Sankhyā : The Indian Journal of Statistics
2003, Volume 65, Part 1, pp 179-206

BAYESIAN CALCULUS FOR GAMMA PROCESSES WITH APPLICATIONS TO SEMIPARAMETRIC INTENSITY MODELS

By LANCELOT F. JAMES
Hong Kong University of Science and Technology

SUMMARY. Explicit calculus for the posterior distribution of convolution mixtures of weighted gamma processes on Polish spaces are derived. This serves to extend the results of Lo and Weng (1989) to a semiparametric setting on arbitrary spaces. The result of this study is applied to two different types of general semiparametric multiplicative intensity models. One in which a prior is constructed based on q conditionally independent weighted gamma measures given a Euclidean parameter and a second dependent model where different hazard rates are based on a common mixing measure. The latter model seems natural for some types of deconvolution models or regression models. As an example, it is shown how this provides a full (implementable) posterior analysis of the Cox regression model. The results also provide the explicit posterior distribution for the Poisson/Gamma random field model considered by Wolpert and Ickstadt (1998a).

1. Introduction

This article presents rigorous development of the necessary prior to posterior calculus for Bayesian non/semiparametric models incorporating convolution mixtures of weighted gamma priors on arbitrary Polish spaces. That is, arbitrary complete and separable metric spaces $\mathcal{V}$. Weighted gamma processes on arbitrary Polish spaces are discussed in Lo (1982) and are shown to be conjugate to the family of inhomogeneous Poisson processes when the parameter of interest is the cumulative hazard or intensity. The extended gamma process proposed in Dykstra and Laud (1981) is a special case of the weighted Gamma process. Tsilevich, Vershik and Yor (2000, 2001) provide some interesting results for the (unweighted) gamma process in a somewhat different but relevant context.

Paper received November 2000; revised July 2001.
AMS (2000) *subject classifications.* Primary 62G05; secondary 62F15.
Key words and phrases. Convolution mixtures of hazards, weighted Chinese restaurant process, inhomogeneous Poisson process, multiplicative intensity models, weighted gamma processes, dependent mixtures of weighted gamma processes.

The approach of Lo and Weng (1989) is analogous to the case of density and mixture model estimation in Lo (1984). Lindsay (1995) gives an excellent discussion on the richness of the mixture model framework. The present work extends Lo and Weng (1989) in two directions: One the important extension to the semiparametric setting which allows for hierarchical modelling and two the extension from the real line to arbitrary complete and separable metric spaces. Thus for instance one can consider a spatial intensity over an abstract geometric space $\mathcal{V}$ as follows

$$\lambda(t|\mu, \theta) = \int_{\mathcal{V}} k(t, v, \theta)\mu(dv).$$

The extension to the semiparametric setting provides full Bayesian calculus and hence a clear path for Bayesian estimation of, for instance, semiparametric multiplicative hazards models and other multiplicative intensity models first made popular by Aalen (1975, 1978) and discussed in the text by Andersen, Borgan, Gill and Keiding (1993). The use of arbitrary kernels k provides the user with a great deal of flexibility in modelling. For instance, as noted in Lo and Weng (1989)[see section 6], the choice of kernel

$$k(t, v) = I\{v \leq t\}, \tag{1}$$

yields a family of nonincreasing hazard rates first considered from a Bayesian viewpoint by Dykstra and Laud (1981). Note that the choice of (1) also provides a Bayesian alternative to estimation of a monotone hazard under right censoring considered in Huang and Wellner (1995). The choice,

$$k(t, v) = e^{-tv} \tag{2}$$

can be used for Bayesian estimation for a hazard rate which is modelled as a mixture of exponentials. This corresponds to a class of completely monotone hazard rates. An interesting discussion of mixed-exponential hazards may be found in Saunders and Myhre (1983)[see also Jewell (1982)]. The kernel,

