This is to certify that the

dissertation entitled

Some properties and characterizations
of neutral to the right priors
and beta processes

presented by

Jyotirmoy Dey

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

R. V. Ramamoorthi
Major professor

Date August 23, 1999

SOME PROPERTIES AND CHARACTERIZATIONS OF NEUTRAL TO THE RIGHT PRIORS AND BETA PROCESSES

By

Jyotirmoy Dey

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1999

ABSTRACT

SOME PROPERTIES AND CHARACTERIZATIONS OF NEUTRAL TO THE RIGHT PRIORS AND BETA PROCESSES

By

Jyotirmoy Dey

The nonparametric Bayesian paradigm requires us to consider probability measures - priors - on the infinite dimensional space $\mathcal{F}$ of all cumulative distribution functions. This dissertation is a study of one such class of measures introduced by Doksum [8] known as Neutral to the right (NR) priors. The material is organized in three chapters.

Chapter 1 is a survey of NR priors and serves as a prelude to the subsequent chapters and results. We conclude the chapter by observing that these priors can be chosen to have all of $\mathcal{F}$ as support and thus satisfies one of the desirable requirements of nonparametric prior distributions as laid down by Ferguson [10].

In chapter 2 we provide necessary and sufficient conditions on a sequence of exchangeable random variables such that the prior distribution obtained from it via de Finetti's Theorem is NR. En route to such characterization we also obtain characterizations in terms of the posterior distributions.

Chapter 3 contains a study of beta and beta-Stacy processes. Hjort [14] and Walker and Muliere [26], respectively, developed beta processes and beta-Stacy processes as concrete examples of NR prior distributions. We first give a construction of beta process priors directly on $\mathcal{F}$ and then prove the following:

1. the posterior corresponding to a beta process is weakly consistent,

2. beta processes with distinct parameters are mutually singular,

3. carefully chosen Dirichlet process priors on the space of right-censored observations induce beta priors on the space of lifetime and censoring-time distributions,

4. a beta-Stacy prior is a simple reparameterization of a beta process.

To the memory of my father.

ACKNOWLEDGEMENTS

My deepest regards and sincerest thanks go to my dissertation advisor, Professor R. V. Ramamoorthi. I have benefited immensely from his insightful advice, brilliant vision, constant encouragement and fathomless patience. His profound knowledge of Statistics, clarity of thought and exemplary work ethic has been a constant source of inspiration to me. I will always cherish the care and interest with which he oversaw my academic and professional progress.

I thank Professors R. Erickson, H. Koul and C. Weil for serving on my dissertation committee. I greatly appreciate the contribution made by Professor Koul in carefully reading previous drafts of this document and suggesting several improvements to the text. Special thanks go to Professors C. Page and J. Stapleton for their involvement and guidance in my development as a graduate instructor and statistical consultant. Their interest and assistance in my search for a professional career is gratefully acknowledged.

It is impossible to express in words how much I owe my academic achievements to the encouragement, affection, support and sacrifice of my mother and my sisters. Without them, my pursuit of a higher education would never be possible.

I must also thank my colleagues and friends for their company, help and encouragement. I will refrain from furnishing a complete list here as it is too long. I must, however, mention that I will always retain fond memories of my associations with Ms. M. Geraldes, Mr. N. Choudhury and Mr. D. Biswas.

Contents

Introduction

Let θ be an unknown parameter and X be an observable random variable whose distribution F_θ depends on θ. The goal of statistical investigation is to make inference on θ based on the observed value of X. In the Bayesian paradigm θ itself is endowed with a distribution Π, called the prior distribution, and the inference essentially consists of updating the prior Π to Π_X - the conditional distribution of θ given X - commonly known as the posterior distribution of θ given X. The prior distribution Π reflects, often as an approximation, the investigators knowledge of the parameter θ prior to observing X.

In the parametric case θ is generally taken to be an element of $\mathbb{R}^k$ and the map $\theta \to F_\theta$ is a smooth parameterization. In other words, the distribution of X is assumed to be among $\{F_\theta : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$. In the nonparametric case the restriction to a finite dimensional parametric family is removed. The set of permissible distributions for X is typically the set of all distributions or a large subset thereof.

The model that we consider consists of

1. $\mathcal{F}$ - the set of probability distributions on $\mathbb{R}^+$,

2. n independent identically distributed random variables $X_1, \ldots, X_n$ whose com-

mon distribution F is, of course, an element of $\mathcal{F}$, and

3. Π - a probability measure (probability distribution) on $\mathcal{F}$.

In the nonparametric case the Bayesian paradigm requires us to consider probability measures Π on the infinite dimensional space $\mathcal{F}$. It is thus necessary to develop and study probability measures on $\mathcal{F}$ which would be analytically tractable and which would also be interpretable.

This thesis is devoted to a study of a class of priors called neutral to the right (NR) priors. A prior Π is said to be NR if, for all $k \geq 1$ and all $t_1 < \cdots < t_k$, $\frac{\bar{F}(t_i)}{\bar{F}(t_{i-1})}, i = 1, \ldots, k$, are independent, where $\bar{F}(\cdot) = 1 - F(\cdot)$. These priors were introduced by Doksum [8], who also showed that if Π is NR, then the posterior given n observations is also NR. This result was extended to the case of right-censored data by Ferguson and Phadia [12]. While these authors considered NR priors in abstraction, Hjort [14] and Walker and Muliere [26], respectively, developed beta processes and beta-Stacy processes which provide concrete and useful classes of NR priors. These priors, which are analogous to the beta prior for the Bernoulli(θ), are analytically tractable and are flexible enough to incorporate a wide variety of prior beliefs.

The map $F \mapsto \phi_D(F) = -\log(1 - F)$ maps $\mathcal{F}$ into the space of increasing functions. Doksum showed that a prior Π on $\mathcal{F}$ is NR if and only if the induced measure $\Pi \circ \phi_D^{-1}$ gives rise to independent increment processes. Hjort proved a similar result by considering the map $\phi_H(F)(\cdot) = \int_{(0,\cdot]} \frac{dF(s)}{F[s,\infty)}$. When F is continuous the two images $\phi_D(F)$ and $\phi_H(F)$ are the same.

Since independent increment processes are well understood, this connection provides a powerful tool for studying NR priors. In particular, independent increment processes have a cannonical structure, the so-called Lévy representation. The associated Lévy measure can be used to elucidate properties of NR priors. For instance Hjort provides an explicit expression for the posterior given n independent observations in terms of the Lévy representation when the Lévy measure is of a specific form.

In Chapter 1 we provide a brief introduction to NR priors and related independent increment processes. We then show that a NR prior with support $\mathcal{F}$ can be obtained by choosing a Lévy measure with full support. This result shows, in particular, that with a proper choice of parameters the beta and beta-Stacy processes can have all of $\mathcal{F}$ as support.

In Chapter 2 we first recall Doksum's result and the fact that if Π is NR, then the posterior distribution of $F(t)$, given $X_1, \cdots, X_n$, depends only on $\{N_n(s) : s \leq t\}$ where $N_n(\cdot) = \sum_{i=1}^n I_{\{X_i \leq \cdot\}}$. In other words, the posterior distribution of $F(t)$ does not depend on the exact values of the observations larger than t, but only on how many there are. We then show that this property of the posterior actually characterizes NR priors. This characterization is then used to provide yet another characterization of NR priors via de Finetti's Theorem. The chapter concludes with an extension of the results to the case of right censored data.

Chapter 3 is devoted to a study of beta and beta-Stacy processes. We first construct beta processes directly on $\mathcal{F}$. Then we show that beta processes yield consistent posteriors, i.e. if F_0 is indeed the true distribution then, as more and more observations accumulate, the posterior converges to F_0 with probability 1.

Hjort has shown that beta processes possess many pleasing properties in the context of right-censored data such as easy updating of the prior parameters. In the same spirit we show that if (Z, Δ) is a right-censored observation arising from a survival time X and an independent censoring time Y, then under a Dirichlet process prior for the distribution of (Z, Δ) the distributions of X and Y marginally have beta priors. We then use a result of Brown [1] on mutual singularity of Poisson processes to show that any two beta process priors are mutually singular. We conclude the chapter by observing that any beta-Stacy process is just a reparameterization of a beta process.

Chapter 1

Neutral to the Right Priors

Neutral to the right(NR) priors is a specific class of nonparametric priors that was introduced by Doksum [8]. Historically, the concept of neutrality is due to Connor and Mosimann [3] who considered it in the multinomial context. Doksum extended it to distributions on the real line in the form of neutral and neutral to the right priors. Subsequent papers by Ferguson [11], Ferguson and Phadia [12], Hjort [14] and Walker and Muliere [26] have made significant contributions to their theory. As mentioned earlier, the theory of independent increment processes provides a powerful tool to understand these priors.

The purpose of this chapter is to give a summary of the basic properties of NR priors. In Section 1.1 the Bayesian nonparametric set-up is formalized. Definition and examples of NR priors are provided in Section 1.2. In Section 1.3, we discuss independent increment processes and their Lévy representation. Section 1.4 then provides a description of the connections between NR priors and independent increment processes. Section 1.5 discusses the support of NR priors.

1.1 The Space $\mathcal{F}$

Consider the measurable space $([0,\infty), \mathcal{B}_{[0,\infty)})$ where $\mathcal{B}_{[0,\infty)}$ denotes the collection of all Borel subsets of $[0,\infty)$. Let $\mathbf{M}(\mathbb{R}^+)$ denote the space of all probability measures on $([0,\infty), \mathcal{B}_{[0,\infty)})$ and $\mathcal{F}$ denote the space of all cumulative distribution functions on $[0,\infty)$. Our goal is to investigate certain properties of probability measures on $\mathbf{M}(\mathbb{R}^+)$. However, since there is a 1-1 correspondence between $\mathbf{M}(\mathbb{R}^+)$ and $\mathcal{F}$, it suffices to consider probability measures on $\mathcal{F}$. To further simplify notation, we will let F denote both a distribution function and its corresponding probability measure.

Equip $\mathbf{M}(\mathbb{R}^+)$ with the Borel σ-algebra under the topology of weak convergence. Since weak convergence on $\mathbf{M}(\mathbb{R}^+)$ is equivalent to convergence in distribution on $\mathcal{F}$ (also called weak convergence for this reason), we will equip $\mathcal{F}$ with the topology of weak convergence too. Formally, a sequence of distribution functions $\{F_n : n \geq 1\} \in \mathcal{F}$ converges to $F \in \mathcal{F}$ weakly if $F_n(t) \to F(t)$ as $n \to \infty$ for all continuity points t of F. We write this fact as $F_n \xrightarrow{w} F$. One of the properties that makes such convergence useful is the fact that, under the weak topology, $\mathcal{F}$ is a complete, separable metric space.

Let $\Sigma_{\mathcal{F}}$ denote the σ-algebra of Borel subsets of $\mathcal{F}$. It is also the smallest σ-algebra with respect to which all coordinate maps $F \mapsto F(t), t \geq 0$ are measurable. A prior is defined to be a probability measure on the space $(\mathcal{F}, \Sigma_{\mathcal{F}})$. It is common practice to view a prior as the process measure corresponding to a stochastic process with paths almost surely in $\mathcal{F}$. A prior is thus the distribution of a random $\mathcal{F}$-valued function.

The standard nonparametric Bayesian set-up consists of a random distribution function F with a prior distribution Π, and given $F \in \mathcal{F}$, a random sequence of observations $\mathbf{X} = (X_1, X_2, \ldots)$ which are independent and identically distributed as F. In short, one writes

$$F \sim \Pi, \quad \text{and given } F, \quad X_1, X_2, \ldots \overset{i.i.d.}{\sim} F \tag{1.1.1}$$

Formally, consider the probability space $(\Omega, \Sigma, \mathbf{P}_\Pi)$ where $\Omega = \mathcal{F} \times [0, \infty)^{\mathbf{N}}$, $\Sigma = \Sigma_{\mathcal{F}} \otimes \mathcal{B}_{[0,\infty)}^{\mathbf{N}}$ and the measure $\mathbf{P}_\Pi$ is defined by

$$\mathbf{P}_\Pi(F \in D, X_1 \in B_1, \ldots, X_n \in B_n) = \mathcal{E}_\Pi \left[I_D(F) \prod_{i \leq n} F(B_i) \right]$$

where $D \in \Sigma_{\mathcal{F}}$; $B_1, \ldots, B_n \in \mathcal{B}_{[0,\infty)}$, for every $n \geq 1$.

For each $n \geq 1$, a version of the conditional distribution of F given $X_1, \ldots, X_n$ is called a posterior distribution of F or, sometimes, a posterior corresponding to the prior distribution Π of F. When there is no ambiguity about the prior, we will refer to it simply as the posterior.

1.2 Definition and Examples

For any $F \in \mathcal{F}$, let $\bar{F}(\cdot) = 1 - F(\cdot)$. $\bar{F}$ is commonly known as the survival function corresponding to F. Let $\bar{F}(0) = 1$.

Definition 1.1. *A prior Π on $\mathcal{F}$ is said to be Neutral to the right (NR) if, under Π, for all $k \geq 1$ and all $0 < t_1 < \ldots < t_k$,*

$$\bar{F}(t_1), \frac{\bar{F}(t_2)}{\bar{F}(t_1)}, \ldots, \frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})}$$

are independent.

If Π is NR, we will also refer to a random cdf F with distribution Π as being neutral to the right. Note that $(0/0)$ is defined here and throughout to be one.

For a fixed F, if X is a random variable distributed as F, then, for every $0 \leq s < t$, $\bar{F}(t)/\bar{F}(s)$ is simply the conditional probability $F(X > t|X > s)$. For $t > 0$, $\bar{F}(t)$ is viewed as the conditional probability $F(X > t|X > 0)$.

Example 1. Consider a finite ordered set $\{t_1, \ldots, t_n\}$ of points in $(0, \infty)$. To construct a NR prior on the set $\mathcal{F}_{t_1,\ldots,t_n}$ of distribution functions supported by the points $t_1, \ldots, t_n$, we only need to specify $(n-1)$ independently distributed $[0,1]$-valued random variables $V_1, \ldots, V_{n-1}$, and then set $\bar{F}(t_i)/\bar{F}(t_{i-1}) = 1 - V_i$ for $1 \leq i \leq n-1$. Finally, set $\bar{F}(t_n)/\bar{F}(t_{n-1}) = 0$. Observe that, $\bar{F}(t_n) = 0$ and, for $1 \leq i \leq n-1$,

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

Example 2. In a very similar fashion we can construct a NR prior on the space $\mathcal{F}_{\underline{\underline{T}}}$ of all distribution functions supported by a countable subset $\underline{\underline{T}} = \{t_1 < t_2 < \ldots\}$ of $(0, \infty)$.

Let $\{V_i\}_{i \geq 1}$ be a sequence of independent $[0,1]$-valued random variables such that, for some $\eta > 0$,

$$\sum_{i \geq 1} \mathbf{P}(V_i > \eta) = \infty.$$

This happens, for instance, when V_i's are identically distributed with $\mathbf{P}(V_i > \eta) > 0$. As before, for $i \geq 1$, set $\bar{F}(t_i)/\bar{F}(t_{i-1}) = 1 - V_i$. In other words, $\bar{F}(t_k) = \prod_{i=1}^{k}(1 - V_i)$, for all $k \geq 1$. By the second Borel-Cantelli lemma, we have

$$\mathbf{P}\left(\prod_{i \geq 1}(1 - V_i) = 0\right) = 1.$$

This defines a NR prior Π on $\mathcal{F}$ because

$$\lim_{t\to\infty} \bar{F}(t) = \lim_{k\to\infty} \prod_{i=1}^{k}(1 - V_i) = 0, \qquad \text{a.s.} \quad \Pi.$$

Other non-trivial examples of NR priors are also available.

Example 3. Dirichlet process priors, introduced by Ferguson [10], provide a ready example of a family of NR priors.

Doksum [8] suggested construction of general classes of such priors via independent increment processes. Examples of suitable independent increment processes are:

Example 4. Beta processes developed by Hjort [14]. These are a family of independent increment processes that correspond to NR prior distributions on $\mathcal{F}$. We will refer to the induced NR priors also as beta priors.

Example 5. Log-beta processes developed by Walker and Muliere [26] are another family of independent increment processes that lead to NR priors on $\mathcal{F}$ for suitable choice of parameters. Priors on $\mathcal{F}$ constructed via suitable log-beta processes were named beta-Stacy prior processes by Walker and Muliere.

We defer further discussion of the last two prior processes and their construction to Chapter 3.

Consider, briefly, the problem of specifying a general NR prior on $\mathcal{F}$. Let $\mathbb{Q}$ be a dense subset of $[0, \infty)$ and let $\{t_1, t_2, \dots\}$ be an enumeration of $\mathbb{Q}$. As seen in Example 1, it is easy to specify a NR prior Π_n on $\mathcal{F}_{t_1,\dots,t_n}$.

