The Annals of Statistics
1999, Vol. 27, No. 2, 589–599

A CHARACTERIZATION OF A NEUTRAL TO THE RIGHT PRIOR VIA AN EXTENSION OF JOHNSON'S SUFFICIENTNESS POSTULATE

BY STEPHEN WALKER[1] AND PIETRO MULIERE

Imperial College and Università degli Studi di Pavia

In this paper we present a new characterization and perspective on a neutral to the right prior. This characterization is based on a sequence of predictive laws which provides explicitly the posterior parameters and Bayes estimators for such a prior.

1. Introduction. Let $X_1, X_2, \ldots$ be an exchangeable sequence of random variables defined on $(0, \infty)$. From de Finetti's representation theorem [de Finetti (1937)] there exists a random distribution F conditional on which $X_1, X_2, \ldots$ are iid from F. That is, there exists a probability (or de Finetti) measure, defined on the space of probability measures on $(0, \infty)$, such that the joint distribution of $X_1, \ldots, X_n$, for any n, can be written as

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

where μ is the de Finetti (or prior) measure. The problem is how to select the prior. One approach is to select μ by appealing to prior information about F and attempting to incorporate this information into μ. This is often a difficult task for nonparametric priors. Alternatively, we may assume the sequence of predictive laws, $\mathscr{L}(X_{n+1} \mid X_1, \ldots, X_n)$, obeys or exhibits some characteristic or property. In practical applications it may be that the form of the predictive law may be an adequate description of our state of knowledge. We will consider an example.

EXAMPLE 1. In the 1920s the English philosopher W. E. Johnson discovered a characterization of the Dirichlet distribution and process [Zabell (1982)]. This was arrived at via a form of predictive, on discrete cells, given by

$$P(X_{n+1} = k \mid X_1, \ldots, X_n) = f_k(n_k);$$

that is, the conditional probability of an outcome in cell k only depends on the number of previous outcomes in that cell. This form of predictive is a natural way of thinking nonparametrically.

More recent characterizations in the continuous framework are provided by Regazzini (1978) and Lo (1991); let $X_1, \ldots, X_n$ be an exchangeable sequence

Received December 1996; revised December 1998.
[1]Supported by an EPSRC ROPA grant.
AMS 1991 *subject classifications.* Primary 62A15; secondary 62M20.
Key words and phrases. Exchangeability, neutral to the right process, prediction.

defined on some space Ω and assume, for every $n \geq 1$ and set A,

$$P(X_{n+1} \in A \mid X_1, \dots, X_n) = \frac{\alpha(A) + \sum_{i=1}^n \delta_{X_i}(A)}{\alpha(\Omega) + n},$$

where α is a finite measure on Ω. The prevision is given by a mixture of the empirical measure and the prior measure α. Regazzini (1978) and Lo (1991) prove this prevision is a characteristic property of the Dirichlet process prior. A similar characterization has been given for Pólya trees by Walker and Muliere (1997b).

In the present note, Johnson's result is extended to the case of a neutral to the right exchangeable sequence. We show that if $X_1, X_2, \dots$ is a sequence of random variables, with each X_i defined on $(0, \infty)$, such that

$$P(X_{n+1} > t \mid X_1, \dots, X_n) = \prod_0^t [1 - d\Lambda(s, n(s), m(s))],$$

where $n(s) = \sum_i I(X_i = s)$, $m(s) = \sum_i I(X_i > s)$,

$$d\Lambda(s, n, m)[1 - d\Lambda(s, n+1, m)] = d\Lambda(s, n, m+1)[1 - d\Lambda(s, n, m)],$$

for all $s > 0$ and nonnegative integers n, m, and $\prod_0^t$ represents a product integral, then the sequence is exchangeable with de Finetti measure or prior a *neutral to the right* process. Note that here we are counting $n(\cdot)$ on the hazard $d\Lambda(\cdot)$ rather than on the density and including $m(\cdot)$. This would appear to be appropriate for survival models where there is often censored data and so $n(\cdot)$ will not adequately capture all the information on its own. Hence the form of the predictive is intuitive for modelling survival data. The use of product integrals is now well established within nonparametric survival analysis [Gill and Johansen (1990), Andersen, Borgan, Gill and Keiding (1993)].

