

Poisson Calculus for Spatial Neutral to the Right Processes

LANCELOT F. JAMES¹

The Hong Kong University of Science and Technology
(June 26, 2005)

In this paper we consider classes of nonparametric priors on spaces of distribution functions and cumulative hazard measures that are based on extensions of the neutral to the right (NTR) concept. In particular, spatial neutral to the right processes that extend the NTR concept from priors on the class of distributions on the real line to classes of distributions on general spaces are discussed. Representations of the posterior distribution of the spatial NTR processes are given. A different type of calculus than traditionally employed in the Bayesian literature, based on Poisson process partition calculus methods described in James (2002), is provided which offers a streamlined proof of posterior results for NTR models and its spatial extension. The techniques are applied to progressively more complex models ranging from the complete data case to semiparametric multiplicative intensity models. Refinements are then given which describes the underlying properties of spatial NTR processes analogous to those developed for the Dirichlet process. The analysis yields accessible moment formulae and characterizations of the posterior distribution and relevant marginal distributions. An EPPF formula and additionally a distribution related to the risk and death sets is computed. In the homogeneous case, these distributions turn out to be connected and overlap with recent work on regenerative compositions defined by suitable discretisation of subordinators. The formulae we develop for the marginal distribution of spatial NTR models provide clues on how to sample posterior distributions in complex settings. In addition the spatial NTR is further extended to the mixture model setting which allows for applicability of such processes to much more complex data structures. A description of a species sampling model derived from a spatial NTR model is also given.

1 Introduction

Doksum (1974) considered a nonparametric Bayesian analysis based on the neutral to the right priors. For these priors, if P is a distribution on the real line, then for each partition B_1, \dots, B_k with $B_j = (s_{j-1}, s_j]$, $j = 1, \dots, k$, $s_0 = -\infty$, $s_k = \infty$, $s_i < s_j$ for $i < j$; $P(B_1), \dots, P(B_k)$ is such that each $P(B_i)$ has the same distribution as $V_i \prod_{j=1}^{i-1} (1 - V_j)$, where V_1, V_2, \dots is a collection of independent non-negative random variables. Each choice of distributions for the V_i gives a new NTR prior. If V_i is chosen to be a beta variable with parameter (α_i, β_i) and $\beta_i = \sum_{j=i+1}^{k-1} \alpha_j$, then this gives the Dirichlet process [see Ferguson (1973)] as described in Doksum (1974). The NTR concept allows for flexibility in the choice of prior. Doksum (1974) shows that if P is a NTR distribution, then the posterior distribution of P given a sample X_1, \dots, X_n is also an NTR. The NTR serves as one of the important classes of models in Bayesian nonparametric statistics. In particular there have been numerous applications to models arising in survival analysis subject to possible right censoring. On the other hand, unlike the Dirichlet process, the larger class of NTR processes have not been used in a wider range of statistical applications. The notions of the Dirichlet and NTR processes are related to the work of Freedman (1963), Dubins and Freedman (1966) and Fabius (1964). See Doksum (1974) for a comparison of the various notions.

This article discusses a previously unknown extension of NTR processes from the positive real line to spatial NTR processes on $\mathcal{S} = \mathcal{R}^+ \times \mathcal{X}$, for \mathcal{X} an arbitrary Polish space. This construction is originally proposed in the manuscript of James (2002, section 7). Our interest will be on the

¹AMS 2000 subject classifications. Primary 62G05; secondary 62F15.

Keywords and phrases. Bayesian Nonparametrics, Inhomogeneous Poisson process, Lévy processes, Neutral to the right processes, Regenerative compositions, Survival analysis

posterior analysis of such models under a variety of data structures. Primarily, we are interested in describing the structural features of spatial NTR models, analogous to those for the Dirichlet process, which facilitate their possible usage in more complex settings such as mixture models. The bulk of those results are not known even in the univariate case and are given in section 5. Traditionally, in the univariate case, posterior analysis of NTR models has been formidable. Here, a different type of calculus than traditionally employed in the Bayesian literature, using the *Poisson process partition calculus* methods discussed in James (2002), is provided which offers a streamlined proof of posterior results for NTR models and its spatial extension to a wide range of data structures. In fact the present work is a refinement and extension of the results given in James (2002, section 7). A general outline is given as follows, Section 2 describes the background for NTR processes and their definition in terms of Lévy Z -processes on $(0, \infty)$ and Lévy cumulative hazards Λ with jumps in $[0, 1]$. Here a map between Z and Λ is identified and related to a common Poisson random measure. That is to say all relevant processes are functionals of a single Poisson random measure. This formulation represents one of the key elements in the manuscript. As shall be shown, the mapping allows the usage of a Laplace functional change of measure argument which reveals how one easily updates over events such as $\{T > t\}$. Moreover, in terms of posterior analysis, the mapping allows us to discuss Z and Λ interchangeably. Section 3 defines a spatial NTR process and a corresponding Lévy cumulative hazard measure. Section 4 describes how the techniques used in James (2002) are applied to progressively more complex models ranging from the complete data case to semiparametric multiplicative intensity models. Section 4.2, describes the relevant results from James (2002). Refinements are then given, for the complete data case, in section 5 which describes the underlying properties of spatial NTR processes analogous to those developed for the Dirichlet process. The analysis yields accessible moment formulae and characterizations of the posterior distribution, such as the prediction rule, and relevant marginal distributions. An EPPF formula and additionally a distribution related to the risk and death sets is computed. These formulae provide clues on how to sample posterior distributions in complex settings. In the homogeneous case, these distributions turn out to be connected to and overlap with recent work by Gneden and Pitman (2002) on *regenerative compositions* defined by suitable discretisation of subordinators. We discovered these connections through a mutual exchange of manuscripts in progress. We use some of their results to highlight some interesting spatial NTR processes in section 5. We also note that their work is in a non-Bayesian context, which serves as another nice indication of the richness of NTR models or exponential functionals of subordinators. Our results also have general connections to other related work on exponential functionals of subordinators. In particular, we relate our analysis to the moment formula developed by Carmona, Petit and Yor (1997) and Epifani, Lijoi and Pruenster (2002). In section 6, analysis of the spatial NTR process is further extended to the mixture model setting which allows for applicability of such processes to much more complex data structures. A description of a class of species sampling model derived from a spatial NTR model is also given in section 5. Sections 5 and 6 may be read independently of section 4.3-4.6. Some of the basic elements of this work, also appear in Doksum and James (2003), where an application to Barlow's total time on test transform is discussed.

2 NTR, Lévy and other processes

Doksum (1974, Theorem 3.1) gives an alternate characterization of NTR processes via positive Lévy processes, Z , on \mathcal{R}^+ as follows;

$$(1) \quad 1 - F(t) = S(t) = e^{-Z(t)}$$

where S denotes the survival distribution associated with F , Z is an increasing independent increment process satisfying $Z(0) = 0$ and $\lim_{t \rightarrow \infty} Z(t) = \infty$, *a.s.* From the proof of Theorem 3.1 in Doksum (1974), it follows that $1 - V_j = e^{-[Z(s_j) - Z(s_{j-1})]}$. NTR survival processes essentially coincide with the class of exponential functionals, of possibly inhomogeneous, non-negative Lévy processes.

An important special case is when Z is a subordinator, that is a process with homogeneous increments, which is defined in more detail below. Such objects, and more general exponential functionals of Lévy processes, such as Brownian motion, have been extensively studied by probabilists with applications, for instance, to finance. See in particular Bertoin and Yor (2001) and Carmona, Petit and Yor (1997). Noting some of these connections, Epifani, Lijoi and Pruenster (2002) apply techniques from those manuscripts to obtain expressions for the moments of mean functionals of NTR models. NTR models also arise in coalescent theory as seen for example in Pitman (1999, Proposition 26).

The analysis here will mainly consider processes Z without a drift component. In that case, Z may be represented as

$$(2) \quad Z(ds) = \int_0^\infty yN(dy, ds)$$

where $N(dy, ds)$ is a Poisson random measure on $(0, \infty) \times (0, \infty)$, with mean measure,

$$E[N(dy, ds)] = \tau(dy|s)\eta(ds)$$

chosen such that, $\int_0^\infty \min(1, y)\tau(dy|s) < \infty$ for almost all s relative to η , and

$$S_0(t) = E[S(t)] = E[e^{-Z(t)}] = \exp\left(-\int_0^t \int_0^\infty (1 - e^{-y}) \tau(dy|s) \eta(ds)\right).$$

$S_0 = 1 - F_0$ represents a prior specification of the survival distribution. Note the notation $E[\cdot]$ will typically denote expectation with respect to a random measure unless otherwise specified. The Dirichlet process with shape parameter θF_0 is specified by choosing

$$(3) \quad \tau_{\theta S_0}(dy|s)\eta(ds) = \frac{1}{1 - e^{-y}} e^{-y\theta S_0(s-)} dy \theta F_0(ds).$$

In the case where $\tau(dy|s) = \tau(dy)$, the NTR process will be called *homogeneous*. If in addition, $\eta(ds) = ds$, then Z is called a *subordinator* without a drift component. While, as we shall see, the characterization of NTR in terms of Z is useful from the point of view of calculations, it is difficult to interpret statistically. In particular, since Z and F are discrete, Z cannot be interpreted as the cumulative hazard of F , say Λ . Instead of modelling F directly, Hjort (1990) suggested to model the associated cumulative hazard, Λ , by a Lévy process with jumps restricted to $[0, 1]$. That is, Λ may be represented as

$$(4) \quad \Lambda(ds) = \int_0^1 uN(du, ds)$$

where $N(du, ds)$ is a Poisson random measure on $[0, 1] \times (0, \infty)$, with mean measure,

$$(5) \quad E[N(du, ds)] = \rho(du|s)\eta(ds).$$

Note that the Poisson random measures N associated with Z and Λ are different. The measures ρ and η are chosen such that,

$$E[\Lambda(ds)] = \int_0^1 u\rho(du|s)\eta(ds) = \Lambda_0(ds),$$

where

$$\Lambda_0(ds) = \frac{F_0(ds)}{S_0(s-)},$$

represents a prior specification for the cumulative hazard. The choice of,

$$(6) \quad \rho(du|s)\eta(ds) = \rho_c(du|s)\Lambda_0(ds, dx) = u^{-1}(1 - u)^{c(s)-1} duc(s)\Lambda_0(ds)$$

for $c(s)$ a positive function corresponds to a beta process with parameters c and Λ_0 . Using this framework, a neutral to the right process can be characterised in terms of the product integral representation, [see Gill and Johansen(1990)], of a survival function as follows;

$$(7) \quad 1 - F(t) = S(t) = \prod_{u \leq t} (1 - \Lambda(du)).$$

Importantly, using the interpretation of a cumulative hazard, it is much more natural to write,

$$F(dt) = S(t-) \Lambda(dt).$$

Hjort (1990) shows that when Λ is a beta process with $c(s) := \theta F_0([s, \infty)) = \theta S_0(s-)$ then the F defined by (7) is a Dirichlet process with shape θF_0 . The case of more general beta processes generates a class of generalized Dirichlet processes, see Hjort (1990, section 7A), which is essentially equivalent in distribution to the Beta-Neutral processes in Lo (1993) and the beta-Stacy process discussed in Walker and Muliere (1997). The homogeneous NTR process considered by Ferguson and Phadia (1979) arises when $c(s) = \theta$. These processes are equivalently defined in terms of Z by the τ Lévy measure,

$$(8) \quad \tau_c(dy|s) \Lambda_0(ds) = \frac{1}{1 - e^{-y}} e^{-yc(s)} dy c(s) \Lambda_0(ds).$$

The expressions (1) and (7) suggests an important functional 1-1 relationship between a particular Z and Λ . Dey (1999) and Dey, Dragichi and Ramamoorthi (2000) establish that the Lévy measure $\tau(dy|s) \eta(ds)$ for Z is the image of the Lévy measure of a Λ with Lévy measure $\rho(du|s) \eta(ds)$, via the map (u, s) on $(0, 1) \times (0, \infty)$ to $(-\log(1-u), s)$. This type of correspondence is actually noted, albeit less explicitly, in Hjort (1990). In particular, writing $\tau(dy|s) := \tau(y|s) dy$ and $\rho(du|s) := \rho(u|s) du$, the relationship between the Lévy measures of Z and Λ is described by

$$\tau(y|s) := e^{-y} \rho(1 - e^{-y}|s) \text{ for } y \in (0, \infty) \text{ or } \rho(u|s) = (1 - u)^{-1} \tau(-\log(1 - u)|s) \text{ for } u \in [0, 1].$$

An important consequence of this result, noted and exploited by James (2002), is the fact one can write

$$Z(ds) = \int_0^1 [-\log(1 - u)] N(du, ds)$$

where N is the Poisson random measure associated with Λ with mean measure specified by (5). Hence, Λ and Z are related in distribution via a common Poisson random measure. Specifically,

$$(9) \quad S(t-) = e^{-Z(t-)} = e^{-\int_0^\infty \int_0^1 I_{\{s < t\}} [-\log(1-u)] N(du, ds)} = \prod_{u < t} (1 - \Lambda(du))$$

The representations allows one to by-pass working directly with product integrals. As a consequence of (9) the event $T = t|F$ can be represented in distribution as,

$$F(dt) = e^{-Z(t-)} \Lambda(dt).$$

3 Spatial neutral to the right processes

Suppose that (T, X) denotes a marked pair of random variables on a general Polish space $\mathcal{S} = \mathcal{R}^+ \times \mathcal{X}$, with distribution $F(ds, dx)$. As an example, T could denote a random variable representing time and X could be a random vector in \mathcal{R}^k . A more specific example, is the important problem of estimation of the total medical costs, say X , for a patient over a period of time T when there is possible univariate right censoring of T . The time T is usually assumed to have maximum value $L < \infty$. Here, one uses information in the possibly censored sample of size n . A particular quantity of interest is the mean cost,

$$(10) \quad E[X] = \int_0^\infty x F(L, dx).$$

Bang and Tsiatis (2000) give for example the following 10-year cost model

$$X = X_0 + bT + \sum_{j=1}^{10} \tau_j (\min\{T - (j - 1)\}^+, 1) + dI(T \leq 10)$$

where X_0 is the initial diagnostic cost, b is the deterministic annual cost, τ_j is the random annual cost for the j th year, d is the terminal death cost and $a^+ = \max(0, a)$.

In this section we discuss a spatial extension, proposed by James (2002), of the class of NTR processes from \mathcal{R}^+ to more general Polish spaces. This provides, for instance, a new class of Bayesian models for multivariate survival and reliability models. In particular, similar to NTR models, the posterior distribution of spatial NTR models is again a spatial NTR model when T is subjected to possible univariate right censoring, left-truncation or more general Aalen filtering.

While indeed it is easy to extend Z or Λ to more abstract spaces, the representations in (9) do not immediately suggest an obvious extension for F . The Dirichlet process which is defined over quite arbitrary spaces is a notable exception. However, James and Kwon (2000) recently proposed a method which extends the Beta-Neutral prior of Lo (1993), and by virtue of the equivalences, the beta-Stacy process in Muliere and Walker (1997) and beta distribution function discussed in Hjort (1990, section 7A), to a spatial setting. James (2002) deduced a general definition from elements of their construction. A definition for F on \mathcal{S} is facilitated by the usage of its associated *hazard measure* on \mathcal{S} . From Last and Brandt (1995, A5.3), it follows that such a measure always exists and is defined by,

$$(11) \quad \Lambda(ds, dx) := I\{s > 0\} \frac{F(ds, dx)}{S(s-)}.$$

In particular, $\Lambda(ds, \mathcal{X}) := \Lambda(ds)$ and hence

$$(12) \quad S(s-) := \prod_{u < s} (1 - \Lambda(du, \mathcal{X})).$$

First let N denote a Poisson random measure on $[0, 1] \times \mathcal{S}$ with mean intensity

$$E[N(du, ds, dx)] = \rho(du|s)\eta(ds, dx),$$

and which is characterized by its Laplace functional for every bounded positive measurable function f as

$$E[e^{-N(f)}] := \mathcal{L}_N(f|\rho, \eta) = \exp\left(-\int_{\mathcal{S}} \int_0^1 (1 - e^{-f(u, s, x)}) \rho(du|s) \eta(ds, dx)\right).$$

Denote the law of N as $\mathcal{P}(dN|\rho, \eta)$. A random *Lévy hazard measure*, Λ can be defined as in (4) by

$$\Lambda(ds, dx) = \int_0^1 uN(du, ds, dx).$$

Furthermore, $\rho(du|s)\eta(ds, dx)$ is chosen such that,

$$(13) \quad E[\Lambda(ds, dx)] = \int_0^1 uE[N(du, ds, dx)] = \Lambda_0(ds, dx)$$

where Λ_0 is a (prior) hazard measure. That is,

$$\Lambda_0(ds, dx) = \left[\int_0^1 u\rho(du|s) \right] \eta(ds, dx) = \frac{F_0(ds, dx)}{S_0(s-)},$$

where F_0 and S_0 are prior specifications for F and S . Denote the marginal cumulative hazard as $\Lambda_0(ds) = \Lambda_0(ds, \mathcal{X})$. In addition, $F_0(ds, dx) = P_0(dx|s)F_0(ds) = F_0(ds|x)P_0(dx)$, where P_0 denotes a prior distribution of X . Note also that

$$\int_0^1 u\rho(du|s) = \int_0^\infty (1 - e^{-y})\tau(dy|s) = \phi(1|s)$$

where for $a > 0$, $\phi(a|s) = \int_0^\infty (1 - e^{-ay})\tau(dy|s)$ or $\phi(a) = \int_0^\infty (1 - e^{-ay})\tau(dy)$, is called the Laplace exponent evaluated at a . Λ is an example of a completely random measure which can be characterized by its Laplace functional for a positive measurable function g as

$$E \left[e^{-\Lambda(g)} \right] := \mathcal{L}_\Lambda(g|\rho, \eta) = \exp \left(- \int_{\mathcal{S}} \int_0^1 \left(1 - e^{-g(s,x)u} \right) \rho(du|s) \eta(ds, dx) \right).$$

Similarly a corresponding Z process is obtained by the definition

$$(14) \quad Z(ds, dx) = \int_0^1 [-\log(1 - u)]N(du, ds, dx).$$

Its Laplace functional can be represented in terms of the Lévy measure ρ as,

$$E \left[e^{-Z(g)} \right] = \mathcal{L}_N(f_g|\rho, \eta) = \exp \left(- \int_{\mathcal{S}} \int_0^1 \left(1 - (1 - u)^{g(s,x)} \right) \rho(du|s) \eta(ds, dx) \right),$$

where $f_g(u, s, x) = g(s, x) [-\log(1 - u)]$. In regards to the specification of ρ or τ and η , hereafter, unless otherwise stated, we will go with the convention in the univariate case and choose

$$\eta(ds, dx) = \Lambda_0(ds, dx) \text{ and } \int_0^1 u\rho(du|s) = 1.$$

In order to arrange for the specification of ρ we simply let $\rho(du|s) := [\phi^*(1|s)]^{-1} \rho^*(du|s)$ for Lévy measures ρ^* and τ^* . The law of Λ will be denoted as $\mathcal{P}(d\Lambda|\rho, \Lambda_0)$. Now a spatial neutral to the right process is defined as,

Definition 3.1 (*Spatial Neutral to the Right Process*) Let Λ denote a completely random measure with specifications given by (13), then a Spatial Neutral to the Right process on \mathcal{S} , with parameters ρ , or τ , and $F_0(ds, dx)$, is defined for $t \geq 0$ and each B , an arbitrary measurable set in \mathcal{X} , by

$$(15) \quad F(t, B) := \int_0^t S(s-) \Lambda(ds, B)$$

where, as in (12), $S(s-)$ is a neutral to the right survival process. In particular, $F(ds, dx) := S(s-) \Lambda(ds, dx)$. The random quantities $S(s-)$ and $\Lambda(ds, dx)$ are independent for each s and arbitrary x and

$$E[F(ds, dx)] := E[S(s-)]E[\Lambda(ds, dx)] := e^{-\Lambda_0(s)} \Lambda_0(ds, dx) := F_0(ds, dx).$$

The class of spatial beta Lévy hazard measures and distribution functions, including Dirichlet processes, are defined by replacing $\Lambda_0(ds)$ with $\Lambda_0(ds, dx)$ in (3), (6) or (8). The explicit construction of spatial Beta-Neutral cumulative hazard and survival processes, via ratios of two independent gamma processes, by James and Kwon (2000) is described in James (2002) and Doksum and James (2003). Lo (1993, Section 6) gives the construction for the univariate case. Other interesting examples are now described.