Let $t_1^{(n)} < \cdots < t_n^{(n)}$ be an ordering of $\{t_1, \dots, t_n\}$. We wish to specify the distributions of the independent $[0,1]$-valued variables $\bar{F}(t_1^{(n)}) = V_1^{(n)}$, $\frac{\bar{F}(t_2^{(n)})}{\bar{F}(t_1^{(n)})} = V_2^{(n)}$,

... $\frac{\bar{F}(t_n^{(n)})}{\bar{F}(t_{n-1}^{(n)})} = V_{n-1}^{(n)}$ in such a fashion that the sequence of priors thus generated, $\{\Pi_n\}$, is weakly convergent to some measure Π, which will then be NR. The difficulty is that we need to know the convergence of a whole family of finite dimensional distributions and their limits for this kind of specification, which is equivalent to knowing Π beforehand.

However, given a prior Π on $\mathcal{F}$, we can, in above fashion, construct a sequence of priors Π_n with support $\mathcal{F}_{t_1,\ldots,t_n}$, $n \geq 1$, which will converge to Π weakly.

We will refer to priors that give probability one to $\mathcal{F}_A$, where A is at most an ordered countable set, as time-discrete prior processes. Thus, concisely, any NR prior on $\mathcal{F}$ can be obtained as a weak limit of time-discrete NR prior processes. Later we will construct the beta prior along these lines.

1.3 Independent Increment Processes

The theory of NR priors owes much of its development and analytic elegance to its connection with independent increment processes. The principal examples of general families of NR priors have been constructed via this connection. Naturally, no discussion of NR priors can be complete without reference to this phenomenon. This connection also leads to special statistical significance for NR priors.

In the next section, we will establish the relationship between NR priors and independent increment processes with non-decreasing paths. For now we briefly discuss the relevant theory of these processes in terms of a representation due to P. Lévy [18]. Here is a brief description of the representation.

The following facts are well-known and may be found in Ito [15] and/or Kallenberg [16].

Definition 1.2. *A stochastic process* $\{A(t)\}_{t\geq 0}$ *is said to be an independent increment process if* $A(0) = 0$ *almost surely and if, for every* k *and every* $\{t_0 < t_1 < \cdots < t_k\} \subset [0, \infty)$, *the family* $\{A(t_i) - A(t_{i-1})\}_{i=1}^{k}$ *is independent.*

Let $\mathcal{H}$ be a space of functions defined by

$$\mathcal{H} = \{H \mid H : [0, \infty) \mapsto [0, \infty], H(0) = 0, H \text{ non-decreasing, right-continuous}\}. \tag{1.3.2}$$

Let $\mathbb{B}_{(0,\infty)\times[0,\infty]}$ be the Borel σ-algebra on $(0, \infty) \times [0, \infty]$.

Theorem 1.1 (Ito). *Let* Π^* *be a probability on* $\mathcal{H}$. *Under* Π^*, $\{A(t) : t > 0\}$ *is an independent increment process if and only if the following three conditions hold: there exists*

1. *a finite or countable set* $\mathbf{M} = \{t_1, t_2, \ldots\}$ *of points in* $(0, \infty)$ *and, for each* $t_i \in \mathbf{M}$, *a positive random variable* Y_i *defined on* $\mathcal{H}$ *with density* f_i;
2. *a non-random continuous non-decreasing function* b; *and*
3. *a measure* λ *on* $\left((0, \infty) \times [0, \infty], \mathbb{B}_{(0,\infty)\times[0,\infty]}\right)$ *which, for all* $t > 0$, *satisfies*
   (a) $\lambda(\{t\} \times [0, \infty]) = 0$,
   (b) $\displaystyle\iint_{\substack{0<s\leq t\\ 0\leq u\leq\infty}} \frac{u}{1+u}\lambda(ds\, du) < \infty$;

such that

$$A(t) = b(t) + \sum_{t_i \leq t} Y_i(A) + \iint_{\substack{0<s\leq t\\ 0\leq u\leq\infty}} u\, \mu(ds\, du, A) \tag{1.3.3}$$

where, for each $A \in \mathcal{H}$, $\mu(\cdot, A)$ *is a measure on* $\left((0,\infty) \times [0,\infty], \mathbb{B}_{(0,\infty)\times[0,\infty]}\right)$ *such that, under* Π^*, $\mu(\cdot,\cdot)$ *is a Poisson process with parameter* $\lambda(\cdot)$, *i.e. for arbitrary disjoint Borel subsets* $E_1, \ldots, E_k$ *of* $(0,\infty)\times[0,\infty]$, $\mu(E_1,\cdot), \ldots, \mu(E_k,\cdot)$ *are independent, and*

$$\mu(E_i, \cdot) \sim \text{Poisson}(\lambda(E_i)) \qquad \text{for } 1 \le i \le k.$$

Note the following facts about independent increment processes which will be useful to us later and facilitate understanding of the remaining subject matter.

1.3.1. The measure λ on $(0,\infty) \times [0,\infty]$ is often expressed as a family of measures $\{\lambda_t : t > 0\}$ where $\lambda_t(A) = \lambda((0,t] \times A)$ for Borel sets A.

1.3.2. The above representation may be expressed equivalently in terms of the moment generating function of $A(t)$ as

$$\mathcal{E}(e^{-\theta A(t)}) = e^{-b(t)} \left[\prod_{t_i \le t} \mathcal{E}(e^{-\theta Y_i})\right] \exp\left[-\iint_{\substack{0<s\le t\\ 0\le u\le\infty}} (1 - e^{-\theta u})\, \lambda(ds\, du)\right]$$

1.3.3. The random variables Y_i occuring in the decomposition arise from the jumps of the process at fixed points. Say that t is a fixed jump-point of the process if $\Pi^*(A\{t\} > 0) > 0$. It is known that there are at most countably many such fixed jump-points, that the set $\mathbf{M}$ is precisely the set of such points and that $Y_i = A\{t_i\}$.

1.3.4. The random measure $A \mapsto \mu(\cdot, A)$ also has an explicit description. For any Borel subset E of $(0,\infty) \times [0,\infty]$,

$$\mu(E, A) = \#\left\{(t, A\{t\}) \in E \mid A\{t\} > 0\right\}.$$

1.3.5. Let $A^c(t) = A(t) - b(t) - \sum_{t_i \le t} A\{t_i\}$. Then

$$A^c(t) = \iint\limits_{\substack{0<s\le t \\ 0\le u\le\infty}} u\,\mu(du\,ds, A).$$

1.3.6. The countable set $\mathbf{M}$, the set of densities $\{f_i : i \ge 1\}$, the measure λ and the non-random function b are known as the four components of the process $\{A(t) : t > 0\}$, or, equivalently, of the measure Π^*. The measure λ is known as the Lévy measure of Π^*.

1.3.7. A Lévy process Π^* without any non-random component, i.e., for which $b(t) = 0$, for all $t > 0$, has sample paths that increase only in jumps almost surely Π^*. All Lévy processes that we will encounter will be of such type.

1.4 NR Priors via Lévy Processes

Consider the map $\phi_D(F) = -\log \bar{F}$ of $\mathcal{F}$ into $\mathcal{H}$. Doksum [8] showed that a prior distribution Π on $\mathcal{F}$ is NR if and only if the measure $\tilde{\Pi}$ induced by the map ϕ_D gives rise to an independent increment process. On the other hand, Hjort [14] used the function $\phi_H(F)(\cdot) = A_F(\cdot) = \int_0^{\cdot} \frac{dF(s)}{F[s,\infty)}$ towards the same end.

These maps lend special statistical significance to NR priors in the context of survival analysis. Several authors define cumulative hazard as $H_F(t) \equiv \phi_D(F)(t) = -\log \bar{F}(t)$. For us, however, cumulative hazard would be the map ϕ_H for the following reasons. First, when F is discrete $A_F\{t\} = \frac{F\{t\}}{F[t,\infty)}$ corresponds to a measure of the rate of occurence of an event at time t given it has not occurred before. The cumulative hazard, then, is just the sum of these rates. Thus, while ϕ_D is mathematically simple, the function A_F is a more natural choice for the cumulative hazard function.

Second, as noted earlier, the two definitions coincide when F is continuous. However, in estimating a survival function or a cumulative hazard function one typically employs a discontinuous estimate. The nature of the map, therefore, plays an important role in inference about lifetime distributions and hazard rates. Also, since all independent increment processes we will consider increase only in jumps, distributions sampled from the corresponding NR priors are discrete.

We will now proceed with the formal details about these maps.

Let $F \in \mathcal{F}$ and let $T_F = \inf\{t : F(t) = 1\}$. Note that T_F may be ∞. Define a transform ϕ_H of F as follows:

$$\phi_H(F)(t) = A_F(t) = \begin{cases} \int_{(0,t]} \frac{dF(s)}{F[s,\infty)}, & \text{for } t \leq T_F; \\ A_F(T_F) & \text{for } t > T_F. \end{cases}$$

1.4.1. The integral in the definition of A_F, for $t \leq T_F$, is a Stieltjes integral. Let $\{s_1, s_2, \ldots\}$ be a dense subset of $(0, \infty)$. For each n, let $s_1^{(n)} < \cdots < s_n^{(n)}$ be an ordering of $\{s_1, \ldots, s_n\}$. Let $s_0^{(n)} = 0$ and define

$$A_F^n(t) = \begin{cases} \sum_{s_i^{(n)} < t} \frac{F(s_i^{(n)}, s_{i+1}^{(n)}]}{F(s_i^{(n)}, \infty)}, & \text{for } t \leq T_F; \\ A_F^n(T_F) & \text{for } t > T_F. \end{cases}$$

Then, for all t, $A_F^n(t) \to A_F(t)$ as $n \to \infty$.

1.4.2. A_F is non-decreasing and right-continuous. The fact that A_F is non-decreasing follows trivially since F is non-decreasing. To see that A_F is right-continuous, fix a point t and note that, if $j = \max\{i \le n : s_i^{(n)} < t\}$, then

$$A_F(t+) - A_F(t) = \lim_{n\to\infty} \frac{F(s_{j+1}^{(n)}, s_{j+2}^{(n)}]}{F(s_{j+1}^{(n)}, \infty)}.$$

where, both $\{s_{j+1}^{(n)}\}$ and $\{s_{j+2}^{(n)}\}$ are non-decreasing sequences converging to t from above. Thus $F(s_{j+1}^{(n)}, s_{j+2}^{(n)}] \to 0$ as $n \to \infty$.

If $t < T_F$, then the denominator of the R.H.S. of the above equation is positive for some n. Hence right-continuity follows. For $t \ge T_F$ it follows from the definition.

It is easy to see that $A_F(t) < \infty$ for every $t < T_F$. As with F, we will think of A_F simultaneously as a function and a measure. Thus the measure of any interval $(s,t]$ under A_F will be defined as $A_F(s,t] = A_F(t) - A_F(s)$. For $T_F < s < t$, define $A_F(s,t] = 0$. One may now uniquely extend it to a σ-finite measure on Borel sets.

1.4.3. For any t, A_F has a jump at t iff F has a jump at t, i.e. $\{t : A_F\{t\} > 0\} = \{t : F\{t\} > 0\}$.

1.4.4. The map A_F has an explicit expression. Let

$$F^c(t) = F(t) - \sum_{s \le t} F\{s\}$$

be the continuous part of F. Then

$$A_F(t) = \sum_{s \le t} \frac{F\{s\}}{F[s,\infty)} - \log(1 - F^c(t)).$$

1.4.5. It follows from 1.4.4 above that

(a) $T_F = \inf\{t : A_F(t) = \infty \text{ or } A_F\{t\} = 1\}$,

(b) $A_F\{t\} \leq 1$ for all t,

(c) $A_F(T_F) = \infty$ if T_F is a continuity point of F,

(d) $A_F\{T_F\} = 1$ if $F\{T_F\} > 0$.

These and other properties of ϕ_H and details may be found in Gill and Johansen [13] and Hjort [14].

Now, let $\mathcal{A}'$ be the space of all functions on $[0, \infty)$ that are non-decreasing, right-continuous and may, at any finite point, be infinity, but has jumps no greater than one, i.e.

$$\mathcal{A}' = \{B \in \mathcal{H} \mid B\{t\} \leq 1 \text{ for all } t\}.$$

Equip $\mathcal{A}'$ with the smallest σ-algebra under which, the maps $\{A \mapsto A(t), t \geq 0\}$ are measurable. From 1.4.1 on page 14 it follows that ϕ_H is measurable with respect to this σ-algebra. Also note, from 1.4.2 on page 14, that ϕ_H maps $\mathcal{F}$ into $\mathcal{A}'$. The actual range, which we will now describe, is smaller.

For $A \in \mathcal{A}'$, let $T_A = \inf\{t : A(t) = \infty \text{ or } A\{t\} = 1\}$. Let $\mathcal{A}$ be the space of all cumulative hazard functions on $[0, \infty)$. Formally define $\mathcal{A}$ as

$$\mathcal{A} = \{A \in \mathcal{A}' \mid A(t) = A(T_A) \text{ for all } t \geq T_A\}.$$

Endow $\mathcal{A}$ with the σ-algebra which is the restriction of the σ-algebra on $\mathcal{A}'$ to $\mathcal{A}$. The map ϕ_H is a 1-1 measurable map from $\mathcal{F}$ onto $\mathcal{A}$, and, in fact, has an inverse (see Gill and Johansen [13]). We consider this inverse map next and briefly summarize its properties.

Let $A \in \mathcal{A}'$. Let $\{s_1, s_2, \dots\}$ be dense in $(0, \infty)$. For each n, let $s_1^{(n)} < \cdots < s_n^{(n)}$ be as before. Fix $s < t$. If $A(t) < \infty$, define the product integral of A by

$$\prod_{(s,t]}(1 - dA) = \lim_{n\to\infty} \prod_{s<s_i^{(n)}\le t} (1 - A(s_{i-1}^{(n)}, s_i^{(n)}])$$

where $A(a,b] = A(b) - A(a)$ for $a < b$. If $A(t) = \infty$ and $A(s) < \infty$, set $\prod_{(s,t]}(1-dA) = 0$. If $A(s) = \infty$, set $\prod_{(s,t]}(1 - dA) = 1$.

Theorem 1.2 (Gill and Johansen, 1990). *Let $A \in \mathcal{A}$. Then F given by*

$$F(t) = 1 - \prod_{(0,t]}(1 - dA)$$

is an element of $\mathcal{F}$. Furthermore,

$$A(t) = \int_{(0,t]} \frac{dF(s)}{F[s, \infty)}.$$

This result explains why $\mathcal{A}$ is referred to as the space of all cumulative hazard functions on $[0, \infty)$.

The following properties of the product integral are included to lend the reader a better understanding of the nature of the map ϕ_H and will be useful later. Details may be found in Gill and Johansen [13] and Hjort [14].

1.4.6. Like ϕ_H, the product integral also has an explicit expression.

$$\prod_{(0,t]}(1 - dA) = \prod_{s\le t}(1 - A\{s\}) \exp(-A^c(t))$$

where A^c is the continuous part of A.

1.4.7. Let Π^* be an independent increment process measure with sample paths almost surely in $\mathcal{A}'$. Then, for all t,

$$\mathcal{E}_{\Pi^*}\left(\prod_{(0,t]}(1 - dA)\right) = \prod_{(0,t]} (1 - d(\mathcal{E}_{\Pi^*}A)).$$

1.4.8. Let ρ_S denote the Skorokhod metric on $D[0,\infty)$ and let $\{A_n\}$ be a sequence in $\mathcal{A}$. Say that $\rho_S(A_n, A) \to 0$ for some $A \in \mathcal{A}$ as $n \to \infty$, if $\rho_S(A_n^T, A^T) \to 0$ for all $T > 0$ where A_n^T and A^T are restrictions of A_n and A to $[0,T]$. It may be shown, following Hjort ([14], Lemma A.2, pp. 1290-91), that if $\{A_n\}, A \in \mathcal{A}$ and $\rho_S(A_n, A) \to 0$, then $\phi_H^{-1}(A_n) \xrightarrow{w} \phi_H^{-1}(A)$. Thus, if $\mathcal{A}$ is endowed with the Skorokhod metric, then ϕ_H^{-1} is a continuous map.

The next result establishes the connection between NR priors and independent increment processes with non-decreasing paths via the map ϕ_H.

Theorem 1.3. *Let Π be a NR prior on $\mathcal{F}$. Then, under the measure Π^* on $\mathcal{A}$ induced by the map ϕ_H, $\{A(t) : t > 0\}$ has independent increments. Conversely, if Π^* is a probability measure on $\mathcal{A}$ such that the process $\{A(t) : t > 0\}$ has independent increments, then the measure induced on $\mathcal{F}$ by the map*

$$\phi_H^{-1} : A \mapsto 1 - \prod_{(0,t]}(1 - dA).$$

is neutral to the right.

