(s)\, dG(s)\, dv.
\end{aligned}$$

By a similar argument it can also be shown, for any $0 = t_0 < t_1 < \cdots < t_k < \infty$, that

$$\log E \exp\biggl(-\sum_{j=1}^{k} \phi_j Z_n[t_{j-1}, t_j)\biggr) \to \sum_{j=1}^{k} \int_0^\infty (\exp(-v\phi_j) - 1)\, dN_{[t_{j-1}, t_j)}(v),$$

which ensures that all the finite-dimensional distributions of $\{Z_n\}$ converge as they should. It also needs to be shown that the sequence $\{Z_n\}$ is *tight*. If

$$A_n(0) = 0 \quad \text{and} \quad A_n(t) = \sum_{k/n \le t} V_{n,k} \quad \text{for } t \ge 0,$$

then $\{A_n\}$ converges in distribution to a (Lévy) beta process A and, from Hjort (1990), $\{A_n\}$ is tight. Since $dZ_n = -\log(1 - dA_n)$ it follows straightforwardly that $\{Z_n\}$ is also tight. This completes the proof. □

COROLLARY 1. *If* $F(t) = 1 - \exp(-Z(t))$, *then*, *infinitesimally speaking*,

$$dF(t) | F(t) \sim \mathscr{C}(d\alpha(t), \beta(t), 1 - F(t))$$

and, *with probability* 1, $F \in \mathscr{F}$.

PROOF. Consider the dependent random variables $\{Y_{n,k}\}$ given by

$$\begin{aligned} Y_{n,1} &= V_{n,1}, \\ Y_{n,2} &= V_{n,2}(1 - V_{n,1}), \\ &\ \vdots \\ Y_{n,k} &= V_{n,k} \prod_{j=1}^{k-1} (1 - V_{n,j}), \end{aligned}$$

with the $\{V_{n,k}\}$ defined as in Theorem 2, so that, from Lemma 4, for any $m > 1$,

$$\mathscr{L}(Y_{n,1}, \dots, Y_{n,m}) = \mathscr{G}(\alpha_{n,1}, \beta_{n,1}, \dots, \alpha_{n,m}, \beta_{n,m}),$$

and also

$$-\log(1 - F_n(t)) = -\sum_{k/n \le t} \log(1 - V_{n,k}) = \sum_{k/n \le t} W_{n,k} = Z_n(t),$$

where

$$F_n(0) = 0 \quad \text{and} \quad F_n(t) = \sum_{k/n \le t} Y_{n,k} \quad \text{for } t \ge 0,$$

implying that $\{F_n\}$ is a discrete time beta-Stacy process and $F_n(t) = 1 - \exp(-Z_n(t))$. Clearly $\{F_n\}$ satisfies, for all $k \ge 1$,

$$\begin{aligned} &\mathscr{L}\bigl(F_n[(k-1)/n, k/n) \big| F_n[0, (k-1)/n)\bigr) \\ &\qquad = \mathscr{C}\bigl(\alpha_{n,k}, \beta_{n,k}, 1 - F_n[0, (k-1)/n)\bigr). \end{aligned}$$

Also $\{F_n\}$ converges in distribution to F. That $F \in \mathscr{F}$ a.s. follows from $\int_0^\infty dG(s)/G[s,\infty) = \infty$, completing the proof. □

4.2. *Posterior distributions.* The main aim in this section will be to find the posterior distribution of F given a set of possibly right censored observations. Posterior representations are typically achieved through the posterior Lévy measures [Ferguson (1974), Ferguson and Phadia (1979)]. Here we give the result for the posterior of the more general neutral to the right process.

Let $F(t) = 1 - \exp(-Z(t))$ be a neutral to the right process. The prior distribution for $Z(t)$ is characterized by

$$M = \{t_1, t_2, \dots\}, \{f_{t_1}, f_{t_2}, \dots\},$$

the set of fixed points of discontinuity with corresponding densities for the jumps, and $N_t(\cdot)$, the Lévy measure for the part of the process without fixed points of discontinuity. We now give the posterior characterization given a single observation Θ (the case for n observations can then be obtained by repeated application). In the following we assume the Lévy measure to be of the type $dN_t(v) = dv \int_{(0,t]} K(v,s)\,ds$ [the beta-Stacy process with parameters $\alpha(\cdot)$ and $\beta(\cdot)$ arises when $K(v,s)\,ds = (1-\exp(-v))^{-1}\exp\{-v(\beta(s) + \alpha\{s\})\}\,d\alpha(s)$].