An important consequence of our characterization is that we are able to obtain Bayesian nonparametric estimators of a survival function without recourse to Lévy theory. For example, the estimator derived from the beta-Stacy process [Walker and Muliere (1997a)] arises when

$$d\Lambda(s, n(s), m(s)) = (d\alpha(s) + n(s))/(\beta(s) + n(s) + m(s))$$

for suitable α and β. Here α and β are the parameters of the neutral to the right prior and therefore our characterization provides immediately the mechanism for the updating of the parameters in the light of data. Interpretation is provided by the fact that $d\alpha(s)/\beta(s)$ represents the prior hazard rate function. This ease of updating and interpretation is not a feature of alternative representations of a neutral to the right prior [Doksum (1974); Ferguson (1974)].

Characterizations of a *neutral to the right* process and connections between neutrality to the right and general concepts of neutrality and tail-freeness are discussed in Doksum (1974). Doksum (1974) and later Ferguson (1974) define neutral to the right processes in terms of Lévy processes [Lévy (1936)], which turns out to be the most convenient definition. Ferguson and Phadia (1979) use such theory to obtain Bayesian nonparametric estimators for survival functions, generalizing the work of Susarla and van Ryzin (1976) who focused their attention on a particular neutral to the right process, the well-known Dirichlet process [Ferguson (1973)]. Walker and Muliere (1997a) considered a more general neutral to the right prior, the beta-Stacy process, which is particularly suitable for the Bayesian nonparametric analysis of censored survival times.

The beta-Stacy process is derivable from the beta process of Hjort (1990) and the beta-neutral process of Lo (1993). All three are defined via a Lévy process of some kind: the beta-Stacy on a log-beta process; the beta obviously on a beta process and the beta-neutral process is constructed from two independent gamma processes. Doksum's original definition of a neutral to the right prior does not use Lévy theory but does not shed light on how to update the prior. In fact, the update presented in Doksum (1974) is complicated.

DEFINITION 1 [Doksum (1974)]. A random distribution function $F(t)$ on $(0, \infty)$ is said to be neutral to the right if for every m and $0 < t_1 < t_2 < \cdots < t_m$, there exist independent random variables $V_1, V_2, \dots, V_m$, such that $(1 - F(t_1), 1 - F(t_2), \dots, 1 - F(t_m))$ has the same distribution as $(V_1, V_1 V_2, \dots, \prod_1^m V_i)$.

If F is neutral to the right then $Z(t) = -\log[1 - F(t)]$ has independent increments and this provides an alternative characterization of a neutral to the right prior in terms of a Lévy process.

DEFINITION 2 [Doksum (1974)]. Let $Z(t)$ be a Lévy process such that:

(i) $Z(t)$ has nonnegative independent increments;
(ii) $Z(t)$ is nondecreasing a.s.;
(iii) $Z(t)$ is right continuous a.s.;
(iv) $Z(t) \to \infty$ a.s. as $t \to \infty$;
(v) $Z(0) = 0$ a.s.

A neutral to the right process is defined by $F(t) = 1 - \exp[-Z(t)]$ and as such defines a probability distribution (prior) on the space of cumulative distribution functions on $(0, \infty)$.

The fundamental result for processes neutral to the right is the following theorem.

THEOREM 1 [Doksum (1974); Ferguson (1974); Ferguson and Phadia (1979)]. *If F is neutral to the right, and $X_1, \dots, X_n$ is a random sample from F, including the possibility of random right censored observations, then the posterior distribution of F given $X_1, \dots, X_n$ is also neutral to the right.*

The purpose of this paper is to give a new characterization of a neutral to the right process by extending Johnson's sufficientness postulate [Zabell (1982)]. An appropriate extension of Johnson's sufficientness postulate to the case of recurrent Markov exchangeable sequence is introduced by Zabell (1995).

The paper is organized as follows. First, in Section 2 we will consider the discrete case when each $X_i \in \Omega = \{1, 2, \ldots\}$. To develop the theory we study the consequences of the following assumption:

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

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)]. We show that (1) combined with the constraint on the $\{f_k\}$ given by

$$\begin{aligned} f_k(n_1, \ldots, n_j + 1, \ldots, n_k, m_k) f_j(n_1, \ldots, n_j, m_j) \\ = f_k(n_1, \ldots, n_k, m_k) f_j(n_1, \ldots, n_j, m_j + 1), \end{aligned} \tag{2}$$

for all $j < k$, where $n_1 + \cdots + n_j + m_j = n_1 + \cdots + n_k + m_k$, implies the exchangeability of the sequence and a neutral to the right prior. Section 3 develops the theory for $\Omega = (0, \infty)$.