A formal proof of this simple result does not appear anywhere in print, and hence, we provide one here.

Proof. First suppose that Π is NR on $\mathcal{F}$ and let $t_1 < \cdots < t_k$ be arbitrary points in $(0,\infty)$. Consider, as before, a dense set $\{s_1, s_2, \ldots\}$ in $(0,\infty)$. Let, for each n, $s_1^{(n)} < \cdots < s_n^{(n)}$ be as before.

Suppose n is large enough such that $s_n^{(n)} \geq t_k$. Then, for each $1 \leq i \leq k$, we have, with A_F^n as in 1.4.1 on page 14,

$$A_F^n(t_i) - A_F^n(t_{i-1}) = \sum_{t_{i-1} < s_j^{(n)} \le t_i} \frac{F(s_{j-1}^{(n)}, s_j^{(n)}]}{F(s_{j-1}^{(n)}, \infty)}$$

$$= \sum_{t_{i-1} < s_j^{(n)} \le t_i} \left(1 - \frac{\bar{F}(s_j^{(n)})}{\bar{F}(s_{j-1}^{(n)})}\right).$$

Since, for each n, $\bar{F}(s_1^{(n)})$, $\frac{\bar{F}(s_2^{(n)})}{\bar{F}(s_1^{(n)})}$, ..., $\frac{\bar{F}(s_n^{(n)})}{\bar{F}(s_{n-1}^{(n)})}$ are independent, $A_F^n(t_1)$, $A_F^n(t_2) - A_F^n(t_1)$, ..., $A_F^n(t_k) - A_F^n(t_{k-1})$ are also independent. Letting $n \to \infty$, we get that $A_F(t_1)$, $A_F(t_2) - A_F(t_1)$, ..., $A_F(t_k) - A_F(t_{k-1})$ are independent.

For the converse, suppose Π^* on $\mathcal{A}$ be such that, under Π^*, $\{A(t) : t > 0\}$ is an independent increment process. Again, let $t_1 < \cdots < t_k$ be arbitrary points in $(0, \infty)$. Then with $s_1^{(n)} < \cdots < s_n^{(n)}$ as before, let, for $1 \le i \le k$,

$$\bar{F}_A^n(t_i) = \prod_{s_j^{(n)} \le t_i} (1 - A(s_{j-1}^{(n)}, s_j^{(n)}]).$$

If $F_A = \phi_H^{-1}(A)$, then it follows from the definition of the product integral that $\bar{F}_A^{(n)}(t) \to \bar{F}_A(t)$ for all t, as $n \to \infty$. Now, observe that, for $1 \le i \le k$,

$$\frac{\bar{F}_A^n(t_i)}{\bar{F}_A^n(t_{i-1})} = \prod_{t_{i-1} < s_j^{(n)} \le t_i} (1 - A(s_{j-1}^{(n)}, s_j^{(n)}]).$$

Since $A(s_{j-1}^{(n)}, s_j^{(n)}], 1 \le j \le n$ are independent for each n, then so are $\frac{\bar{F}_A^n(t_i)}{\bar{F}_A^n(t_{i-1})}, 1 \le i \le k$. Consequently, we have independence in the limit, i.e. $\bar{F}_A(t_1)$, $\frac{\bar{F}_A(t_2)}{\bar{F}_A(t_1)}$, ..., $\frac{\bar{F}_A(t_k)}{\bar{F}_A(t_{k-1})}$ are independent. ■

We have noted earlier that Doksum was the first to observe a connection between NR priors and independent increment processes. He, however, considered a map slightly different from the product integral. Here we briefly turn our attention to his map.

Recall the function space $\mathcal{H}$ defined in (1.3.2).

Theorem 1.4 (Doksum, 1974). *A prior* Π *on* $\mathcal{F}$ *is NR if and only if* $\tilde{\Pi} \equiv \Pi \circ \phi_D^{-1}$ *is an independent increment process measure such that* $\tilde{\Pi}\{H \in \mathcal{H} : \lim_{t\to\infty} H(t) = \infty\} = 1$.

Let Π be a NR prior on $\mathcal{F}$. From what we have seen so far, the maps ϕ_D and ϕ_H yield independent increment process measures $\tilde{\Pi}$ and Π^*, respectively. Let the Lévy measures of $\tilde{\Pi}$ and Π^* be denoted $\tilde{\lambda}$ and λ^* respectively. The next proposition establishes a simple relationship between $\tilde{\lambda}$ and λ^*.

Proposition 1.1. *Suppose* $\tilde{\lambda}$ *and* λ^* *are as above. Then*

(1). For each t, $\tilde{\lambda}_t$ *is the distribution of* $x \mapsto -\log(1-x)$ *under the measure* λ_t^*, *and*

(2). For each t, λ_t^* *is the distribution of* $x \mapsto 1 - e^{-x}$ *under* $\tilde{\lambda}_t$,

where $\tilde{\lambda}_t$ *and* λ_t^* *are defined following 1.3.1 on page 12.*

Proof. The proposition is an easy consequence of the following, again easy, fact.

If $\omega \mapsto \mu(\cdot, \omega)$ is a $\mathbf{M}(\mathcal{X})$-valued random measure which is a Poisson process with parameter measure λ, then for any measurable function $g : \mathcal{X} \to \mathcal{X}$, the random measure $\omega \mapsto \mu(g^{-1}(\cdot), \omega)$ is a Poisson process with parameter measure $\lambda \circ g^{-1}$.

Now note that,

$$\begin{aligned}
\phi_D(F)(t) - \phi_D(F)(t-) &= -\log \frac{F(t,\infty)}{F[t,\infty)} \\
&= -\log\left\{1 - \frac{F\{t\}}{F[t,\infty)}\right\} \\
&= -\log[1 - (\phi_H(F)(t) - \phi_H(F)(t-))].
\end{aligned}$$

■

1.5 Support of NR Priors

Since Π on $\mathcal{F}$ corresponds, via ϕ_H, to an independent increment process measure Π^* and, since Π^* is described by its Lévy measure λ, it is natural to expect that properties of Π would be describable in terms of λ. The next theorem is a case in point.

Theorem 1.5. *Let Π be a NR prior on $\mathcal{F}$ and let Π^* be the measure on $\mathcal{A}$ induced by the map ϕ_H. If λ, the Lévy measure of Π^*, has support $[0,\infty) \times [0,1]$, then the support of Π is $\mathcal{F}$.*

Proof. Let $F_0 \in \mathcal{F}$ and let $A_0 = \phi_H(F_0)$. Let, for $\epsilon > 0$, $U = \{A : \rho_S^T(A_0, A) < \epsilon\}$ where ρ_S^T is the Skorokhod metric on $D[0,T]$, the space of all right-continuous functions with left limits on $[0,T]$. By 1.4.8 on page 18, every weak open neighborhood of F_0 contains a set of the form $\phi_H^{-1}(U)$. Hence it is enough to show that $\Pi^*(U) > 0$.

Recall, $\rho_S^T(A_0, A)$ is the infimum of all $\epsilon' > 0$ such that there exists a strictly increasing function α from $[0,T]$ onto $[0,T]$ for which

$$\sup_{t \le T} |\alpha(t) - t| < \epsilon'$$

and

$$\sup_{t \le T} |A_0(t) - A(\alpha(t))| < \epsilon'.$$

For any given $\delta > 0$ and T, let $0 = t_0 < t_1 < \cdots < t_n = T$ be chosen such that $A_0(t_i, t_{i+1}) < \delta$ for all $i = 0, \ldots, n-1$.

Let $a_i < t_i < b_i$ be continuity points of A_0 such that

$$a_i > t_i - \delta, \qquad A_0(a_i) > A_0(t_i-) - \delta/2, \quad i = 1, \ldots, n;$$

$$b_i < t_i + \delta, \qquad A_0(b_i) > A_0(t_i) + \delta/2, \quad i = 0, \ldots, n-1.$$

Set $a_0 = 0$ and $b_n = T$.

Consider $A \in \mathcal{A}$ such that, for every $0 \le i \le n$,

(i) $|A(a_i) - A_0(a_i)| < \delta$,

(ii) $|A(b_i) - A_0(b_i)| < \delta$,

(iii) there exists $a_i < s_i < b_i$ such that $|s_i - t_i| < \delta$ and $|A\{s_i\} - A\{t_i\}| < \delta$.

Suppose W be the set of all such $A \in \mathcal{A}$, i.e. let

$$W = \{A \in \mathcal{A} : A \text{ satisfies (i), (ii) and (iii) above}\}.$$

We will argue that $W \subseteq U$.

Let $A \in W$ and let $s_0, \ldots, s_n$ as in (iii). Let α be the function defined by

$$\alpha(x) = \begin{cases} s_i, & \text{for } x = t_i; \\ x, & \text{for } b_{i-1} \le x \le a_i; \\ \text{linear} & \text{for } a_i \le x \le t_i; \\ \text{linear} & \text{for } t_i \le x \le b_i. \end{cases}$$

Clearly $|\alpha(t) - t| < \delta$ for all $t \le T$.

Consider $t \in [b_{i-1}, a_i)$. Then

$$A(\alpha(t)) > A(b_{i-1}) > A_0(b_{i-1}) - \delta, \text{ and}$$

$$A_0(t) < A_0(a_i) < A_0(b_{i-1}) + \delta.$$

Therefore $A_0(t) - A(\alpha(t)) < 2\delta$.

Similarly

$$A(\alpha(t)) < A(a_i) < A_0(a_i) + \delta, \text{ and}$$

$$A_0(t) > A_0(b_{i-1}) > A_0(a_i) - \delta$$

thus, $A(\alpha(t)) - A_0(t) < 2\delta$.

From above, it follows that $|A(\alpha(t)) - A_0(t)| < 2\delta$ for $t \in [b_{i-1}, a_i)$.

Now suppose $t \in [a_i, t_i)$. Then, since

$$A(\alpha(t)) \geq A(a_i) > A_0(a_i) - \delta, \text{ and}$$

$$A_0(t) \leq A_0(t_i-) < A_0(a_i) + \delta/2,$$

we have $A_0(t) - A(\alpha(t)) < \frac{3}{2}\delta$. On the other hand,

$$A(\alpha(t)) \leq A(s_i-) \leq A(b_i) - A\{s_i\} < A_0(b_i) + \delta - A_0\{t_i\} + \delta$$
$$< (A_0(t_i) + \delta/2) + \delta - A_0\{t_i\} + \delta = A_0(t_i-) + \frac{5}{2}\delta$$

and $A_0(t) \geq A_0(a_i) > A_0(t_i-) - \delta/2$. Hence $A(\alpha(t)) - A_0(t) < 3\delta$.

It follows that $|A(\alpha(t)) - A_0(t)| < 3\delta$ for $t \in [a_i, t_i)$.

Finally, let $t \in [t_i, b_i)$. Observe that

$$A(\alpha(t)) \geq A(s_i) \geq A(a_i) - A\{s_i\} > A_0(a_i) + A_0\{t_i\} - 2\delta$$

and

$$A_0(t) \leq A_0(b_i) < A_0(t_i) + \delta/2 < A_0(t_i-) + A_0\{t_i\} + \delta/2 < A_0(a_i) + A_0\{t_i\} + \delta.$$

Therefore, $A_0(t) - A(\alpha(t)) < 3\delta$.

Also, since

$$A(\alpha(t)) < A(b_i) < A_0(b_i) + \delta, \text{ and}$$

$$A_0(t) \geq A_0(t_i) > A_0(b_i) - \delta/2,$$

we conclude $A(\alpha(t)) - A_0(t) < \frac{3}{2}\delta$.

Combining the last two inequalities we get $|A(\alpha(t)) - A_0(t)| < 3\delta$ for $t \in [t_i, b_i)$.

Now let $\delta \leq \epsilon/3$ and note that, for all $0 \leq t \leq T$,

$$|A(\alpha(t)) - A_0(t)| < \epsilon.$$

Consequently $A \in U$. This shows that $W \subseteq U$. We will now find a subset W_0 of W such that $\Pi^*(W_0) > 0$.

To this end, for δ as above and $i = 1, ..., n$, define the sets

$$E_i = [b_{i-1}, a_i) \times (A_0(t_{i-1}, t_i) - \frac{\delta}{4n}, A_0(t_{i-1}, t_i) + \frac{\delta}{4n})$$
$$G_i = [a_i, b_i) \times (A_0\{t_i\} - \frac{\delta}{4n}, A_0\{t_i\} + \frac{\delta}{4n})$$

and subsequently define

$$W_0 = \{A : \mu(E_i, A) = 1 \text{ and } \mu(G_i, A) = 1, i = 1, ..., n\}.$$

W_0 is clearly measurable and has positive Π^* probability since $\mu(E_i, A)$ and $\mu(G_i, A)$, $i = 1, ..., n$, are all independent Poisson with positive parameters. Hence, to conclude the proof, all we need to show is that $W_0 \subseteq W$. To see this, let $A \in W_0$ and observe that,

(a) for $i = 1, \ldots, n$,

$$|A(a_i) - A_0(a_i)| < |A(a_i) - A_0(t_i)| + \delta/2$$

$$\leq \sum_{j=1}^{i} |A(b_{j-1}, a_j) - A_0(t_{j-1}, t_j)| + \sum_{j=0}^{i-1} |A[a_j, b_j) - A_0\{t_j\}| + \delta/2$$

$$< \delta/4 + \delta/4 + \delta/2 = \delta.$$

(b) for $i = 0, \ldots, n-1$,

$$|A(b_i) - A_0(b_i)| < |A(b_i) - A_0(t_i)| + \delta/2$$

$$\leq \sum_{j=1}^{i} |A(b_{j-1}, a_j) - A_0(t_{j-1}, t_j)| + \sum_{j=0}^{i} |A[a_j, b_j) - A_0\{t_j\}| + \delta/2$$

$$< \delta/4 + \delta/4 + \delta/2 = \delta.$$

and, since $\mu(G_i, A) = 1$, letting $\{s_i\} = \{t \in [a_i, b_i) : A\{t\} > 0\}$, we get

(c) $A[a_i, b_i) = A\{s_i\}$ and consequently

$$|A\{s_i\} - A_0\{t_i\}| = |A[a_i, b_i) - A_0\{t_i\}| < \frac{\delta}{4n} < \delta$$

which concludes the proof.

■

Chapter 2

Posterior Properties and Characterizations

In discussing the specification of prior distributions on $\mathcal{F}$ for nonparametric problems, Ferguson [11] stated that a good prior should be such that: (1) its support, with respect to some suitable topology on $\mathcal{F}$, should be "large", and (2) the posterior distribution given a sample from the true distribution should be analytically manageable. In Section 1.5 we addressed the issue of support for NR priors. In the present chapter we focus attention on their posterior.

Doksum [8] demonstrated that NR priors are a conjugate family of priors on $\mathcal{F}$. Later Ferguson and Phadia [12] established that these priors are also conjugate in the presence of right-censored data. Section 2.1 presents Doksum's result on the conjugacy of NR priors and its extension to the case of right-censored observations due to Ferguson and Phadia. An explicit expression for the posterior, in terms of the Lévy representation, due to Hjort, is provided. In Section 2.2 we observe that, if Π is

NR, then the posterior distribution of $F(t)$ given n observations depends on the exact observations preceeding t and just the number of observations beyond t. We show that this property of the posterior actually characterizes NR priors. We then use the result to provide another characterization via de Finetti's Theorem. The result is then extended to the case of right-censored data in Section 2.3.

2.1 Posterior Distribution

Consider the standard Bayesian set-up considered in Section 1.1, i.e. let Π be a prior and let F be a random element of $\mathcal{F}$ distributed as Π. Given F, let $X_1, X_2, \ldots$ be i.i.d. F. For each $n \geq 1$, denote by $\Pi_{X_1,\ldots,X_n}$ a version of the posterior distribution, i.e. the conditional distribution of F given $X_1, \ldots, X_n$. We will now state Doksum's result on the conjugacy of NR priors and provide an alternate proof for it.

Theorem 2.1 (Doksum, 1974). *Let Π be NR. Then $\Pi_{X_1,\ldots,X_n}$ is also NR.*

Remark 1. Hereafter, all equalities involving conditional probabilities, in particular posterior probabilities, and conditional expectations are to be interpreted in the almost sure sense.

Notation 1. For $n \geq 1$, define the observation process $N_n(.)$ as follows:

$$N_n(t) = \sum_{i \leq n} I_{(0,t]}(X_i) \quad \text{for all } t > 0.$$

For every $n \geq 1$, let $N_n(0) \equiv 0$. Observe that $N_n(.)$ is right-continuous on $[0, \infty)$. Let

$$\mathcal{G}_{t_1 \ldots t_k} = \sigma \left\{ \bar{F}(t_1), \frac{\bar{F}(t_2)}{\bar{F}(t_1)}, \ldots, \frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})} \right\}.$$

Thus $\mathcal{G}_{t_1 \ldots t_k}$ denotes the collection of all sets of the form

$$D = \left\{ (\bar{F}(t_1), \frac{\bar{F}(t_2)}{\bar{F}(t_1)}, \ldots, \frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})}) \in C \right\}$$

where $C \in \mathcal{B}^k_{[0,1]}$.