3.0.1 Two parameter extended Beta process

Based on the recent work of Gneden and Pitman (2002) and Pitman (1997), the following class of NTR processes defined by

$$\rho(du) = \frac{\Gamma(a + \theta)}{\Gamma(a)\Gamma(\theta)} u^{a-1} (1-u)^{\theta-1} du$$

for $a > -1$ and $\theta > 0$ takes on new interest. In particular the case of $-1 < a = -\alpha < 0$ and $\theta = \alpha$, which generates renewal compositions connected to the zero set of a Bessel bridge with parameter $2 - 2\alpha$ [see Pitman (1997)], is particularly attractive and remarkable. In particular, it yields a quite tractable NTR process highly related to the important two parameter Poisson-Dirichlet process with parameters (α, α) . We will discuss these points in section 5. Call the corresponding spatial NTR process a beta $(-\alpha, \alpha)$ NTR process. Substituting $c(s)$ for θ and $a(s)$ for a leads to a *extended beta process model* introduced by Kim and Lee (2001). That is the NTR process defined by the Lévy measure,

$$\rho(du) = \frac{\Gamma(a(s) + c(s))}{\Gamma(a(s))\Gamma(c(s))} u^{a(s)-1} (1-u)^{c(s)-1} du.$$

The model now contains the Dirichlet process and more generally Hjort's (1990) beta process.

REMARK 1. Kim and Lee (2001) show that if the true underlying distribution is continuous then the posterior distribution of these models under right censoring are consistent if and only if $a(s) = 0$. However, based on other considerations, such as the work of Pitman (1997), the models for $a(s) \neq 0$ are clearly of interest. Moreover, we note that it is not surprising that most NTR models are inconsistent with respect to observed data from a continuous distribution. However, the behaviour of such models can be quite different in say mixture models where we believe that they potentially have the most utility, in a statistical sense. In such cases the number of multiplicities is governed by the marginal distribution, or true urn distribution, induced by the NTR process. The finite dimensional Dirichlet priors, albeit not NTR models, are an important example of random processes which are not consistent under continuous observed data but exhibit consistent behaviour under a variety of more complex models. See Ishwaran, James and Sun (2001) for a nice application. We expect similar behaviour for some NTR processes. For instance, it is simple to show that under right censoring the posterior distribution of the entire class of extended beta processes are consistent when the number of unique values, say $n(\mathbf{p})$, in a sample of size n is such that $n(\mathbf{p})/n \rightarrow 0$ as $n \rightarrow \infty$. That is to say the true distribution must be discrete.

3.0.2 Generalized Gamma models

While it is true that modelling in terms of the Lévy measure Λ is a more natural object from a statistical point of view, it is noted that the majority of known Lévy measures have jumps in $(0, \infty)$. That is to say, many of the specific processes used in the literature look more like Z than Λ . Another interesting class of measures are the family of generalized gamma random measures discussed in Brix (1999). Using the description of Brix (1999), these are Z processes with Levy measure

$$(16) \quad \tau_{\alpha,b}(dy)\Lambda_0(ds, dx) = \frac{1}{\phi_{\alpha,b}(1)\Gamma(1-\alpha)} y^{-\alpha-1} \exp(-by) dy \Lambda_0(ds, dx)$$

where $\phi_{\alpha,b}(a) = \frac{1}{\alpha}[(b+a)^\alpha - (b)^\alpha]$. The values for α and b are restricted to satisfy $0 < \alpha < 1$ and $0 \leq b < \infty$ or $-\infty < \alpha \leq 0$ and $0 < b < \infty$. Different choices for α and b in $\rho_{\alpha,b}$ yield various subordinators. These include the stable subordinator when $b = 0$, the gamma process subordinator when $\alpha = 0$ and the inverse-Gaussian subordinator when $\alpha = 1/2$ and $b > 0$. When $\alpha < 0$ this results in a class of gamma compound Poisson processes [see Aalen (1992)]. A weighted gamma process, see Lo (1982), is an inhomogeneous process defined by setting $b = b(y)$ and $\alpha = 0$. The corresponding Lévy hazard measure Λ is defined by the choice

$$\rho_{\alpha,b}(du|s) = [\phi_{\alpha,b}(1)]^{-1} (1-u)^{-(b-1)} [-\log(1-u)]^{-\alpha-1} du.$$

Generalized gamma NTR processes with $b > 0$ are discussed in Epifani, Lijoi and Pruenster (2002).

REMARK 2. The definition of F can be adjusted to include prior points of discontinuity say $\{(s_1, w_1), \dots, (s_k, w_k)\}$, in \mathcal{S} , by replacing Λ with the Lévy process,

$$(17) \quad \tilde{\Lambda}_k(ds, dx) = \Lambda(ds, dx) + \sum_{l=1}^k U_l \delta_{s_l, w_l}(ds, dx)$$

where Λ is $\mathcal{P}(d\Lambda|\rho, \Lambda_0)$ and independent of Λ , U_j are independent random variables on $[0, 1]$ with distribution H_j for $j = 1, \dots, k$.

4 Posterior Analysis

The next sections will describe how to systematically obtain the posterior distribution of spatial NTR processes and hence NTR processes in a straightforward manner. The approach amplifies and refines the discussion in James (2002, section 7) and is applied to progressively more complex models. That is we start with the posterior distribution given a completely observed pair in section 4.1 and end with semiparametric models subject to general *Aalen filtering*. The materials in section 5 can be read independent of sections 4.3-4.6.

4.1 Posterior distributions given $\{T=t, X=x\}$

The question of how to find the posterior distribution of Λ and F given the event $\{T_i = t_i, X_i = x_i\}$ for $i = 1, \dots, n$ is a formidable one. Here we start with the case $n = 1$. In the case of NTR processes on \mathcal{R} , Doksum (1974, Example 4.1) gave results that provide the moment generating function of the posterior distribution of $Z(t)$. From these results, Ferguson (1974), describes the posterior distribution of Z via Lévy measures. Ferguson and Phadia (1979) subsequently gave extensions to right censored data which involves a further description of the posterior distribution of F given an event $T > t$ or $T \geq t$. Under such data structures the resulting posterior distribution is again neutral to the right. Analogously, Hjort (1990) derived the posterior distribution of the cumulative hazard Lévy process Λ under possible right censoring and shows that the resulting posterior distribution is again a Lévy process. Kim (1999) using a semi-martingale framework for Λ extended Hjort's result to include more general censoring mechanisms. However, it is fair to say that, in all cases, the techniques used are not trivial and not easily understood. Moreover, they do not provide an obvious way to extend to more general spaces. Here we describe a method of proof to obtain the posterior distribution of F and Λ given $\{T = t, X = x\}$ based on the Poisson process partition calculus framework given in James (2002, 2003). This also provides a simple proof for the univariate case, and as shall be shown essentially contains the result for T subject to right censoring and more general censoring mechanisms. We will then address the case of progressively more complex structures in the forthcoming sections.

The method relies on an exponential change of measure formula derived in James (2002) for Poisson random measures. The formula is analogous to, but much more general than, Proposition 3.1 of Lo and Weng (1989) in the case of the weighted gamma process. The result is now described below in Proposition 1. The proof of Proposition 1 given in James (2002, 2003) is also provided to demonstrate its simple derivation. Here \mathcal{W} denotes an arbitrary Polish space, $BM(\mathcal{W})$ denotes a space of boundedly measurable functions as defined in Daley and Vere-Jones(1988), ν is an arbitrary non-atomic measure.

PROPOSITION 1. (James (2002)) For each non-negative function $f \in BM(\mathcal{W})$ and each g on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$

$$\int_{\mathcal{M}} g(N) e^{-N(f)} \mathcal{P}(dN|\nu) = \mathcal{L}_N(f|\nu) \int_{\mathcal{M}} g(N) \mathcal{P}(dN|e^{-f}\nu),$$

where $\mathcal{P}(dN|e^{-f}\nu)$ is the law of a Poisson process with intensity $e^{-f(w)}\nu(dw)$. In other words, the following absolute continuity result holds, $e^{-N(f)}\mathcal{P}(dN|\nu) = \mathcal{L}_N(f|\nu)\mathcal{P}(dN|e^{-f}\nu)$.

PROOF. By the unicity of Laplace functionals for random measures on \mathcal{X} it suffices to check this result for the case $g(N) = e^{-N(h)}$. It follows that

$$\int_{\mathcal{M}} e^{-N(f+h)} \mathcal{P}(dN|\nu) = \mathcal{L}_N(f|\nu) \int_{\mathcal{M}} e^{-N(h)} \mathcal{P}_f(dN),$$

where for the time being \mathcal{P}_f denotes some law on N . Simple algebra shows that

$$\int_{\mathcal{M}} e^{-N(h)} \mathcal{P}_f(dN) = \frac{\mathcal{L}_N(f+h|\nu)}{\mathcal{L}_N(f|\nu)},$$

and hence $\mathcal{P}_f(dN) = \mathcal{P}(dN|e^{-f}\nu)$. \square

After a few preliminary statements, it will be shown how Proposition 1 is used to provide a simple and clear derivation of the posterior distribution. Note that the joint law of $\{T, X, N\}$ is given by

$$F(dt, dx)\mathcal{P}(dN|\rho, \Lambda_0) = S(t-)\Lambda(dt, dx)\mathcal{P}(dN|\rho, \Lambda_0)$$

and since F and Λ are functionals of N a description of the posterior distribution of N given $\{T = t, X = x\}$ will easily yield the corresponding posterior distributions of F and Λ . Furthermore choosing $f_{t-}(u, s, x) = -I\{s < t\} \log(1 - u)$, we have the crucial (in distribution) representation,

$$S(t-)\Lambda(dt, dx)\mathcal{P}(dN|\rho, \Lambda_0) = e^{-N(f_{t-})} \left[\int_0^1 uN(du, dt, dx) \right] \mathcal{P}(dN|\rho, \Lambda_0).$$

The posterior distribution is described below.

THEOREM 1. Suppose that N is $\mathcal{P}(dN|\rho, \Lambda_0)$ and $T, X|N$ has distribution F , then the posterior distribution of $N|T, X$ is equivalent to the distribution of the random measure

$$N_1^* = N_1 + \delta_{J, T, X}$$

where N_1 is a Poisson random measure with mean measure

$$E[N_1(du, ds, dx)] = (1 - u)^{I\{T > s\}} \rho(du|s)\Lambda_0(ds, dx)$$

and conditionally independent of N_1 , J is a random variable with distribution

$$(18) \quad \mathbb{P}\{J \in du|T\} \propto u\rho(du|T).$$

PROOF. The task is to identify the posterior distribution denoted as $\pi(dN|T, Y)$. This is equivalent to identifying the conditional Laplace functional N given $\{T = t, X = x\}$, say $\mathcal{L}_N^*(h|t, x)$, which must satisfy

$$(19) \quad \int_{S \times \mathcal{M}} g(t, x) e^{-N(h)} S(t-)\Lambda(dt, dx)\mathcal{P}(dN|\rho, \Lambda_0) = \int_{S \times \mathcal{M}} g(t, x) \mathcal{L}_N^*(h|t, x) E[S(t-)\Lambda(dt, dx)]$$

for some arbitrary integrable function g and bounded positive measurable h . By Proposition 1,

$$S(t-)\mathcal{P}(dN|\rho, \Lambda_0) = \mathcal{P}(dN|e^{-f_{t-}}\rho, \Lambda_0) E[S(t-)]$$

with $f_{t-}(u, s, x) = -I\{s < t\} \log(1 - u)$. Note that this operation identifies $\mathcal{P}(dN|e^{-f_{t-}}\rho, \Lambda_0)$ as the posterior distribution of $N|\{T \geq t\}$. Another application of Proposition 1 gives,

$$e^{-N(h)}\mathcal{P}(dN|e^{-f_{t-}}\rho, \Lambda_0) = \mathcal{P}(dN|e^{-h-f_{t-}}\rho, \Lambda_0)\mathcal{L}_N(h|e^{-f_{t-}}\rho, \Lambda_0).$$

Now the expectation of $\Lambda(dt, dx)$ with respect to the law $\mathcal{P}(dN|e^{-h-f_{t-}}\rho, \Lambda_0)$ is given by,

$$(20) \quad \left[\int_0^1 e^{-h(u,t,x)}(1-u)^{I\{t>t\}} u \rho(du|t) \right] \Lambda_0(dt, dx).$$

Recalling that $E[\Lambda(dt, dx)] = \left[\int_0^1 u \rho(du|t) \right] \Lambda_0(dt, dx)$, the statements above imply that the expressions in (19) are equal to,

$$\int_{\mathcal{S} \times \mathcal{M}} g(t, x) \mathcal{L}_N(h|e^{-f_{t-}}\rho, \Lambda_0) \left[\int_0^1 e^{-h(u,t,x)} \mathbb{P}\{J \in du|t\} \right] E[S(t-)] E[\Lambda(dt, dx)]$$

which implies the desired result,

$$\mathcal{L}_N^*(h|t, x) = \mathcal{L}_N(h|e^{-f_{t-}}\rho, \Lambda_0) \left[\int_0^1 e^{-h(u,t,x)} \mathbb{P}\{J \in du|t\} \right]$$

almost everywhere. \square

The result in Theorem 1 now yields readily the following description of the posterior distributions of Λ and F and Z ,

Corollary 4.1 *Let F be a spatial NTR and Λ is its corresponding Lévy hazard measure with law $\mathcal{P}(d\Lambda|\rho, \Lambda_0)$. Suppose that $T, X|F$ has distribution F , then the posterior distribution of Λ given T, X is equivalent to the law of the Lévy hazard measure,*

$$\Lambda_1^*(ds, dx) = \int_0^1 u N_1^*(du, ds, dx) = \Lambda_1(ds, dx) + J \delta_{T, X}(ds, dx)$$

where Λ_1 is a Lévy hazard measure with Lévy mean measure $(1-u)^{I\{T>s\}} \rho(du|s) \Lambda_0(ds, dx)$ and J has distribution described in (18). The law of Λ_1 is equivalent to the posterior distribution of Λ given $T \geq t$. The posterior distribution of F is a spatial NTR whose law is equivalent to the law of a random measure F_1^* , defined by replacing Λ with Λ_1^* in (15). Additionally the distribution of its corresponding Z process, as in (14), is equivalent to the law of the random measure

$$Z_1^*(ds, dx) = Z_1(ds, dx) - \log(1-J) \delta_{T, X}(ds, dx) = \int_0^1 [-\log(1-u)] N_1^*(du, ds, dx),$$

with $Z_1(ds, dx) = \int_0^1 [-\log(1-u)] N_1(du, ds, dx)$. These result imply that the posterior spatial NTR process of F is representable as

$$F_1^*(ds, dx) = e^{-Z_1^*(s-)} \Lambda_1^*(ds, dx) = e^{-Z_1(s-)} (1-J)^{I\{s>T\}} \Lambda_1^*(ds, dx)$$

REMARK 3. Note that the representation in (20) shows clearly why in the homogeneous case, as remarked prior to Theorem 5 in Ferguson and Phadia (1979), the jump J is independent of the jump time T . In general, we see that the distribution of the J can be described as,

$$\mathbb{P}\{J \in du|T\} \propto u(1-u)^{I\{T>T\}} \rho(du|T),$$

where of course $I\{T > T\} = 0$.