2. Result in the discrete case. When (1) and (2) hold, the following result is obtained.

LEMMA 1. *There exists a function $\lambda_k(n_k, m_k)$ such that*

$$\frac{P(X_{n+1} = k \mid X_1, \ldots, X_n)}{P(X_{n+1} \ge k \mid X_1, \ldots, X_n)} = \lambda_k(n_k, m_k) \quad \text{for all } k. \tag{3}$$

PROOF. Using (2), it is possible to see that

$$\frac{f_l(n_1 + 1, \ldots, n_l, m_l)}{f_k(n_1 + 1, \ldots, n_k, m_k)} = \frac{f_l(n_1, \ldots, n_l, m_l)}{f_k(n_1, \ldots, n_k, m_k)},$$

for all $1 < k < l$. The LHS of (3) can be written as

$$\left\{1 + \frac{\sum_{l>k} f_l(n_1 + 1, \ldots, n_l, m_l)}{f_k(n_1 + 1, \ldots, n_k, m_k)}\right\}^{-1}$$

and therefore an observation at $\{1\}$, that is, $n_1 \to n_1 + 1$, has no contribution to the LHS of (3). A similar argument shows that no observation from the set $\{1, \ldots, k-1\}$ has any contribution to

$$\frac{P(X_{n+1} = k \mid X_1, \ldots, X_n)}{P(X_{n+1} \ge k \mid X_1, \ldots, X_n)}. \tag{4}$$

Therefore, (4) depends only on n_k and m_k, completing the proof. □

LEMMA 2. *Conditions* (2) *and* (3) *imply*

$$\lambda_k(n, m)\bar{\lambda}_k(n + 1, m) = \bar{\lambda}_k(n, m)\lambda_k(n, m + 1), \tag{5}$$

for each k, where $\bar{\lambda} = 1 - \lambda$.

PROOF. Since $\sum_k f_k(n_1, \dots, n_k, m_k) = 1$, we deduce, by replacing n_k by $n_k + 1$, that

$$\sum_{l<k} f_l(n_1, \dots, n_l, m_l + 1) + \sum_{l\geq k} f_l(n_1, \dots, n_k + 1, \dots, n_l, m_l) = 1. \tag{6}$$

Using (3) and (6) we obtain

$$\begin{aligned}\lambda_k(n_k, m_k + 1)\bar{\lambda}(n_k, m_k) \\ = \frac{f_k(n_1, \dots, n_k, m_k + 1)}{\sum_{l\geq k} f_l(n_1, \dots, n_k + 1, \dots, n_l, m_l)} \frac{\sum_{l>k} f_l(n_1, \dots, n_l, m_l)}{\sum_{l\geq k} f_l(n_1, \dots, n_l, m_l)}\end{aligned}$$

and, using (2), this is identical to

$$\begin{aligned}\frac{f_k(n_1, \dots, n_k, m_k)}{\sum_{l\geq k} f_l(n_1, \dots, n_l, m_l)} \frac{\sum_{l>k} f_l(n_1, \dots, n_k + 1, \dots, n_l, m_l)}{\sum_{l\geq k} f_l(n_1, \dots, n_k + 1, \dots, n_l, m_l)} \\ = \lambda_k(n_k, m_k)\bar{\lambda}(n_k + 1, m_k),\end{aligned}$$

completing the proof. □

LEMMA 3. *Let $Z_1, Z_2, \dots$ be a $\{0, 1\}$ sequence such that*

$$P(Z_{i+1} = 0 \mid Z_1, \dots, Z_i) = \lambda(n_i, m_i),$$

where $n_i = \sum_l I(Z_l = 0)$ and $n_i + m_i = i$. If $\lambda(n, m)\bar{\lambda}(n+1, m) = \bar{\lambda}(n, m)\lambda(n, m+1)$ for all $(n, m) \in \tilde{\Omega} \times \tilde{\Omega}$, where $\tilde{\Omega} = \{0\} \cup \Omega$, then the sequence $Z_1, Z_2, \dots$ is exchangeable,

$$L^{-1} \sum_{l=1}^{L} I(Z_l = 0) \to V \ a.s.$$

and $E(V^n) = M(n, 0)$ where $M(0, 0) = 1$, $M(n+1, m) = M(n, m)\lambda(n, m)$ and $M(n, m + 1) = M(n, m)\bar{\lambda}(n, m)$.