$$k(t, v) = I\{|t - a| \geq v\}, \tag{3}$$

$v \in [0, \infty)$; corresponds to the class of U-shaped or bath-tub shaped hazards with minimum at a. [See Lo and Weng (1989) for further details]. More recently, Wolpert and Ickstadt (1998a) discuss a Bayesian hierarchical spatial Poisson/Gamma random field model which fits in our framework. They give as examples intensities based on semiparametric mixtures of multivariate normal kernels. In addition, Ishwaran (1998) provides an interesting discussion and important consistency results for semiparametric mixture models characterized by general Pólya urn sequences. This includes for instance models which may be derived from the Dirichlet process. It is possible to include these models in this framework as well. See Remark 1 below. Some other related Bayesian works and discussions include, Kalbfleisch (1978), Hjort (1990), Damien, Laud and Smith (1996), Sinha and Dey (1997), Sinha and Dey (1998), Ibrahim and Sinha (1998), Laud, Damien and Smith (1998), Wolpert and Ickstadt (1998b), Walker, Damien, Laud and Smith (1999), Ibrahim, Chen and MacEachern (1999), Best, Ickstadt and Wolpert (2000), Gasbarra and Karia (2000), and Hayakawa, Paul and Vignaux (2000). A recent survey, with additional references, is given in Gelfand (1999). It is also evident that the Bayesian approach discussed in this section may be used as a Bayesian alternative to maximum pseudo-likelihood methods for spatial point patterns on abstract spaces based on Besag's (1977) pseudo-likelihood approach as discussed in Baddeley and Turner (2000).

The layout of the paper is as follows. In section 2, the weighted gamma measure is described and some important connections to the Dirichlet process are mentioned [see Ferguson (1973, 1974) and Antoniak (1974)]. Section 3 focuses on obtaining the analogues of Proposition 3.1. and Lemma 3.1 in Lo and Weng (1989) for general Polish spaces $\mathcal{V}$. It follows that once this is established all the resulting arguments in Lo and Weng (1989) may be carried over to this more general setting. In section 4.1, applications to an extension of a q-component nonparametric multiplicative intensity model are discussed. The priors for these models are based on q-component mixtures of weighted gamma processes and are analogous to mixtures of Dirichlet processes considered in Antoniak (1974). In section 4.2, the results in section 3 are applied to models where the hazard or intensities rates share a common mixing measure. Such models are not treated in Lo and Weng (1989).

In section 5, important partition representations for the posterior quantities discussed in section 4.2. are derived. Lo (1984) and Lo and Weng (1989), show that these representations reduce complex integrations with respect to a Blackwell-MacQueen (1973) type urn to simpler averages over partitions of the data. Lo, Brunner and Chan (1996) [see also Brunner, Chan, James and Lo (2001)] discuss iid and Gibbs sampling based procedures for convolution mixtures of Dirichlet and gamma processes specifically utilizing the partition structure. They dub these computational procedures Weighted Chinese Restaturant (WCR) processes. In particular, the iid sequential importance sampling (SIS) procedure is referred to as the iidWCR. Brunner, Chan, James and Lo (2001) point out that the iidWCR when specialized to the case of a binomial kernel mixture of Dirichlet process models coincides operationally with an SIS procedure in MacEachern, Clyde and

Liu (1999). Lo and Weng (1989), in section 6 of that paper, actually apply a Chinese Restuarant (CR) approximation procedure [see Kuo (1986) and Aldous (1985) and Pitman (1996)] to the mixture of exponentials model mentioned above. While the CR algorithm of Kuo (1986) samples from a prior partition distribution the WCR procedures samples from a "posterior" partition distribution connected with the observed data and hence is more efficient. Hayakawa, Paul, Vignaux (2000) and Hayakawa, Zukerman, Paul and Vignaux (2001) apply the iidWCR to implement a test for detecting increasing failure rate (IFR) data. Ho and Lo (2001) apply a Gibbs sampler based on partitions in the case of a monotone hazard rate. Recently Ishwaran and James (2000) develop extensions of the iidWCR and discuss several novel Gibbs sampling procedures to fit the semiparametric mixture of hazards gamma process models. These procedures provide methods for practical application of the results discussed here. What is evident in all these works is that efficient approximation procedures for the Dirichlet process may be adapted, with some adjustments, to models based on the gamma process. More subtle features of the partition representations are discussed in Brunner, Chan, James and Lo (2001). In section 6, a complete posterior characterization of the Cox regression model is given. In addition, it is shown how one may apply the iidWCR to approximate posterior quantities for this model.