THEOREM 3 [Ferguson (1974), Ferguson and Phadia (1979)]. *Let F be neutral to the right and let Θ be a random sample from F.*

(i) *Given $\Theta > \theta$ the posterior parameters (denoted by an asterisk) are $M^* = M$,*

$$f^*_{t_j}(v) = \begin{cases} \kappa e^{-v} f_{t_j}(v), & \text{if } t_j \le \theta, \\ f_{t_j}(v), & \text{if } t_j > \theta, \end{cases}$$

and $K^*(v,s) = \exp\{-vI(\theta \ge s)\}K(v,s)$ (here κ is the normalizing constant).

(ii) *Given $\Theta = \theta \in M$ the posterior parameters are $M^* = M$,*

$$f^*_{t_j}(v) = \begin{cases} \kappa e^{-v} f_{t_j}(v), & \text{if } t_j < \theta, \\ \kappa(1-e^{-v}) f_{t_j}(v), & \text{if } t_j = \theta, \\ f_{t_j}(v), & \text{if } t_j > \theta, \end{cases}$$

and, again, $K^*(v,s) = \exp\{-vI(\theta \ge s)\}K(v,s)$.

(iii) *Given $\Theta = \theta \notin M$ the posterior parameters are $M^* = M \cup \{\theta\}$, with $f_\theta(v) = \kappa.(1-e^{-v})K(v,\theta)$,*

$$f^*_{t_j}(v) = \begin{cases} \kappa e^{-v} f_{t_j}(v), & \text{if } t_j < \theta, \\ f_{t_j}(v), & \text{if } t_j > \theta, \end{cases}$$

and, again, $K^*(v,s) = \exp\{-vI(\theta \ge s)\}K(v,s)$.

COROLLARY 2. *Let Z be a log-beta process, with parameters $\alpha(\cdot)$ and $\beta(\cdot)$ satisfying* (7), *and let $F(t) = 1 - \exp(-Z(t))$. For an observation Θ from F*

with $\Theta=\theta$ or $\Theta>\theta$ the posterior Lévy measure for $Z(t)$ is given by

$$dN_t(v|\theta)=\frac{dv}{(1-\exp(-v))}\int_0^t \exp\big(-v\{\beta(s)+\alpha\{s\}+I(s\le\theta)\}\big)\,d\alpha_c(s). \tag{15}$$

Additionally, for the observation $\Theta=\theta$, there is a fixed point of discontinuity at $t=\theta$ in the posterior Lévy process with jump J_θ, where $1-\exp(-J_\theta)\sim\mathscr{B}(\alpha\{\theta\}+1,\beta(\theta))$.

PROOF. The proof is immediate from Theorem 3. □

COROLLARY 3. *Given a sample of size n, $(\Theta_1,\dots,\Theta_n)$, then $Z|data$ is a log-beta process with parameters $\alpha^*(t)=\alpha(t)+N(t)$ and $\beta^*(t)=\beta(t)+M(t)-N\{t\}$, where $N(\cdot)$ is the counting process for uncensored observations and $M(t)=\sum_i I(\Theta_i\ge t)$.*

PROOF. The posterior Lévy measure

$$dN_t(v|\text{data})=\frac{dv}{(1-\exp(-v))}\int_0^t \exp\big(-v\{\beta(s)+\alpha\{s\}+M(s)\}\big)\,d\alpha_c(s) \tag{16}$$

follows from repeated application of Theorem 3. For the jump J, corresponding to exact observation(s) at θ, it needs to be shown that $J=-\log(1-S)$ in distribution with $S\sim\mathscr{B}(\alpha^*\{\theta\},\beta^*(\theta))$. From Theorem 3 it is easily seen that the density function for J is given by

$$f_J(x)\propto\exp\big(-x(\beta(\theta)+M(\theta)-N\{\theta\})\big)(1-\exp(-x))^{\alpha\{\theta\}+N\{\theta\}-1}. \tag{17}$$