PROOF. Suppose that after observing $Z_1, \dots, Z_i$ we have n 0's and m 1's ($n + m = i$). We can think of a random walk in $\tilde{\Omega} \times \tilde{\Omega}$ starting at $(0, 0)$ and after the ith move has reached (n, m). We need to demonstrate that the probabilities associated with the possible paths are all equal. This is quite straightforward to show with the condition imposed on λ. The probability for such a path is given, for example, by

$$\prod_{j=0}^{n-1} \lambda(j, 0) \prod_{k=0}^{m-1} \bar{\lambda}(n, k),$$

which is equal to $M(n, m)$. See also Zabell (1995), Lemma 1.1, for a more general result than this one. □

Next we consider the sequence $\{Y_1^{(k)}, Y_2^{(k)}, \dots\}$, for $k = 1, 2, \dots$, which is the sequence $\{X_1, X_2, \dots\}$ with all the $X_i < k$ removed. Then we construct the sequence $\{Z_1^{(k)}, Z_2^{(k)}, \dots\}$ where $Z_i^{(k)} = 0$ if $Y_i^{(k)} = k$ and $Z_i^{(k)} = 1$ if $Y_i^{(k)} > k$. Let $\mathcal{Z}_k = \{Z_1^{(k)}, Z_2^{(k)}, \dots\}$ and $\mathcal{Z} = \{\mathcal{Z}_1, \mathcal{Z}_2, \dots\}$.

LEMMA 4. *For each k, $\mathscr{Z}_k$ is an exchangeable sequence, and $\mathscr{Z}$ is an independent sequence.*

PROOF. That $\mathscr{Z}_k$ is an exchangeable sequence is obvious from Lemma 3. The independence of $\mathscr{Z}$ follows from the fact that $\mathscr{Z}_{k+1}$ is obtained from $\mathscr{Z}_k$ via only those $\{Z_i^{(k)}\}$ which are equal to 1. □

THEOREM 2. *A sequence has a neutral to the right prior if, and only if, conditions* (1) *and* (2) *are satisfied.*

PROOF. If the sequence is neutral to the right then then conditions (1) and (2) are surely satisfied.

Lemmas 3 and 4 imply that there exist independent random variables $\{V_1, V_2, \dots\}$, with each V_k defined on $[0, 1]$, such that

$$n^{-1} \sum_{i=1}^{n} I\left(Z_i^{(k)} = 0\right) \to V_k \quad \text{a.s.}$$

Additionally, it is well known that $P(Z_i^{(k)} = 0 | V_k) = V_k$. That is, given V_k, $\mathscr{Z}_k$ is a collection of independent Bernoulli (V_k) random variables.

We can characterize the distribution of V_k via λ_k. It is convenient at this point to introduce the maps $M_k \colon \tilde{\Omega} \times \tilde{\Omega} \to [0, 1]$, defined by

$$M_k(0, 0) = 1,$$
$$M_k(n+1, m) = M_k(n, m)\lambda_k(n, m)$$

and

$$M_k(n, m+1) = M_k(n, m)\bar{\lambda}_k(n, m).$$

That M_k is well defined is a consequence of (5). The following are now obtained:

$$E(V_k^n) = M_k(n, 0) = \prod_{i=0}^{n-1} \lambda_k(i, 0),$$
$$E\{V_k^n(1 - V_k)^m\} = M_k(n, m),$$
$$E\{V_k^{n+1}(1 - V_k)^m\}/E\{V_k^n(1 - V_k)^m\} = \lambda_k(n, m).$$

We can now write

$$P(X_{n+1} = k \mid X_1, \dots, X_n) = \lambda_k(n_k, m_k) \prod_{j<k} \bar{\lambda}_j(n_j, m_j)$$

or

$$P(X_{n+1} = k \mid X_1, \dots, X_n) = \frac{E\{V_k^{n_k+1}(1 - V_k)^{m_k} \prod_{j<k} V_j^{n_j}(1 - V_j)^{m_j+1}\}}{E\{V_k^{n_k}(1 - V_k)^{m_k} \prod_{j<k} V_j^{n_j}(1 - V_j)^{m_j}\}}. \tag{7}$$

Now define $T_1 = V_1$ and, for $k = 2, 3, \ldots$, define $T_k = V_k(1 - V_{k-1}) \cdots (1 - V_1)$ so that (7) can be written as

$$P(X_{n+1} = k \mid X_1, \ldots, X_n) = \frac{E\{T_k^{n_k+1} \prod_{j \neq k} T_j^{n_j}\}}{E\{T_k^{n_k} \prod_{j \neq k} T_j^{n_j}\}},$$