1.1 *Cox proportionals hazards regression model.* As a motivational example the well known Cox regression model, Cox (1972), is presented in this section. Let $(T_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$ be random variables with T_i non-negative (lifetimes) and each $\mathbf{Z}_i$ is a m-dimensional vector of covariates. The Cox proportional hazards (partial) likelihood may be written as,

$$\prod_{j=1}^{n} \lambda(T_j|\mu) \exp\{\eta^t \mathbf{Z}_j\} \exp\left\{-\int_{\mathcal{R}} \lambda(u|\mu) \sum_{l=1}^{n} \exp\{\eta^t \mathbf{Z}_l\} Y_l(u) du\right\}, \quad (4)$$

where $Y_l(u) = I\{T_l > u\}$, $\lambda(t|\mu)$ represents the unknown baseline hazard rate and $\eta = (\eta_1, \ldots \eta_m)$ is a vector of unknown regression coefficients. Here the baseline hazard rate is modelled as a mixture,

$$\lambda(t|\mu) = \int_{\mathcal{R}} k(t, v)\mu(dv).$$

The semiparametric hazard rate is then expressed as,

$$\lambda(T_j|\mu, \eta) = \int_{\mathcal{R}} k_j(T_j, v, \eta)\mu(dv)$$

where

$$k_j(T_j|v,\eta) = k(T_j, v) \exp\{\eta^t \mathbf{Z}_j\}.$$

The Bayesian framework is completed by specifying μ as a weighted gamma process with shape α not depending on η, and choosing a prior for η, say $\pi(d\eta)$. A convenient choice for $\pi(d\eta)$ is a multivariate normal distribution.

Bayesian treatments of the the Cox model, (4), and its variations have been discussed in the works of Kalbfleisch (1978), Hjort (1990), Clayton (1991), Sinha and Dey (1997), Sinha and Dey (1998), Ibrahim and Sinha (1998), Laud, Damien and Smith (1998), among others. The mixture model framework presented here allows one to model various shapes for the baseline hazard and in that sense is more general than the above mentioned works. In section 6, it will be shown how the results in sections 4 and 5 yield the explicit posterior characterization of these models. In addition, it is shown how this facilitates application of the iid computational procedures discussed in Brunner, Chan, James and Lo (2001) [see also Lo, Brunner and Chan (1996) and Ishwaran and James (2000)].

REMARK 1. The above framework is chosen for simplicity, but certainly illustrates the main points. Using the results in this paper one may also treat the case of time dependent covariates $\mathbf{Z}_j(t)$. In general, the results in this paper also apply to the case where α and k depend on an additional parameter θ with prior $\pi(d\theta)$. For instance one may use the choices of kernels discussed in Ishwaran (1998). This includes,

$$k(x, v, \theta) = v\theta \exp(\theta x - v \exp(\theta x)) \tag{5}$$

or

$$k(x_1, x_2, v, \theta) = v \exp(-vx_1)\theta v \exp(-\theta v x_2) \tag{6}$$

which correspond to mixtures of Weibull and Paired Exponential models respectively. Another potentially interesting choice for k and α may be deduced from the class of exponential mixture models proposed in Lindsay (1986)

2. Weighted Gamma Random Measures

It is well known that gamma and Dirichlet processes may be defined over arbitrary measurable spaces, see Kingman (1975) and Ferguson (1973) for details. Here, the case of gamma random measures defined over arbitrary Polish spaces is considered. In order to allow for possibly σ-finite parameters the concept of *boundedly finite measures* is used which is defined in Daley and Vere-Jones (1988) [see their Definition 6.1.1] as, DEFINITION 1. A Borel measure μ on a Polish space is boundedly finite if $\mu(A) < \infty$ for every bounded Borel set A.