Proof of the theorem. Fix $k \geq 1$ and let $t_1 < t_2 < \cdots < t_k$ be arbitrary points in $(0, \infty)$. Denote by $\mathbb{Q}$ the set of all rationals in $(0, \infty)$ and let $\mathbb{Q}' = \mathbb{Q} \cup \{t_1, \ldots, t_k\}$. Let $\{s_1, s_2, \ldots\}$ be an enumeration of $\mathbb{Q}'$. Observe that, for large enough m, $\{t_1, \ldots, t_k\} \subset \{s_1, \ldots, s_m\}$.

For such an m, let $s_1^{(m)} < \cdots < s_m^{(m)}$ be an ordering of $\{s_1, \ldots, s_m\}$. Let $Y_i^{(m)} = \frac{\bar{F}(s_i^{(m)})}{\bar{F}(s_{i-1}^{(m)})}$ and, under Π, let $\Pi_i^{(m)}$ denote the distribution of $Y_i^{(m)}$.

Let $n_1 \leq \cdots \leq n_m$. Then, given $\{N_n(s_1^{(m)}) = n_1, \ldots, N_n(s_m^{(m)}) = n_m\}$, the posterior density of $(Y_1^{(m)}, \ldots, Y_m^{(m)})$ is written as

$$\begin{aligned} f_{Y_1^{(m)}, \ldots, Y_m^{(m)}}(y_1, \ldots, y_m) &= \frac{\prod_{i=1}^m (1 - y_i)^{n_i - n_{i-1}} y_i^{n - n_i}}{\int \prod_{i=1}^m (1 - y_i)^{n_i - n_{i-1}} y_i^{n - n_i} d\Pi_i^{(m)}(y_i)} \\ &= \prod_{i=1}^m \frac{(1 - y_i)^{n_i - n_{i-1}} y_i^{n - n_i}}{\int (1 - y_i)^{n_i - n_{i-1}} y_i^{n - n_i} d\Pi_i^{(m)}(y_i)}. \end{aligned}$$

This shows that $(Y_1^{(m)}, \ldots, Y_m^{(m)})$ are independent under the posterior given $\{N_n(s_1^{(m)}), \ldots, N_n(s_m^{(m)})\}$. Hence,

$$\frac{\bar{F}(t_i)}{\bar{F}(t_{i-1})} = \prod_{t_{i-1} < s_j^{(m)} \leq t_i} \frac{\bar{F}(s_j^{(m)})}{\bar{F}(s_{j-1}^{(m)})}, \qquad i = 1, \ldots, k$$

are also independent under the posterior given the same information.

Now, by the right-continuity of $N_n(\cdot)$ we have, as $n \to \infty$,

$$\sigma\{N_n(s_j), j \leq m\} \uparrow \sigma\{N_n(t), t \geq 0\} \equiv \sigma(X_1, \ldots, X_n).$$

Hence, for any $A \in \mathcal{G}_{t_1 \dots t_k}$, by the Martingale Convergence theorem, we have

$$\Pi(A \mid N_n(s_1^{(m)}), \dots, N_n(s_m^{(m)})) \to \Pi(A \mid X_1, \dots, X_n) \quad \text{almost surely.}$$

In other words, one concludes that, the distribution of ($\bar{F}(t_1), \frac{\bar{F}(t_2)}{\bar{F}(t_1)}, \dots, \frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})}$) under the posterior given $N_n(s_1^{(m)}), \dots, N_n(s_m^{(m)})$ converges to its distribution under the posterior given $X_1, \dots, X_n$, as $m \to \infty$.

Since $\bar{F}(t_1), \frac{\bar{F}(t_2)}{\bar{F}(t_1)}, \dots, \frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})}$ are independent given $\sigma(N_n(s_1^{(m)}), \dots, N_n(s_m^{(m)}))$, independence also holds in the limit. ■

Doksum provides a representation of the posterior $\Pi_{X_1, \dots, X_n}$ of a general NR prior and shows that, unlike the Dirichlet, the distribution of $\frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})}$, under the posterior, depends on the exact values of the observations in $(0, t_k]$ and not just on the number of observations in that interval.

Ferguson and Phadia [12] extend Doksum's result in the case of inclusively and exclusively right-censored observations. Let x be a real number in $(0, \infty)$. Given a cdf $F \in \mathcal{F}$, an observation X from F is said to be exclusively right-censored if we only know $X \geq x$ and inclusively right-censored if we know $X > x$. We state their result next and reproduce the proof for the sake of completeness.

Theorem 2.2 (Ferguson and Phadia, 1979). *Let F be a random cdf neutral to the right. Let X be a sample of size one from F, and let x be a number in $(0, \infty)$.*

(a) the posterior distribution of F given $X > x$ is NR,

(b) the posterior distribution of F given $X \geq x$ is NR.

Proof. Let $t_1 < \cdots < t_k$ be arbitrary points in $(0, \infty)$ including x. Let $t_j = x$ and let $t_0 = 0$. Define

$$Y_l = \begin{cases} \frac{\bar{F}(t_{l+1})}{\bar{F}(t_l)} & \text{for } l = 0, \ldots, j-2; \\ \frac{\bar{F}(t_l)}{\bar{F}(t_{l-1})} & \text{for } l = j+1, \ldots, k; \\ \frac{\bar{F}(t_{l+1}-)}{\bar{F}(t_l)} & \text{for } l = j-1; \\ \frac{\bar{F}(t_l)}{\bar{F}(t_l-)} & \text{for } l = j. \end{cases}$$

Under the prior, $Y_0, \ldots, Y_k$ are independent random variables with joint density, say,

$$f_{Y_0 \ldots Y_k}(y_0, \ldots, y_k) = \prod_{i=0}^{k} f_{Y_i}(y_i)$$

with respect to some convenient product measure. Given F, the probability of $X > x$ is written as

$$\bar{F}(x) = \prod_{i=0}^{j} Y_i$$

hence, the posterior density of $Y_0, \ldots, Y_k$ given $X > x$ is written as

$$f_{Y_0 \ldots Y_k}(y_0, \ldots, y_k \mid X > x) = C. \left(\prod_{i=0}^{j} y_i \, f_{Y_i}(y_i) \right) \cdot \left(\prod_{i=j+1}^{k} f_{Y_i}(y_i) \right)$$

where C is a normalizing constant depending on x.

Thus Y_i are independent under the posterior as well, and hence, the posterior is NR.

In the same way, since, given F, the probability of $X \geq x$ is

$$\bar{F}(x-) = \prod_{i=0}^{j-1} Y_i,$$

the posterior density of $Y_0, \ldots, Y_k$ given $X \geq x$ is

$$f_{Y_0 \ldots Y_k}(y_0, \ldots, y_k \mid X \geq x) = C. \left(\prod_{i=0}^{j-1} y_i f_{Y_i}(y_i) \right) \cdot \left(\prod_{i=j}^{k} f_{Y_i}(y_i) \right)$$

where C is again a normalizing constant depending on x.

Thus Y_i are again independent under the posterior given $X \geq x$ and hence it is NR. ■

Suppose Π is a NR prior for which the corresponding independent increment process has $b \equiv 0$ and a Lévy measure of the form

$$d\lambda = a(s, u)\, ds\, dA_0(u) \tag{2.1.1}$$

for some non-decreasing function A_0 and a positive function a. In view of Theorems 2.1 and 2.2, the posterior, given some possibly right-censored observations, is also NR. The following theorem, taken from Hjort [14] and due to Ferguson and Phadia, provides an explicit expression for the components of the independent increment process corresponding to the posterior.

Let $A_0 \in \mathcal{A}$ and let $\mathbf{M} = \{t_1, \ldots, t_k\}$ be the set of fixed jumps of A_0.

Theorem 2.3 (Hjort, 1990). *Let F be a NR random cdf such that A_F is an independent increment process with components $b \equiv 0$, Lévy measure given by equation (2.1.1), fixed points of jumps belonging to $\mathbf{M}$ and density for the jump $A_F\{t_j\}$ denoted by f_j, $j = 1, \ldots, k$. Denote by $\mathbf{M}^*$, $\{f_j^*\}$ and a^* the posterior parameters. Then the following hold:*

(1). Given $X > x$, the posterior parameters are given by

(a) $\mathbf{M}^* = \mathbf{M}$

(b)

$$f_j^*(s) = \begin{cases} c(1-s)f_j(s) & \text{if } t_j \leq x; \\ \\ f_j(s) & \text{if } t_j > x. \end{cases}$$

where c is a normalizing constant; and

(c)

$$a^*(s,u) = \begin{cases} (1-s)a(s,u) & \text{if } u \leq x; \\ \\ a(s,u) & \text{if } u > x. \end{cases}$$

(2). Given $X = x, x \in \mathbf{M}$, *the posterior parameters are given by*

(a) $\mathbf{M}^* = \mathbf{M}$

(b)

$$f_j^*(s) = \begin{cases} c(1-s)f_j(s) & \text{if } t_j \leq x; \\ \\ csf_i(s) & \text{if } t_i = x; \\ \\ f_j(s) & \text{if } t_j > x. \end{cases}$$

where c is again a normalizing constant; and

(c) $a^*(s,u)$ *as in (1).*

(3). Given $X = x, x \notin \mathbf{M}$, *the posterior parameters are given by*

(a) $\mathbf{M}^* = \mathbf{M}$

(b)

$$f_j^*(s) = \begin{cases} c(1-s)f_j(s) & \text{if } t_j < x; \\ f_j(s) & \text{if } t_j > x. \end{cases}$$

The new jump $A\{x\}$ has density $f^*(s) = c\,s\,a(x,s)$. where the c's are normalizing constants; and

(c) $a^*(s,u)$ *as in (1).*

Of interest to us is the following consequence.

Theorem 2.4. *Let A be a Lévy process with no fixed jumps, i.e. $\mathbf{M} = \phi$. Suppose its Lévy measure is given by $d\lambda = a(s,u)ds\,dA_0(u)$ for some $A_0 \in \mathcal{A}$. Then the posterior distribution given $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ is a Lévy process with parameters*

(a) $\mathbf{M}^* = \{s_1 < \cdots < s_p\}$, *the distinct elements of* $\{x_1, \ldots, x_n\}$;

(b) $f_j^*(s) = c\,s^{N_n(s_j)-N_n(s_{j-1})}(1-s)^{n-N_n(s_j)}a(s_j,u)$ *where N_n is as before; and*

(c) $a^*(s,u) = (1-s)^{n-N_n(u-)}a(s,u)$.

The proof follows by repeated application of Theorem 2.3.

2.2 Characterizations of NR Priors

For each T in $\mathbb{R}$, let $N_n^T = \{N_n(t) : t \leq T\}$. As before, for a prior Π, $\Pi_{X_1,\ldots,X_n}$ stands for its posterior given $X_1, \ldots, X_n$, and $\Pi_{N_n^T}$ for its posterior given $\{N_n(t) : t \leq T\}$. In general, for any family of random variables $\tilde{N}$ depending on $X_1, \ldots, X_n$, and any mea

ν_S

Th

the

(1

(2

(3

Re

the

eve

eve

Pr

Thi

L

given

measure ν, the conditional distribution corresponding to ν, given $\tilde{N}$, will be denoted $\nu_{\tilde{N}}$.

Theorem 2.5. *Suppose* Π *is a prior such that,* $\Pi\{F : 0 < F(t) < 1 \; \forall \, t\} = 1$. *Then the following are equivalent:*

(1). Π *is NR.*

(2). For all $T \in \mathbb{R}$, *the distributions of* $\{F(t) : t \leq T\}$ *under* $\Pi_{X_1,\ldots,X_n}$ *and* $\Pi_{N_n^T}$ *are the same.*

(3). For all $t_1 < \cdots < t_k < t_{k+1}$, *the distributions of* $(F(t_1),\ldots,F(t_k))$ *under* $\Pi_{N_n(t_1),\ldots,N_n(t_{k+1})}$ *and* $\Pi_{N_n(t_1),\ldots,N_n(t_k)}$ *are the same.*

Remark 2. Note that (2) is a statement that holds almost everywhere with respect to the marginal distribution of $X_1,\ldots,X_n$. However, under our assumption, (3) holds everywhere. Interpret (2) as, there exists a version of the L.H.S. equal to the R.H.S. everywhere.

Proof. (1) $\Rightarrow$ (2) is well known. See for instance Doksum [8].

To see (3) $\Rightarrow$ (1), we will show that, for fixed $0 = t_0 < t_1 < \cdots < t_{k+1}$,

$$\{\bar{F}(t_1),\ldots,\bar{F}(t_k)\} \quad \text{is independent of} \quad \frac{\bar{F}(t_{k+1})}{\bar{F}(t_k)}.$$

This would then show that

$$\left\{\bar{F}(t_1), \frac{\bar{F}(t_2)}{\bar{F}(t_1)},\ldots,\frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})}\right\} \quad \text{is independent of} \quad \frac{\bar{F}(t_{k+1})}{\bar{F}(t_k)}.$$

Let n be a positive integer. Since the posterior densities of $(F(t_1),\ldots,F(t_k))$, given $\{N_n(t_1) = 0,\ldots,N_n(t_k) = 0\}$, and $\{N_n(t_1) = 0,\ldots, N_n(t_k) = 0, N_n(t_{k+1}) = 0\}$, are equal, the same holds for the posterior densities of $(\bar{F}(t_1), \frac{\bar{F}(t_2)}{\bar{F}(t_1)}, \ldots, \frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})})$ given the same information. This gives

$$\frac{\prod_{i=1}^{k}[\frac{\bar{F}(t_i)}{\bar{F}(t_{i-1})}]^n}{\int_{\mathcal{F}}\prod_{i=1}^{k}[\frac{\bar{F}(t_i)}{\bar{F}(t_{i-1})}]^n d\Pi(F)} = \frac{\mathcal{E}_\Pi\left[\prod_{i=1}^{k+1}[\frac{\bar{F}(t_i)}{\bar{F}(t_{i-1})}]^n \mid \bar{F}(t_1),\ldots,\bar{F}(t_k)\right]}{\int_{\mathcal{F}}\prod_{i=1}^{k+1}[\frac{\bar{F}(t_i)}{\bar{F}(t_{i-1})}]^n d\Pi(F)}$$

and hence that

$$\mathcal{E}_\Pi\left[[\frac{\bar{F}(t_{k+1})}{\bar{F}(t_k)}]^n \mid \bar{F}(t_1),\ldots,\bar{F}(t_k)\right] = constant, \qquad a.e. \quad \Pi.$$

Since this holds for all n, $\frac{\bar{F}(t_{k+1})}{\bar{F}(t_k)}$ is independent of $\{\bar{F}(t_1),\ldots,\bar{F}(t_k)\}$. Repeating the argument with k replaced by $j = k-1, k-2, \ldots, 1$ the result follows.

We now prove (2) $\Rightarrow$ (3). For this, note that since, by (2),

$$\{F(t_1),\ldots F(t_k)\} \underset{\{N_n(t):t\le t_k\}}{\perp} \{N_n(t) : t \geq 0\},$$

it follows that

$$\{F(t_1),\ldots F(t_k)\} \underset{\{N_n(t):t\le t_k\}}{\perp} \{N_n(t_1),\ldots N_n(t_{k+1})\}.$$

Now, for any measurable function g of $(F(t_1),\ldots F(t_k))$, the conditional expectation $\mathcal{E}[g \mid N_n(t_1),\ldots,N_n(t_{k+1})]$ is measurable both with respect to the σ-algebra generated by $\{N_n(t_1),\ldots,N_n(t_{k+1})\}$ and $\sigma\{N_n(t) : t \leq t_k\}$ and is hence measurable with respect to their intersection. But this intersection is precisely $\sigma\{N_n(t_1),\ldots,N_n(t_k)\}$ which concludes the proof. ■

We now provide a characterization of NR priors via de Finetti's Theorem, which states that, if $X_1, X_2, \ldots$ is a sequence of exchangeable random variables under a measure μ, then there is a unique prior Π on $\mathcal{F}$ such that, for all n,

$$\mu\{X_1 \in B_1,\ldots,X_n \in B_n\} = \int_{\mathcal{F}}\prod_{i=1}^{n} F(B_i)d\Pi(F).$$

The following result gives conditions on μ such that the corresponding de Finetti prior Π is NR.