Now it is possible to show that $J=-\log(1-S)$ in distribution with $S\sim\mathscr{B}(\alpha\{\theta\}+N\{\theta\},\beta(\theta)+M(\theta)-N\{\theta\})$ and this completes the proof. □

Let the location of the exact observations be the set $\{t_i\}$ with corresponding jumps $\{J_i\}$. Then, given the data,

$$\begin{aligned}\log E\exp\{-\phi Z(t)\}&=\sum_{t_i\le t}\log E\exp(-\phi J_i)\\&\quad+\int_0^\infty(\exp(-\phi v)-1)\,dN_t(v|\text{data}),\end{aligned} \tag{18}$$

where

$$1-\exp(-J_i)\sim\mathscr{B}\big(\alpha\{t_i\}+N\{t_i\},\beta(t_i)+M(t_i)-N\{t_i\}\big). \tag{19}$$

Typically it would be the case that $N\{t_i\}=1$ and $\alpha\{t_i\}=0$, which gives $J_i\sim\mathscr{E}(\beta(t_i)+M(t_i)-1)$, where $\mathscr{E}(\beta)$ is the exponential distribution with mean $1/\beta$.

The next theorem gives the parameters of the posterior process, which is also a beta-Stacy process, given possibly right censored observations.

THEOREM 4. *Let $\Theta_1,\dots,\Theta_n$ be an iid sample, possibly with right censoring, from an unknown cdf F on $[0,\infty)$ and $F\sim\mathscr{S}(c(\cdot),G)$. Then, given the data,*

F is a beta-Stacy process with parameters $c^*(\cdot)$ and G^*, where

$$G^*(t) = 1 - \prod_{[0,\,t]} \left\{ 1 - \frac{c(s)\,dG(s) + dN(s)}{c(s)G[s,\infty) + M(s)} \right\},$$

$$c^*(t) = \frac{c(t)G[t,\infty) + M(t) - N\{t\}}{G^*[t,\infty)}.$$

PROOF. That $F|$data is a beta-Stacy process follows from $Z|$data being a log-beta process. For the parameters of the posterior beta-Stacy process, using (8) and (9),

$$G^*(t) = 1 - \prod_{[0,\,t]} \left\{ 1 - d\alpha^*(s)/(\beta^*(s) + \alpha^*\{s\}) \right\}$$

and $c^*(t) = \beta^*(t)/G^*[t,\infty)$, where $\alpha^*(t) = \alpha(t) + N(t)$ and $\beta^*(t) = \beta(t) + M(t) - N\{t\}$, which completes the proof. □

The result generalizes the work of Susarla and Van Ryzin (1976), who take a Dirichlet process prior for F. This follows since if F is from a Dirichlet process then $F|$data is from a beta-Stacy process if the data includes right censored observations.

4.3. *Bayes estimates and simulation.* Here the Bayes estimate for F is given. If $F \sim \mathscr{S}(c(\cdot), G)$, then, given an iid sample from F, with possible right censoring, the Bayes estimate of $F(t)$, with quadratic loss function, is given by

$$\hat{F}(t) = E\big(F(t)|\text{data}\big) = 1 - \prod_{[0,\,t]} \left\{ 1 - \frac{c(s)\,dG(s) + dN(s)}{c(s)G[s,\infty) + M(s)} \right\}. \tag{20}$$

This is the same nonparametric estimate of F as that obtained from the beta process of Hjort (as was expected from Remark 2). The asymptotic distribution of $\hat{F}$ is discussed by Hjort [(1990), page 1286]. The standard nonparametric estimate of Kaplan and Meier is obtained as $c(\cdot) \to 0$. The beta-Stacy process can, however, be seen as more of a natural generalization of the Dirichlet process than is the beta process, since the attention is focused on the cdf.