Let γ_α denote a gamma random measure on the measure space $(\mathcal{V}, \mathcal{B}(\mathcal{V}))$ with σ-finite shape measure α on $\mathcal{V}$, where $\mathcal{V}$ is an arbitrary Polish space and $\mathcal{B}(\mathcal{V})$ is an appropriately defined Borel sigma field. That is, γ_α is a boundedly finite random measure such that for each bounded Borel subset A of $\mathcal{V}$, $\gamma_\alpha(A) = \int_A \gamma_\alpha(du)$ has a gamma $(\alpha(A); 1)$ distribution. Moreover, for any disjoint sets $B_1, \ldots, B_m$, the $\gamma_\alpha(B_i)$ are independent gamma $(\alpha(B_i); 1)$. Now, let β be a non-negative α integrable function on $\mathcal{V}$, and define as in Lo (1982) [see also Dykstra and Laud (1981) and Lo and Weng (1989)], a weighted gamma measure,

$$\mu_{\alpha,\beta}(A) = \int_A \beta(y)\gamma_\alpha(dy).$$

If μ denotes a measure on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$, a suitably defined Borel measurable space of boundedly finite measures, the weighted gamma distribution of μ is denoted by $\mathcal{G}(d\mu|\alpha, \beta)$. This means that μ is a random measure that may be expressed as $\mu_{\alpha,\beta}$. The notation $\mathcal{G}_{\alpha,\beta}$ will be used interchangeably. It follows from Lo (1982), that the mean and Laplace functional of μ are,

$$\int_{\mathcal{M}} \int_{\mathcal{V}} g(v)\mu(dv)\mathcal{G}(d\mu|\alpha, \beta) = \int_{\mathcal{V}} g(v)\beta(v)\alpha(dv)$$

and

$$\begin{aligned} L_{\mathcal{G}_{\alpha,\beta}}(f) &= \int_{\mathcal{M}} \exp\left\{-\int_{\mathcal{V}} f(v)\mu(dv)\right\} \mathcal{G}(d\mu|\alpha, \beta) \\ &= \exp\left\{-\int_{\mathcal{V}} \ln\left[1 + f(v)\beta(u)\right] \alpha(du)\right\} \end{aligned}$$

for non-negative functions $f \in BM(\mathcal{V})$ on $(\mathcal{V}, \mathcal{B}(\mathcal{V}))$ where $BM(\mathcal{V})$ denotes the collection of measurable functions of bounded support on $\mathcal{V}$.

An important point to note is that if P_α is a Dirichlet process with finite shape measure $\alpha(\cdot)$ then,

$$P_\alpha(\cdot) = \frac{\gamma_\alpha(\cdot)}{\gamma_\alpha(\mathcal{V})},$$

and moreover $\gamma_\alpha(\mathcal{V})$ is independent of $P_\alpha(\cdot)$. This result may be found in Kingman (1975)[see also Ferguson (1973, 1974)] but follows essentially from a well-known elementary result about gamma random variables. An implication is that many properties of a Dirichlet process may be deduced from that of a gamma process and vice versa. Also many computational procedures devised for the Dirichlet process have natural analogues in the gamma process setting. One will note that the derivations in the forthcoming sections will bear many similarities to perhaps more familiar expressions obtained via Dirichlet process calculus. A weighted Dirichlet process as defined in Lo (1993) may be expressed as follows,

$$P_{\alpha,\beta}(\cdot) = \frac{\mu_{\alpha,\beta}(\cdot)}{\mu_{\alpha,\beta}(\mathcal{V})}.$$

3. Fubini Type Disintegration Results for General Polish Spaces

A key to the development of a systematic approach for the calculation of posterior quantities in nonparametric and semiparametric settings rests on the development of a Fubini-type theorem which characterizes the posterior distribution. The approach used here is first to obtain an extension of Proposition 3.1. and Lemma 3.1. of Lo and Weng (1989) to arbitrary Polish spaces $\mathcal{V}$.