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

$$(8) \qquad P(X_1 = k_1, \ldots, X_n = k_n) = E\left\{\prod_k T_k^{n_k}\right\}.$$

Clearly $T = (T_1, T_2, \ldots)$ represents a neutral to the right prior provided we have $\sum_k T_k = 1$ a.s. This is satisfied if $\prod_k \{1 - EV_k\} = 0$; that is, if $\prod_k \bar{\lambda}_k(0,0) = 0$. Note that $\sum_k f_k(n_1, \ldots, n_k, m_k) = 1$ for all n and therefore in particular $\sum_k f_k(0, \ldots, 0, 0) = 1$. Therefore, we must have $1 - \prod_k \bar{\lambda}_k(0,0) = 1$.

We have shown, (8), that given T, the X_i's are iid and $P(X_1 = k \mid T) = T_k$ where T is derived from a neutral to the right prior; by construction, if F_k is the random mass assigned to $\{1, \ldots, k\}$, then

$$1 - F_k = \prod_{j \leq k} \{1 - V_j\}.$$

and $T_k = F_k - F_{k-1}$ with $F_0 = 0$, completing the proof. □

We can obtain the posterior representation of the neutral to the right prior. The prior predictive probabilities are

$$P(X_1 = k) = \tau_k \prod_{j=1}^{k-1} (1 - \tau_j),$$

where $\tau_k = \lambda_k(0,0)$, are based on $\{\lambda_k\}$. The posterior predictive probabilities, given a single observation $X = x$, are also based on $\{\lambda_k\}$, where

$$P(X_2 = k \mid X_1 = x) = \tau_k^* \prod_{j=1}^{k-1} (1 - \tau_j^*)$$

and

$$\tau_k^* = \begin{cases} \lambda_k(0,1), & \text{if } x > k, \\ \lambda_k(1,0), & \text{if } x = k, \\ \lambda_k(0,0), & \text{if } x < k. \end{cases}$$

This gives a nice representation of the neutral to the right process in terms of $\{\lambda_k\}$. So $\{\lambda_k(0,0)\}$ define the prior and, given n observations, $\{\lambda_k(n_k, m_k)\}$ define the posterior, where $\{n_k, m_k\}$ are defined in (1). Note also that if π_k is the prior for V_k then the posterior is given by $\pi_k^*(v) \propto v^{n_k}(1-v)^{m_k}\pi_k(v)$ so the beta distribution is going to lead to conjugacy.

REMARK 1. In practical applications the condition (1) on the predictive may not be an adequate description of our state of knowledge. A fundamental assumption concerning the sequence, that is, (1), is hard to justify. When an observation is greater than k, why should it not matter where it occurs [as far as $P(X_{n+1} = k \mid X_1, \dots, X_n)$ is concerned] when this is not the case for an observation less than k. Perhaps an intuitive justification is possible for censored data and a desire for conjugacy.

Condition (2) is equivalent to

$$\frac{f_k(n_1, \dots, n_j + 1, \dots, n_k, m_k)}{f_k(n_1, \dots, n_k, m_k)} = \frac{f_j(n_1, \dots, n_j, m_j + 1)}{f_j(n_1, \dots, n_j, m_j)}. \tag{9}$$

Therefore, the multiplicative factor for updating f_k given an observation at $j < k$ is equal to the multiplicative factor for updating f_j given an observation greater than j. Also, rearranging (9), in an obvious notation, $f_j f_{k|j} = f_k f_{j|l}$ for any $l, k > j$. Using (1), we have $f_{j|l} = f_{j|k}$ and so (9) is equivalent to $f_j f_{k|j} = f_k f_{j|k}$. So (2) can be seen as an exchangeability requirement.

Actually, (2) $\Leftrightarrow$ (9) $\Leftrightarrow$ $f_{j|k} = f_{j|l}$ for all $k, l > j \Rightarrow$ (1) since $f_{j|k} = f_{j|l}$ for all $k, l > j$ implies it does not matter where an observation greater than j occurs, with respect to updating f_j. Therefore we have the corollary.