Theorem 2.6. *Assume that, for every $t > 0$, as $n \to \infty$*

$$\mu\{X_1 < t, \ldots, X_n < t\} \to 0 \quad \textit{and} \quad \mu\{X_1 > t, \ldots, X_n > t\} \to 0.$$

Then the following are equivalent:

(1). Π is NR.

(2). For all $t \in \mathbb{R}$ and $n \geq 1$,

$$\mu_{X_1,\ldots,X_n}\{X_{n+1} > t\} = \mu_{N_n^t}\{X_{n+1} > t\}.$$

Proof. In view of Theorem 2.5, (2) immediately follows from (1), since

$$\mu_{X_1 \ldots X_n}\{X_{n+1} > t\} = \int_{\mathcal{F}} \bar{F}(t) d\Pi_{X_1,\ldots,X_n}(F)$$

and

$$\mu_{N_n^t}\{X_{n+1} > t\} = \int_{\mathcal{F}} \bar{F}(t) d\,\Pi_{N_n^t}(F).$$

To see (2) $\Rightarrow$ (1), fix $S \in \mathbb{R}$ and define $T(x) = xI_{(0,S]}(x) + (S+1)I_{(S,\infty)}(x)$. Then (2) implies, for every positive integer m, that

$$X_1, \ldots, X_n \perp_{N_n^S} T(X_{n+1}), \ldots, T(X_{n+m}). \tag{2.2.2}$$

To prove this claim, consider $0 = t_0 < t_1 < \cdots < t_k < t_{k+1} = S$. Since $\sigma\{N_n^{t_1}\} \subseteq \cdots \subseteq \sigma\{N_n^{t_{k+1}}\} = \sigma\{N_n^S\}$, by (2) each of the events $\{X_{n+1} > t_i\}$, $i = 1, \ldots, k+1$, is conditionally independent of $X_1, \ldots, X_n$, given N_n^S, under the measure μ. Hence the random variable $T_k(X_{n+1}) = \sum_{i=1}^{k+1} t_i I_{(t_{i-1},t_i]}(X_{n+1}) + (S+1)I_{(S,\infty)}(X_{n+1})$ is conditionally independent of $X_1, \ldots, X_n$, given N_n^S, under μ. Letting $t_1, \ldots, t_k$ run through a countable dense set in $(0, S]$, we get that

$$X_1, \ldots, X_n \underset{N_n^S}{\perp} T(X_{n+1}).$$

A simple induction argument then yields the claim (2.2.2).

Now, fix $t_1 < \cdots < t_k$. Given integers $n_1, \ldots, n_k$, set $m = n_1 + \cdots + n_k$ and, given $X_1, \ldots, X_n$, consider the predictive probability of the event: "of the next m observations, n_i fall in $(t_{i-1}, t_i]$ for $i = 1, \ldots, k$."

Since the event $\{X_{n+j} \in (t_{i-1}, t_i]\}$ is the same as the event $\{T(X_{n+j}) \in (t_{i-1}, t_i]\}$, for all $i = 1, \ldots, k$, and $S \geq t_k$, we have

$$\int_{\mathcal{F}} [1 - \bar{F}(t_1)]^{n_1} [\bar{F}(t_1) - \bar{F}(t_2)]^{n_2} \ldots [\bar{F}(t_{k-1}) - \bar{F}(t_k)]^{n_k} d\Pi_{X_1,\ldots,X_n}(F)$$
$$= \int_{\mathcal{F}} [1 - \bar{F}(t_1)]^{n_1} [\bar{F}(t_1) - \bar{F}(t_2)]^{n_2} \ldots [\bar{F}(t_{k-1}) - \bar{F}(t_k)]^{n_k} d\Pi_{N_n^S}(F).$$

This shows that the distribution of $\bar{F}(t_1), \bar{F}(t_1) - \bar{F}(t_2), \ldots, \bar{F}(t_{k-1}) - \bar{F}(t_k)$, given $X_1, \ldots, X_n$, is the same as that given N_n^S, for all $S \geq t_k$. Since $\bar{F}(t_1), \ldots, \bar{F}(t_k)$ is a function of these quantities, the same will be true for $\bar{F}(t_1), \ldots, \bar{F}(t_k)$. Hence (1) follows from Theorem 2.5. ■

In a recent work, yet to appear, Walker and Muliere [27] also obtained a similar result. Their condition on μ is expressed in terms of the expected instantaneous hazard rate under the de Finetti prior and corresponding posteriors. When the set of values for the observations is finite, they provide an explicit condition on the predictive distributions corresponding to μ. The above results and a similar characterization for Tailfree priors may be found in Dey, Draghici and Ramamoorthi [5].

2.3 NR Priors from Censored Observations

We will now extend Theorem 2.5 and provide a characterization of neutral to right priors in terms of their posterior distributions in the presence of right-censored observations.

Suppose, as before, that $F \sim \Pi$ and given F, $X_1, X_2, \ldots$ are independent and identically distributed as F. Here $X_1, X_2, \ldots$ are thought of as survival times. Let $c_1, c_2, \ldots$ be constants. These are our censoring times. For each i, we only get to observe:

$$Z_i = \min(X_i, c_i) \quad \text{and} \quad \Delta_i = I_{\{X_i \le c_i\}}.$$

Therefore, $\Delta_i = 0$ means Z_i is a right-censored observation.

Define the observation processes:

$$N_i^{(n)}(t) = \sum_{j=1}^{n} I_{\{Z_j \le t, \Delta_j = i\}}, \qquad i = 0, 1$$

and let $N^{(n)}(t) = \left(N_1^{(n)}(t), N_0^{(n)}(t)\right)$ and also let $A_n^T = \{t \le T : N_0^{(n)}(t) - N_0^{(n)}(t-) > 0\}$.

Under this set-up, the following theorem characterizes NR priors.

Theorem 2.7. *Suppose* Π *is a prior such that,* $\Pi\{F : 0 < F(t) < 1 \;\forall\, t\} = 1$. *Then the following are equivalent:*

(1). Π *is NR.*

(2). For all $T \in \mathbb{R}$ *and* $n \geq 1$, *the distributions of* $\{F(t) : t \leq T\}$ *under* $\Pi_{\{N^{(n)}(t):t>0\}}$ *and* $\Pi_{\{N^{(n)}(t):t\leq T\}}$ *are the same.*

(3). For all $n \geq 1$, $k \geq 1$ *and* $t_1 < \cdots < t_k < t_{k+1}$, *the distribution of* $\bar{F}(t_1), \ldots, \bar{F}(t_k)$ *under the posterior given* $\{N_1^{(n)}(t_1), \ldots, N_1^{(n)}(t_{k+1})\}$, $\{N^{(n)}(t) : t \leq t_k\}$ *is the same as that under the posterior given* $\{N_1^{(n)}(t_1), \ldots, N_1^{(n)}(t_k)\}$, $\{N^{(n)}(t) : t \leq t_k\}$.

Proof. (1) $\Rightarrow$ (2) is due to Ferguson and Phadia [12]. We omit the proof here.

To prove (2) $\Rightarrow$ (3) note that, by (2), $\{F(t_1), \ldots F(t_k)\}$ is conditionally independent of $\{N^{(n)}(t) : t \geq 0\}$ given $\{N^{(n)}(t) : t \leq t_k\}$. Therefore $\{F(t_1), \ldots F(t_k)\}$ is conditionally independent of $\{N_1^{(n)}(t_1), \ldots, N_1^{(n)}(t_{k+1})\}$ and $\{N_0^{(n)}(t) : t \leq t_k\}$ given $\{N^{(n)}(t) : t \leq t_k\}$.

Then, for any measurable function g of $(F(t_1), \ldots F(t_k))$,

$$\mathcal{E}[g \mid N_1^{(n)}(t_1), \ldots, N_1^{(n)}(t_{k+1})]$$

is measurable both with respect to the $\sigma\{N_1^{(n)}(t_1), \ldots, N_1^{(n)}(t_{k+1})\} \vee \sigma\{N_0^{(n)}(t) : t \leq t_k\}$ and $\sigma\{N_n(t) : t \leq t_k\}$. Hence it is also measurable with respect to their intersection, which is $\sigma\{N_1^{(n)}(t_1), \ldots, N_1^{(n)}(t_k)\} \vee \sigma\{N_0^{(n)}(t) : t \leq t_k\}$.

To see (3) $\Rightarrow$ (1), we will show that for arbitrary $t_1 < \cdots < t_{k+1}$,

$$\left\{\bar{F}(t_1), \ldots, \bar{F}(t_k)\right\} \quad \text{is independent of} \quad \frac{\bar{F}(t_{k+1})}{\bar{F}(t_k)}.$$

Fix $y_1 < \cdots < y_r \leq t_k$. Let $A_n^{t_k} = \{y_1, \ldots, y_r\}$ and let $s_1 < \cdots < s_{k+r+1}$ be points such that

$$A_n^{t_k} \cup \{t_1, \ldots, t_{k+1}\} = \{s_1 < \cdots < s_{k+r+1}\}$$

Note that $s_{k+r} = t_k$ and $s_{k+r+1} = t_{k+1}$. Now, for integers $m_1 \leq \cdots \leq m_r$, consider the set

$$B = \{N_0^{(n)}(t) = \sum_{j=1}^{r}(m_j - m_{j-1})I_{[y_j,\infty)}(t), t \leq t_k\}$$

Thus, on B, $(m_j - m_{j-1})$ observations are censored at the point y_j, $j = 1, \ldots, r$. Let $C = \{N_1^{(n)}(s_j) = 0, j = 1, \ldots, k+r\}$ and let $C' = \{N_1^{(n)}(s_j) = 0, j = 1, \ldots, k+r+1\}$.

Since the posterior density of $(F(t_1), \ldots, F(t_k))$ given $C \cap B$ and the posterior density given $C' \cap B$ are equal, then so are the densities of $\left(\bar{F}(t_1), \frac{\bar{F}(t_2)}{\bar{F}(t_1)}, \ldots, \frac{\bar{F}(t_k)}{\bar{F}(t_{k-1})}\right)$. This gives,

$$\frac{\prod_{i=1}^{k+r}[\frac{\bar{F}(s_i)}{\bar{F}(s_{i-1})}]^{n-m_{i-1}}}{\int_{\mathcal{F}}\prod_{i=1}^{k+r}[\frac{\bar{F}(s_i)}{\bar{F}(s_{i-1})}]^{n-m_{i-1}}d\Pi(F)} = \frac{\mathcal{E}_\Pi\left[\prod_{i=1}^{k+r+1}[\frac{\bar{F}(s_i)}{\bar{F}(s_{i-1})}]^{n-m_{i-1}} \mid \bar{F}(s_1), \ldots, \bar{F}(s_{k+r})\right]}{\int_{\mathcal{F}}\prod_{i=1}^{k+r+1}[\frac{\bar{F}(s_i)}{\bar{F}(s_{i-1})}]^{n-m_{i-1}}d\Pi(F)}$$

and hence that

$$\mathcal{E}_\Pi\left[[\frac{\bar{F}(s_{k+r+1})}{\bar{F}(s_{k+r})}]^{n-m_k} \mid \bar{F}(s_1), \ldots, \bar{F}(s_{k+r})\right] = constant \qquad a.e. \quad \Pi.$$

Since this holds for all $n - m_k$, $\frac{\bar{F}(t_{k+1})}{\bar{F}(t_k)}$ is independent of $\{\bar{F}(t_1), \ldots, \bar{F}(t_k)\}$. Repeating the argument with k replaced by $k-1, k-2, \ldots, 1$ the result follows. ■

Chapter 3

Beta Processes

Beta processes, introduced by Hjort [14], are independent increment processes with sample paths almost surely in $\mathcal{A}$. As such, the process measures for beta processes, hereafter referred to as beta process priors on $\mathcal{A}$, induce NR prior distributions on $\mathcal{F}$ via the ϕ_H^{-1} transform. We will call the induced priors beta priors on $\mathcal{F}$.

Apart from Ferguson's well known Dirichlet process priors, beta priors constitute the first specific examples of a family of NR priors on $\mathcal{F}$. Another family of NR priors are the beta-Stacy priors introduced and named by Walker and Muliere [26]. However, as we shall see in this chapter, these two families are identical.

The organization of this chapter is as follows. Section 3.1 serves as an introduction to beta processes and discusses the construction of beta priors on $\mathcal{F}$. Section 3.2 provides some posterior properties and establishes weak consistency of beta priors. Section 3.3 considers Dirichlet process priors for distributions on the space of right-censored observations. We show that suitable such Dirichlet process priors induce independent beta priors on the space of lifetime and censoring time distributions.

Section 3.4 then shows that beta processes with distinct parameters tend to be mutually singular. We conclude the chapter in Section 3.5 by discussing the beta-Stacy priors and establishing that they are simple reparameterizations of beta priors on $\mathcal{F}$.

3.1 Definition and Construction

Let A_0 be a hazard function with finitely many jumps. Let $t_1, \ldots, t_k$ be the jump-points of A_0. Let $c(\cdot)$ be a piecewise continuous non-negative function on $[0, \infty)$ and let A_0^c denote the continuous part of A_0. Let $A_0(t) < \infty$ for all t.

Definition 3.1. *An independent increment process A is said to be a beta process with parameters $c(.)$ and $A_0(.)$, written $A \sim beta(c, A_0)$, if the following holds: A has Lévy representation as in Theorem 1.1 with*

(1). $\mathbf{M} = \{t_1, \ldots, t_k\}$ and the jump-size at any t_j given by

$$Y_j \equiv A\{t_j\} \sim beta(c(t_j)A_0\{t_j\}, c(t_j)(1 - A_0\{t_j\}));$$

(2). Lévy measure given by

$$\lambda(ds\, du) = c(s)u^{-1}(1-u)^{c(s)-1} du\, dA_0^c(s)$$

for $0 \leq s < \infty, 0 < u < 1$; and for which

(3). $b(t) \equiv 0$ for all $t > 0$.

The existence of such a process with sample paths almost surely in $\mathcal{H}$ is guaranteed by the Lévy representation theorem (Theorem 1.1), because

$$0 < \iint\limits_{\substack{0<s\le t\\ 0<u<1}} \frac{u}{1+u}\lambda(ds\,du)$$

$$= \int_0^t \int_0^1 \frac{u}{1+u} c(s) u^{-1}(1-u)^{c(s)-1} du\, dA_0^c(s) < \infty.$$

Hjort provides a way to construct these priors as weak limits of time-discrete processes on $\mathcal{A}$ and shows that the sample paths are almost surely in $\mathcal{A}$. However, our focus is on $\mathcal{F}$ and hence we provide a construction of the induced prior on $\mathcal{F}$ as a weak limit of priors sitting on a discrete set of points on $(0,\infty)$. As mentioned above a Lévy process with the given Lévy measure exists with paths in $\mathcal{H}$. The important task is to ensure that the time-discrete priors converge to a prior on $\mathcal{F}$ and that the limiting prior corresponds to the given Lévy process. The construction is as follows.

Let $F_0 \in \mathcal{F}$ and, to start with, assume that it is continuous. Let $A_0 = \phi_H(F_0)$ be the cumulative hazard function corresponding to F_0.

Let $\mathbb{Q}$ be a countable dense set in $(0,\infty)$, enumerated as $\{s_1, s_2, \dots\}$. For each $n \ge 1$, let $\{\, s_1^{(n)} < \cdots < s_n^{(n)} \,\}$ be an ordering of $s_1, \dots, s_n$. Construct a prior Π_n on $\mathcal{F}_{s_1,\dots,s_n}$ as in Example 1 by requiring that, under Π_n,

$$V_i^{(n)} \sim \text{beta}\left(c(s_{i-1}^{(n)}) \frac{\bar{F}_0(s_i^{(n)})}{\bar{F}_0(s_{i-1}^{(n)})}, c(s_{i-1}^{(n)}) \left(1 - \frac{\bar{F}_0(s_i^{(n)})}{\bar{F}_0(s_{i-1}^{(n)})}\right)\right) \quad \text{for } 1 \le i \le n-1. \quad (3.1.1)$$

Let $V_n^{(n)} \equiv 1$ and let F be a random cdf, such that, under Π_n,

$$\mathcal{L}(\bar{F}(t)) = \mathcal{L}\left(\prod_{s_i^{(n)} \le t} (1 - V_i^{(n)}) \right) \qquad \text{for all } t > 0.$$

Theorem 3.1. $\{\Pi_n\}_{n\ge 1}$ *converges weakly to a NR prior* Π *on* $\mathcal{F}$ *which corresponds to a beta process.*

Proof. First observe that, as $n \to \infty$,

$$\begin{aligned}
\mathcal{E}_{\Pi_n}(\bar{F}(t)) &= \prod_{s_i^{(n)} \leq t} \mathcal{E}_{\Pi_n}(1 - V_i^{(n)}) \\
&= \prod_{s_i^{(n)} \leq t} \left(1 - \frac{F_0(s_{i-1}^{(n)}, s_i^{(n)}]}{F_0(s_{i-1}^{(n)}, \infty)}\right) \\
&\to \prod_{(0,t]} (1 - d\phi_H(F_0)) \\
&= \prod_{(0,t]} (1 - dA_0) = \bar{F}_0(t)
\end{aligned}$$

for all $t \geq 0$. Thus $\mathcal{E}_{\Pi_n}(F) = F_n \stackrel{w}{\to} F_0$ as $n \to \infty$. Hence, by a result due to Sethuraman, $\{\Pi_n\}$ is tight.