LEMMA 1. *For non-negative functions $f \in BM(\mathcal{V})$ on $(\mathcal{V}, \mathcal{B}(\mathcal{V}))$ and g on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$*

$$\int_{\mathcal{M}} g(\mu) \exp\left\{-\int_{\mathcal{V}} f(v)\mu(dv)\right\} \mathcal{G}_{\alpha,\beta}(d\mu) = L_{\mathcal{G}_{\alpha,\beta}}(f) \int_{\mathcal{M}} g(\mu)\mathcal{G}(d\mu|\alpha,\beta^*),$$

where $\beta^* = \beta/(1+\beta f)$.

PROOF. The arguments proceed exactly as in Lo and Weng (1989) with $\mathcal{R}$ replaced by a general Polish space $\mathcal{V}$. Those arguments combined with the unicity of Laplace functionals for random measures on $\mathcal{V}$ validate the Lemma. □

The next result gives an extension of the Fubini-type disintegration Lemma 3.1 in Lo and Weng (1989) to arbitrary Polish spaces $\mathcal{V}$, which provides a key tool for many applications, including missing data models and density/hazard estimation and estimation in spatial point proceses. Note that the arguments used in Lo and Weng (1989) cannot be used for the arbitrary extension.

LEMMA 2. (Bayes Disintegration Lemma) *Let h be any non-negative function defined on $(\mathcal{V} \times \mathcal{M}, \mathcal{B}(\mathcal{V}) \otimes \mathcal{B}(\mathcal{M}))$, then*

$$\int_{\mathcal{M}} \int_{\mathcal{V}} h(v,\mu)\mu(dv)\mathcal{G}_{\alpha,\beta}(d\mu) = \int_{\mathcal{V}} \int_{\mathcal{M}} h(v,\mu)\mathcal{G}_{\alpha+\delta_v,\beta}(d\mu)\beta(v)\alpha(dv).$$

PROOF. One may argue in a standard fashion by choosing h to be indicators of measurable subsets of $\mathcal{B}(\mathcal{V}) \otimes \mathcal{B}(\mathcal{M})$. However, the proof follows directly from an application of Lemma 1 and the unicity property of Laplace functionals. It suffices to show that the result holds for

$$h(v,\mu) = s(v) \exp\left\{-\int_{\mathcal{V}} f(y)\mu(dy)\right\},$$

for any non-negative s,f in $BM(\mathcal{V})$. That is,

$$\int_{\mathcal{M}}\int_{\mathcal{V}} s(v) \exp\left\{-\int_{\mathcal{V}} f(y)\mu(dy)\right\} \mu(dv)\mathcal{G}_{\alpha,\beta}(d\mu)$$
$$= \int_{\mathcal{V}} s(v) \left[\int_{\mathcal{M}} \exp\left\{-\int_{\mathcal{V}} f(y)\mu(dy)\right\} \mathcal{G}_{\alpha+\delta_v,\beta}(d\mu)\right] \beta(v)\alpha(dv).$$

An application of Lemma 1, setting

$$g(\mu) = \int_{\mathcal{V}} s(v)\mu(dv)$$

gives that the l.h.s. is,

$$\exp\left\{-\int_{\mathcal{V}} \ln\left[1+\beta(y)f(y)\right]\alpha(dy)\right\} \int_{\mathcal{M}} \left[\int_{\mathcal{V}} s(v)\mu(dv)\right] \mathcal{G}(d\mu|\alpha,\beta^*)$$
$$= L_{\mathcal{G}_{\alpha,\beta}}(f) \int_{\mathcal{V}} s(v)[1+\beta(v)f(v)]^{-1}\beta(v)\alpha(dv).$$