A full Bayesian analysis is available by sampling from the posterior distribution of $Z(\Delta)$, where Δ represents an interval $[t, t+h)$ for $t, h > 0$. Let $Z_c(\Delta) = Z(\Delta) - J(\Delta)$, where $J(\Delta)$ is the total jump from the fixed points of discontinuity in Δ, then

$$\log E \exp\big\{ -\phi Z_c(\Delta) \big\} = \int_0^\infty (\exp(-z\phi) - 1)\, dN_\Delta(z|\text{data}).$$

Damien, Laud and Smith (1995) describe an algorithm which can be adapted to obtain a sample from $Z_c(\Delta)$: take X_l: $l = 1, \dots, L$, iid from the density proportional to $(1 - e^{-v})\, dN_\Delta(v|\text{data})$. Let $\gamma = \int_0^\infty (1 - e^{-v})\, dN_\Delta(v|\text{data})$ and take, for $l = 1, \dots, L$,

$$Q_l \sim \mathscr{P}\big(\gamma/[L\{1 - \exp(-X_l)\}]\big),$$

where $\mathscr{P}$ represents the Poisson distribution. Finally put $Z_c^L = \sum_{l=1}^{L} X_l Q_l$ with the result that $Z_c^L \to Z_c(\Delta)$ in distribution, which can be proven by showing that $\log E\exp(-\phi Z_c^L) \to \log E\exp(-\phi Z_c(\Delta))$.

4.4. *Numerical example.* We use the beta-Stacy to model survival data. A numerical example is taken from the paper of Ferguson and Phadia (1979). The data consist of exact observations at 0.8, 3.1, 5.4 and 9.2 months and right censored data at 1.0, 2.7, 7.0 and 12.1 months. We follow Ferguson and Phadia (1979) and take $\beta(s) = \exp(-0.1s)$ and $d\alpha(s) = 0.1\exp(-0.1s)\,ds$. The prior therefore is a Dirichlet process but, with the inclusion of censored observations within the data set, the posterior process is beta-Stacy. We obtain samples from the posterior distribution of $F(0,1)$. This will involve sampling $Z_c^*(0,0.8)$, $Z_c^*[0.8,1)$, which we do using the algorithm of Damien, Laud and Smith (1995), and sampling $J_{0.8}$, the jump at 0.8, which has an exponential distribution with mean $1/(\exp(-0.08)+7)$. A required sample from the posterior distribution of $F(0,1)$ is then given by $1-\exp(-Z_c^*(0,0.8)-Z_c^*[0.8,1)-J_{0.8})$.

Here we briefly consider the sampling of $Z_c^*(0,0.8)$ using the algorithm of Damien, Laud and Smith (1995). This will involve sampling from the density given up to a constant of proportionality by

$$f(v) \propto \exp(-8v)\int_0^{0.8} \exp(-v\exp(-0.1s))\exp(-0.1s)\,ds.$$

We define the joint density of v and w, a latent variable defined on $(0,1)$, by

$$f(v,w) \propto \exp(-8v - vw)I(w \in (\exp(-0.08),1)).$$

Clearly the marginal distribution for v is as required. We can now use a Gibbs sampling scheme [Smith and Roberts (1993)] to generate iid samples (using L chains) from the required density. Straightforward calculations give $\gamma = \log 9 - \log(8+\exp(-0.08))$. We collected 1000 samples from the posterior distribution of $F(0,1)$ and the resulting histogram estimate is given in Figure 1.

5. A generalization of the Pólya-urn scheme. It is well known that the Pólya-urn sampling scheme (on a finite sample space) leads to an exchangeable sequence of variables. From the representation theorem of de Finetti this guarantees the existence of a random probability measure F conditional on which the variables of the sequence are iid with distribution F. These random probability measures can be shown to be from a Dirichlet process (with finite support).