4.2 Updating in the general case via Poisson partition calculus

The results in the previous section already hint at the basic framework of how to proceed for more complex models. As shown, a key role is played by the exponential change of measure formula in Proposition 1. In fact this result, in combination with a moment calculation, is all that is needed to obtain Theorem 1. The use of Proposition 1 is facilitated by the functional relationship between the product integrals based on Lévy hazard measures and corresponding Lévy processes Z via the map (u, s, x) to $(-\log(1-u), s, x)$. In general, because of this functional relationship and the characteristics of general multiplicative intensity models, this allows one to represent the joint structure of such Bayesian models in the form

$$(21) \quad L(\text{data}|\Lambda)\mathcal{P}(dN|\rho, \Lambda_0) = e^{-N(f_m)} \left[\prod_{i=1}^n \int_0^1 h_i(u)N(du_i, dT_i, X_i) \right] \mathcal{P}(N|\rho, \Lambda_0).$$

In this case, m denotes the total sample size, n denotes the number of completely observed points, and h_i for $i = 1, \dots, m$ are strictly positive functions taking values in $[0, 1]$. The quantity $L(\text{data}|\Lambda)$ denotes the likelihood of the data depending on the Lévy hazard measure Λ . The framework also allows for further dependence on covariates and Euclidean parameters, say β , which, for the moment, is suppressed. Similar to the case of Theorem 1, a posterior analysis for the various functionals of N follows from the identification of the posterior distribution of $N|\text{data}$, say $\pi(dN|\text{data})$, which must satisfy the following disintegration of (21),

$$L(\text{data}|\Lambda)\mathcal{P}(dN|\rho, \eta) = \pi(dN|\text{data})E[L(\text{data}|\Lambda)],$$

where

$$E[L(\text{data}|\Lambda)] = \int_{\mathcal{M}} e^{-N(f_m)} \prod_{i=1}^n \int_0^1 h_i(u)N(du_i, dT_i, dX_i)\mathcal{P}(N|\rho, \Lambda_0).$$

is the marginalized likelihood of the data. We see already from an application of Proposition 1, that identification of the form of f_m will dictate how the Lévy measure, ρ , associated with Λ and N is updated. Specifically we can apply Proposition 1 to easily describe an exponential updating formula for Λ .

Proposition 4.1 *Let N denote a Poisson process with law $\mathcal{P}(dN|\rho, \Lambda_0)$ over \mathcal{S} . Let f denote a positive function on $[0, 1] \times \mathcal{S}$. Suppose that $w(\Lambda)$ is a positive integrable function of Λ representable in distribution as $w(\Lambda) = e^{-N(f)}$. Then,*

$$w(\Lambda)\mathcal{P}(d\Lambda|\rho, \Lambda_0) = \mathcal{L}_N(f|\rho, \Lambda_0)\mathcal{P}(d\Lambda|e^{-f}\rho, \Lambda_0),$$

where $\mathcal{P}(d\Lambda|e^{-f}\rho, \Lambda_0)$ denotes the law of Λ with intensity $e^{-f(u,s,x)}\rho(du|s)\Lambda_0(ds, dx)$.

REMARK 4. This result not only gives easy updating formulae but also establishes various absolute continuity relationships for classes of measures.

We present another interesting variation of Proposition 1, that illustrates in more detail how Bayesian updating works over products of exponential functionals and additionally yields otherwise nonobvious moment formulae for such expressions. This will be used in section 5.

Proposition 4.2 *Let $\tilde{f}_1, \dots, \tilde{f}_n$ denote functions in $BM(\mathcal{W})$. Furthermore define for $k = 1, \dots, n$, $f_k = \sum_{i=1}^k \tilde{f}_i$. Then for each n ,*

$$e^{-N(f_n)}\mathcal{P}(dN|\rho, \eta) = \left[\prod_{j=k+1}^n e^{-N(\tilde{f}_j)} \right] \mathcal{P}(dN|e^{-f_k}\rho, \eta)\mathcal{L}_N(f_1|\rho, \eta) \prod_{i=2}^k \mathcal{L}_N(\tilde{f}_i|e^{-f_{i-1}}\rho, \eta)$$

which implies for any n , $\mathcal{L}_N(f_n|\rho, \eta) = \prod_{i=1}^n \mathcal{L}_N(\tilde{f}_i|e^{-f_{i-1}}\rho, \eta)$, with $f_0 = 0$.

Notice that from an application of Proposition 1, it follows that the expressions in (21) are equivalent to

$$(22) \quad \mathcal{L}_N(f_m|\rho, \Lambda_0) \left[\prod_{i=1}^n \int_0^1 h_i(u) N(du_i, dT_i, X_i) \right] \mathcal{P}(dN|e^{-f_m} \rho, \Lambda_0).$$

We now present results from James (2002, 2003), and show how they can be used to easily obtain descriptions of the posterior distributions and moment formulae, such as the evaluation of $E[L(\text{data}|\Lambda)]$. Consider the joint measure,

$$(23) \quad \prod_{i=1}^n N(dW_i) \mathcal{P}(dN|\nu)$$

where W_i for $i = 1, \dots, n$ are random elements on an arbitrary polish space \mathcal{W} and $\mathcal{P}(dN|\nu)$ denotes the law of a Poisson random measure with non-atomic mean intensity, ν . The results of James (2002, 2003) provide a partition based representation of the disintegration of (23) say

$$\prod_{i=1}^n N(dW_i) \mathcal{P}(dN|\nu) = \mathcal{P}(dN|\nu, \mathbf{W}) M(d\mathbf{W})$$

where, $M(d\mathbf{W}) = \int_{\mathcal{M}} \prod_{i=1}^n N(dW_i) \mathcal{P}(dN|\nu)$, denotes a joint marginal measure of $\mathbf{W} = \{W_1, \dots, W_n\}$, which behaves in many respects like an exchangeable urn distribution. It follows that one can always represent \mathbf{W} as $(\mathbf{W}^*, \mathbf{p})$, where $\mathbf{W}^* = \{W_1^*, \dots, W_{n(\mathbf{p})}^*\}$ denotes the unique values and $\mathbf{p} = \{E_1, \dots, E_{n(\mathbf{p})}\}$ denotes a partition of $\{1, \dots, n\}$ of size $n(\mathbf{p})$ recording which variables are equal. That is, \mathbf{p} is induced according to the relationship $W_i = W_j^*$ if and only if $i \in E_j$. The number of elements in each cell E_j is denoted as e_j or $e_{j:n}$. The partition mechanism is ideally suited to easily handle the posterior analysis of random measures when observed data contains ties. In particular, this framework eases the description of posterior quantities when W_i are possibly missing values in a Bayesian mixture model analogous to results for the Dirichlet process in Antoniak (1974) and more specifically Lo (1984). The quantity $\mathcal{P}(dN|\nu, \mathbf{W})$ denotes a conditional distribution of N given the points \mathbf{W} , which is described by the Laplace functional

$$\int_{\mathcal{M}} e^{-N(f)} \mathcal{P}(dN|\nu, \mathbf{W}) = \mathcal{L}_N(f|\nu) \prod_{j=1}^{n(\mathbf{p})} e^{-f(W_j^*)}.$$

That is, $\mathcal{P}(dN|\nu, \mathbf{W})$ corresponds to the law of the random measure $N + \sum_{j=1}^{n(\mathbf{p})} \delta_{W_j^*}$, where N is $\mathcal{P}(dN|\nu)$ independent of the points \mathbf{W} . These results coupled with Proposition 1 pave the way for quite straightforward analysis, in the tradition of known results for the Dirichlet process, for a variety of challenging problems arising in Bayesian Nonparametrics. In fact the results described above, and the forthcoming Proposition 2, follow essentially from Proposition 1, see James (2002, 2003) for an elementary proof using induction. For our present purposes, it suffices to use the following form of the results contained in James (2002, 2003);

PROPOSITION 2. (*James (2002)*) *For each non-negative function $f \in BM(\mathcal{W})$, the following change of measure formulae holds,*

$$e^{-N(f)} \left[\prod_{i=1}^n N(dW_i) \right] \mathcal{P}(dN|\nu) = \mathcal{P}(dN|e^{-f} \nu, \mathbf{W}) \mathcal{L}_N(f|\nu) \prod_{j=1}^{n(\mathbf{p})} e^{-f(W_j^*)} \prod_{j=1}^{n(\mathbf{p})} \nu(dW_j^*).$$

Note that the moment measure is expressible via conditional moment measures as,

$$M(d\mathbf{W}) = \prod_{j=1}^{n(\mathbf{p})} \nu(dW_j^*) = \nu(dW_1) \prod_{i=2}^n \left[\nu(dW_i) + \sum_{j=1}^{n(\mathbf{p}_{i-1})} \delta_{X_j^*}(dW_i) \right]$$

where $n(\mathbf{p}_{i-1})$ is the size of the partition of $\{1, \dots, i-1\}$ encoding the ties between W_1, \dots, W_{i-1} .

In the forthcoming applications, $W_i = \{U_i, T_i, X_i\}$ for $i = 1, \dots, n$, and for $w = (u, s, x)$, $\nu(dw) = \rho(du|s)\Lambda_0(ds, dx)$. In addition, $W_j^* = \{U_j^*, T_j^*, X_j^*\}$ for $j = 1, \dots, n(\mathbf{p})$, denotes the $n(\mathbf{p})$ unique triples. For notational convenience we will often write $u_j^* := u_j$. Posterior analysis is then facilitated by proper identification of f in each case.

4.3 Posterior distributions for complete data

We now focus on the posterior distribution of Λ and F given iid pairs, (T_i, X_i) for $i = 1, \dots, n$. Notice that in this case the (conditional) joint distribution of the data takes the form,

$$(24) \quad L(d\mathbf{T}, d\mathbf{X}|\Lambda) := \left[\prod_{i=1}^n S(T_i-) \right] \prod_{i=1}^n \Lambda(dT_i, dX_i).$$

Now, assuming that Λ is $\mathcal{P}(d\Lambda|\rho, \Lambda_0)$, we have the following, in distribution, equivalences,

$$\left[\prod_{i=1}^n S(T_i-) \right] = e^{-\int Y_n(s)Z(ds)} = \prod (1 - \Lambda(du))^{Y_n(s)}$$

for $Y_n(s) = \sum_{i=1}^n I\{T_i > s\}$. Hence, analogous to the case of $n = 1$, one can represent the joint distribution of $\mathbf{T}, \mathbf{X}|N$ as

$$(25) \quad L(d\mathbf{T}, d\mathbf{X}|\Lambda) = e^{-N(f_{Y_n})} \prod_{i=1}^n \int_0^1 u_i N(du_i, dT_i, dX_i)$$

where $f_{Y_n}(u, s, x) = -Y_n(s) \log(1 - u)$. It follows that the posterior distribution of Λ and F are determined by the posterior distribution of N , say $\pi(dN|\mathbf{T}, \mathbf{X})$, which must satisfy the following disintegration,

$$e^{-N(f_{Y_n})} \prod_{i=1}^n \Lambda(dT_i, dX_i) \mathcal{P}(dN|\rho, \Lambda_0) = \pi(dN|\mathbf{T}, \mathbf{X}) L(d\mathbf{T}, d\mathbf{X})$$

where using Proposition 1 on (25),

$$(26) \quad L(d\mathbf{T}, d\mathbf{X}) = \mathcal{L}_N(f_{Y_n}|\rho, \Lambda_0) \int_{\mathcal{M}} \left[\prod_{i=1}^n \int_0^1 u_i N(du_i, dT_i, dX_i) \right] \mathcal{P}(dN|e^{-f_{Y_n}}\rho, \Lambda_0),$$

is the marginal distribution of $\{(T_1, X_1), \dots, (T_n, X_n)\}$. Notice that, $E[\prod_{i=1}^n S(t_i-)] = \mathcal{L}_N(f_{Y_n}|\rho, \eta)$ is an expression for the marginal joint distribution of $\{T_1 \geq t_1, \dots, T_n \geq t_n\}$. An application of Proposition 4.2 yields an important refined version which will be given in section 5. In addition $\mathcal{P}(dN|e^{-f_{Y_n}}\rho, \Lambda_0)$ is the posterior distribution of $N|\{T_1 \geq t_1, \dots, T_n \geq t_n\}$ which indicates that N is a Poisson random measure with mean measure,

$$(27) \quad E[N(du, ds, dx)|e^{-f_{Y_n}}\rho, \eta] = (1 - u)^{Y_n(s)} \rho(du|s)\Lambda_0(ds, dx).$$

This, as indicated in Proposition 4.1, also identifies the posterior distributions of Λ , F and Z given $\{T_1 \geq t_1, \dots, T_n \geq t_n\}$. Using Proposition 2 and the fact that $\prod_{i=1}^n u_i = \prod_{j=1}^{n(\mathbf{p})} \prod_{i \in E_j} u_j^* = \prod_{j=1}^{n(\mathbf{p})} u_j^{e_{j,n}}$, it follows that the marginal joint distribution in (26) is expressible as,

$$(28) \quad L(d\mathbf{T}, d\mathbf{X}) = \mathcal{L}_N(f_{Y_n}|\rho, \Lambda_0) \prod_{j=1}^{n(\mathbf{p})} \kappa_{e_{j,n}}(e^{-f_{Y_n}}\rho|T_j^*)\Lambda_0(dT_j^*, dX_j^*),$$

where

$$(29) \quad \kappa_{e_{j,n}}(e^{-f_{Y_n}} \rho | T_j^*) = \int_0^1 u^{e_{j,n}} (1-u)^{Y_n(T_j^*)} \rho(du | T_j^*).$$

A description of the posterior distribution is now given.

THEOREM 2. *Let F be a spatial NTR process and Λ is its corresponding Lévy hazard measure with law $\mathcal{P}(d\Lambda | \rho, \Lambda_0)$. Suppose that $T_i, X_i | F$ are iid F for $i = 1, \dots, n$, then the posterior distribution of Λ given $\{(T_1, X_1), \dots, (T_n, X_n)\}$ is equivalent to the law of the Lévy hazard measure,*

$$\Lambda_n^* = \Lambda_n + \sum_{j=1}^{n(\mathbf{p})} J_{j,n} \delta_{T_j^*, X_j^*}$$

where Λ_n is a Lévy hazard measure with Lévy measure as in (27), and the $J_{j,n}$ are conditionally independent with distributions

$$\mathbb{P}\{J_{j,n} \in du | T_j^*\} := H_j^*(du) \propto (1-u)^{Y_n(T_j^*)} u^{e_{j,n}} \rho(du | T_j^*)$$

The posterior distribution of F is a spatial NTR whose law is equivalent to the law of a random measure F_n^* , defined by replacing Λ with Λ_n^* in (15). The posterior distribution of N is equivalent to the random measure $N_n^* = N_n + \sum_{j=1}^{n(\mathbf{p})} \delta_{J_{j,n}, T_j^*, X_j^*}$, where N_n is $\mathcal{P}(dN | e^{-f_{Y_n}} \rho, \Lambda_0)$.

PROOF. The proof is similar in spirit to the case of $n = 1$. First, noting the arguments used to derive the marginal distribution, an application of Proposition 2 yields the following disintegration,

$$(30) \quad \left[\prod_{i=1}^n u_i N(dW_i) \right] \mathcal{P}(dN | e^{-f_{Y_n}} \rho, \Lambda_0) = \mathcal{P}(dN | e^{-f_{Y_n}} \rho, \Lambda_0, \mathbf{W}) \prod_{j=1}^{n(\mathbf{p})} u_j^{e_{j,n}} (1-u_j)^{Y_n(T_j^*)} \nu(dW_j^*),$$

where $\nu(dW_j^*) = \rho(du_j | T_j^*) \Lambda_0(dT_j^*, dX_j^*)$, and $\prod_{j=1}^{n(\mathbf{p})} e^{-f_{Y_n}(W_j^*)} = \prod_{j=1}^{n(\mathbf{p})} (1-u_j)^{Y_n(T_j^*)}$. In addition,

$$\int_{\mathcal{M}} e^{-N(h)} \mathcal{P}(dN | e^{-f_{Y_n}} \rho, \Lambda_0, \mathbf{W}) = \mathcal{L}_N(h | e^{-f_{Y_n}} \rho, \Lambda_0) \prod_{j=1}^{n(\mathbf{p})} e^{-h(u_j, T_j^*, X_j^*)}.$$

Now, combining (30) with the form of the marginal distribution given in (28) implies that the Laplace functional of the posterior distribution of N must be,

$$\mathcal{L}_N(h | e^{-f_{Y_n}} \rho, \Lambda_0) \prod_{j=1}^{n(\mathbf{p})} \int_0^1 e^{-h(u, T_j^*, X_j^*)} H_{j,n}^*(du),$$

which is the Laplace functional of N_n^* . In other words this result is deduced by multiplying and dividing (30) by $\prod_{j=1}^{n(\mathbf{p})} \kappa_{e_{j,n}}(e^{-f_{Y_n}} \rho | T_j^*)$ defined in (29). \square

Corollary 4.2 *Suppose that Λ is $\mathcal{P}(d\Lambda | \rho, \Lambda_0)$, then, it follows from Theorem 2 that the posterior distribution of the corresponding Z process is equivalent to the the law of the random measure*

$$Z_n^*(ds, dx) = Z_n(ds, dx) + \sum_{j=1}^{n(\mathbf{p})} Z_{j,n} \delta_{T_j^*, X_j^*}(ds, dx) = \int_0^1 [-\log(1-u)] N_n^*(du, ds, dx).$$

where $Z_n(ds, dx) = \int_0^1 [-\log(1-u)] N_n(du, ds, dx)$, and each $Z_{j,n} = -\log(1 - J_{j,n})$, with distribution,

$$\mathbb{P}\{Z_{j,n} \in dy | T_j^*\} := H_j^*(d(1 - e^{-y})) \propto (1 - e^{-y})^{e_{j,n}} e^{-Y_n(T_j^*)} \tau(dy | T_j^*)$$

Additionally the posterior distribution of F is equivalent in law to,

$$(31) \quad F_n^*(ds, dx) = e^{-Z_n^*(s^-)} \Lambda_n^*(ds, dx)$$

REMARK 5. Using the representation in (31) it follows that the Bayesian prediction rule is given by $E[F_n^*(ds, dx)]$, which can be expressed in several ways. We discuss this further in section 5.