We shall now follow Hjort's calculatons to show that the finite-dimensional distributions of the process F, under the prior Π_n, converges weakly to those under the prior induced by a beta process with parameters c and A_0 on $\mathcal{H}$.

Consider, for each $n \geq 1$, an independent increment process A_n^c with process measure Π_n^* on $\mathcal{A}$ such that, for each fixed $t > 0$,

$$\mathcal{L}(A_n^c(t)) = \mathcal{L}(\sum_{s_i^{(n)} \leq t} V_i^{(n)}).$$

Thus, for each $n \geq 1$, A_n^c is a purely jump-process with fixed jump-points at $s_1^{(n)}, \ldots, s_{n-1}^{(n)}$ and with random jump sizes given by $V_i^{(n)}, \ldots, V_{n-1}^{(n)}$ at these sites. Clearly Π_n^* induces the prior Π_n on $\mathcal{F}$.

Now, for any fixed $t > 0$, repeating computations as in Hjort ([14], Theorem 3.1, pp. 1270-72) with

$$c_{n,i} = c(s_{i-1}^{(n)}), \qquad b_{n,i} = c_{n,i} \frac{\bar{F_0}^c(s_i^{(n)})}{\bar{F_0}^c(s_{i-1}^{(n)})}, \qquad \text{and} \qquad a_{n,i} = c_{n,i} - b_{n,i}.$$

one concludes that, for each θ, as $n \to \infty$,

$$\mathcal{E}[e^{-\theta A_n^c(t)}] \to \exp\left\{\int_0^1 \int_0^t (1 - e^{-\theta u})\lambda(ds\, du)\right\}$$

and, similarly,

$$\mathcal{E} \exp - \sum_{j=1}^{m} \theta_j A_n^c(a_{j-1}, a_j] \to \exp\left\{-\sum_{j=1}^{m} \int_0^1 \int_{a_{j-1}}^{a_j} (1 - e^{-\theta_j u})\lambda(ds\, du)\right\}.$$

Thus the finite dimensional distributions of the independent increment processes A_n converge to the finite dimensional distributions of an independent increment process with Lévy measure as in Definition 3.1. If the process measure is denoted by Π^* and the corresponding induced measure on $\mathcal{F}$ is denoted by Π, then considering the Skorokhod topology on $\mathcal{A}$ and by the continuity of ϕ_H^{-1} we conclude that, for all $a_1, \ldots, a_m$,

$$\mathcal{L}(\bar{F}(a_1), \ldots, \bar{F}(a_m) \mid \Pi_n) \stackrel{w}{\to} \mathcal{L}(\bar{F}(a_1), \ldots, \bar{F}(a_m) \mid \Pi).$$

Therefore $\{\Pi_n\}$ converges weakly to Π, a NR prior on $\mathcal{F}$. ■

As noted earlier, the existence of a Lévy process with the given Lévy measure is not difficult to establish. It is also not much trouble to show that the finite dimensional distributions under the time-discrete processes converge to the finite dimensional distributions of the given Lévy process. The difficult issue to resolve is the tightness of the sequence of time-discrete processes and that the paths are almost surely in $\mathcal{A}$. By constructing the prior directly on $\mathcal{F}$ one is able to apply Sethuraman's result and thereby resolve the tightness issue very easily. Since explicit expressions for the finite dimensional distributions are hard to obtain, we take recourse to the continuity of the product integral to establish proper convergence of these distributions.

3.2 Properties of Beta Processes

In this section we prove weak consistency of the posteriors of beta process priors on $\mathcal{A}$. First we take a look at some properties of the prior and the posterior.

3.2.1 General Properties

The following properties of beta processes may be found in Hjort [14].

(1). Let $A_0 \in \mathcal{A}$ be a hazard function with finitely many points of discontinuity and let c be a piecewise continuous function on $(0, \infty)$.

If $A \sim \text{beta}(c, A_0)$ then $\mathcal{E}(A(t)) = A_0(t)$. In other words $F = \phi_H^{-1}(A)$ follows a beta(c, F_0) prior distribution and we have $\mathcal{E}(F(t)) = F_0(t)$ where $F_0 = \phi_H^{-1}(A_0)$.

The function c enters the expression for the variance. If $\mathbf{M} = \{t_1, \ldots, t_k\}$ is the set of discontinuity points of A_0 then

$$\mathbf{V}(A(t)) = \sum_{t_j \le t} \frac{A_0\{t_j\}(1 - A_0\{t_j\})}{c(t_j) + 1} + \int_0^t \frac{dA_{0,c}(s)}{c(s) + 1}$$

where $A_{0,c}(t) = A_0(t) - \sum_{t_i \le t} A_0\{t_i\}$.

(2). Let $A \sim \text{beta}(c, A_0)$ where, as before, A_0 has discontinuities at points in $\mathbf{M}$. Let, given F, $X_1, \ldots, X_n$ be i.i.d. F. Then the posterior distribution of F given $X_1, \ldots, X_n$ is again a beta prior, i.e. the corresponding independent increment process is again beta.

To describe the posterior parameters, let $\mathbf{X_n}$ be the set of distinct elements of $\{x_1, \ldots, x_n\}$. Define

$$Y_n(t) = \sum_{i=1}^{n} I_{(X_i \geq t)} \quad \text{and} \quad \bar{Y}_n(t) = \sum_{i=1}^{n} I_{(X_i > t)}.$$

With $N_n(t)$ as before, note that $\bar{Y}_n(t) = n - N_n(t)$ and $Y_n(t) = n - N_n(t-)$.

Using this notation, the posterior beta process has parameters

$$c_{X_1 \ldots X_n}(t) = c(t) + Y_n(t);$$
$$A_{0,X_1 \ldots X_n}(t) = \int_0^t \frac{c(z)\, dA_0(z) + dN_n(z)}{c(z) + Y_n(z)}.$$

More explicitly, $A_{0,X_1 \ldots X_n}$ has discontinuities at points in $\mathbf{M}^* = \mathbf{M} \cup \mathbf{X_n}$, and for $t \in \mathbf{M}^*$,

$$A_{0,X_1 \ldots X_n}\{t\} = \frac{c(t).A_0\{t\} + N_n\{t\}}{c(t) + Y_n(t)};$$
$$A^c_{0,X_1 \ldots X_n}(t) = \int_0^t \frac{c(z)\, dA^c_0(z)}{c(z) + Y_n(z)}.$$

Note that, if $t \in \mathbf{M}^*$,

$$A\{t\} \sim \text{beta}\left(c(t)\, A_0\{t\} + N_n\{t\}, c(t)(1 - A_0\{t\}) + Y_n(t) - N_n\{t\}\right).$$

(3). Our interest is in the following special case of (2). Suppose $A \sim$ beta(c, A_0) and that A_0 is continuous. Then the posterior given $X_1, \ldots, X_n$ is again a beta process with parameters

$$c_{X_1 \ldots X_n}(t) = c(t) + Y_n(t) \quad \text{and} \quad A_{0,X_1 \ldots X_n}(t) = A^d_{0,X_1 \ldots X_n}(t) + A^c_{0,X_1 \ldots X_n}(t)$$

where

$$A^d_{0,X_1\ldots X_n}(t) = \sum_{\substack{t_i \in \mathbf{X}_n \\ t_i \le t}} \frac{N_n\{t_i\}}{c(t_i) + Y_n(t_i)}$$

and

$$A^c_{0,X_1\ldots X_n}(t) = \int_0^t \frac{c(z)\, dA_0(z)}{c(z) + Y_n(z)}.$$

As a consequence, if $t \in \mathbf{X}_n$, then under the posterior $\Pi_{X_1,\ldots,X_n}$ we have

$$A\{t\} \sim \text{beta}(N_n\{t\}, c(t) + \bar{Y}_n(t)).$$

Also note that the Bayes estimates are

$$\mathcal{E}_{\Pi_{X_1,\ldots,X_n}}(A(t)) = A_{0,X_1\ldots X_n}(t)$$

and

$$\mathcal{E}_{\Pi_{X_1,\ldots,X_n}}(\bar{F}(t)) = \prod_{\substack{t_i \in \mathbf{X}_n \\ t_i \le t}} \left(1 - \frac{N_n\{t_i\}}{c(t_i) + Y_n(t_i)}\right) \exp\left\{-\int_0^t \frac{c(z)\, dA_0(z)}{c(z) + Y_n(z)}\right\}. \tag{3.2.2}$$

(4). If A_0 is continuous then, since $A_{0,X_1\ldots X_n}$ has jumps, the prior and posterior are mutually singular.

3.2.2 Weak Consistency of the Posterior

Let Π be a prior on $\mathcal{F}$ and, as before, let $\Pi_{X_1,\ldots,X_n}$ be (a fixed version of) the posterior given $X_1, \ldots, X_n$.

Definition 3.2. *The sequence of posteriors* $\{\Pi_{X_1,\dots,X_n}\}$ *is said to be weakly consistent at* F_0 *if*

$$\lim_{n\to\infty} \Pi_{X_1,\dots,X_n}(U) = 1 \qquad a.s.\ F_0^\infty$$

for every weak neighborhood U *of* F_0.

Weak consistency, as shown in the next proposition, is a weak requirement.

Proposition 3.1. *If, for all continuity points* t *of* F_0 *and all* $\epsilon > 0$,

$$\Pi_{X_1,\dots,X_n}\{F : |F(t) - F_0(t)| < \epsilon\} \to 1 \qquad a.s.\ F_0^\infty,$$

then $\Pi_{X_1,\dots,X_n}$ *is weakly consistent.*

Proof. Since every weak neighborhood of F_0 is of the form

$$\{F : |F(t_i) - F_0(t_i)| < \epsilon_i, i = 1, \dots, k, t_i \text{ continuity points of } F_0\}$$

the result follows easily. ∎

Theorem 3.2. *Let* $A_0 \in \mathcal{A}$ *be continuous and finite for all* $t > 0$. *Let* Π *be a beta(c,* A_0*) prior. then the posterior is weakly consistent at all* $F_0 \in \mathcal{F}$.

Proof. Let $\Omega = \{(x_1, x_2, \dots) : \frac{Y_n(t)}{n} \to \bar{F}_0(t-) \text{ for all } t\}$. By the Glivenko-Cantelli lemma $F_0^\infty(\Omega) = 1$.

Fix $\underset{\sim}{x} = (x_1, x_2, \dots)$ in Ω. Let $\mathbf{X}_n = \{t_1, \dots, t_{k(n)}\}$ be the distinct elements in $\{x_1, \dots, x_n\}$. Fix t, a continuity point of F_0, with $\bar{F}_0(t) > 0$. For notational convenience we will use $N_{n,j} = N_n\{t_j\}$, $Y_{n,j} = Y_n(t_j)$, $\bar{Y}_{n,j} = \bar{Y}_n(t_j)$ and $c_j = c(t_j)$.

Also let $K = \sup_{z \le t} c(z)$. Recall from (3.2.2) that

$$\hat{S}_n(t) \equiv \mathcal{E}_{\Pi_{X_1,\ldots,X_n}}(\bar{F}(t)) = \prod_{\substack{t_j \in \mathbf{X}_n \\ t_j \le t}} \left(1 - \frac{N_{n,j}}{c_j + Y_{n,j}}\right) \exp\left\{-\int_0^t \frac{c(z)\, dA_0(z)}{c(z) + Y_n(z)}\right\}.$$

We will argue that $\hat{S}_n(t) \to \bar{F}_0(t)$ as $n \to \infty$. Since $Y_n(t) \to \infty$ as $n \to \infty$,

$$\int_0^t \frac{c(z)\, dA_0(z)}{c(z) + Y_n(z)} \le K \int_0^t \frac{dA_0(z)}{Y_n(z)} \le K \frac{A_0(t)}{Y_n(t)} \to 0. \tag{3.2.3}$$

Consequently, as $n \to \infty$,

$$\exp\left\{-\int_0^t \frac{c(z)\, dA_0(z)}{c(z) + Y_n(z)}\right\} \to 1.$$

Now, observe that,

$$\prod_{\substack{t_j \in \mathbf{X}_n \\ t_j \le t}} \left(1 - \frac{N_{n,j}}{c_j + Y_{n,j}}\right) = \prod_{\substack{t_j \in \mathbf{X}_n \\ t_j \le t}} \left(\frac{c_j + \bar{Y}_{n,j}}{c_j + Y_{n,j}}\right).$$

Since, for positive numbers a, b, c such that $c > b$, we have $\frac{a+b}{a+c} > \frac{b}{c}$, as $n \to \infty$,

$$\prod_{\substack{t_j \in \mathbf{X}_n \\ t_j \le t}} \left(\frac{c_j + \bar{Y}_{n,j}}{c_j + Y_{n,j}}\right) \ge \prod_{\substack{t_j \in \mathbf{X}_n \\ t_j \le t}} \frac{\bar{Y}_{n,j}}{Y_{n,j}} \ge \frac{\bar{Y}_n(t)}{n} \to \bar{F}_0(t)$$

On the other hand,

$$\prod_{\substack{t_j \in \mathbf{X}_n \\ t_j \le t}} \left(\frac{c_j + \bar{Y}_{n,j}}{c_j + Y_{n,j}}\right) \le \prod_{\substack{t_j \in \mathbf{X}_n \\ t_j \le t}} \frac{K + \bar{Y}_{n,j}}{K + Y_{n,j}} \le \frac{K + \bar{Y}_n(t)}{K + n} \to \bar{F}_0(t).$$

Next we will show that

$$\lim_{n \to \infty} \mathcal{E}_{\Pi_{X_1,\ldots,X_n}}(\bar{F}^2(t)) = \bar{F}_0^2(t) \tag{3.2.4}$$

which will prove that

$$\lim_{n \to \infty} V_{\Pi_{X_1,\ldots,X_n}}(\bar{F}(t)) = 0 \tag{3.2.5}$$

and will establish the result.

For each $n \geq 1$, let $A_{n,c}(t) = A(t) - \sum_{\substack{t_i \in \mathbf{X}_n \\ t_i \leq t}} A\{t_i\}$, then for any A,

$$\bar{F}(t) = \prod_{[0,t]}(1 - dA(s)) = \prod_{\substack{t_i \in \mathbf{X}_n \\ t_i \leq t}} (1 - A\{t_i\}) \prod_{[0,t]}(1 - dA_{n,c}(s)).$$

We will show that

(1). $\lim_{n\to\infty} \mathcal{E}_{\Pi_{X_1,\ldots,X_n}} [\prod_{[0,t]}(1 - d\, A_{n,c}(s))]^2 = 1$; and that

(2).

$$\lim_{n\to\infty} \mathcal{E}_{\Pi_{X_1,\ldots,X_n}} \left[\prod_{\substack{t_i \in \mathbf{X}_n \\ t_i \leq t}} (1 - A\{t_i\})\right]^2 = \lim_{n\to\infty} \mathcal{E}_{\Pi_{X_1,\ldots,X_n}} \left[\prod_{t_i \leq t}(1 - A\{t_i\})\right]^2 = \bar{F}_0^2(t)$$

Since $A\{t\} < 1$ for all A,

$$\begin{aligned}
\left[\prod_{[0,t]}(1 - dA_{n,c}(s))\right]^2 &= \left[\lim_{|t_i - t_j| \downarrow 0} \prod (1 - A_{n,c}(t_i, t_{i+1}])\right]^2 \\
&= \lim_{|t_i - t_j| \downarrow 0} \prod (1 - A_{n,c}(t_i, t_{i+1}])^2 \\
&\leq \lim_{|t_i - t_j| \downarrow 0} \prod (1 - A_{n,c}^2(t_i, t_{i+1}]) \\
&= \prod_{[0,t]}(1 - d(A_{n,c}^2)(s)),
\end{aligned}$$

and since, by (3.2.3), $\mathcal{E}_{\Pi_{X_1,\ldots,X_n}} A_{n,c} \to 0$,

$$\begin{aligned}
\mathcal{E}_{\Pi_{X_1,\ldots,X_n}} \left[\prod_{[0,t]}(1 - dA_{n,c}(s))\right]^2 &\leq \mathcal{E}_{\Pi_{X_1,\ldots,X_n}} \left[\prod_{[0,t]}(1 - dA_{n,c}^2(s))\right] \\
&= e^{-[\mathcal{E}_{\Pi_{X_1,\ldots,X_n}} A_{n,c}^2]} \leq e^{-[\mathcal{E}_{\Pi_{X_1,\ldots,X_n}} A_{n,c}]^2} \to 1. \quad (3.2.6)
\end{aligned}$$

Since, under $\Pi_{X_1,\ldots,X_n}$, $A\{t_i\} \sim \text{beta}(N_{n,i}, c_i + \bar{Y}_{n,i})$ and are independent,

$$\begin{aligned}
\mathcal{E}_{\Pi_{X_1,\ldots,X_n}}\left[\prod_{\substack{t_i\in\mathbf{X}_n\\ t_i\le t}}(1-A\{t_i\})^2\right] &= \prod_{\substack{t_i\in\mathbf{X}_n\\ t_i\le t}}\left[\frac{\bar{Y}_{n,i}+c_i}{Y_{n,i}+c_i}\times\frac{\bar{Y}_{n,i}+c_i+1}{Y_{n,i}+c_i+1}\right] \\
&\le \prod_{\substack{t_i\in\mathbf{X}_n\\ t_i\le t}}\left[\frac{\bar{Y}_{n,i}+K}{Y_{n,i}+K}\times\frac{\bar{Y}_{n,i}+K+1}{Y_{n,i}+K+1}\right] \\
&\le \left(\frac{\bar{Y}_n(t-)+K}{n+K}\right)\left(\frac{\bar{Y}_n(t-)+K+1}{n+K+1}\right) \\
&\to \bar{F}_0^2(t-) = \bar{F}_0^2(t). \qquad (3.2.7)
\end{aligned}$$

Thus, by (3.2.6) and (3.2.7), we have

$$\lim_{n\to\infty}\mathcal{E}_{\Pi_{X_1,\ldots,X_n}}(\bar{F}^2(t)) \le \bar{F}_0^2(t)$$

which, of course, yields (3.2.4) and hence (3.2.5). ■

3.3 The Problem of Right-Censored Observations

Let X and Y be independent positive random variables with distributions F_X and F_Y respectively. Our motivation is to think of X as a lifetime and of Y as a censoring time. In the problem of right-censored data, we get to observe $Z = \min(X, Y)$. We also observe $\Delta = I_{\{X \le Y\}}$, thus being informed whether a lifetime has been observed or was censored. Our intention is to infer about F_X.