Since for each v, the Laplace functional $L_{\mathcal{G}_{\alpha+\delta_v,\beta}}(f)$ equals

$$\begin{aligned}\int_{\mathcal{M}} \exp\left\{-\int_{\mathcal{V}} f(y)\mu(dy)\right\} \mathcal{G}_{\alpha+\delta_v,\beta}(d\mu) &= L_{\mathcal{G}_{\alpha,\beta}}(f)[1+\beta(v)f(v)]^{-1} \\ &= L_{\mathcal{G}_{\alpha,\beta}}(f)L_{\mathcal{G}_{\delta_v,\beta}}(f),\end{aligned}$$

the proof is concluded. □

Lemma 2 is labelled Bayes Disintegration Lemma since an explicit Fubini result of this type is fundamental in characterizing the posterior distribution of infinite-dimensional random measures. That is, cases where the elementary Bayes rule does not apply. Hence analogous to Lo's (1984) Lemma 1 result for the Dirichlet process, Lemma 2 provides a master tool for Bayesian models based on the gamma process. In fact, as noted in Lo and Weng (1989, Corollary 3.2), Lemma 2 for the gamma process implies Lo's (1984) Lemma 1 for the Dirichlet process on $\mathcal{V}$. It however should be pointed out that

Lemma 1, the extension of Proposition 3.1 of Lo and Weng (1989), also plays a fundamental role. These points will be illustrated in the forthcoming sections. Note in particular that Lemmas 1 and 2 show that for arbitrary Polish spaces $\mathcal{V}$ the following disintegrations holds,

$$\mu(dv)\mathcal{G}_{\alpha,\beta}(d\mu) = \mathcal{G}_{\alpha+\delta_v,\beta}(d\mu)\beta(v)\alpha(dv),$$

and for any $f \in BM(\mathcal{V})$,

$$\exp\left\{-\int_{\mathcal{V}} f(y)\mu(dy)\right\}\mu(dv)\mathcal{G}_{\alpha,\beta}(d\mu) = L_{\mathcal{G}_{\alpha,\beta}}(f)\mathcal{G}(d\mu|\alpha+\delta_v,\beta^*)\beta(v)\alpha(dv).$$

$\mathcal{G}_{\alpha+\delta_v,\beta}(d\mu)$ has an interpretation as a conditional distribution of μ given v, and $\beta(v)\alpha(dv)$ is a marginal measure associated with v. Le Cam (1986, ch.12) gives a general discussion of this phenomena in a Bayesian context.

REMARK 2. In the case where $\mathcal{V} = \mathcal{R}^d$ a proof may be carried out along the lines of the proof of Lemma 3.1 Lo and Weng (1989) by setting

$$h(v,\mu) = e^{-i<v,t>}\exp\left\{-\int_{\mathcal{V}} f(y)\mu(dy)\right\},$$

where $< v,t >$ denotes inner product on $\mathcal{R}^d$. That is, by using the unicity of characteristic functions in $\mathcal{R}^d$ and the unicity of Laplace functionals on $\mathcal{V}$. This may be extended to other inner product spaces such as Hilbert spaces. However, this argument breaks down for arbitrary Polish spaces $\mathcal{V}$.

Now let

$$\pi(d\mathbf{v}|\mu) = \prod_{i=1}^{n}\mu(dv_i)$$

be a conditional measure of $\mathbf{v} = (v_1,\ldots,v_n)$ given μ where μ is $\mathcal{G}_{\alpha,\beta}(d\mu)$. The following implication of Lemma 1 and Lemma 2 is now stated.

THEOREM 1. *Let $g(\mathbf{v},\mu)$ be a non-negative or integrable function, then for each $n \geq 1$, and $f \in BM(\mathcal{V})$,*

$$\int_{\mathcal{M}} g(\mathbf{v},\mu)\exp\left\{-\int_{\mathcal{V}} f(s)\mu(ds)\right\}\pi(d\mathbf{v}|\mu)\mathcal{G}_{\alpha,\beta}(d\mu)$$
$$= L_{\mathcal{G}_{\alpha,\beta}}(f)\int_{\mathcal{M}} g(\mathbf{v},\mu)\mathcal{G}(d\mu|\alpha+\sum_{i=1}^{n}\delta_{v_i},\beta^*)\pi(d\mathbf{v}),$$

where

$$\pi(d\mathbf{v}) = \int_{\mathcal{M}}\pi(d\mathbf{v}|\mu)\mathcal{G}(d\mu|\alpha,\beta^*) = m(d\mathbf{v})\prod_{i=1}^{n}\beta^*(v_i),$$

and

$$m(d\mathbf{v}) = \prod_{i=1}^{n} \left(\alpha + \sum_{l=1}^{i-1} \delta_{v_l}\right)(dv_i).$$