4.4 Censoring, Filtering and Truncation

In the previous section, we established the posterior distribution of F , Z and Λ in the complete data setting. We now focus on the case where $\{T_1, \dots, T_m\}$ are subject to general univariate independent censoring, filtering and truncation which can be expressed via an *Aalen-filter*. The Aalen-filter concept is described in Andersen, Borgan, Gill and Keiding (1993, Section III), and includes multiplicative intensity models subject to right censoring, left truncation and left filtering. In addition, we will assume that, similar to the type of models discussed in Huang and Louis (1998), the marks X_i are observed only when the corresponding T_i are observed. Specifically, again suppose that Λ is a Lévy hazard measure with law $\mathcal{P}(d\Lambda|\rho, \Lambda_0)$, and now conditional on Λ consider $\{(T_1, X_1), \dots, (T_m, X_m)\}$ iid random pairs with hazard measure Λ , and define their corresponding filtered hazard measures,

$$\Lambda_{c,i}(ds, dx) = C_i(s)\Lambda(ds, dx)$$

where $C_i(s) = I\{s \in B_i\}$, for a random set B_i independent of (T_i, X_i) , is called an Aalen-filter. Here we assume that the pair is observed only if $C_i(T_i) = 1$, that is $T_i \in B_i$. The right censoring model corresponds to the choice of $B_i = [0, U_i]$ for U_i a positive random variable. That is to say T_i is right censored by U_i if $T_i > U_i$, otherwise it is observed. Univariate right censoring and left truncation can be represented by the random set $B_i = (V_i, U_i]$ where $V_i < U_i$ almost surely. Let $Y_i^c(s) = C_i(s)I\{T_i > s\}$, then the likelihood of the data (depending on) given Λ may be represented as

$$\left[\prod_{i=1}^m \prod (1 - \Lambda_{c,i}(ds))^{Y_i^c(s)} \right] \prod_{\{i: C_i(T_i)=1\}} \Lambda(dT_i, dX_i).$$

Hereafter, for $n \leq m$, we will let $\{(T_1, X_1), \dots, (T_n, X_n)\}$ denote the observed pairs. Hence the multiplicative intensity likelihood can be re-expressed as

$$(32) \quad L_c(d\mathbf{T}, d\mathbf{X}|\Lambda) = \left[\prod_{i=1}^m \prod (1 - \Lambda_{c,i}(ds))^{Y_i^c(s)} \right] \prod_{i=1}^n \Lambda(dT_i, dX_i)$$

We will now show that subject to the model (32), the posterior distributions of F and Λ are still spatial NTR and Lévy hazard measures, which is analogous to the univariate results of Ferguson and Phadia (1979) and Hjort (1990) for right censoring. Additionally our results are generalizations of Kim's (1999) results, who seems to be the first to notice that the posterior distribution of the cumulative hazard Lévy process Λ on \mathcal{R}^+ is again a Levy process under more general multiplicative intensity models with hazards of the form $\Lambda_{c,i}$. Now notice that $-\log(1 - C_i(s)u) = -C_i(s)\log(1 - u)$, and hence for each $\Lambda_{c,i}$ there exist filtered processes $Z_{c,i}(ds, dx) = C_i(s)Z(ds, dx)$ for $i=1, \dots, m$. Moreover, similar to the complete data case, we have the following in distribution results.

$$\left[\prod_{i=1}^m \prod (1 - \Lambda_{c,i}(ds))^{Y_{i,c}(s)} \right] = e^{-\int Y_{c,m}(s)Z(ds)} = e^{-N(f_{Y_{c,m}})},$$

where $Y_{c,m}(s) = \sum_{i=1}^m Y_i^c(s)$, and N is $\mathcal{P}(dN|\rho, \Lambda_0)$. It follows that,

$$(33) \quad L_c(d\mathbf{T}, d\mathbf{X}|\Lambda) := e^{-N(f_{Y_{c,m}})} \prod_{i=1}^n \int u_i N(du_i, dT_i, dX_i).$$

Now, similar to (28), an application of Proposition 2 yields the marginal likelihood,

$$(34) \quad E[L_c(d\mathbf{T}, d\mathbf{X}|\Lambda) = \mathcal{L}_N(f_{Y_{c,m}}|\rho, \eta) \prod_{j=1}^{n(\mathbf{p})} \kappa_{e_{j,n}}(e^{-f_{Y_{c,m}}} \rho|T_j^*) \Lambda_0(dT_j^*, dX_j^*),$$

where $\kappa_{e_{j,n}}(e^{-f_{Y_{c,m}}} \rho|T_j^*)$ is defined similar to (29). The form of the model in (33), and (34), show that using the arguments for the complete data case, with obvious minor adjustments, immediately yields the following description of the posterior distribution;

THEOREM 3. *Let Λ be a Lévy hazard measure with law $\mathcal{P}(d\Lambda|\rho, \Lambda_0)$. Then subject to the data structure described by (33), the posterior distribution of Λ is equivalent to the law of the Lévy hazard measure, $\Lambda_{c,m}^* = \Lambda_{c,m} + \sum_{j=1}^{n(\mathbf{p})} J_{j,n} \delta_{T_j^*, X_j^*}$ where $\Lambda_{c,m}$ is a Lévy hazard measure with Lévy measure $(1-u)^{Y_{c,m}(s)} \rho(du|s) \Lambda_0(ds, dx)$ and the $J_{j,n}$ are conditionally independent with with distributions*

$$\mathbb{P}\{J_{j,n} \in du|T_j^*\} := H_{j,c}^*(du) \propto (1-u)^{Y_{c,m}(T_j^*)} u^{e_{j,n}} \rho(du|T_j^*)$$

The posterior distribution of F is a spatial NTR whose law is equivalent to the law of a random measure F_n^* , defined by replacing Λ with Λ_n^* in (15). Descriptions for Z and N are similar to Theorem 2, with obvious adjustments relative to $\Lambda_{c,m}$.

The next corollary describes the posterior distribution of the spatial beta process, and hence of a Dirichlet process and generalized spatial Dirichlet processes on \mathcal{S} , subject to general Aalen-filtering,

Corollary 4.3 *Let Λ be a spatial beta process on \mathcal{S} with parameters, $c(s)$ and $\Lambda_0(ds, dx)$, then subject to the data structure described by (33), the posterior distribution of Λ is a spatial beta process with parameters $c(s) + Y_{c,m}(s)$ and $[c(s)/(c(s) + Y_{c,m}(s))] \Lambda_0(ds, dx)$. That is to say, $\Lambda_{c,m}$ is a beta process with intensity $u^{-1}(1-u)^{Y_{c,m}(s)+c(s)} c(s) \Lambda_0(ds, dx)$ and $(J_{j,n})$ are independent beta random variables with parameters $e_{j,n}$ and $c(T_j^*) + Y_{c,m}(T_j^*)$.*

4.5 Multiplicative intensity semiparametric models

In this section analysis for classes of semiparametric multiplicative intensity models subject to the various *censoring* mechanisms described in the previous section and depending possibly on covariates is given. The framework captures as a special case the Bayesian Cox proportional hazards models subject to right censoring considered in Hjort (1990) and recently discussed and extended to left truncation and right censoring in Kim and Lee (2003). Although semiparametric models are in some sense much more complex than the models discussed in the previous section, we will show that our techniques easily yield explicit posterior analysis in this setting with only mild adjustments to the arguments. Again, a key feature is the proper identification of mappings between hazard functions and their corresponding Z functions. First, for $i = 1, \dots, m$ and for $\beta \in \mathcal{R}^d$ we define (conditional on β) Lévy hazard measures

$$(35) \quad \Lambda_{c,i}(ds, dx|\beta) = C_i(s) \int_0^1 h_i(u, \beta) N(du, ds, dx)$$

where N is $\mathcal{P}(dN|\rho, \Lambda_0)$, and h_i are integrable functions taking values in $[0, 1]$. The restriction to $[0, 1]$ is necessary to ensure that these processes are hazard measures. Accordingly, corresponding to each $\Lambda_{c,i}(\cdot|\beta)$, there exists

$$Z_{c,i}(ds, dx|\beta) = \int_0^1 [-C_i(s) \log(1 - h_i(u, \beta))] N(du, ds, dx),$$

where, $\prod_{u < t} (1 - \Lambda_{c,i}(du|\beta)) = e^{-Z_{c,i}(t-\beta)}$. The definition in (35) indicates that the Laplace functional of each $\Lambda_{c,i}(\cdot|\beta)$ takes the form,

$$E \left[e^{-\Lambda_{c,i}(g|\beta)} \right] := \exp \left(- \int_{\mathcal{S}} \int_0^1 \left(1 - e^{-g(s,x)C_i(s)h_i(u,\beta)} \right) \rho(du|s) \Lambda_0(ds, dx) \right).$$

As intended, this shows that they are functionals of Λ , which is $\mathcal{P}(d\Lambda|\rho, \Lambda_0)$, and β . Now consider the filtered likelihood of the data,

$$L_c(d\mathbf{T}, d\mathbf{X}|\Lambda, \beta) = \left[\prod_{i=1}^m \prod (1 - \Lambda_{c,i}(ds|\beta))^{Y_{i,c}(s)} \right] \prod_{i=1}^n \Lambda_{c,i}(dT_i, dX_i|\beta).$$

Note that in this case we have the distributional equality,

$$(36) \quad \left[\prod_{i=1}^m \prod (1 - \Lambda_{c,i}(ds|\beta))^{Y_{i,c}(s)} \right] = \left[\prod_{i=1}^m e^{-\int Y_i^c(s)Z_{c,i}(ds|\beta)} \right] = e^{-N(f_{\beta,m})},$$

where $f_{\beta,m}(u, s, x) = -\sum_{i=1}^m Y_i^c(s) \log(1 - h_i(u, \beta))$. The likelihood takes the form

$$(37) \quad L_c(d\mathbf{T}, d\mathbf{X}|\Lambda, \beta) := e^{-N(f_{\beta,m})} \prod_{i=1}^n \int h_i(u_i, \beta) N(du_i, dT_i, dX_i).$$

Placing a prior on β , say $\pi(d\beta)$, the joint structure of $\mathbf{T}, \mathbf{X}, N, \beta$ takes the form

$$(38) \quad L_c(dt, dx|\Lambda, \beta) \mathcal{P}(dN|\rho, \Lambda_0) \pi(d\beta).$$

Typically, one is interested in obtaining the posterior distribution of Λ or F when applicable which requires a description also of the posterior distribution of β given the data. Again, it is easiest to do this by obtaining the posterior distribution of N , which then yields the posterior distribution for a variety of functionals. We first obtain the marginal likelihood of the data given β by another application of Proposition 2. The variation on the previous models is that we now use the fact that

$$(39) \quad \prod_{i=1}^n h_i(u_i, \beta) = \prod_{j=1}^{n(\mathbf{p})} \left[\prod_{i \in E_j} h_i(u_j, \beta) \right].$$

It follows from Proposition 1 that the marginal likelihood given β , say $L_c(d\mathbf{T}, d\mathbf{X}|\beta)$ is equivalent to,

$$\mathcal{L}_N(f_{\beta,m}|\rho, \Lambda_0) \int_{\mathcal{M}} \left[\prod_{i=1}^n \int_0^1 h_i(u_i, \beta) N(du_i, dT_i, dX_i) \right] \mathcal{P}(dN|e^{-f_{\beta,m}}\rho, \Lambda_0),$$

where for fixed β , $\mathcal{P}(dN|e^{-f_{\beta,m}}\rho, \Lambda_0)$ denotes the law of a Poisson random measure, say $N_{c,m,\beta}$, with Lévy measure,

$$(40) \quad e^{-f_{\beta,m}(u,s,x)} \rho(du|s) := \rho_{\beta,m}(du|s) = \left[\prod_{i=1}^m (1 - h_i(u, \beta))^{Y_i^c(s)} \right] \rho(du|s).$$

Again this implies that conditional on the information contained in (36), Λ given β is a Lévy hazard measure with Lévy measure $\rho_{\beta,m}$. Using Proposition 2, and (39), an expression for the marginal likelihood of the data given β is,

$$(41) \quad L_c(d\mathbf{T}, d\mathbf{X}|\beta) = \mathcal{L}_N(f_{\beta,m}|\rho, \Lambda_0) \prod_{j=1}^{n(\mathbf{p})} \left[\int_0^1 \left[\prod_{i \in E_j} h_i(u, \beta) \right] \rho_{\beta,m}(du|T_j^*) \right] \Lambda_0(dT_j^*, dX_j^*).$$

Additionally the expectation of the expressions given in (36) is

$$E \left[e^{-N(f_{\beta,m})} \right] = \mathcal{L}_N(f_{\beta,m} | \rho, \Lambda_0) = \prod_{i=1}^m E \left[e^{-Z_{c,i}(Y_i^c | \beta)} | \rho_{\beta,i-1} \right]$$

where for $l = 1, \dots, m$, with $\rho_{\beta,0} = \rho$,

$$E \left[e^{-Z_{c,i}(Y_i^c | \beta)} | \rho_{\beta,i-1} \right] = \exp \left(- \int_0^\infty \int_0^1 \left(1 - [1 - h_i(u, \beta)]^{Y_i^c(s)} \right) \rho_{\beta,i-1}(du | s) \Lambda_0(ds) \right),$$

which can be verified using Proposition 4.2.

THEOREM 4. *Let N be $\mathcal{P}(dN | \rho, \Lambda_0)$ and let Λ be $\mathcal{P}(d\Lambda | \rho, \Lambda_0)$ on \mathcal{S} . Then subject to the model described in (38), the posterior distribution of Λ given the data and β is equivalent to the law of the Lévy hazard measure,*

$$\Lambda_{c,m,\beta}^*(ds, dx) = \Lambda_{m,\beta}(ds, dx) + \sum_{j=1}^{n(\mathbf{p})} J_{j,n} \delta_{T_j^*, X_j^*}(ds, dx) = \int_0^1 u N_{c,m,\beta}^*(du, ds, dx)$$

where conditional on β , $\Lambda_{m,\beta}$ is a Lévy hazard measure with Lévy measure $\rho_{m,\beta}$ as in (40) and given the data and β , the $(J_{j,n})$ are conditionally independent with distributions

$$\mathbb{P}\{J_{j,n} \in du | T_j^*, \beta\} := H_{j,c}^*(du | \beta) \propto \left[\prod_{i=1}^m (1 - h_i(u, \beta))^{Y_i^c(T_j^*)} \right] \left[\prod_{i \in E_j} h_i(u, \beta) \right] \rho(du | T_j^*)$$

The law of $N_{c,m,\beta}^* = N_{c,m,\beta} + \sum_{j=1}^{n(\mathbf{p})} \delta_{J_{j,n}, T_j^*, X_j^*}$, is equivalent to the posterior distribution of N . The posterior distribution of $F | \beta$, data, is a spatial NTR whose law is equivalent to the law of a random measure $F_{c,m,\beta}^*$, defined by replacing Λ with $\Lambda_{c,m,\beta}^*$ in (15). Conditional on β , the posterior distribution of Z is equivalent to the law of $Z_{c,m,\beta}^*(ds, dx) = \int_0^1 [-\log(1-u)] N_{c,m,\beta}^*(du, ds, dx)$. The posterior distribution of β takes the form $\pi(d\beta | \mathbf{T}, \mathbf{X}) \propto \pi(d\beta) L_c(d\mathbf{T}, d\mathbf{X} | \beta)$

PROOF. Similar to (30), an application of Proposition 2 yields the following disintegration for fixed β ,

$$\left[\prod_{i=1}^n h_i(u_i, \beta) N(dW_i) \right] \mathcal{P}(dN | e^{-f_{Y_n}} \rho, \Lambda_0) = \mathcal{P}(dN | e^{-f_{Y_n}} \rho, \Lambda_0, \mathbf{W}) \prod_{j=1}^{n(\mathbf{p})} \left[\prod_{i \in E_j} h_i(u, \beta) \right] e^{-f_{\beta,m}(W_j^*)} \nu(dW_j^*),$$

where $\nu(dW_j^*) = \rho(du_j | T_j^*) \Lambda_0(dT_j^*, dX_j^*)$, and $e^{-f_{\beta,m}(W_j^*)} = \left[\prod_{i=1}^m (1 - h_i(u, \beta))^{Y_i^c(T_j^*)} \right]$. Now since the form of $L_c(d\mathbf{T}, d\mathbf{X} | \beta)$ has been obtained in (41), the result follows, similar to the proof of Theorem 2, by multiplying and dividing appropriately by $\prod_{j=1}^{n(\mathbf{p})} \left[\int_0^1 \left[\prod_{i \in E_j} h_i(u, \beta) \right] \rho_{\beta,m}(du | T_j^*) \right]$. \square

4.5.1 Cox regression model

We now show that the Bayesian Cox regression model considered by Hjort (1990) and Kim and Lee (2003) falls within our framework. In that setup, one has, conditional on β , $i = 1, \dots, m$ NTR survival functions

$$S_i(t | \beta) := S(t)^{\exp \beta^T \mathbf{M}_i},$$

where β is a vector parameter of interest and \mathbf{M}_i is a covariate vector. As shown in Hjort (1990), the corresponding Lévy cumulative hazard is of the form

$$(42) \quad \Lambda_i(ds|\beta) := 1 - (1 - \Lambda(ds))^{\exp \beta^T \mathbf{M}_i}.$$

The representation (42), as noted in Kim and Lee (2003), imply that Λ_i and Λ are related by the the choice of $h_i(u, \beta) := 1 - (1 - u)^{\exp \beta^T \mathbf{M}_i}$. The choice of $f_{\beta, m}$, in the case of the proportional hazards model in Hjort (1990) and Kim and Lee (2003) is, $f_{\beta, H}(u, s) = - [\sum_{i=1}^m \exp(\beta^T \mathbf{M}_i) Y_i^c(s)] \log(1 - u)$. Using Theorem 4, a description of the posterior distribution, alternate to that of Kim and Lee (2003), via Z rather than Λ subject to general Aalen filtering, is given by the specifications,

$$\mathbb{P}\{Z_{j,n} \in dy|T_j^*, \beta\} \propto \left[\prod_{i \in E_j} \left(1 - e^{-y \exp(\beta^T \mathbf{M}_i) Y_i^c(T_j^*)}\right) \right] \tau_{\beta, m}(dy|T_j^*),$$

where now, $\tau_{\beta, m}(dy|s) = \exp\left(-\left[\sum_{i=1}^m \exp(\beta^T \mathbf{M}_i) Y_i^c(s)\right] y\right) \tau(dy|s)$. For example, in the case of a generalized gamma process, the resulting posterior is such that $Z_{m, \beta}$ is a *weighted* generalized gamma process with Lévy measure,

$$\tau_{\alpha, b, \beta}(dy|s) := \frac{1}{\phi_{\alpha, b}(1) \Gamma(1 - \alpha)} y^{-\alpha-1} \exp\left(-\left[b + \sum_{i=1}^m \exp(\beta^T \mathbf{M}_i) Y_i^c(s)\right] y\right) dy$$

The corresponding $(Z_{j,n})$ do not have a simple distribution.