A description of the Pólya-urn sampling scheme and its connections with the Dirichlet process is given in Blackwell and MacQueen [(1973), page 354]. It is reviewed and generalized here. Let $\theta_1,\dots,\theta_m$ represent m different colors in an urn. Let c be a positive number. On the first draw a ball of color j is drawn with probability q_j. The color drawn is replaced in the urn and a ball of its color is added. On the second draw color j is now drawn with probability

$$p_1(j|j_1) = \{cq_j + I(j=j_1)\}/(c+1),$$

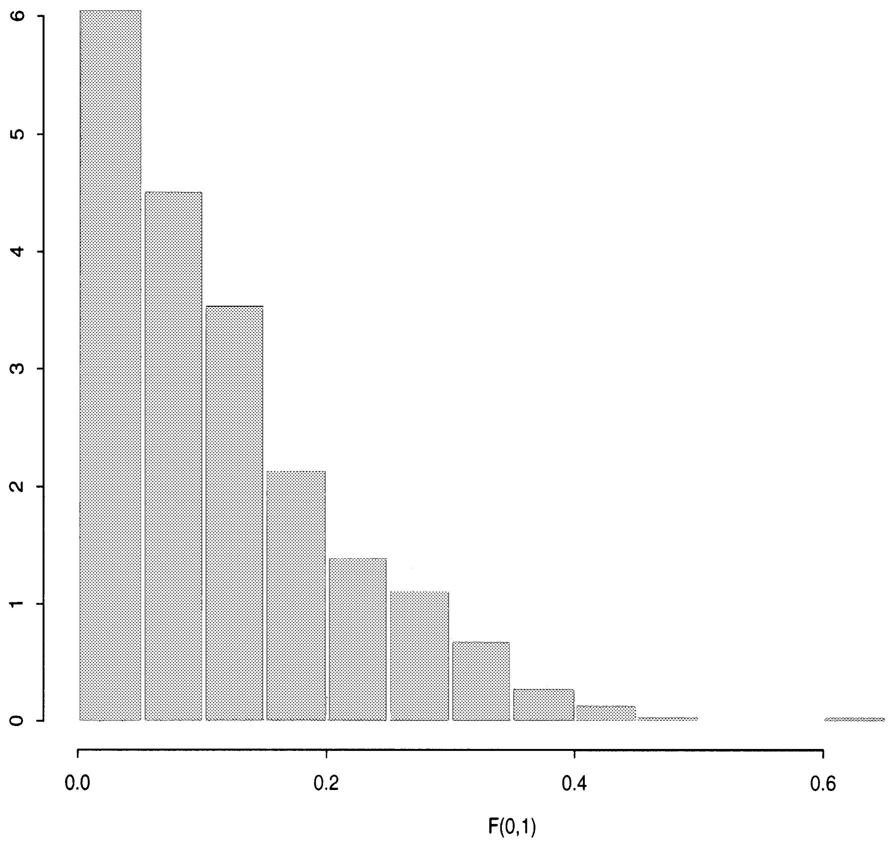

FIG. 1. *Histogram estimate of posterior density of* $F(0, 1)$.

where j_1 was the first color drawn. The second color drawn is replaced and a ball of its color added to the urn. In general at the $(n+1)$th draw the color j is taken with probability $p_n(j|j_1, \dots, j_n)$, where

$$p_n(j|j_1, \dots, j_n) = \frac{\{cq_j + \sum_{i=1}^n I(j = j_i)\}}{c+n}. \tag{21}$$

The connection with the Dirichlet process is that if F has finite support on $\{\theta_j\colon j = 1, \dots, m\}$ and is from a Dirichlet process, with precision parameter c and discrete location parameter G given by weights q_j at θ_j, then

$$p_n(j|j_1, \dots, j_n) = E\big(F(\theta_j)|j_1, \dots, j_n\big),$$

where now $F(\theta)$ represents the random weight assigned to θ. Let $\tilde{\alpha}_j = cq_j$, for $j = 1, \dots, m$, and $\tilde{\beta}_1 = c(1 - q_1)$ so that

$$p_n(j|j_1, \dots, j_n) = \frac{\tilde{\alpha}_j + n_j}{\tilde{\alpha}_1 + \tilde{\beta}_1 + n}, \tag{22}$$

where $n_j = \sum_{i=1}^n I(j = j_i)$.

The beta-Stacy process suggets the existence of a more general scheme for generating an exchangeable sequence from m different colors. This generalized scheme allows a generalization of the finite population Bayesian bootstrap [Lo (1988)], which uses the Pólya-urn scheme, to include censored observations [Muliere and Walker (1995)].