PROOF. The result follows by first applying Lemma 1 to deal with the exponential term and then repeated application of Lemma 2. Naturally, once one has obtained the extensions to arbitrary $\mathcal{V}$ the arguments are then more or less the same as in Lo and Weng (1989). □

REMARK 3. Note that the Laplace functional of $\mathcal{G}(d\mu|\alpha + \sum_{i=1}^{n} \delta_{v_i}, \beta)$ is,

$$L_{\mathcal{G}_{\alpha,\beta}}(f) \prod_{i=1}^{n} [1 + \beta(v_i) f(v_i)]^{-1}.$$

REMARK 4. The random measure

$$\mu_{\alpha,n}(\cdot) = \int_{\cdot} \beta^*(s)\gamma_\alpha(ds) + \sum_{i=1}^{n} \beta^*(v_i) E_i I\{v_i \in \cdot\},$$

where $E_1, \ldots, E_n$ are iid $exp(1)$ random variables, has precisely the distribution $\mathcal{G}(d\mu|\alpha + \sum_{i=1}^{n} \delta_{v_i}, \beta^*)$. In many applications

$$\int_{\mathcal{V}} \beta^*(v)\gamma_\alpha(dv)$$

is of order n. When this is true, the second term

$$\mu_n(\cdot) = \sum_{i=1}^{n} \beta^*(v_i) E_i I\{v_i \in \cdot\},$$

may be used as a Bayesian bootstrap type approximation to the posterior law $\mathcal{G}(d\mu|\alpha + \sum_{i=1}^{n} \delta_{v_i}, \beta^*)$. See Lo (1992) and references therein. This follows since,

$$\sup_{C \in \mathcal{C}} |\mu_{\alpha,n}(C) - \mu_n(C)| \leq \int_{\mathcal{V}} \beta^*(v)\gamma_\alpha(dv)$$

for $\mathcal{C}$ a suitably rich V-C class of sets.

4. Semiparametric Hazard Rate Mixture Models

4.1 *Conditionally independent models.* Suppose that one has the following model

$$\begin{aligned} L(\mu,\theta) &= \prod_{j=1}^{q}\left[\prod_{i=1}^{n(j)}\int_{\mathcal{V}} k_j(T_{ji},v_{ji},\theta)\mu_j(dv_{ji})\right] \\ &\quad \times \exp\left\{-\int_{\mathcal{V}}\left[\int_{\mathcal{S}} Y_j(s)k_j(s,v,\theta)\tau(ds)\right]\mu_j(dv)\right\}, \end{aligned}$$

where k_j, $j=1,\ldots,q$ are known (τ,α)-integrable kernels depending on a Euclidean parameter $\theta\in\Theta$ and $(\mu_1,\ldots,\mu_q)$ are unknown sigma-finite measures, T_{ji} are observations in a region and v_{ji} can be viewed as missing observations. $Y_j(s)$ are non-negative predictable functions which for many applications in event history analysis denotes the number of observed individuals still at risk just before time s. An important point is that the strucutural form of $L(\mu,\theta)$ remains the same under right censoring and left filtering [see Andersen, Borgan, Gill and Keiding (1993)]. If $Y_j(s)=1$ then the model may correspond to the likelihood of q conditionally independent inhomogeneous Poisson processes with respective intensity rates

$$\lambda_j(t|\mu,\theta)=\int_{\mathcal{V}} k_j(t,v,\theta)\mu_j(dv) \tag{7}$$

for $j=1,\ldots,q$. If in addition $q=1$, then this essentially corresponds to a Poisson/Gamma random field model proposed in Wolpert and Ickstadt (1998a). The model $L(\mu,\theta)$, with the exclusion of the Euclidean parameter θ and $\mathcal{V}$ restricted to the real line $\mathcal{R}$ is proposed in Lo and Weng (1989) to model the hazard rates (intensity) λ_j as a mixture of k_j and the unknown measure μ_j. That is, if k_j is an integrable kernel then

$$\lambda_j(t|\mu)=\int_{\mathcal{R}} k_j(t,v)\mu_j(dv).$$

The authors in addition model $(\mu_1,\ldots,\mu_q)$ as *a priori* independent weighted gamma measures.