4.6 Remarks on prior fixed points of discontinuity

We have so far omitted any discussion on the form of the posterior distribution when there are prior points of discontinuity as in $\tilde{\Lambda}_k$ defined in (17). In fact the analysis is essentially already contained in our results. Recall that for $n \geq 1$, the posterior process for Λ in the complete data is, $\Lambda_n^* = \Lambda_n + \sum_{j=1}^{n(\mathbf{p})} J_{j,n} \delta_{T_j^*, X_j^*}$ where the $(J_{j,n})$ are conditionally independent of Λ_n . Using this fact one can simply replace $n(\mathbf{p}) = k$, and let $\{U_l, s_l, w_l\}$ play the role of $\{J_{j,n}, T_j^*, X_j^*\}$. Let $n_l = |\{i : (T_i, X_i) = (s_l, w_l)\}|$ for $l = 1, \dots, k$. In addition let (T_j^*, X_j^*) denote $n(\mathbf{p})$ unique values distinct from $\{(s_1, w_1), \dots, (s_k, w_k)\}$. Then it is easy to see that the posterior distribution of $\tilde{\Lambda}_k$ is of the form,

$$\tilde{\Lambda}_n^* = \Lambda_n + \sum_{l=1}^k U_{l:n} \delta_{s_l, w_l} + \sum_{j=1}^{n(\mathbf{p})} J_{j,n} \delta_{T_j^*, X_j^*}$$

where $\mathbb{P}\{U_{l:n} \in du|s_l\} \propto u^{n_l} (1 - u)^{Y_n(s_l)} H_l(du)$, for $l = 1, \dots, k$. Similar statements hold for the models considered in sections 4.4 and 4.5.

5 EPPF, moment formulae, posterior characterizations and species sampling models

In this section, refinements are made to the description of the various characteristics of the spatial NTR processes, as described in Theorem 2. Most of these details do not appear in the Bayesian literature on univariate NTR processes and may also be of independent interest to those interested in properties of exponential functionals of subordinators. The results are a refinement of the material which appears in section 7, in particular 7.2 and 7.3, of James (2002). Our interest is both of a theoretical and practical nature. In particular, we discuss methods to simplify general calculations with respect to the marginal distributon, $L(d\mathbf{T}, d\mathbf{X})$, of (\mathbf{T}, \mathbf{X}) and to characterize the posterior

distribution, including new descriptions of the prediction rule. These characterizations become particularly important when one is interested in sampling from (\mathbf{T}, \mathbf{X}) , such as applications to mixture models. With the exception of the Dirichlet process, we obtain previously unknown expressions for *exchangeable partition probability functions* (EPPF) derived from NTR processes. However, see Remark 6 below on the related work of Gnedin and Pitman (2002). The EPPF, $\pi(\mathbf{p}) = p(e_1, \dots, e_{n(\mathbf{p})})$, is the distribution of \mathbf{p} which is exchangeable in its arguments and only depend on the cardinality of the cells $E_1, \dots, E_{n(\mathbf{p})}$. This leads to a decomposition of the marginal distribution of (\mathbf{T}, \mathbf{X}) into $\mathbf{T}, \mathbf{X}|\mathbf{p}$ and \mathbf{p} , which is governed by the EPPF $\pi(\mathbf{p})$. It is known from Pitman (1996), that the EPPF plus knowledge of F_0 uniquely characterizes the distribution of the class of species sampling random probability measures, which includes the Dirichlet process. In principal, this fact allows simplification of moment formulae and also allows one to apply a Chinese restaurant sampling scheme to approximate posterior quantities. Ishwaran and James (2003) apply these ideas to the class of species sampling mixture models, which generalizes the work of Brunner, Chan and Lo (1996) for Dirichlet process mixture models. However, because NTR processes additionally depend on Y_n , they are in general not species sampling models. It seems that in fact the Dirichlet process may be the only such model among NTR processes. The dependence on Y_n , which is not completely determined by \mathbf{p} , increases the difficulty in terms of calculations. Here we introduce a further conditioning on a vector $\mathbf{m} = \{E_{(1)}, \dots, E_{(n(\mathbf{p}))}\}$, where $E_{(j)}$ is the collection of values equal to the j th largest unique value. This leads to a manageable and interesting characterizations of the distribution of $\mathbf{T}, \mathbf{X}|\mathbf{m}, \mathbf{p}$, and a joint distribution of (\mathbf{m}, \mathbf{p}) , say $\pi(\mathbf{m}, \mathbf{p})$, which in the homogeneous case has a nice product form. A sampling scheme to generate (\mathbf{m}, \mathbf{p}) , which can be derived from the prediction rule is discussed. Additionally, we illustrate, some not entirely expected connections, to the work of Carmona, Petit and Yor (1997) and Epifani, Lijoi and Pruenster (2002). Moreover, we identify a new class of species sampling models induced by the spatial NTR processes. For general discussions and applications of the EPPF and related quantities such as the Chinese restaurant process we refer the reader to Pitman (2002a) and closely related work on Poisson-Kingman models in Pitman (2002b).

Let $\infty = T_{(0)} > T_{(1)} > T_{(2)} > \dots > T_{(n(\mathbf{p}))}$, denote an ordering of the unique values $\{T_1^*, \dots, T_{n(\mathbf{p})}^*\}$. In addition let $\mathbf{m} = \{E_{(1)}, \dots, E_{(n(\mathbf{p}))}\}$ denote the collection of sets $E_{(j)} = \{i : T_i = T_{(j)}\}$ for $j = 1, \dots, n(\mathbf{p})$. That is, $E_{(j)}$ is the collection of values equal to the j th largest unique value. Then it follows that,

$$\left[\prod_{i=1}^n S(T_i-) \right] = \left[\prod_{j=1}^{n(\mathbf{p})} S(T_j^*-)^{e_{j,n}} \right] = \left[\prod_{j=1}^{n(\mathbf{p})} S(T_{(j)}-)^{m_j} \right],$$

where $m_j = |E_{(j)}|$ is the number of values equal to the j th largest unique value. Note that the collection $\{m_1, \dots, m_{n(\mathbf{p})}\}$, respectively (\mathbf{m}) , given $\mathbf{p} = \{E_1, \dots, E_{n(\mathbf{p})}\}$, takes its values over the symmetric group, say $S_{n(\mathbf{p})}$, of all $n(\mathbf{p})!$ permutations of $\{e_1, \dots, e_{n(\mathbf{p})}\}$, (of $\{E_1, \dots, E_{n(\mathbf{p})}\}$). For integers, (i, k) let,

$$\psi_{i,k}(s) = \int_0^\infty (1 - e^{-y^i}) e^{-yk} \tau(dy|s) \text{ and } \psi_{i,k} = \int_0^\infty (1 - e^{-y^i}) e^{-yk} \tau(dy)$$

and recall that for each j

$$\phi(j) = \psi_{j,0} = \int_0^\infty (1 - e^{-jy}) \tau(dy).$$

Now for each j define the process $Y_{(j-1)}(s) = \sum_{l=1}^{j-1} m_l I\{T_{(l)} > s\}$. Then for each j , $r_{j-1} = Y_{(j-1)}(T_{(j)}) := Y_n(T_{(j)}) = \sum_{l=1}^{j-1} m_l$ denotes the number larger than the j th largest unique value. Note that $r_0 = 0$ and $r_{n(\mathbf{p})} = n$. An application of Proposition 4.2 yields the following identity,

Lemma 5.1 *Let N be $\mathcal{P}(dN|\Lambda_0)$, then an application of Proposition 4.2 using for $j = 1 \dots, n(\mathbf{p})$,*

$f_j(u, s) = -Y_{(j-1)}(s) \log(1 - u)$, yields the formula,

$$(43) \quad E \left[\prod_{j=1}^{n(\mathbf{p})} S(T_{(j)}-)^{m_j} \right] = \prod_{j=1}^{n(\mathbf{p})} \exp \left(- \int_0^{T_{(j)}} \psi_{m_j, r_{j-1}}(s) \Lambda_0(ds) \right).$$

In particular when there are no ties, that is $n(\mathbf{p}) = n$, it follows that

$$(44) \quad E \left[\prod_{i=1}^n S(T_i-) \right] = \prod_{j=1}^n \exp \left(- \int_0^{T_{(j)}} \psi_{1, j-1}(s) \Lambda_0(ds) \right).$$

Lemma 5.1 implies that the marginal joint distribution can be written as,

$$L(d\mathbf{T}, d\mathbf{X}) = \prod_{j=1}^{n(\mathbf{p})} e^{-\int_0^{T_{(j)}} \psi_{m_j, r_{j-1}}(s) \Lambda_0(ds)} \kappa_{m_j, r_{j-1}}(\rho|T_{(j)}) \Lambda(dT_j^*, dX_j^*)$$

where $\kappa_{m_j, r_{j-1}}(\rho) = \int_0^1 u^{m_j} (1-u)^{r_{j-1}} \rho(du|s)$. For each $\mathbf{m} \in S_{n(\mathbf{p})}$ and integrable function $g(\mathbf{T})$, define

$$L(g; \mathbf{m}) = \int_0^\infty \int_{t_{n(\mathbf{p})}}^\infty \cdots \int_{t_2}^\infty g((t, \mathbf{m})) \prod_{j=1}^{n(\mathbf{p})} e^{-\int_0^{t_j} \psi_{m_j, r_{j-1}}(s) \Lambda_0(ds)} \kappa_{m_j, r_{j-1}}(\rho|t_j) \Lambda_0(dt_j),$$

where $t_1 > t_2 > \cdots > t_{n(\mathbf{p})}$ denotes one of $n(\mathbf{p})!$ orderings of the unique values. With some abuse of notation, the vector $(\mathbf{t}, \mathbf{m}) = (\mathbf{t})$ denotes the collection of n points whose $n(\mathbf{p})$ unique values are ordered according to \mathbf{m} .

Lemma 5.2 Assume that the functional $I(g) = \int g(\mathbf{t}) \prod_{i=1}^n F(dt_i)$ is integrable, where F is a NTR process specified by the Poisson law $\mathcal{P}(dN|\rho, \Lambda_0)$, then it follows from Theorem 2 that,

$$(45) \quad E[I(g)] = \sum_{\mathbf{p}} \left[\sum_{\mathbf{m} \in S_{n(\mathbf{p})}} L(g; \mathbf{m}) \right].$$

In the homogeneous case, $\rho(du|s) = \rho(du)$, the expression reduces to,

$$\sum_{\mathbf{p}} \left[\sum_{\mathbf{m} \in S_{n(\mathbf{p})}} \left[\prod_{j=1}^{n(\mathbf{p})} \kappa_{m_j, r_{j-1}}(\rho) \right] \int_0^\infty \int_{t_{n(\mathbf{p})}}^\infty \cdots \int_{t_2}^\infty g(t, \mathbf{m}) \prod_{j=1}^{n(\mathbf{p})} e^{-\Lambda_0(t_j) \psi_{m_j, r_{j-1}}} \Lambda_0(dt_j) \right].$$

PROOF. The result follows from an application of Theorem 2 and Fubini's theorem which yields

$$\int_{\mathcal{M}} I(g) \mathcal{P}(dN|\rho, \Lambda_0) = \int g(\mathbf{t}) L(d\mathbf{t}, d\mathbf{x})$$

□

5.1 Exponential functionals and means of NTR processes

We now relate Lemma 5.1 with the results of Epifani, Lijoi and Pruenster (2002) and Carmona, Petit, and Yor (1997) concerning moment formulae for means of NTR processes. Briefly, using the relationship,

$$(46) \quad I = \int_0^\infty t F(dt) = \int_0^\infty S(t) dt = \int_0^\infty e^{-Z(t)} dt,$$

Epifani, Lijoi, and Pruenster (2002, Proposition 5), establish the following moment formulae, expressed in our notation, which characterizes the distribution of I ,

$$(47) \quad E[I^n] = n! \int_0^\infty \int_{t_n(\mathbf{p})}^\infty \cdots \int_{t_2}^\infty \prod_{j=1}^n \exp\left(-\int_0^{t_j} \int_0^\infty (1 - e^{-y}) e^{-y(j-1)} \tau(dy|s) \Lambda_0(ds)\right) dt_j.$$

The authors also provide conditions under which the moments exist, which amounts to the finiteness of the moment of order n of F_0 . That is $\int_0^\infty t^n F_0(dt) < \infty$. In addition, when $\rho(du|s) = \rho(du)$ and $\Lambda_0(t) = t$, the expression in (47) reduces to the interesting formulae of Carmona, Petit and Yor (1997, Proposition 3.3), viewed within the context of exponential functionals of a subordinator,

$$(48) \quad E[I^n] = \frac{n!}{\prod_{j=1}^n \phi(j)}.$$

Notice that the specification $\Lambda_0(t) = t$ is equivalent to specifying F_0 as an exponential(1) distribution. In addition, Carmona, Petit and Yor (1997, Proposition 3.1) establish the following result for any $\lambda \geq 1$, and more general Lévy processes,

$$E[I^\lambda] = \frac{\lambda}{\phi(\lambda)} E[I^{\lambda-1}].$$

A proof of the expression (47), somewhat different from the approaches of Epifani, Lijoi and Pruenster (2002) and Carmona, Petit and Yor (1997) can be obtained by using (44) in Lemma 5.1, and noting by Fubini's Theorem that,

$$E[I^n] = \int E\left[\prod_{i=1}^n S(t_i-)\right] dt_1 \dots dt_n.$$

Lemma 5.2 offers a complementary result to theirs in that one can express $E[I^n]$ in terms of sums over partitions \mathbf{p} . Apparently, for NTR processes, a result of this type is only widely known in the case of the Dirichlet process, which follows as a special case of Lo (1984). The result is as follows,

Corollary 5.1 *Let I be defined as in (46), then for $g(\mathbf{t}) = \prod_{i=1}^n t_i = \prod_{j=1}^{n(\mathbf{p})} t_j^{m_j}$,*

$$E[I^n] = \sum_{\mathbf{p}} \left[\sum_{\mathbf{m} \in S_n(\mathbf{p})} L(g; \mathbf{m}) \right].$$

In particular, in the case where, $\rho(du|s) = \rho(du)$ and $\Lambda_0(t) = t$, Lemma 5.2 combined with the result of Carmona, Petit and Yor (1997) yields the identity,

$$\sum_{\mathbf{p}} \left[\sum_{\mathbf{m} \in S_n(\mathbf{p})} \left[\prod_{j=1}^{n(\mathbf{p})} \kappa_{m_j, r_{j-1}}(\rho) \right] \int_0^\infty \int_{t_n(\mathbf{p})}^\infty \cdots \int_{t_2}^\infty \prod_{j=1}^{n(\mathbf{p})} t_j^{m_j} e^{-t_j \psi_{m_j, r_{j-1}}} dt_j \right] = \frac{n!}{\prod_{j=1}^n \phi(j)}$$

5.2 EPPF and related formulae derived from NTR processes

Lemma 5.2 now allows us to find a form of the *exchangeable partition probability function* (EPPF), $\pi(\mathbf{p}) = p(e_1, \dots, e_{n(\mathbf{p})})$, associated with NTR processes. Since the marginal distribution $(\mathbf{T}, \mathbf{X}) = (\mathbf{T}^*, \mathbf{X}^*, \mathbf{p})$ is exchangeable it follows that the EPPF can always be calculated by integrating $L(d\mathbf{T}, d\mathbf{X})$, over the unique values. This in turn yields the decomposition of the distribution of (\mathbf{T}, \mathbf{X}) into $\mathbf{T}, \mathbf{X}|\mathbf{p}$ and \mathbf{p} , which is governed by the EPPF $\pi(\mathbf{p})$. Interestingly, another connection is made to the expressions for $E[I^n]$, given in (48), obtained by Carmona, Petit and Yor (1997).

Proposition 5.1 *Let F be a spatial NTR process determined by a Poisson random measure with law $\mathcal{P}(dN|\rho, \Lambda_0)$, the EPPF derived by iid sampling from F is expressible as,*

$$\pi(\mathbf{p}) = \sum_{\mathbf{m} \in S_n(\mathbf{p})} L(1; \mathbf{m}),$$

The representations imply the existence of a joint distribution of (\mathbf{m}, \mathbf{p}) given by $\pi(\mathbf{m}, \mathbf{p}) = L(1; \mathbf{m})$. Additionally in the case where $\rho(du|s) = \rho(du)$, the formulae reduce to,

$$(49) \quad \pi(\mathbf{p}) = \sum_{\mathbf{m} \in S_n(\mathbf{p})} \frac{\prod_{j=1}^{n(\mathbf{p})} \kappa_{m_j, r_{j-1}}(\rho)}{\prod_{j=1}^{n(\mathbf{p})} \phi(r_j)} \quad \text{and} \quad \pi(\mathbf{m}, \mathbf{p}) = \frac{\prod_{j=1}^{n(\mathbf{p})} \kappa_{m_j, r_{j-1}}(\rho)}{\prod_{j=1}^{n(\mathbf{p})} \phi(r_j)}.$$

PROOF. The proof in the general case follows from Lemma 5.2 with $g := 1$. In the case of $\rho(du|s) = \rho(du)$, $\pi(\mathbf{p})$ is equivalent to

$$\sum_{\mathbf{m} \in S_n(\mathbf{p})} \left[\prod_{j=1}^{n(\mathbf{p})} \kappa_{m_j, r_{j-1}}(\rho) \right] \int_0^\infty \int_{t_n(\mathbf{p})}^\infty \cdots \int_{t_2}^\infty \prod_{j=1}^{n(\mathbf{p})} e^{-\Lambda_0(t_j) \psi_{m_j, r_{j-1}}} \Lambda_0(dt_j)$$

The result is concluded by evaluating, $\int_0^\infty \int_{t_n(\mathbf{p})}^\infty \cdots \int_{t_2}^\infty \prod_{j=1}^{n(\mathbf{p})} e^{-\Lambda_0(t_j) \psi_{m_j, r_{j-1}}} \Lambda_0(dt_j)$. This is done by noting that for any positive C , $\int_t^\infty e^{-C\Lambda_0(u)} \Lambda_0(du) = C^{-1} e^{-C\Lambda_0(t)}$. In addition, $\psi_{m_1, 0} = \phi(m_1)$ and for each j , $\phi(r_{j-1}) + \psi_{m_j, r_{j-1}} = \phi(r_j)$. \square

A closer relationship to the formula for $E[I^n]$, (48), given in Carmona, Petit and Yor (1997), is seen in the next corollary which describes formula for the case where all cells are of the same size.