One of the many Bayesian approaches to this problem is to put a suitable prior on the joint distribution of (Z, Δ). We will refer to the space of all such joint distributions as the observational distribution space and denote it by $\mathcal{F}_{(Z,\Delta)}$. Peterson [20] provides explicit expressions for recovering F_X and F_Y from the joint distribution of (Z, Δ)

under certain mild conditions. We will use this to study the distributions of F_X and F_Y for suitably chosen priors on $\mathcal{F}_{(Z,\Delta)}$.

Specifically we will consider a Dirichlet process prior for the observational distribution space. Our goal is to obtain the induced distributions of F_X and F_Y under such a prior. We begin by defining a Dirichlet process on $\mathbb{R}^+ \equiv (0,\infty)$.

Definition 3.3. *Let α be positive measure on $\mathbb{R}^+$. A Dirichlet process prior on $\mathbb{R}^+$ with parameter α, denoted Dir(α), is such that for any $k \geq 1$ and for arbitrary points $0 < t_1 < t_2 < \ldots < t_k$, under Dir($\alpha$), $(F(t_1), F(t_2) - F(t_1), \ldots, F(t_k) - F(t_{k-1})$ has a Dirichlet distribution with parameters $(\alpha((0,t_1]), \alpha((t_1,t_2]), \ldots, \alpha((t_{k-1},t_k]))$.*

As before, let $\bar{F}_X(t) = 1 - F_X(t)$ and $\bar{F}_Y(t) = 1 - F_Y(t)$ be the survival functions for the lifetime and the censoring time respectively. Let

$$F_1(t) = \mathbf{P}(Z \leq t \mid \Delta = 1); \quad F_0(t) = \mathbf{P}(Z \leq t \mid \Delta = 0); \text{ and } p = \mathbf{P}(\Delta = 1).$$

Let α be a positive measure on $\mathbb{R}^+ \times \{0,1\}$ such that

- $\alpha_0(\cdot) = \alpha(\cdot \times \{0\})$ and $\alpha_1(\cdot) = \alpha(\cdot \times \{1\})$ are positive measures on $\mathbb{R}^+$ with full support.
- $\{t : \alpha_0\{t\} > 0\}$ and $\{t : \alpha_1\{t\} > 0\}$ are disjoint.

Definition 3.4. *A Dirichlet process prior Π with parameter α on $\mathcal{F}_{(Z,\Delta)}$ is a prior such that, under Π,*

(1). $F_1 \sim$ Dir(α_1);

(2). $F_0 \sim$ Dir(α_0); and

(3). $p \sim beta(\alpha_1(\mathbb{R}^+), \alpha_0(\mathbb{R}^+))$,

and, F_0, F_1 *and* p *are independent.*

Our goal is to obtain the distribution of F_X using Peterson's map. The definition of the hazard function is crucial to the map. Peterson uses Doksum's definition for this purpose. However, we intend to use the product integral, i.e. the definition used by Hjort. Hence a slight modification of the map is necessary.

First let us observe the following properties of F_0, F_1 and p. These properties and more may be found in Peterson [20].

(1). $p\bar{F}_1(t) = -\int_t^\infty \bar{F}_Y(s-)d\bar{F}_X(s)$, and $(1-p)\bar{F}_0(t) = -\int_t^\infty \bar{F}_X(s-)d\bar{F}_Y(s)$.

(2). With obvious notation $p\bar{F}_1(t) + (1-p)\bar{F}_0(t) = \bar{F}_X(t)\,\bar{F}_Y(t) = \bar{F}_Z(t)$.

(3). Jump-points of F_1 are jump-points of F_X and vice versa. Similarly, jump-points of F_0 are jump-points of F_Y and vice versa.

Under our assumption on α_1 and α_2, if Π is a Dirichlet process with parameter α, then F_1 and F_0 does not share jump-points almost surely Π. Also F_1 and F_0 have full support almost surely Π.

As before, we will treat each of the distributions F_X, F_Y, F_Z, F_0 and F_1 both as distribution functions and the measures corresponding to them.

Proposition 3.2. *Let* F_1 *and* F_0 *be distribution functions with full support and no jump-points in common. Let* $p > 0$. *Then* F_0, F_1 *and* p *uniquely determines* F_X *and*

F_Y. The map for F_X is given by:

$$\bar{F}_X(t) = \prod_{s \le t} \left(1 - \frac{p\,F_1\{s\}}{F_Z[s,\infty)}\right) \exp\left\{-\int_0^t \frac{p\,dF_1^c(s)}{F_Z[s,\infty)}\right\}$$

where $F_1^c(t) = F_1(t) - \sum_{s \le t} F_1\{t\}$. The map for F_Y is similar.

Observe that here

$$F_Z[s,\infty) = p\,F_1[s,\infty) + (1-p)\,F_0[s,\infty).$$

Proof. We know that if A_X is the cumulative hazard function corresponding to F_X, then

$$\bar{F}_X(t) = \prod_{[0,t]}(1 - dA_X(s)) = \prod_{s \le t}(1 - A_X\{s\}) \exp[-A_X^c(t)].$$

where A_X^c is the continuous part of A_X.

Note that

$$\int_0^t \frac{p\,dF_1^c(s)}{F_Z[s,\infty)} = -\int_0^t \frac{\bar{F}(s-)\,d\bar{F}_X^c(s)}{\bar{F}_X(s-)\,\bar{F}_Y(s-)} = -\int_0^t \frac{d\bar{F}_X^c(s)}{\bar{F}_X(s-)} = A_X^c(t)$$

and also that, for a jump-point s of F_X (and hence of F_1 and A_X),

$$\begin{aligned}\frac{p\,F_1\{s\}}{F_Z[s,\infty)} &= \frac{F_Z(s,\infty)}{F_Z[s,\infty)} = 1 - \frac{F_Z(s,\infty)}{F_Z[s,\infty)} \\ &= 1 - \frac{F_X(s,\infty)\,F_Y(s,\infty)}{F_X[s,\infty)\,F_Y[s,\infty)} = 1 - \frac{F_X(s,\infty)}{F_X[s,\infty)} = A_X\{s\}.\end{aligned}$$

■

The above result shows that

$$\bar{F}_X(t) = \prod_{[0,t]} \left(1 - \frac{p\,dF_1(s)}{F_Z[s,\infty)}\right).$$

Or, in other words, $A_X(t) = \int_0^t \frac{p\,dF_1(s)}{F_Z[s,\infty)}$.

Let

$$A_\alpha(t) = \int_0^t \frac{d\alpha_1(s)}{\alpha([s,\infty) \times \{0,1\})} \tag{3.3.8}$$

and suppose $A_\alpha \in \mathcal{A}$.

Let $\mathbb{Q}$ be a dense subset of $\mathbb{R}^+$ with enumeration $\{t_1, t_2, \ldots\}$. For each n, let $t_1^{(n)} < \cdots < t_n^{(n)}$ be an ordering of $\{t_1, \ldots, t_n\}$. Now let

$$A_n(t) \equiv \phi_n(F_1, F_0, p)(t) = \sum_{t_i^{(n)} \le t} \frac{p\,F_1[t_i^{(n)}, t_{i+1}^{(n)})}{F_Z[t_i^{(n)}, \infty)}$$

Then, for all t and all F_1, F_0, p, $\phi_n(F_1, F_0, p)(t) \to A_X(t)$ as $n \to \infty$.

Let $F_{(Z,\Delta)}$ denote the joint distribution of (Z, Δ) specified by F_1, F_0 and p. Since $F_{(Z,\Delta)} \sim \Pi$, the Dir(α) distribution,

$$\frac{p\,F_1[t_i^{(n)}, t_{i+1}^{(n)})}{F_Z[t_i^{(n)}, \infty)} \sim \text{beta}(\alpha(A_i^{(n)}), \alpha(B_i^{(n)}))$$

where $M_i^{(n)} = [t_i^{(n)}, \infty) \times \{0,1\}$, $A_i^{(n)} = [t_i^{(n)}, t_{i+1}^{(n)}) \times \{1\}$ and $B_i^{(n)} = M_i^{(n)} - A_i^{(n)}$.

Rewrite this as,

$$\frac{p\,F_1[t_i^{(n)}, t_{i+1}^{(n)})}{F_Z[t_i^{(n)}, \infty)} \sim \text{beta}\left(\alpha(M_i^{(n)})\frac{\alpha(A_i^{(n)})}{\alpha(M_i^{(n)})}, \alpha(M_i^{(n)})\left(1 - \frac{\alpha(A_i^{(n)})}{\alpha(M_i^{(n)})}\right)\right)$$

Further $X_{n,i} \equiv \frac{p\,F_1[t_i^{(n)}, t_{i+1}^{(n)})}{F_Z[t_i^{(n)}, \infty)}$, $i \ge 1$ are all independent.

Assume that α_1 is continuous. We will now show that, under Π, the distribution of A_n converges in distribution to a beta process with parameters c and A_α, where

$$c(t) = \alpha([t,\infty) \times \{0,1\}). \tag{3.3.9}$$

Since $A_n(t) \to A_X(t)$ for all t as $n \to \infty$, the distribution of A_n under Π converges to the distribution of A_X under Π, i.e.

$$\mathcal{L}(\phi_n(F_1, F_0, p) \mid \Pi) \xrightarrow{w} \mathcal{L}(A_X \mid \Pi).$$

We may then conclude that, $\mathcal{L}(A_X \mid \Pi)$ is a beta process with parameters as stated above.

Theorem 3.3. *$A_X \sim beta(c, A_\alpha)$ where c and A_α are as in (3.3.9) and (3.3.8) respectively.*

Proof. In view of the above discussion it suffices to show that

$$\mathcal{E}exp\{-\theta\, A_n(t)\} \to exp\left\{-\int_0^1 (1 - e^{-\theta s})\, d\lambda_t(s)\right\}$$

where $d\lambda_t(s) = \left(\int_0^t c(z)\, s^{-1}(1-s)^{c(z)-1} dA_\alpha(z)\right) ds$, and that $\{A_n\}$ is tight.

Letting $c_{n,i} = \alpha([t_i^{(n)}, \infty) \times \{0,1\})$,

$$a_{n,i} = c_{n,i} \frac{\alpha(A_i^{(n)})}{\alpha(M_i^{(n)})} \qquad \text{and} \qquad b_{n,i} = c_{n,i} \frac{\alpha(B_i^{(n)})}{\alpha(M_i^{(n)})}$$

and observing that $X_{n,i} \sim$ beta$(a_{n,i}, b_{n,i})$ and $A_n(t) = \sum_{t_i^{(n)} \le t} X_{n,i}$, the proof follows easily by mimicking Hjort's construction of the beta process ([14], Theorem 3.1, pp. 1270-72). ■

3.4 Mutual Singularity of Beta Processes

Let Π_1^* and Π_2^* be two independent increment processes on $\mathcal{A}$ with no fixed jump-points. Then, as we saw from the Lévy representation, the random measure $A \mapsto$

$\mu(\cdot, A)$ defined by

$$\mu(E, A) = \#\left\{(t, A\{t\}) \in E : A\{t\} > 0\right\},$$

for any Borel subset E of $(0,\infty) \times [0,\infty]$, is a Poisson process under Π_1^* and Π_2^* with mean measures say λ_1 and λ_2.

If the Poisson processes induced by Π_1^* and Π_2^* are mutually singular then so are Π_1^* and Π_2^*. Conditions on λ_1 and λ_2 can be given which will ensure that the corresponding Poisson processes are mutually singular. We quote below a theorem due to Brown [1] in this context.

Theorem 3.4 (M. Brown, 1971). *Let P_{λ_1} and P_{λ_2} be Poisson processes over a measurable space $(\mathcal{X}, \mathcal{A})$ with σ-finite mean measures λ_1 and λ_2. let $\lambda_1 = \mu + \nu$ be the Lebesgue decomposition of λ_1 with respect to λ_2 ($\mu \ll \lambda_2, \nu \perp \lambda_2$). Then $P_{\lambda_1} \perp P_{\lambda_2}$ if and only if one of the following conditions hold:*

(1). $\nu(\mathcal{X}) = \infty$,

(2). $\int_{B_c} |f - 1| d\lambda_2 = \infty$ *for some* $c > 0$

where $f = \frac{d\lambda_1}{d\lambda_2}$ *and* $B_c = \{|f - 1| > c\}$,

(3). $\int_{\bar{B}_c} (f - 1)^2 d\lambda_2 = \infty$ *for all* $c > 0$.

We apply the above theorem and show that, in general, beta processes tend to be singular.

Theorem 3.5. *Let Π_1^* and Π_2^* be two beta processes with parameters (c_1, A_1) and (c_2, A_2). Assume that A_1 and A_2 are continuous. Then $(c_1, A_1) \neq (c_2, A_2)$ implies $\Pi_1^* \perp \Pi_2^*$.*

Proof. Recall that the Lévy measures corresponding to Π_1^* and Π_2^* are given by

$$\lambda_1(ds\,dz) = c_1(z)s^{-1}(1-s)^{c_1(s)-1}dA_1(z)ds$$

$$\lambda_2(ds\,dz) = c_2(z)s^{-1}(1-s)^{c_2(s)-1}dA_2(z)ds.$$

As before, we will continue to use A_1 and A_2 for the measures generated by these functions.

(1) Suppose A_1 and A_2 are not mutually absolutely continuous, so that there exists a Borel set $B \subset \mathbb{R}^+$ such that $A_1(B) > 0$ but $A_2(B) = 0$. Then

$$\lambda_1(B \times (0,1)) = \int_B \int_0^1 s^{-1}\,(1-s)^{c_1(z)-1}\,c_1(z)\,ds\,dA_1(z) = \infty$$

because the inner integral is ∞ for each fixed z, and $\lambda_2(B \times (0,1)) = 0$.

Consequently, by condition ((1)) of Theorem 3.4, the result follows.

(2) Now suppose that A_1 and A_2 are mutually absolutely continuous. Let $g(z) = \frac{dA_1}{dA_2}(z)$. Let $\delta > 0$ and consider the sets

$$E_\delta = \left\{ z : \frac{c_1(z)}{c_2(z)} g(z) > 1 - \delta \right\} \qquad \text{and} \qquad D_\delta = \left\{ z : \frac{c_1(z)}{c_2(z)} g(z) > 1 + \delta \right\}.$$

Let $G = \{z : \frac{c_1(z)}{c_2(z)} g(z) = 1\}$.

Case 1. Suppose $\lambda_2(D_\delta \times (0,1)) > 0$.