Let F be from a finite discrete beta-Stacy process with parameters given by $\{\alpha_j, \beta_j \colon j = 1, \dots, m-1\}$ and support on $\{\theta_1, \dots, \theta_m\}$. Then, from (13), for any $n \ge 1$ and $j = 1, \dots, m-1$,

$$E\big(F(\theta_j) | j_1, \dots, j_n\big) = \frac{\alpha_j + n_j}{\alpha_j + \beta_j + n_j + m_j} \prod_{l=1}^{j-1} \frac{\beta_l + m_l}{\alpha_l + \beta_l + n_l + m_l}, \tag{23}$$

where $m_j = \sum_{i=1}^n I(j < j_i)$. Note that $n_1 + m_1 = n$, $n_j + m_j = m_{j-1}$ and that if $\alpha_j + \beta_j = \beta_{j-1}$, for all $j = 2, \dots, m-1$, then (23) reduces to (22).

A generalization of Pólya's urn scheme is now given which leads to a $p_n(j | j_1, \dots, j_n)$ given by (23). Consider m Pólya-urns: the first urn has the different colors $\theta_1, \dots, \theta_m$ and the parameters of the urn are $c_1 > 0$ and $q_1, \dots, q_m$. The second urn has the colors $\theta_2, \dots, \theta_m$ and parameters $c_2 > 0$ and $q_2/(1-q_1), \dots, q_m/(1-q_1)$. The third urn has the colors $\theta_3, \dots, \theta_m$ and parameters $c_3 > 0$ and $q_3/(1-q_1-q_2), \dots, q_m/(1-q_1-q_2)$. Continue in this fashion up to the mth urn, which only has the color θ_m. The scheme is now described:

1. Start at urn $j = 1$.
2. Sample urn j once according to Pólya's urn scheme.
3. If the color sampled is θ_j then go to 4 else $j = j + 1$ and go to 2.
4. θ_j is a single sample from the generalized Pólya-urn scheme.

THEOREM 5. *The sequence of variables from the generalized Pólya-urn scheme are exchangeable.*

PROOF. We first note that if

$$\mathscr{L}(y_1, \dots, y_{m-1}) = \mathscr{G}\big(\alpha_1^*, \beta_1^*, \dots, \alpha_{m-1}^*, \beta_{m-1}^*\big),$$

where $\alpha_j^* = \alpha_j + n_j$ and $\beta_j^* = \beta_j + m_j$, where $m_j = \sum_{k=j+1}^m n_k$, then the density function for $y_1, \dots, y_{m-1}$ is given, up to a constant of proportionality, by

$$y_1^{n_1} \cdots y_{m-1}^{n_{m-1}} (1 - y_1 - \cdots - y_{m-1})^{n_m} p(y_1, \dots, y_{m-1}),$$

where $p(y_1, \dots, y_{m-1})$ is given by (1). Therefore $E(x_1^{n_1} \cdots x_j^{n_j} \cdots x_m^{n_m}) = C(\alpha^*, \beta^*)$, where $C(\alpha, \beta)$ is the normalizing constant for (1), given by

$$C(\alpha, \beta) = \prod_{j=1}^{m-1} \Gamma(\alpha_j)\Gamma(\beta_j)/\Gamma(\alpha_j + \beta_j).$$

Therefore if

$$\mathscr{L}(x_1, \dots, x_{m-1}) = \mathscr{G}(\alpha_1, \beta_1, \dots, \alpha_{m-1}, \beta_{m-1}),$$

$x_m = 1 - x_1 - \cdots - x_{m-1}$ and $n_1 + \cdots + n_m = n$, then

$$\frac{E(x_1^{n_1} \cdots x_j^{n_j+1} \cdots x_m^{n_m})}{E(x_1^{n_1} \cdots x_j^{n_j} \cdots x_m^{n_m})} = \frac{\alpha_j + n_j}{\alpha_j + \beta_j + n_j + m_j} \prod_{l=1}^{j-1} \frac{\beta_l + m_l}{\alpha_l + \beta_l + n_l + m_l}.$$