Corollary 5.2 *Suppose that $\rho(du|s) = \rho(du)$ and $n = kn(\mathbf{p})$, then with respect to the EPPF given in (49), the probability of the event $\mathbf{p} = \{E_1, \dots, E_{n(\mathbf{p})}\}$, such that the size of each cell is k , is*

$$\pi(\mathbf{p}) = \frac{n(\mathbf{p})! \prod_{j=1}^{n(\mathbf{p})} \int_0^1 u^k (1-u)^{(j-1)k} \rho(du)}{\prod_{j=1}^{n(\mathbf{p})} \phi(jk)}$$

As special cases, when $n(\mathbf{p}) = n$, the probability of no ties in the sample, corresponds to the probability of the event $p = \{\{1\}, \{2\}, \dots, \{n\}\}$, given by,

$$\pi(\mathbf{p}) = \frac{n! \prod_{j=1}^n \int_0^1 u(1-u)^{j-1} \rho(du)}{\prod_{j=1}^n \phi(j)} = E[I^n] \prod_{j=1}^n \int_0^1 u(1-u)^{j-1} \rho(du),$$

for $E[I^n]$ given in (48). When $n(\mathbf{p}) = 1$, $\mathbf{p} = \{1, 2, \dots, n\}$, corresponds to the event that all the values in the sample are the same, the probability is given by,

$$\pi(\mathbf{p}) = \frac{\int_0^1 u^n \rho(du)}{\phi(n)} = \frac{\int_0^\infty (1 - e^{-y})^n \tau(dy)}{\int_0^\infty (1 - e^{-ny}) \tau(dy)}.$$

REMARK 6. As mentioned in the introduction, for the case $\rho(du|s) = \rho(du)$, Gneden and Pitman (2002), independent of this work and by different arguments, obtain formulae for what are called regenerative compositions that contain our results in (49). Their formulae are derived from a discretization of subordinators. In fact, the authors show that all such regenerative compositions are determined uniquely by their construction via subordinators. The authors' result is more general, in the homogeneous case, as they include the result for subordinators with drift components and

allow τ to have mass at infinity. It is however a simple matter to adjust Proposition 1 to allow for a drift[see James (2002, Remark 28)]. They do not cover the inhomogeneous cases we consider. The authors description via a *decrement function* and *composition structure* contain additional binomial coefficients. Explicitly in terms of our notation, their composition structure is expressed as

$$(50) \quad \frac{n!}{\prod_{j=1}^{n(\mathbf{p})} e_j!} \pi(\mathbf{m}, \mathbf{p})$$

We use their work to identify some interesting spatial NTR models. See Donnelly and Joyce (1991), Gnedin (1997) and Pitman (1997) for relevant references.

Some examples are now presented.

5.2.1 Dirichlet process

The classical example of $\pi(\mathbf{p})$ in the inhomogeneous case is when F is a Dirichlet process with parameter θF_0 or equivalently Λ is a beta process with $c(s) = \theta S_0(s-)$, then the EPPF is

$$(51) \quad \pi(\mathbf{p}) = PD(\theta) = \frac{\Gamma(\theta)\theta^{n(\mathbf{p})} \prod_{j=1}^{n(\mathbf{p})} (e_j - 1)!}{\Gamma(\theta + n)},$$

which is a variant, mod the factor $n!/\prod_{j=1}^{n(\mathbf{p})} e_j!$, of the Ewens sampling formula derived by Ewens (1972) and Antoniak (1974). While the formula (51) is well-known, it is certainly interesting to see how it arises via the description in Proposition 5.1. Here we present arguments that will yield $\pi(\mathbf{p}, \mathbf{m}) = L(1, \mathbf{m})$, which turns out to have a very simple form. Notice that for the Dirichlet process, $\psi_{m_j, r_{j-1}}(s) = \sum_{l=1}^{m_j} \theta S_0(s-)/[\theta S_0(s-) + l - 1 + r_{j-1}]$ and hence for each j , by a change of variable,

$$\int_0^{T_{(j)}} \psi_{m_j, r_{j-1}}(s) \Lambda_0(ds) = \sum_{l=1}^{m_j} \log \left[\frac{\theta + l - 1 + r_{j-1}}{\theta S_0(T_{(j)}-) + l - 1 + r_{j-1}} \right].$$

Notice that, $\kappa_{m_j, r_{j-1}}(\rho_{\theta S_0} | T_{(j)}) = \theta S_0(T_{(j)}-) \Gamma(m_j) \Gamma(r_{j-1} + S_0(T_{(j)}-)) / \Gamma(r_j + \theta S_0(T_{(j)}-))$. By obvious cancellations, it follows that

$$L(1, \mathbf{m}) = \pi(\mathbf{m}, \mathbf{p}) = \frac{\theta^{n(\mathbf{p})}}{n(\mathbf{p})!} \prod_{j=1}^{n(\mathbf{p})} \prod_{l=1}^{m_j} \frac{1}{\theta + l - 1 + r_{j-1}} = \frac{1}{n(\mathbf{p})!} PD(\theta).$$

Of course, this argument also shows, the known fact, that $L(d\mathbf{T}, d\mathbf{X}) := PD(\theta) \prod_{j=1}^{n(\mathbf{p})} F_0(dT_j^*, dX_j^*)$.

5.2.2 Ferguson and Phadia's homogeneous process

For the homogeneous process considered by Ferguson and Phadia (1979), defined by ρ_c or τ_c with $c(s) = \theta$ and Λ_0 we have, $\psi_{m_j, r_{j-1}} = \sum_{l=1}^{m_j} \theta / (\theta + l - 1 + r_{j-1})$,

$$\kappa_{m_j, r_{j-1}}(\rho) = \frac{\theta \Gamma(m_j) \Gamma(r_{j-1} + \theta)}{\Gamma(r_j + \theta)} \text{ and } \phi(r_j) = \sum_{l=1}^{r_j} \left[\frac{\theta}{\theta + l - 1} \right],$$

which yields the formula, $\pi(\mathbf{m}, \mathbf{p}) = PD(\theta) \prod_{j=1}^{n(\mathbf{p})} \frac{1}{\sum_{l=1}^{r_j} \theta / (\theta + l - 1)}$. Gnedin (2002) appears to be the first to derive this formula.

5.2.3 Renewal compositions: The beta $(-\alpha, \alpha)$ NTR process

Perhaps the most striking example of a composition structure, in the homogeneous case, is due to Pitman (1997),

$$\frac{n!}{\prod_{j=1}^{n(\mathbf{p})} e_j!} \pi(\mathbf{m}, \mathbf{p}) = \frac{n!}{\prod_{j=1}^{n(\mathbf{p})} e_j!} \frac{\alpha^{n(\mathbf{p})} \Gamma(\alpha) \prod_{j=1}^{n(\mathbf{p})} \prod_{l=1}^{e_j-1} (l - \alpha)}{\Gamma(\alpha + n)}$$

which was originally derived from the zero-set of a Bessel bridge of dimension $2 - 2\alpha$. Gnedin and Pitman (2002) call this the class of renewal compositions. As indicated in Gnedin and Pitman (2002), the result surprisingly coincides with composition structure of the beta $(-\alpha, \alpha)$ process where,

$$\kappa_{m_j, r_{j-1}}(\rho) = \frac{\Gamma(m_j - \alpha) \Gamma(\alpha + r_{j-1})}{\Gamma(r_j) \Gamma(\alpha) \Gamma(1 - \alpha)} \text{ and } \phi(r_j) = \frac{\Gamma(\alpha + r_j)}{\Gamma(r_j) \alpha \Gamma(\alpha)}.$$

Moreover, the EPPF derived from the beta $(-\alpha, \alpha)$ NTR process is the EPPF of the two parameter Poisson-Dirichlet family[see Pitman (1996, 2002a, 2002b)] with parameters (α, α) . That is,

$$(52) \quad \pi(\mathbf{p}) = \frac{n(\mathbf{p})! \alpha^{n(\mathbf{p})} \Gamma(\alpha) \prod_{j=1}^{n(\mathbf{p})} \prod_{l=1}^{e_j-1} (l - \alpha)}{\Gamma(\alpha + n)}.$$

Similar to the Dirichlet process, the conditional distribution of $\mathbf{m}|\mathbf{p}$ is $\pi(\mathbf{m}|\mathbf{p}) = 1/[n(\mathbf{p})!]$. Note that while the EPPF is easy to sample, the corresponding beta $(-\alpha, \alpha)$ spatial NTR process, unlike the two parameter Poisson-Dirichlet process, is not a species sampling model. As such it is of practical interest to identify specifically the marginal distribution of $(\mathbf{T}, \mathbf{X})|\mathbf{m}, \mathbf{p}$. We will do so in section 5.3.

5.2.4 Generalized gamma process

The last example presented is for the case of the generalised gamma process, here, $\psi_{m_j, r_{j-1}} = [(r_j + b)^\alpha - (r_{j-1} + b)^\alpha] / [(1 + b)^\alpha - b^\alpha]$,

$$\phi(r_j) = \frac{[(r_j + b)^\alpha - b^\alpha]}{[(1 + b)^\alpha - b^\alpha]} \text{ and } \kappa_{m_j, r_{j-1}}(\rho) = \frac{\sum_{l=0}^{m_j} (-1)^{l+1} \binom{m_j}{l} (b + r_{j-1} + l)^\alpha}{[(1 + b)^\alpha - b^\alpha]}$$

Hence,

$$\pi(\mathbf{m}, \mathbf{p}) = \frac{\prod_{j=1}^{n(\mathbf{p})} \sum_{l=0}^{m_j} \binom{m_j}{l} (-1)^{l+1} (b + r_{j-1} + l)^\alpha}{\prod_{j=1}^{n(\mathbf{p})} [(b + r_j)^\alpha - b^\alpha]}.$$

5.3 Marginal distributions

We now proceed to describe the marginal distributions given \mathbf{p} and \mathbf{m}, \mathbf{p} . The event of no ties, $n(\mathbf{p}) = n$, corresponds to the common assumption in the literature for observed data. Analogous to Antoniak (1974) for the Dirichlet process, it follows that the distribution of $\mathbf{T}, \mathbf{X}|\mathbf{p}$, in the homogeneous case is,

$$n! E[I^n]^{-1} \left[\prod_{i=1}^n e^{-\Lambda_0(t_{(i)}) \psi_{1, i-1}} \right] \prod_{j=1}^n \Lambda_0(dt_j, dx_j).$$

In general the definition of $L(d\mathbf{T}, d\mathbf{X})$ coupled with Proposition 5.1, yields the following description of a distribution of $\mathbf{T}, \mathbf{X}|\mathbf{m}, \mathbf{p}$ distribution

Lemma 5.3 *Subject to the distributions $\pi(\mathbf{m}, \mathbf{p})$, and $L(d\mathbf{T}, d\mathbf{X})$, there exists a marginal distribution of $\mathbf{T}, \mathbf{X}|\mathbf{m}, \mathbf{p}$, given by*

$$\pi(d\mathbf{T}, d\mathbf{X}|\mathbf{m}, \mathbf{p}) \propto \left[\prod_{j=1}^{n(\mathbf{p})} e^{-\int_0^{T(j)} \psi_{m_j, r_{j-1}}(s) \Lambda_0(ds)} \kappa_{m_j, r_{j-1}}(\rho|T(j)) \right] \prod_{l=1}^{n(\mathbf{p})} \Lambda_0(dT_{(j)}, dX_j^*),$$

and $T_{(1)} > T_{(2)} > \dots > T_{(n(\mathbf{p}))}$ denotes an ordering of the unique values. In the homogeneous case the result reduces to,

$$(53) \quad \pi(d\mathbf{T}, d\mathbf{X}|\mathbf{m}, \mathbf{p}) = \left[\prod_{j=1}^{n(\mathbf{p})} \phi(r_j) \right] \left[\prod_{j=1}^{n(\mathbf{p})} e^{-\psi_{m_j, r_{j-1}} \Lambda_0(T(j))} \right] \prod_{l=1}^{n(\mathbf{p})} \Lambda_0(dT_{(j)}, dX_j^*)$$

In both cases there exists a conditional distribution of $\mathbf{X}|\mathbf{T}, \mathbf{m}$ given by $\prod_{j=1}^{n(\mathbf{p})} P_0(dX_j^*|T_{(j)})$.

Lemma 5.3, (53), can be used to deduce an explicit Markov property in the homogeneous case which has the interpretation that the distribution of the next death time only depends on the previous death time.

Proposition 5.2 *Given \mathbf{m}, \mathbf{p} , let $T_{(1)}, \dots, T_{(n(\mathbf{p}))}$ be distributed according to (53), moreover set $\Lambda_0(t) = t$, then the conditional distribution of $T_{(j)}$ given $T_{(j+1)}, \dots, T_{(n(\mathbf{p}))}$ only depends on $T_{(j+1)} = t_{j+1}$ and is given by the truncated exponential distribution with density,*

$$(54) \quad \phi(r_j) e^{-\phi(r_j)[t_j - t_{j+1}]} dt_j$$

for $t_j > t_{j+1}$. In particular, the smallest value, or equivalently the first of $n(\mathbf{p})$ death times, $T_{((n(\mathbf{p})))}$ has a marginal distribution which is exponential with parameter $\phi(n)$, that is

$$\mathbb{P}\{T_{(n(\mathbf{p}))} \in dy\} = \phi(n) e^{-\phi(n)y} dy$$

The result is similar for general Λ_0 and is given by replacing each t_j with $\Lambda_0(t_j)$ in (54).

Results for Ferguson and Phadia's homogeneous process, the beta $(-\alpha, \alpha)$ spatial NTR process and the generalized gamma process can be deduced from the calculations in 5.2.2, 5.2.3, and 5.2.4. In particular, when $b = 0$ and $\Lambda_0(t) = t$, the density corresponding to the stable process with index $0 < \alpha < 1$ is

$$\mathbb{P}\{T_{(j)} \in dt_j | T_{(j+1)} = t_{j+1}\} = r_j^\alpha e^{-r_j^\alpha [t_j - t_{j+1}]} \text{ for } t_j > t_{j+1}.$$

REMARK 7. From the characterization of the V_j in Theorem 3.1 of Doksum, one can relate the densities, (54), in Proposition 5.2 to the V_j as follows; Let $(1 - V_j) = e^{-[Z(t_j) - Z(t_{j+1})]}$, then,

$$E[(1 - V_j)^{r_j}] = \int_{t_{j+1}}^{t_j} \phi(r_j) e^{-\phi(r_j)[\Lambda_0(s) - \Lambda_0(t_{j+1})]} \Lambda_0(ds) = e^{-\phi(r_j)[\Lambda_0(t_j) - \Lambda_0(t_{j+1})]}$$

The results, coupled with Theorem 2, lead to an alternate characterization of the posterior distribution given in Theorem 2 presented below,

THEOREM 5. *Let F be a spatial NTR process determined by N which is $\mathcal{P}(dN|\rho, \Lambda_0)$ If g is a measurable non-negative or integrable function on $\mathcal{S}^n \times \mathcal{M}$, then*

$$\begin{aligned} & \int_{\mathcal{S}^n \times \mathcal{M}} g(\mathbf{T}, \mathbf{X}, N) \left[\prod_{i=1}^n F(dT_i, dX_i) \right] \mathcal{P}(dN|\rho, \Lambda_0) \\ &= \sum_{\mathbf{p}} \sum_{\mathbf{m} \in \mathcal{S}_n(\mathbf{p})} \left[\int_{\mathcal{S}^n(\mathbf{p})} \left[\int_{\mathcal{M}} g((\mathbf{T}, \mathbf{X}), N) \pi(dN|\mathbf{T}, \mathbf{X}) \right] \pi(d\mathbf{T}, d\mathbf{X}|\mathbf{m}, \mathbf{p}) \right] \pi(\mathbf{m}, \mathbf{p}). \end{aligned}$$

Note that we use the fact that the random measure corresponding to the law $\pi(dN|\mathbf{T}, \mathbf{X})$ can be written as, $N_n^* = N_n + \sum_{j=1}^{n(\mathbf{p})} \delta_{J_j, T_{(j)}, X_j^*}$ where X_j^* is paired with the j th largest unique value $T_{(j)}$. Moreover the distribution of $J_j|T_{(j)}$ for $j = 1, \dots, n(\mathbf{p})$ are conditionally independent with distribution

$$(55) \quad \mathbb{P}\{J_j \in du|T_{(j)}\} = \frac{u^{m_j}(1-u)^{T_{(j)}-1} \rho(du|T_{(j)})}{\kappa_{m_j, r_{j-1}}(\rho|T_{(j)})}.$$