(a) If, in addition, $\lambda_2\left(D_\delta \cap \{\frac{c_1(z)}{c_2(z)} \geq 1\} \times (0,1)\right) > 0$, then let $\epsilon < \frac{\delta}{2(1+\delta)}$, and note that, for each fixed $z \in D_\delta \cap \{\frac{c_1(z)}{c_2(z)} \geq 1\}$, there exists $t_0(z)$ such that $(1-s)^{c_1(z)-c_2(z)} > 1 - \epsilon$ for all $s < t_0(z)$.

Let $D = \{(z, s) : z \in D_\delta, s < t_0(z)\}$. On D

$$f(z,s) \equiv \frac{d\lambda_1}{d\lambda_2}(z,s) = \frac{c_1(z)}{c_2(z)} g(z)\,(1-s)^{c_1(z)-c_2(z)} > (1+\delta)(1-\epsilon) > 1 + \frac{\delta}{2}.$$

Letting $c = \delta/2$ in Theorem 3.4, we have

$$\int_{B_c} |f - 1|\, d\lambda_2 \geq \int_D (f-1)\, d\lambda_2 \geq \frac{\delta}{2}\lambda_2(D).$$

Since, for each z,

$$\int_0^{t_0(z)} s^{-1}(1-s)^{c_1(z)-c_2(z)} ds = \infty$$

we have $\lambda_2(D) = \infty$. Thus, condition ((2)) of Brown's Theorem is satisfied and, hence, we have the result.

(b) If, on the other hand, $\lambda_2 \left(D_\delta \cap \{\frac{c_1(z)}{c_2(z)} \geq 1\} \times (0,1) \right) = 0$, then

$$\lambda_2 \left(D_\delta \cap \{\frac{c_1(z)}{c_2(z)} < 1\} \times (0,1) \right) > 0.$$

Since, for each z, $(1-s)^{c_1(z)-c_2(z)} \to \infty$ as $s \uparrow 1$, there exists $t_1(z)$ such that $(1-s)^{c_1(z)-c_2(z)} > 1 + \delta$ for all $s > t_1(z)$.

Noting that $\int_{t_1(z)}^1 s^{-1}(1-s)^{c_1(z)-c_2(z)} ds = \infty$, for each z, an argument similar to the one in (a) yields the result.

Case 2. Suppose $\lambda_2(D_\delta) = 0$, for all δ, but $\lambda_2(E_\delta) > 0$ for some δ. Then the above argument goes through by reversing the roles of λ_1 and λ_2.

Case 3. Suppose $\lambda_2(D_\delta) = 0$ and $\lambda_2(E_\delta) = 0$ for all δ. Then $\frac{c_1(z)}{c_2(z)} g(z) = 1$ a.e. λ_2.

In this case $f(z,s) = (1-s)^{c_1(z)-c_2(z)}$. Since, on G, at least one of $\{c_1(z)-c_2(z) > \delta\}$ and $\{c_2(z) - c_1(z) > \delta\}$ has positive measure for some δ, it is easy to see that, at least one of $\int_{B_c} |f-1|\, d\lambda_2$ and $\int_{B_c} |\frac{1}{f} - 1|\, d\lambda_1$ is ∞.

■

3.5 Beta-Stacy Process Priors

The family of beta-Stacy process priors was introduced by Walker and Muliere [26] as examples of NR priors on $\mathcal{F}$. We will see here that this family is, however, identical to the family of beta priors on $\mathcal{F}$.

Definition 3.5. *Let* $c(.)$ *be a piecewise continuous positive function on* $(0,\infty)$. *Let* $F_0 \in \mathcal{F}$ *have finitely many jumps at the points* $t_1,\ldots,t_k$. *A random cdf* F *is said to have a beta-Stacy process prior with parameters* c *and* F_0, *written* $F \sim$ *beta-Stacy*(c, F_0), *if for all* $t \geq 0$, $\bar{F}(t) = e^{-H(t)}$ *where* $H(t)$ *is an independent increment process with Lévy representation such that*

(1). $t_1,\ldots,t_k$ *are fixed jump-points of* H *and jump-sizes are given by*

$$1 - e^{-H\{t_j\}} \sim beta(c(t_j)F_0\{t_j\}, c(t_j)F_0(t_j,\infty));$$

(2). the Lévy measure is given by

$$\lambda(ds\, du) = c(s)\frac{e^{-u\, c(s)\, F_0(s,\infty)}}{1-e^{-u}} dF_0^c(s)\, du$$

where $0 \leq s < \infty$, $0 < u < \infty$; *and for which*

(3). $b(t) \equiv 0$.

Walker and Muliere referred to the Lévy process H as a Log-beta process. A discussion of such processes is unnecessary for our purpose. The distribution of F as above will be called a beta-Stacy(c, F_0) distribution.

We will now show that any beta-Stacy(c, F_0) prior distribution on $\mathcal{F}$ is simply a reparameterization of a beta prior. For this we take a careful look at the following construction of beta-Stacy process priors given in Walker and Muliere [26].

To avoid notational complications, let us again assume that $F_0 \in \mathcal{F}$ is continuous. Let $\mathbb{Q}$ be a dense subset of $(0, \infty)$. As before, let $\{s_1, s_2, \dots\}$ be an enumeration of $\mathbb{Q}$ and for each $n \geq 1$, let $s_1^{(n)} < \cdots < s_n^{(n)}$ denote an ordering of $s_1, \dots, s_n$. Now, for $1 \leq i \leq n-1$, define

$$W_i^{(n)} \sim beta(c(s_{i-1}^{(n)}).F_0(s_i^{(n)}, \infty), c(s_{i-1}^{(n)}).F_0(s_{i-1}^{(n)}, s_i^{(n)}]). \tag{3.5.10}$$

Define an independent increment process H_n as

$$H_n(t) = -\sum_{s_i^{(n)} \leq t} \log W_i^{(n)} \quad \text{for all } t \geq 0$$

and denote its process measure by Π'_n.

If Π' is the process measure of the beta-Stacy process, then Walker and Muliere shows that the finite dimensional distributions under Π'_n converge to the finite dimensional distributions under Π' in the following manner.

Let $c_{n,i} = c(s_{i-1}^{(n)})$, $a_{n,i} = c_{n,i}F_0(s_i^{(n)}, \infty)$ and $b_{n,i} = c_{n,i}F_0(s_{i-1}^{(n)}, s_i^{(n)}]$. Fix $t > 0$ and observe that

$$\mathcal{E}[e^{-\theta H_n(t)}] = \prod_{s_i^{(n)}\le t} \mathcal{E}(W_i^{(n)})^{\theta}$$

$$= \prod_{i:s_i^{(n)}\le t} \frac{\Gamma(c_{n,i})}{\Gamma(a_{n,i})\Gamma(b_{n,i})} \int_0^1 x_i^{\theta+b_{n,i}-1}(1-x_i)^{a_{n,i}-1}dx_i$$

$$= \prod_{i:s_i^{(n)}\le t} \frac{\Gamma(c_{n,i})\Gamma(b_{n,i}+\theta)}{\Gamma(c_{n,i}+\theta)\Gamma(b_{n,i})}.$$

Therefore

$$\log \mathcal{E} \exp(-\theta H_n(t)) = \sum_{i:s_i^{(n)}\le t} \log \frac{\Gamma(c_{n,i})\Gamma(b_{n,i}+\theta)}{\Gamma(c_{n,i}+\theta)\Gamma(b_{n,i})}.$$

Using the fact $\Gamma(x) = \Gamma(x+1)/x$ repeatedly, one obtains

$$\frac{\Gamma(c_{n,i})\Gamma(b_{n,i}+\theta)}{\Gamma(c_{n,i}+\theta)\Gamma(b_{n,i})} = \left\{\prod_{l=0}^{m-1} \frac{(c_{n,i}+\theta+l)(b_{n,i}+l)}{(c_{n,i}+l)(b_{n,i}+\theta+l)}\right\} \times \frac{\Gamma(c_{n,i}+m)\Gamma(b_{n,i}+\theta+m)}{\Gamma(c_{n,i}+\theta+m)\Gamma(b_{n,i}+m)}.$$

Now, using Sterling's formula $\Gamma(x) \approx \sqrt{2\pi x}(\frac{x}{e})^x$ for large x, we get

$$\lim_{m\to\infty} \frac{\Gamma(c_{n,i}+m)\Gamma(b_{n,i}+\theta+m)}{\Gamma(c_{n,i}+\theta+m)\Gamma(b_{n,i}+m)}$$

$$= \lim_{m\to\infty} \frac{[(1+\frac{c_{n,i}}{m})^m e^{-c_{n,i}}][(1+\frac{b_{n,i}+\theta}{m})^m e^{-(b_{n,i}+\theta)}]}{[(1+\frac{b_{n,i}}{m})^m e^{-b_{n,i}}][(1+\frac{c_{n,i}+\theta}{m})^m e^{-(c_{n,i}+\theta)}]}$$

$$\times \lim_{m\to\infty} \frac{(1+\frac{c_{n,i}}{m})^{c_{n,i}+1/2}(1+\frac{b_{n,i}+\theta}{m})^{b_{n,i}+\theta+1/2}}{(1+\frac{b_{n,i}}{m})^{b_{n,i}+1/2}(1+\frac{c_{n,i}+\theta}{m})^{c_{n,i}+\theta+1/2}} = 1.$$

It follows that

$$\log \mathcal{E} \exp(-\theta H_n(t)) = \sum_{i:s_i^{(n)}\le t} \sum_{l=0}^{\infty} \log \frac{(c_{n,i}+\theta+l)(b_{n,i}+l)}{(c_{n,i}+l)(b_{n,i}+\theta+l)}$$

$$= \int_0^\infty \frac{e^{-u\theta}-1}{1-e^{-u}} \sum_{i:s_i^{(n)}\le t} \frac{e^{-b_{n,i}u}(1-e^{-a_{n,i}u})}{u} du.$$

Now, letting $n \to \infty$, conclude that

$$\log \mathcal{E} \exp(-\theta H_n(t)) \to \int_0^\infty \int_0^t \frac{e^{-u\theta} - 1}{1 - e^{-u}} e^{-u.c(s).F_0(s,\infty)} \, c(s) \, dF_0(s) \, du.$$

Similarly, for any $0 = a_0 < a_1 < \cdots < a_r < \infty$, as $n \to \infty$

$$\log \mathcal{E} \exp\left\{ -\sum_{j=1}^{r} \theta_j (H_n(a_j) - H_n(a_{j-1})) \right\} \to \sum_{j=1}^{r} \int_{a_{j-1}}^{a_j} \int_0^\infty (1 - e^{-\theta_j u}) \lambda(ds\, du).$$

The following result now explicitly provides the reparameterization that yields a beta-Stacy prior from a beta prior.

Theorem 3.6. Π *is a beta-Stacy(c,* F_0*) prior if and only if* Π *is a beta(*c'*,* F_0*) prior, where* $c'(s) = c(s)F_0(s, \infty)$.

Proof. Let Π denote a beta(c', F_0) prior and $\tilde{\Pi}$ denote a beta-Stacy(c, F_0) prior on $\mathcal{F}$. Since $c'(s) = c(s)F_0(s, \infty)$, it is easy to see that $V_i^{\prime(n)}$ in equation (3.1.1) and $W_i^{\prime(n)}$ in equation (3.5.10) have the same distribution for all $n \geq 1$ and all $1 \leq i \leq n - 1$.

Also, if $\bar{F}_n(t) = e^{-H_n(t)}$, then, for each fixed t,

$$\bar{F}_n(t) \stackrel{d}{=} \prod_{i:s_i^{(n)} \leq t} W_i^{\prime(n)} \stackrel{d}{=} \prod_{i:s_i^{(n)} \leq t} V_i^{\prime(n)}$$

where $\stackrel{d}{=}$ denotes equality in distribution.

Let $\{\Pi_n\}$ be the approximating sequence of time-discrete prior distrubtions converging weakly to any one of Π and $\tilde{\Pi}$. Then it converges to the other, as well, and by the uniqueness of the limit we conclude that Π and $\tilde{\Pi}$ are the same. ■

Another way to see the above connection between the two prior processes is in terms of the Lévy measure of the two corresponding independent increment processes.

Let λ^{BS} denote the Lévy measure for the independent increment process corresponding to the beta-Stacy(c, F_0) distribution. Let λ^H denote the Lévy measure for the beta(c', A_0) process where $A_0 = \phi_H(F_0)$.

If F is a random cdf such that F follows a beta-Stacy(c, F_0) distribution, then $H(t) = \phi_D(F)(t)$ is an independent increment process with Lévy measure

$$\lambda^{BS}(ds\,du) = \frac{c(s)\,e^{-u\,c(s)\,F_0(s,\infty)}}{1-e^{-u}} dF_0(s)\,du. \tag{3.5.11}$$

On the otherhand, if F follows a beta(c', F_0) distribution, then $A(t) = \phi_H(F)(t)$ is an independent increment process with Lévy measure

$$\lambda^H(ds\,du) = c'(s)\,u^{-1}(1-u)^{c'(s)-1} dA_0(s)\,du. \tag{3.5.12}$$

Making the transformation $g : x \mapsto 1 - e^{-x}$ in (3.5.11) and using the relationship between c' and c we get, for any Borel set B,

$$\begin{aligned}
\lambda_t^{BS}(g^{-1}(B)) &= \int_{g^{-1}(B)} \int_0^t \frac{c(s)\,e^{-u\,c(s)\,F_0(s,\infty)}}{1-e^{-u}}\, dF_0(s)\,du \\
&= \int_B \int_0^t c'(s)\,v^{-1}(1-v)^{c'(s)-1} \frac{dF_0(s)}{F_0[s,\infty)}\, dv \\
&= \int_B \int_0^t c'(s)\,v^{-1}(1-v)^{c'(s)-1} dA_0(s)\,dv \\
&= \lambda_t^H(B)
\end{aligned}$$

That the two priors are the same now follows from Proposition 1.1. Thus, if we consider a particular sample path of the beta(c', F_0) process with jumps at the points $\{t_1, t_2, \dots\}$ (note that such a path only increases in jumps), and replace the corresponding jump-sizes $\frac{F\{t_i\}}{F[t_i,\infty)}$ with the jump-amount $-\log\left(1 - \frac{F\{t_i\}}{F[t_i,\infty)}\right)$, then we obtain a sample path of the independent increment process corresponding to a beta-Stacy(c,F_0) prior.

Bibliography

[1] Brown, M. (1971). Discrimination of Poisson processes. *Ann. Math. Statist.* **42**, 773-776.

[2] Cox, D. R.; Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall, London.

[3] Connor, R. J.; Mosimann, J. E. (1969). Concepts of independence of proportions with a generalization of the Dirichlet distribution. *J. Amer. Statist. Assoc.* **64**, 194-206.

[4] Damien, P.; Laud, P. W.; Smith, A. F. M. (1996). Implementation of Bayesian non-parametric inference based on beta processes. *Scand. J. Statist.* **23**, 27–36.

[5] Dey, J.; Draghici, L.; Ramamoorthi, R. V. (1999). Characterizations of Tailfree and neutral to the right Priors. *MSU Tech. Report.*

[6] Diaconis, P.; Freedman, D. (1986). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14**, 1-67.

[7] Diaconis, P.; Freedman, D. (1986). On inconsistent Bayes estimates of location. *Ann. Statist.* **14**, 68-87.

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

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

[10] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 615-629.

[11] Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 209-230.

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

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

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

[15] Ito, K. (1969). *Stochastic Processes.* Lecture notes series, **16**.

[16] Kallenberg, O. (1997). *Foundations of Modern Probability.* Springer-Verlag, New York.

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

[18] Lévy, P. (1937). *Théorie de l'Addition des Variables Aléatoire.* Gauthier-Villars, Paris.

[19] Muliere, P.; Walker, S. A. (1997). Bayesian nonparametric approach to determining a maximum tolerated dose. *J. Statist. Plann. Inference.* **61**, 339–353.

[20] Peterson, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72**, 854-858.

[21] Schervish, M.J. (1995). *Theory of Statistics.* Springer Series in Statistics, Springer-Verlag, New-York.

[22] Susarla, V.; Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.* **71**, 897-902.

[23] Tsai, W. (1986). Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian estimation application. *Ann. Statist.* **25**, 1762-1780.

[24] Walker, S. A. (1998) Characterisation of Hjort's discrete time beta process. *Statist. Probab. Lett.* **37**, 351–355.

[25] Walker, S. A.; Damien, P. (1998). A full Bayesian nonparametric analysis involving a neutral to the right process. *Scand. J. Statist.* **25**, 669-680.

[26] Walker, S. A.; Muliere, P. (1997). Beta-Stacy processes and a generalization of the Polya-urn scheme. *Ann. Statist.* **25**, 1762-1780.

[27] Walker, S. A.; Muliere, P. (1999). A characterisation of a neutral to the right prior via an extension of Johnson's sufficientness postulate. *Ann. Statist.* **27**.