From the generalized Pólya-urn scheme the probability of taking the color j at the $(n+1)$th iteration is given by

$$p_n(j|j_1, \ldots, j_n) = \frac{c_j q_j^* + n_j}{c_j + n_j + m_j} \prod_{l=1}^{j-1} \left(1 - \frac{c_l q_l^* + n_l}{c_l + n_l + m_l}\right), \tag{24}$$

which implies

$$p_n(j|j_1, \ldots, j_n) = \frac{E(x_1^{n_1} \cdots x_j^{n_j+1} \cdots x_m^{n_m})}{E(x_1^{n_1} \cdots x_j^{n_j} \cdots x_m^{n_m})},$$

where $n_j = \sum_{i=1}^n I(j = j_i)$, $\alpha_j = c_j q_j^*$, $\beta_j = c_j(1 - q_j^*)$ and $q_j^* = q_j/(1 - q_1 - \cdots - q_{j-1})$. It is now easy to see that a representation of the distribution of $(j_1, \ldots, j_n)$, for any $n = 1, 2, \ldots,$ is given by

$$p(j_1, \ldots, j_n | x_1, \ldots, x_{m-1}) = x_1^{n_1} \cdots x_m^{n_m}$$

and

$$\mathscr{L}(x_1, \ldots, x_{m-1}) = \mathscr{G}(\alpha_1, \beta_1, \ldots, \alpha_{m-1}, \beta_{m-1}),$$

completing the proof. □

The usual Pólya-urn scheme, with parameters c and $q_1, \ldots, q_m$ arises if the condition $\beta_{j-1} = \beta_j + \alpha_j$ holds for all $j = 2, \ldots, m-1$, that is, if $c_{j-1}(1 - q_{j-1}^*) = c_j$ and $c_1 = c$. This is straightforward to see by noting that the conditions $c_j q_j^* = c_1 q_j$ and $m_{j-1} = n_j + m_j$ hold for all $j = 1, \ldots, m-1$, where $m_0 = n$. In fact for the usual Pólya-urn scheme

$$\mathscr{L}(n_1, \ldots, n_m) = \mathscr{M}(x_1, \ldots, x_m; n),$$

where $\mathscr{M}$ denotes the multinomial distribution, and

$$\mathscr{L}(x_1, \ldots, x_{m-1}) = \mathscr{D}(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_m).$$

For the generalized Pólya-urn scheme the Dirichlet distribution is replaced by the generalized Dirichlet distribution.

A further point to consider is replacing each of the Pólya-urns in the generalized case with generalized Pólya-urns themselves. This could obviously be carried on ad infinitum.

Acknowledgments. The work was completed while the second author was visiting at Imperial College. The authors are grateful to A. F. M. Smith and P. Secchi for comments, and to a referee and Associate Editor for helpful suggestions and critical comments which have greatly improved the paper.

REFERENCES

BLACKWELL, D. and MACQUEEN, J. B. (1973). Ferguson distributions via Pólya-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.

DAMIEN, P., LAUD, P. and SMITH, A. F. M. (1995). Random variate generation approximating infinitely divisible distributions with application to Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **57** 547–564.

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.

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.

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.

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

LAVINE, M. (1992). Some aspects of Pólya tree distributions for statistical modelling. *Ann. Statist.* **20** 1222–1235.

LAVINE, M. (1994). More aspects of Pólya trees for statistical modelling. *Ann. Statist.* **22** 1161–1176.

LÉVY, P. (1936). *Théorie de l'Addition des Variables Aléatoire.* Gauthiers-Villars, Paris.

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

MAULDIN, R. D., SUDDERTH, W. D. and WILLIAMS, S. C. (1992). Pólya 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 WALKER, S. G. (1995). Extending the family of Bayesian bootstraps and exchangeable urn schemes. *J. Roy. Statist. Soc. Ser. B*. To appear.

SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* **55** 3–23.

SUSARLA, V. and VAN RYZIN, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete data. *J. Amer. Statist. Assoc.* **71** 897–902.

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE
180 QUEEN'S GATE
LONDON, SW7 2AZ
UNITED KINGDOM
E-MAIL: s.walker01@ic.ac.uk

DIPARTIMENTO DI ECONOMIA
POLITICA E MQ
UNIVERSITA DI PAVIA
S. FELICE 5, 27100
PAVIA
ITALY