COROLLARY 1. *The statement: "the multiplicative factor for updating f_k given an observation at $j < k$ is equal to the multiplicative factor for updating f_j given an observation $> j$" characterizes an NTR prior.*

PROOF. By assumption, $f_{k|j}/f_k = f_{j|l}/f_j$ for all $k, l > j$ which implies $f_{j|k} = f_{j|l}$ and hence implies condition (1), having started with $f_j = f_j(n_1, n_2, \dots, n_j, n_{j+1}, \dots)$. This completes the proof. □

3. Result in the continuous case. We now consider the characterization which is the continuous version of Theorem 2. If F is chosen from a neutral to the right prior then, by construction, $F(t) = 1 - \exp[-Z(t)]$ where Z is a Lévy process. We prefer to use the notion of a product integral giving $F(t) = 1 - \prod_0^t [1 - dV(s)]$ where $dV = 1 - \exp(-dZ)$ and let $E(dV) = d\Lambda$. Here, however, we will be consistent with previous notation and write $d\Lambda(s) = \lambda_{ds}$.

THEOREM 3. *A sequence $X_1, X_2, \dots$ with each X_i defined on $\Omega = (0, \infty)$ has a neutral to the right process prior if and only if for all n and t $P(X_{n+1} > t \mid X_1, \dots, X_n) = \prod_0^t \bar{\lambda}_{ds}(n_s, m_s)$ where $n_s = \sum_i I(X_i = s)$, $m_s = \sum_i I(X_i > s)$ and $\lambda_{ds}(n, m)\bar{\lambda}_{ds}(n+1, m) = \lambda_{ds}(n, m+1)\bar{\lambda}_{ds}(n, m)$ for all $s > 0$ and nonnegative integers n, m.*

PROOF. Our aim is to show that $(X_1, \dots, X_n)$ is exchangeable for all n. The de Finetti representation theorem will then imply by uniqueness that the prior for the sequence is a neutral to the right process. We write $P(X_1 \in dt_1)$ as $[t_1]$ which is given by $\lambda_{dt_1} \prod_0^{t_1} \bar{\lambda}_{ds}$. Now let us consider $[t_1, \dots, t_n]$ which is given by

$$\prod_{i=1}^{n} P(X_i \in dt_i | X_1 = t_1, \dots, X_{i-1} = t_{i-1}),$$

and let $t_{(1)} \leq \cdots \leq t_{(n)}$ be the t's in increasing order. We show that however the t's are arranged in $[t_1, \dots, t_n]$, the term involving λ_{ds}, for an arbitrary s, is unaltered. This will demonstrate the exchangeability of the sequence. To clarify, we briefly consider the case when $n = 2$. Assume that $t_2 \geq t_1$ so

$$[t_1, t_2] = \lambda_{dt_1}(0,0)\left\{\prod_0^{t_1} \bar{\lambda}_{ds}(0,0)\right\}\lambda_{dt_2}(0,0)\left\{\prod_{t_1}^{t_2} \bar{\lambda}_{ds}(0,0)\right\}\bar{\lambda}_{dt_1}(1,0) \times \left\{\prod_0^{t_1} \bar{\lambda}_{ds}(0,1)\right\}$$

and

$$[t_2, t_1] = \lambda_{dt_2}(0,0)\left\{\prod_0^{t_2} \bar{\lambda}_{ds}(0,0)\right\}\lambda_{dt_1}(0,1)\left\{\prod_0^{t_1} \bar{\lambda}_{ds}(0,1)\right\}.$$

These are clearly identical provided $\lambda_{dt_1}(0,0)\bar{\lambda}_{dt_1}(1,0) = \bar{\lambda}_{dt_1}(0,0)\lambda_{dt_1}(0,1)$ and also note that for $s \notin \{t_1, t_2\}$ the term involving λ_{ds} for both $[t_1, t_2]$ and $[t_2, t_1]$ are equal. For general n it is not hard to see that if $s \notin \{t_1, \dots, t_n\}$ then the term involving λ_{ds} will be the same in $[t_{\pi(1)}, \dots, t_{\pi(n)}]$ for all permutations π on $\{1, \dots, n\}$. Let us consider the case when $s = t(k)$ and first we assume there are no ties in the data. The term involving $\lambda_{dt(k)}$, from now on written as λ_k, will only depend on where $t(k)$ is located in $[t_1, \dots, t_n]$ relative to $t(k+1), \dots, t(n)$. For example, if $t(k)$ precedes all of $\{t(k+1), \dots, t(n)\}$ then the term involving λ_k is given by

$$\lambda_k(0,0)\bar{\lambda}_k(1,0)\bar{\lambda}_k(1,1)\cdots\bar{\lambda}_k(1, m_{t(k)} - 1).$$