5.4 Sequential constructions of (\mathbf{m}, \mathbf{p}) : modified Chinese restaurant processes

In this section it is shown how one might generate (\mathbf{m}, \mathbf{p}) from $\pi(\mathbf{m}, \mathbf{p})$ in the case where $\rho(du|s) = \rho(du)$, via a sequential seating scheme with probabilities derived from the prediction rule given (\mathbf{m}, \mathbf{p}) . The scheme bears similarities to generalized Chinese restaurant processes which can be used to generate general EPPF's, $\pi(\mathbf{p}) = p(e_1, \dots, e_{n(\mathbf{p})})$. The generalized Chinese restaurant process sequentially sits customers entering a chinese restaurant. Using the description in Pitman (2002a, p. 60), an initially empty restaurant has an unlimited number of tables labelled $1, 2, \dots$. Customers numbered $1, 2, \dots$ arrive one by and are seated according to the following scheme: The first customer is seated at the first table. For $n \geq 1$, given the partition \mathbf{p}_n , which is equivalent to the configuration of the the first n customers seated at $n(\mathbf{p})$ tables, the next customer $n+1$ is seated at an occupied table j with probability $p(\dots, e_j+1, \dots)/p(e_1, \dots, e_{n(\mathbf{p})})$, otherwise the customer is placed at a new table with probability $p(e_1, \dots, e_{n(\mathbf{p})}, 1)/p(e_1, \dots, e_{n(\mathbf{p})})$. This scheme for the Dirichlet process generates the original Chinese restaurant process based on the Blackwell and MacQueen (1973) urn scheme which seats customers $n+1$ to an occupied table j with probability $e_{j:n}/(n+\theta)$ for $j = 1, \dots, n(\mathbf{p})$, and otherwise to a new table with probability $\theta/(n+\theta)$. For the beta $(-\alpha, \alpha)$ NTR process or equivalently the two-parameter (α, α) Poisson-Dirichlet process, the rule is seat customer $n+1$ to an occupied table j with probability $(e_{j:n} - \alpha)/(n+\alpha)$, and otherwise to a new table with probability $\alpha(1+n(\mathbf{p}))/(\alpha+n)$. In these two cases, one can then generate (\mathbf{m}, \mathbf{p}) easily by drawing the vector $\mathbf{m}|\mathbf{p}$ from the discrete uniform distribution on $S_{n(\mathbf{p})}$

In general, to generate a (\mathbf{m}, \mathbf{p}) from $\tilde{p}(m_1, \dots, m_{n(\mathbf{p})}) = \pi(\mathbf{m}, \mathbf{p})$, one can use a modified Chinese restaurant process which also records the rank of the entering customers relative to the already seated customers. That is, given a configuration based on n customers seated at $n(\mathbf{p})$ existing tables labelled with ranks from $j = 1, \dots, n(\mathbf{p})$, the next customer $n+1$ is seated at an occupied table j , denoting that customer $n+1$ is equivalent to the j th largest seated customers, with probability,

$$p_j = \frac{\tilde{p}(\dots, m_j+1, \dots)}{\tilde{p}(m_1, \dots, m_{n(\mathbf{p})})} = \frac{\kappa_{m_j+1, r_{j-1}}(\rho) \prod_{l=j+1}^{n(\mathbf{p})} \kappa_{m_l, r_{l-1}+1}(\rho)}{\kappa_{m_j, r_{j-1}}(\rho) \prod_{l=j+1}^{n(\mathbf{p})} \kappa_{m_l, r_{l-1}}(\rho)} \prod_{l=j}^{n(\mathbf{p})} \frac{\phi(r_l)}{\phi(r_l+1)}$$

Customer $n+1$ is seated at a new table with probability $1 - \sum_{j=1}^{n(\mathbf{p})} p_j$. However, if customer $n+1$ is new, it is also necessary to know the customer's rank and as such re-rank by one position all customers smaller than the new customer. Hence the probability that customer $n+1$ is new and is the j th largest among $n(\mathbf{p})+1$ possible ranks is,

$$q_j = \frac{\tilde{p}(\dots, m_{j-1}, 1, m_j, \dots)}{\tilde{p}(m_1, \dots, m_{n(\mathbf{p})})} = \frac{\kappa_{1, r_{j-1}}(\rho) \prod_{l=j}^{n(\mathbf{p})} \kappa_{m_l, r_{l-1}+1}(\rho)}{\phi(r_{j-1}+1) \prod_{l=j}^{n(\mathbf{p})} \kappa_{m_l, r_{l-1}}(\rho)} \prod_{l=j}^{n(\mathbf{p})} \frac{\phi(r_l)}{\phi(r_l+1)},$$

with $q_{n(\mathbf{p})+1} = \kappa_{1, n}(\rho)/\phi(n+1)$. Note that in the calculation of $\kappa_{1, r_{j-1}}(\rho)$, $r_{j-1}+1$ is to be used rather than $r_j = r_{j-1} + m_j$.

As an example, recall from section 5.2.2 that Ferguson and Phadia's homogeneous process has $\phi(r_j) = \sum_{l=1}^{r_j} \theta/(\theta + l - 1)$, it follows that in this case

$$(56) \quad p_j = \frac{m_j}{n + \theta} \prod_{l=j}^{n(\mathbf{p})} \frac{\phi(r_l)}{\phi(r_l + 1)} \text{ and } q_j = \frac{1}{n + \theta} \frac{1}{\sum_{i=1}^{r_{j-1}+1} 1/(\theta + i - 1)} \prod_{l=j}^{n(\mathbf{p})} \frac{\phi(r_l)}{\phi(r_l + 1)}.$$

It turns out that q_j and p_j can be derived from the Bayesian prediction rule. To see the connection to a prediction rule, notice that the posterior spatial NTR process, F_n^* , described in Corollary 4.1, (31), can be written as,

$$(57) \quad F_n^*(ds, dx) = e^{-Z_n(s^-)} \left[\prod_{\{j: T_{(j)} < s\}}^{n(\mathbf{p})} (1 - J_j) \right] \Lambda_n(ds, dx) + \sum_{j=1}^{n(\mathbf{p})} \tilde{P}_{j;n} \delta_{T_{(j)}, X_j^*}(ds, dx)$$

where, $\tilde{P}_{j;n} = e^{-Z_n(T_{(j)}^-)} J_j \prod_{l=j+1}^{n(\mathbf{p})} (1 - J_l)$, and $J_j | T_{(j)}$ are distributed as in (55). Now the probabilities q_j and p_j can be deduced by taking expectations, conditioned on (\mathbf{m}, \mathbf{p}) , over $\tilde{P}_{j;n}$ and $\int_0^\infty e^{-Z_n(s^-)} \left[\prod_{\{j: T_{(j)} < s\}} (1 - J_j) \right] \Lambda_n(ds)$. We will illustrate this below where an expression for the prediction rule is derived.

5.5 Prediction rule

Although one can deduce expressions for the prediction rules based on existing result, we now focus on a refined form, which reveals a bit more about the underlying structure of NTR models. As a by-product, this allows for more explicit expressions for the Bayes estimates of certain functionals. The prediction rule, $\mathbb{P}\{T_{n+1} \in ds, X_{n+1} \in dx | \mathbf{T}, \mathbf{X}\}$ is equal to $E[F_n^*(ds, dx) | \mathbf{T}, \mathbf{X}]$ based on (57). In order to calculate this, using the notation $T_{(0)} = \infty$ and $T_{(n(\mathbf{p})+1)} = 0$, it follows that for each j ,

$$E[J_j \prod_{l=j+1}^{n(\mathbf{p})} (1 - J_l) | \mathbf{T}] = \frac{\kappa_{m_j+1, r_{j-1}}(\rho | T_{(j)}) \prod_{l=j+1}^{n(\mathbf{p})} \kappa_{m_l, r_{l-1}+1}(\rho | T_{(l)})}{\kappa_{m_j, r_{j-1}}(\rho | T_{(j)}) \prod_{l=j+1}^{n(\mathbf{p})} \kappa_{m_l, r_{l-1}}(\rho | T_{(l)})}$$

and $E[e^{-Z_n(T_{(j)}^-)} | \mathbf{T}] = \prod_{l=j}^{n(\mathbf{p})} \exp(-\int_{T_{(l+1)}}^{T_{(l)}} \psi_{1, r_l}(s) \Lambda_0(ds))$. These calculations give

$$p_j^* = \frac{\kappa_{m_j+1, r_{j-1}}(\rho | T_{(j)}) \prod_{l=j+1}^{n(\mathbf{p})} \kappa_{m_l, r_{l-1}+1}(\rho | T_{(l)})}{\kappa_{m_j, r_{j-1}}(\rho | T_{(j)}) \prod_{l=j+1}^{n(\mathbf{p})} \kappa_{m_l, r_{l-1}}(\rho | T_{(l)})} \prod_{l=j}^{n(\mathbf{p})} e^{-\int_{T_{(l+1)}}^{T_{(l)}} \psi_{1, r_l}(s) \Lambda_0(ds)}$$

where $p_j^* = E[\tilde{P}_{j;n} | \mathbf{T}] = \mathbb{P}\{T_{n+1} = T_{(j)} | \mathbf{T}\}$. In the homogeneous case we have,

$$(58) \quad E[e^{-Z_n(T_{(j)}^-)} | \mathbf{T}] = e^{-\psi_{1, r_j} \Lambda_0(T_{(j)})} \prod_{l=j+1}^{n(\mathbf{p})} e^{-[\psi_{m_l, 1+r_{l-1}} - \psi_{m_l, r_{l-1}}] \Lambda_0(T_{(l)})},$$

with $\psi_{m_l, 1+r_{l-1}} - \psi_{m_l, r_{l-1}} = \psi_{1, r_l} - \psi_{1, r_{l-1}}$. From this it follows that

$$(59) \quad p_j^* = E[\tilde{P}_{j;n} | \mathbf{T}] = p_j E[e^{-Z_n(T_{(j)}^-)} | \mathbf{T}] \prod_{l=j}^{n(\mathbf{p})} \frac{\phi(r_l + 1)}{\phi(r_l)}.$$

Furthermore, integrating $E[e^{-Z_n(T_{(j)}^-)} | \mathbf{T}]$, in (58), with respect to the distribution of $\mathbf{T} | \mathbf{m}, \mathbf{p}$ given in Lemma 5.3, and using the relationships $\psi_{1, r_j} + \psi_{m_j, r_{j-1}} = \psi_{m_j+1, r_{j-1}}$, $\psi_{m_j+1, r_{j-1}} + \phi(r_{j-1}) = \phi(r_j + 1)$ and $\phi(r_{l-1} + 1) + \psi_{m_l, 1+r_{l-1}} = \phi(r_l + 1)$ for $l > j$, yields the expression,

$$(60) \quad E[e^{-Z_n(T_{(j)}^-)} | \mathbf{m}, \mathbf{p}] = \prod_{l=j}^{n(\mathbf{p})} \frac{\phi(r_l)}{\phi(r_l + 1)},$$

and hence $E[\tilde{P}_{j,n}|\mathbf{m}, \mathbf{p}] = E[p_j^*|\mathbf{m}, \mathbf{p}] = p_j$. The arguments to obtain expressions for the case when T_{n+1} is distinct from the current unique values are a bit more delicate. Using the the fact that $F_0(ds)$ is non-atomic, express

$$E[e^{-Z_n(s^-)} \left[\prod_{\{j: T_{(j)} < s\}} (1 - J_j) \right] \Lambda_n(ds, dx) | \mathbf{T}, \mathbf{X}] = \sum_{j=1}^{n(\mathbf{p})+1} I_{\{T_{(j)} < s < T_{(j-1)}\}} B_j(s) \Lambda_0(ds, dx),$$

where for $j = 1, \dots, n(\mathbf{p}) + 1$,

$$B_j(s) = e^{-\int_{T_{(j)}}^s \psi_{1,r_{j-1}}(u) \Lambda_0(du)} E[e^{-Z_n(T_{(j)}^-)} | \mathbf{T}] \kappa_{1,r_{j-1}}(\rho|s) \prod_{l=j}^{n(\mathbf{p})} \frac{\kappa_{m_l, r_{l-1}+1}(\rho|T_{(l)})}{\kappa_{m_l, r_{l-1}}(\rho|T_{(l)})}.$$

In particular, $B_{n(\mathbf{p})+1}(s) = \kappa_{1,n}(\rho|s) \exp(-\int_0^s \psi_{1,n}(u) \Lambda_0(du))$. From this one can construct conditional probabilities, $\mathbb{P}\{T_{n+1} \in ds, X_{n+1} \in dx | T_{(j)} < T_{n+1} < T_{(j-1)}, \mathbf{T}\}$, defined as,

$$F_{j:n}(ds, dx) := P_0(dx|s) F_{j:n}(ds) = [q_j^*]^{-1} I_{\{T_{(j)} < s < T_{(j-1)}\}} B_j(s) \Lambda_0(ds, dx).$$

and $q_j^* = \mathbb{P}\{T_{(j)} < T_{n+1} < T_{(j-1)} | \mathbf{T}\} = \int_{T_{(j)}}^{T_{(j-1)}} B_j(s) \Lambda_0(ds)$. Or equivalently,

$$q_j^* = E[e^{-Z_n(T_{(j)}^-)} - e^{-Z_n(T_{(j-1)}^-)} | \mathbf{T}] \prod_{l=j}^{n(\mathbf{p})} \frac{\kappa_{m_l, r_{l-1}+1}(\rho|T_{(l)})}{\kappa_{m_l, r_{l-1}}(\rho|T_{(l)})}$$

In the homogeneous case this can be expressed as

$$(61) \quad q_j^* = q_j \frac{\phi(r_{j-1} + 1)}{\psi_{1,r_{j-1}}} E[e^{-Z_n(T_{(j)}^-)} - e^{-Z_n(T_{(j-1)}^-)} | \mathbf{T}] \prod_{l=j}^{n(\mathbf{p})} \frac{\phi(r_l + 1)}{\phi(r_l)}.$$

Interestingly, $F_{j:n}(ds, dx) := P_0(ds|x) F_{j:n}(ds)$ has a simple form with

$$F_{j:n}(ds) = \frac{\psi_{1,r_{j-1}}(s) e^{-\int_{T_{(j)}}^s \psi_{1,r_{j-1}}(u) \Lambda_0(du)}}{1 - e^{-\int_{T_{(j)}}^{T_{(j-1)}} \psi_{1,r_{j-1}}(u) \Lambda_0(du)}} \Lambda_0(ds)$$

which in the homogeneous case becomes,

$$(62) \quad F_{j:n}(ds) = \frac{\psi_{1,r_{j-1}} e^{-\psi_{1,r_{j-1}}[\Lambda_0(s) - \Lambda_0(T_{(j)})]}}{1 - e^{-\psi_{1,r_{j-1}}[\Lambda_0(T_{(j-1)}) - \Lambda_0(T_{(j)})]}} \Lambda_0(ds)$$

for $T_{(j)} < s < T_{(j-1)}$. Note that $\kappa_{1,r_{j-1}}(\rho|s) = \psi_{1,r_{j-1}}(s)$. Now for the homogeneous case, using (60), and $\psi_{1,r_{j-1}} = \phi(r_{j-1} + 1) - \phi(r_{j-1})$, it follows that $E[q_j^*|\mathbf{m}, \mathbf{p}] = q_j$. In addition let $T_{(1):n+1} > \dots > T_{(n(\mathbf{p})+1):n+1}$ denote the order statistics of the $n(\mathbf{p}) + 1$ unique values in the sample of size $n + 1$. An expression for the prediction rule and related quantities is given,

Proposition 5.3 *Let F denote a spatial NTR process with parameters ρ , and $F_0(ds, dx)$, then based on $L(d\mathbf{T}, d\mathbf{X})$, the Bayesian prediction rule is,*

$$\mathbb{P}\{T_{n+1} \in ds, X_{n+1} \in dx | \mathbf{T}, \mathbf{X}\} = \sum_{j=1}^{n(\mathbf{p})+1} q_j^* P_0(dx|s) F_{j:n}(ds) + \sum_{j=1}^{n(\mathbf{p})} p_j^* \delta_{T_{(j)}, X_j^*}(ds, dx),$$

The probability that the pair (T_{n+1}, X_{n+1}) is new is $1 - \sum_{j=1}^{n(\mathbf{p})} p_j^* = \sum_{j=1}^{n(\mathbf{p})+1} q_j^*$. When $\rho(du|s) = \rho(du)$, p_j^* , q_j^* , and $F_{j:n}(ds)$ simplify to the expressions given in (59), (61) and (62). In addition $E[q_j^*|\mathbf{m}, \mathbf{p}] = q_j$ and $E[p_j^*|\mathbf{m}, \mathbf{p}] = p_j$. Moreover, it can be deduced from Lemma 5.3, that the quantity $E[q_j^* F_{j:n}(ds) | \mathbf{m}, \mathbf{p}] / q_j$ is equivalent to the distribution of $T_{(j):n+1}$ in the case where there is one value, namely T_{n+1} , equal to it.

From the prediction rule one can obtain expressions for posterior quantities such as the Bayes estimate for $E[X]$ in the medical cost model in (10), given as,

$$E[E[X]|\mathbf{T}, \mathbf{X}] = \sum_{j=1}^{n(\mathbf{p})+1} q_j^* \int_{T(j)}^{T(j-1)} \left[\int_0^\infty x P_0(dx|s) \right] F_{j:n}(ds) + \sum_{j=1}^{n(\mathbf{p})} p_j^* X_j^*.$$

We now provide a description of the prediction rule for a spatial generalized Dirichlet processes, process.

5.5.1 Spatial beta processes

As an example in the inhomogeneous case, for the spatial NTR process F on \mathcal{S} defined by the spatial beta process with parameters $c(s)$ and $\Lambda_0(ds, dx)$ the prediction rule is specified by,

$$p_j^* = \frac{m_j}{r_j + c(T(j))} \left[\prod_{l=j+1}^{n(\mathbf{p})} \frac{r_{l-1} + c(T(l))}{r_l + c(T(l))} \right] \prod_{l=j}^{n(\mathbf{p})} e^{-\int_{T(l+1)}^{T(l)} \frac{c(s)}{c(s)+r_l} \Lambda_0(ds)},$$

$$q_j^* = \left[1 - e^{-\int_{T(j)}^{T(j-1)} \frac{c(s)}{c(s)+r_{j-1}} \Lambda_0(ds)} \right] \prod_{l=j}^{n(\mathbf{p})} \frac{r_{l-1} + c(T(l))}{r_l + c(T(l))} e^{-\int_{T(l+1)}^{T(l)} \frac{c(s)}{c(s)+r_l} \Lambda_0(ds)},$$

and

$$F_{j:n}(ds) \propto e^{-\int_{T(j)}^s \frac{c(u)}{c(u)+r_{j-1}} \Lambda_0(du)} \frac{c(s)}{c(s)+r_{j-1}} \Lambda_0(ds) \text{ for } T(j) < s < T(j-1).$$

For Ferguson and Phadia's homogeneous process with $c(s) = \theta$, we have

$$p_j^* = \frac{m_j}{\theta + n} \prod_{l=j}^{n(\mathbf{p})} e^{\frac{-\theta}{\theta+r_l} [\Lambda_0(T(l)) - \Lambda_0(T(l+1))]},$$

$$q_j^* = \frac{\theta + r_{j-1}}{\theta + n} \left[1 - e^{\frac{-\theta}{\theta+r_{j-1}} [\Lambda_0(T(j-1)) - \Lambda_0(T(j))]} \right] \prod_{l=j}^{n(\mathbf{p})} e^{\frac{-\theta}{\theta+r_l} [\Lambda_0(T(l)) - \Lambda_0(T(l+1))]},$$

and

$$F_{j:n}(ds) \propto e^{\frac{-\theta}{\theta+r_{j-1}} [\Lambda_0(s) - \Lambda_0(T(j))]} \frac{\theta}{\theta + r_{j-1}} \Lambda_0(ds) \text{ for } T(j) < s < T(j-1).$$

One can compare these formula with those for p_j and q_j given (56). Notice that for models subject to Aalen Filtering as in section 4.4, the rules are given by replacing $c(s)$ with $c(s) + Y_{c,m}(s) - Y_n(s)$.