Now we can replace $\lambda_k(0,0)\bar{\lambda}_k(1,0)$ by $\bar{\lambda}_k(0,0)\lambda_k(0,1)$ to obtain

$$\bar{\lambda}_k(0,0)\lambda_k(0,1)\bar{\lambda}_k(1,1)\cdots\bar{\lambda}_k(1, m_{t(k)} - 1),$$

which is the term involving λ_k if one observation from $\{t(k+1), \dots, t(n)\}$ (it does not matter which one) precedes $t(k)$ and the rest follow $t(k)$. We can then replace $\lambda_k(0,1)\bar{\lambda}_k(1,1)$ by $\bar{\lambda}_k(0,1)\lambda_k(0,2)$ to obtain

$$\bar{\lambda}_k(0,0)\bar{\lambda}_k(0,1)\lambda_k(0,2)\bar{\lambda}_k(1,2)\cdots\bar{\lambda}_k(1, m_{t(k)} - 1),$$

which is the term involving λ_k if two observations from $\{t(k+1), \dots, t(n)\}$ precede $t(k)$ and the rest follow $t(k)$. We can continue like this "all the way to the end." To consider the case of ties we draw on the connection between the above λ_k's and those in (5) which were used to define the M_k's. For example, if $t(k)$ is repeated twice and precedes all of $\{t(k+1), \dots, t(n)\}$ then the term involving λ_k is given by

$$\lambda_k(0,0)\lambda_k(1,0)\bar{\lambda}_k(2,0)\bar{\lambda}_k(2,1)\cdots$$

and we can move this "along" to the second position (and to the end) since

$$\lambda_k(0,0)\lambda_k(1,0)\bar{\lambda}_k(2,0) = \bar{\lambda}_k(0,0)\lambda_k(0,1)\lambda_k(1,1),$$

concluding the proof. □

REMARK 2. Explicitly, we have a priori $E[F(t)] = 1 - \prod_0^t \bar{\lambda}_{ds}(0,0)$ and a posteriori $E[F(t)] = 1 - \prod_0^t \bar{\lambda}_{ds}(n_s, m_s)$ providing the updating mechanism.

Essentially, to characterize a neutral to the right prior we need to identify a function $\lambda\colon \tilde{\Omega}\times\tilde{\Omega} \to [0,1]$ which satisfies $\lambda(n,m)\bar{\lambda}(n+1,m) = \bar{\lambda}(n,m)\lambda(n,m+1)$. For example, the beta-Stacy process is based on $\lambda(n,m) = (\alpha+n)/(\alpha+\beta+n+m)$ for suitable α and β. In general the best way to generate such λ is via

$$\lambda(n,m) = \frac{E[V^{n+1}(1-V)^m]}{E[V^n(1-V)^m]},$$

where V is a random variable defined on $[0,1]$. The beta-Stacy arises when $V \sim \text{beta}(\alpha,\beta)$.

Acknowledgments. The authors are grateful to two referees, an Associate Editor and P. Secchi for comments and criticisms on an earlier draft of the paper. The paper was completed during a visit by the second author to Imperial College.

REFERENCES

ANDERSEN, P. K., BORGAN, O., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

DE FINETTI, B. (1937). La prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7** 1–68.

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

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 estimates based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294.

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

LO, A. Y. (1991). A characterisation of the Dirichlet process. *Statist. Probab. Lett.* **12** 185–187.

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

REGAZZINI, E. (1978). Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità. *Giornale dell'Istituto Italiano degli Attuari* **1–2** 77–89.

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. and MULIERE, P. (1997a). Beta-Stacy processes and a generalisation of the Pólya-urn scheme. *Ann. Statist.* **25** 1762–1780.

WALKER, S. G. and MULIERE, P. (1997b). A characterisation of Pólya tree distributions. *Statist. Probab. Lett.* **31** 163–168.

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

ZABELL, S. L. (1995). Characterizing Markov exchangeable sequences. *J. Theoret. Probab.* **8** 175–178.

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

DIPARTIMENTO DI ECONOMIA
POLITICA E MQ
UNIVERSITÀ DI PAVIA
S. FELICE 5
27100 PAVIA
ITALY
E-MAIL: pmuliere@eco.unipv.it