5.6 Species sampling models generated by spatial NTR processes

The availability of the EPPF, coupled with Pitman's (1996) theory of species sampling random probability models, implies that there exists a new explicit class of random probability measures of the form,

$$(63) \quad P_F(\cdot) = \sum_{i=1}^{\infty} P_i \delta_{Z_i}(\cdot)$$

where Z_i are iid random elements in \mathcal{X} with some non-atomic law P_0 and, independent of (Z_i) , (P_i) denotes an collection of random probabilities, summing to 1, whose law is completely determined by the EPPF $\pi(\mathbf{p})$ given in Proposition 5.1. We will call P_F an *NTR species sampling model*. The next proposition describes how such a process can always be generated by an F with an independent prior specification, $F_0(ds, dx) = F_0(ds)P_0(dx)$.

Proposition 5.4 *Let N be $\mathcal{P}(dN|\rho, \Lambda_0)$, where Λ_0 is chosen such that, $\Lambda_0(ds, dx) = \Lambda_0(ds)P_0(dx)$, then the corresponding spatial NTR process, F , generates an NTR species sampling model, P_F , given in (63), by the representatons, $P_F(dx) := F((0, \infty), dx) = \int_0^\infty S(s-) \Lambda(ds, dx)$, or equivalently, the marginal distribution of $\mathbf{X} = (\mathbf{X}^*, \mathbf{p})$ is given by $E[\prod_{i=1}^n P_F(dX_i)] = \pi(\mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} P_0(dX_j^*)$. Additionally the posterior distribution of P_F is characterized by Theorem 5.*

PROOF. Under the specifications, $F_0(ds, dx) = F_0(ds)P_0(dx)$, $L(d\mathbf{T}, d\mathbf{X})$ is such that given \mathbf{p} the vectors \mathbf{T}^* and \mathbf{X}^* are independent, where \mathbf{X}^* has joint law $\prod_{j=1}^{n(\mathbf{p})} P_0(dX_j^*)$. The result is concluded by integrating out \mathbf{T}^* . \square

In most cases $\pi(\mathbf{p})$ will not be easy to work with directly, as such one can work with $\pi(\mathbf{m}, \mathbf{p})$. As an example we present an expression for the prediction rule.

Corollary 5.3 *Let P_F denote an NTR species sampling model defined by the choice $\rho(du|s) = \rho(du)$, then one can define a prediction rule for X_{n+1} given \mathbf{X}, \mathbf{m} as follows,*

$$\mathbb{P}\{X_{n+1} \in dx | \mathbf{X}, \mathbf{m}\} = (1 - \sum_{j=1}^{n(\mathbf{p})} p_j) P_0(dx) + \sum_{j=1}^{n(\mathbf{p})} p_j \delta_{X_j^*}(dx).$$

The prediction rule given \mathbf{X} is obtained by $\mathbb{P}\{X_{n+1} \in dx | \mathbf{X}\} = \sum_{\mathbf{m} \in S_{n(\mathbf{p})}} \mathbb{P}\{X_{n+1} \in dx | \mathbf{X}, \mathbf{m}\} \pi(\mathbf{m} | \mathbf{p})$.

It is interesting to note that while the Dirichlet process is an example of P_F , it also arises without the independence specification. The most surprising case is the emergence of the two parameter (α, α) Poisson-Dirichlet process, that arises from the beta $(-\alpha, \alpha)$ spatial NTR process. This fact leads to another representation for the two-parameter (α, α) Poisson-Dirichlet process as follows,

Proposition 5.5 *Let F denote a beta $(-\alpha, \alpha)$ spatial NTR process with independent prior specification, then the process P_F is a two-parameter (α, α) Poisson-Dirichlet process. This implies that using the results in Pitman (1996) and Pitman and Yor (1997), in combination with Proposition 5.4, yields the following representations,*

$$P_F(dx) = \int_0^\infty S(s-) \Lambda(ds, dx) = \sum_{j=1}^{\infty} \tilde{V}_j \prod_{i=1}^{j-1} (1 - \tilde{V}_i) \delta_{Z_j}(dx) = \frac{\mu(dx)}{\mu(\mathcal{X})},$$

where (\tilde{V}_k) are independent beta $(1 - \alpha, \alpha(1+k))$ random variables independent of the (Z_k) which are iid P_0 . The random measure μ , has law, say $\mathcal{PD}_{\alpha, \alpha}(d\mu | P_0)$, determined by the stable law process with index $0 < \alpha < 1$, via the absolute continuity formula, $\mathcal{PD}_{\alpha, \alpha}(d\mu | P_0) \propto \mu(\mathcal{X})^{-\alpha} P(d\mu | \tau_{\alpha, 0}, P_0)$. The quantity, $P(d\mu | \tau_{\alpha, 0}, P_0)$, indicates that μ is a stable law process. That is, if μ has law $P(d\mu | \tau_{\alpha, 0}, P_0)$, then $\mu(dx) = \int_0^\infty y N(dy, dx)$ where N is $\mathcal{P}(dN | \tau_{\alpha, 0}, P_0)$.

6 NTR mixture models

Define,

$$(64) \quad f(y|F) = \int_{\mathcal{S}} K(y|t, x) F(dt, dx) = \int_{\mathcal{S}} K(y|t, x) S(t-) \Lambda(dt, dx),$$

where K is a positive integrable kernel on a polish space \mathcal{Y} . Similar to Lo (1984) for Dirichlet processes, the quantity represents a rich class of spatial NTR mixture models. That is (\mathbf{T}, \mathbf{X}) , represent missing values from an unknown distribution F . This framework potentially allows for the application of NTR models to a much larger collection of complex statistical models. Note also that the class $\int_{\mathcal{X}} K(y|x)P_F(dx)$, generates a new class of species sampling mixture models. Species sampling mixture models are described in Ishwaran and James (2003). Assuming that $Y_1, \dots, Y_n|F$ are iid $f(y|F)$, the likelihood given F is,

$$(65) \quad \prod_{i=1}^n f(Y_i|F) = \prod_{i=1}^n \int_{\mathcal{S}} K(Y_i|t_i, x_i)F(dt_i, dx_i)$$

Notice that from an application of Theorem 5 and Lemma 5.3, the marginal distribution of \mathbf{Y} is given by $\sum_{\mathbf{p}} \sum_{\mathbf{m} \in S_{n(\mathbf{p})}} f(\mathbf{Y}|\mathbf{m}, \mathbf{p})\pi(\mathbf{m}, \mathbf{p})$, where in the homogeneous case,

$$f(\mathbf{Y}|\mathbf{m}, \mathbf{p}) = \left[\prod_{j=1}^{n(\mathbf{p})} \phi(r_j) \right] \int \prod_{j=1}^{n(\mathbf{p})} f(\mathbf{Y}|t_{(j)})e^{-\psi_{m_j, r_{j-1}}\Lambda_0(t_{(j)})}\Lambda_0(dt_{(j)}),$$

with, $f(\mathbf{Y}|t_{(j)}) = \int_{\mathcal{X}} \left[\prod_{i \in E_{(j)}} K(Y_i|t_{(j)}, x) \right] P_0(dx|t_{(j)})$. The formula $f(\mathbf{Y}|t_{(j)})$ remains the same for the inhomogeneous case. Given the results in sections 5.2-5.6 it is evident that, with some additional work, one can devise Gibbs sampling schemes to approximate posterior quantities in a mixture model setting. See also Ishwaran and James (2001) and Ishwaran and James (2003) for a discussion on various techniques for the related classes of stick-breaking priors and species sampling mixture models. The next result presents a characterization of the posterior distribution of $\mathbf{T}, \mathbf{X}, F|\mathbf{Y}$ which is designed so that one might exploit those results.

THEOREM 6. *Suppose that F is a spatial NTR induced by N which is $\mathcal{P}(dN|\rho, \Lambda_0)$. Then the posterior distribution of $F, \mathbf{T}, \mathbf{X}$ given \mathbf{Y} , based on (65) is characterized by the following joint law, $\pi(dN|\mathbf{T}, \mathbf{X})\pi(d\mathbf{T}, d\mathbf{X}|\mathbf{m}, \mathbf{p}, \mathbf{Y})\pi(\mathbf{m}, \mathbf{p}|\mathbf{Y})$ where*

$$\pi(d\mathbf{T}, d\mathbf{X}|\mathbf{m}, \mathbf{p}, \mathbf{Y}) \propto L(d\mathbf{T}, d\mathbf{X}|\mathbf{m}, \mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} \left[\prod_{i \in E_{(j)}} K(Y_i|T_{(j)}, X_j^*) \right]$$

and $\pi(\mathbf{p}, \mathbf{m}|\mathbf{Y}) \propto \pi(\mathbf{m}, \mathbf{p})f(\mathbf{Y}|\mathbf{m}, \mathbf{p})$.

6.0.1 Density estimation

A Bayesian analogue of density estimation using spatial NTR processes can be carried out by random density,

$$(66) \quad \int_{\mathcal{S}} \tau K(\tau(y-x))F(d\tau, dx),$$

where K is a convolution kernel, such as a Normal or Epanechnikov, $T = \tau = 1/\sigma$, where σ plays the role of a bandwidth, $x \in (-\infty, \infty)$ is a mean parameter and F is chosen to be a spatial NTR process with $\Lambda_0(d\tau, dx) = \Lambda_0(d\tau)P_0(dx)$. Note for this problem, it seems ok to specify a prior such that \mathbf{T}, \mathbf{X} are independent. This simple model illustrates how one might envision spatial NTR models outside of the survival context. The posterior random measure corresponding to (66) can be expressed as,

$$\sum_{j=1}^{n(\mathbf{p})+1} \left[\prod_{l=j}^{n(\mathbf{p})} (1 - J_l) \right] \int_{\tau_{(j)}}^{\tau_{(j-1)}} \int_{-\infty}^{\infty} \tau K(\tau(y-x))e^{-Z_n(\tau^-)}\Lambda_n(d\tau, dx) + \sum_{j=1}^{n(\mathbf{p})} \tilde{P}_{j:n}\tau_{(j)}K(\tau_{(j)}(y-x))$$

where the law of the relevant random quantities are given by Theorem 6. Note for instance that a predictive density given \mathbf{T}, \mathbf{X} is given by an application of Proposition 5.3 as,

$$f(y|\mathbf{T}, \mathbf{X}) = \sum_{j=1}^{n(\mathbf{p})+1} q_j^* \int_{\tau_{(j)}}^{\tau_{(j-1)}} \left[\int_{-\infty}^{\infty} \tau K(\tau(y-x)) P_0(dx) \right] F_{j:n}(d\tau) + \sum_{j=1}^{n(\mathbf{p})} p_j^* \tau(j) K(\tau_{(j)}(y - X_j^*))$$

which coupled with Theorem 6 yields expressions for the predictive density given \mathbf{Y} . See James (2003) for the case of smoothing the hazard, which exhibits quite different behaviour.

7 Acknowledgements

I would like to thank Kjell Doksum for early comments which helped in the exposition of this work. Thanks to Jim Pitman for clarifying some nice connections to his and Alexander Gnedin's work.

References

- AALEN, O. O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability* **2** 951-972.
- ANDERSON, P. K., BORGAN, O. , GILL, R. D. AND KEIDING, N. (1993). *Statistical Models Based On Counting Processes*. Springer-Verlag, New York.
- ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152-1174.
- BANG, H., AND TSIATIS, A. A. (2000). Estimating medical costs with censored data. *Biometrika* **87** 329-343.
- BERTOIN, J. AND YOR, M. (2001). On subordinators, self-similar Markov processes and some factorizations of the exponential variable. *Electronic Communications in Probability* **6** 95-106.
- BLACKWELL, D. AND MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353-355.
- BRIX, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. in Appl. Probab.* **31** 929-953.
- CARMONA, P., PETIT, F., AND YOR, M. (1997). On the distribution and asymptotic results for exponential functionals of Levy processes. In: M. Yor (editor) Exponential functionals and principal values related to Brownian motion, pp. 73-121. Biblioteca de la Revista Matemática Iberoamericana..
- DALEY, D. J. AND VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York.
- DEY, J. (1999). Some Properties and Characterizations of Neutral-to-the-Right Priors and Beta Processes. Ph.D. Thesis. Michigan State University.
- DEY, J, ERICKSON, R.V. AND RAMAMOORTHI, R.V. (2000). Neutral to right priors- A review. Preprint. To appear in the *International Statistical Review*.
- DONNELLY, P. AND JOYCE, P. (1991). Consistent ordered sampling distributions: characterization and convergence. *Adv. Appl. Prob.* **23** 229-258.
- DOKSUM, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183-201.
- DOKSUM, K. A. AND JAMES, L. F. (2003). On spatial neutral to the right processes and their posterior distributions. In *Mathematical Reliability: An Expository Perspective*, Editors: Mazzuchi, Singpurwalla and Soyer. Kluwer Academic Publishers, to appear.
- DUBINS, L. E. AND FREEDMAN, D. A. (1966). Random distribution functions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **2**. Univ. of California, Berkeley, 183-214.

- EPIFANI, I., LIJOI, A., AND PRUENSTER, I. (2002). Exponential functionals and means of neutral to the right priors. Technical report Universita di Pavia no. 144.
Available at <http://economia.unipv.it/eco-pol/quaderni.htm>. To appear in *Biometrika*.
- EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3** 87-112.
- FABIUS, J. (1964). Asymptotic behavior of Bayes estimates. *Ann. Math. Statist.* **35** 846-856.
- FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209-230.
- FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615-629.
- FERGUSON, T. S. AND PHADIA, E. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7** 163-186.
- FREEDMAN, D. A. (1963). On the asymptotic behavior of Bayes estimates in the discrete case. *Ann. Math. Statist.* **34** 1386-1403.
- GILL, R.D. AND JOHANSEN, S. (1990). Survey of product-integration with a view towards applications in survival analysis. *Ann. Statist.* **18** 1501-1555.
- GNEDIN, A. V. (2002). Three sampling formulas arXiv:math.PR/0210319 , 2002.
Available at <http://arxiv.org/abs/math.PR/0210319>.
- GNEDIN, A. V. (1997). The representation of composition structures. *Ann. Probab.* **25** 1437-1450..
- GNEDIN, A. V. AND PITMAN, J. (2002). Regenerative compositions. Manuscript in preparation.
- HJORT, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *Ann. Statist.* **18** 1259-1294.
- HUANG, Y. AND LOUIS, T.A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables.. *Biometrika* **85** 785-798.
- ISHWARAN, H. AND JAMES, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statistica Sinica*, to appear.
- ISHWARAN, H. AND JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Assoc.* 161-173.
- ISHWARAN, H., JAMES, L.F. AND SUN, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *J. Amer. Statist. Assoc.* **96** 1316-1332.
- JAMES, L. F. (2003). Poisson process partition calculus with applications to Bayesian Lévy moving averages and shot-noise processes. Manuscript.
- JAMES, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics, arXiv:math.PR/0205093, 2002.
Available at <http://arxiv.org/abs/math.PR/0205093>.
- JAMES, L. F. AND KWON, S. (2000). A Bayesian nonparametric approach for the joint distribution of survival time and mark variables under univariate censoring. Technical Report, Department of Mathematical Sciences, Johns Hopkins University.
- KIM, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* **27** 562-588.
- KIM, Y. AND LEE, J. (2003). Bayesian analysis of proportional hazard models. *Ann. Statist.* **31**.
- KIM, Y. AND LEE, J. (2001). On posterior consistency of survival models. *Ann. Statist.* **29** 666-686.
- LAST, G. AND BRANDT, A. (1995). *Marked point processes on the real line: The dynamic approach*. Springer-Verlag, New York.
- LO, A. Y. (1993). A Bayesian bootstrap for censored data. *Ann. Statist.* **21** 100-123.
- LO, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density Estimates. *Ann. Statist.* **12** 351-357.
- LO, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrsch. verw. Gebiete* **59** 55-66.
- LO, A. Y. AND WENG, C. S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rates estimates. *Ann. Inst. Stat. Math.* **41** 227-245.

- PITMAN, J. (2002a). Combinatorial stochastic processes. Lecture notes for Saint-Flour summer school, July 2002. Available at <http://stat-www.berkeley.edu/users/pitman/bibliog.html>.
- PITMAN, J. (2002b). Poisson-Kingman Partitions. Technical Report 625, Dept. Statistics, U.C. Berkeley.
- PITMAN, J. (1999). Coalescents with multiple collisions. *Ann. Probab.* **27** 1870-1902.
- PITMAN, J. (1997). Partition structures derived from Brownian motion and stable subordinators. *Bernoulli* **3** 79-96.
- PITMAN, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory* T.S. Ferguson, L.S. Shapley and J.B. MacQueen editors, IMS Lecture Notes-Monograph series, Vol 30, pages 245-267.
- PITMAN, J. AND YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855-900.
- WALKER, S. AND MULIERE, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Statist.* **25** 1762-1780.

LANCELOT F. JAMES
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF INFORMATION SYSTEMS AND MANAGEMENT
CLEAR WATER BAY, KOWLOON
HONG KONG
lancelot@ust.hk