Here the vector $\mu=(\mu_1,\ldots,\mu_q)$ is modelled as a mixture of weighted gamma measures. This prior specification is an analogue of the mixture of Dirichlet process models developed in Antoniak (1974). Let $\theta=(\theta_1,\ldots,\theta_d)$ be a Euclidean vector with prior distribution denoted as $\pi(d\theta)$. In addition

$(\mu_1,\ldots,\mu_q)$ are chosen to be conditionally independent given θ with $\mu_j|\theta \sim \mathcal{G}_{\alpha_{j,\theta},\beta_{j,\theta}}(d\mu_j)$ for $j=1,\ldots,q$. Thus, the law of $\mu|\theta$ is,

$$\pi(d\mu|\theta) = \prod_{j=1}^{q} \mathcal{G}_{\alpha_{j,\theta},\beta_{j,\theta}}(d\mu_j).$$

Hence, the (marginal) law of μ is expressible as

$$\pi(d\mu) = \int_{R^d} \prod_{j=1}^{q} \mathcal{G}_{\alpha_{j,\theta},\beta_{j,\theta}}(d\mu_j)\pi(d\theta).$$

Now set, $\mathbf{v}_j = \{v_{ji}, i=1,\ldots,n(j)\}$ and $\mathbf{v} = \{\mathbf{v}_1,\ldots,\mathbf{v}_q\}$ and define the finite measure

$$\pi(d\mathbf{v}|\mu,\theta) = \prod_{j=1}^{q}\prod_{i=1}^{n(j)} \mu_j(dv_{ji}) = \prod_{j=1}^{q} \pi(d\mathbf{v}_j|\mu_j,\theta).$$

A joint measure for $\mathbf{V},\mu,\theta$ is given as,

$$\pi(d\mathbf{v},d\mu,d\theta) = \pi(d\mathbf{v}|\theta,\mu)\pi(d\mu|\theta)\pi(d\theta).$$

To complete the Bayesian model, let $\mathbf{T}_j = \{T_{ji}, i=1,\ldots,n(j)\}$, and $\mathbf{T} = \{\mathbf{T}_1,\ldots,\mathbf{T}_q\}$ where $\mathbf{T}|\mu,\theta$ has (partial) likelihood $L(\mu,\theta)$. Based on $L(\mu,\theta)$, a (partial) likelihood of $\mathbf{T}|\mathbf{V},\mu,\theta$ which incorporates the missing data,$\mathbf{V}$, is given as follows,

$$\begin{aligned} L(\mathbf{V},\mu,\theta) &= \prod_{j=1}^{q}\left[\prod_{i=1}^{n(j)} k_j(T_{ji},V_{ji},\theta)\mu_j(dV_{ji})\right] \\ &\quad \times \exp\left\{-\int_{\mathcal{V}}\left[\int_{\mathcal{S}} Y_j(s)k_j(s,v,\theta)\tau(ds)\right]\mu_j(dv)\right\}. \end{aligned}$$

Now define for $j=1,\ldots,q$,

$$\beta^*_{j,\theta}(v) = \frac{\beta_{j,\theta}(v)}{[1+\beta_{j,\theta}(v)f_j(v|\theta)]},$$

$$f_j(v|\theta) = \int_{\mathcal{S}} Y_j(s)k_j(s,v,\theta)\tau(ds),$$

$$E_j(\theta) = \exp\left\{-\int_{\mathcal{V}} \ln\left[1+\beta_{j,\theta}(v)f_j(v|\theta)\right]\alpha_{j,\theta}(dv)\right\},$$

and,

$$m(d\mathbf{v}_j|\theta) = \prod_{i=1}^{n(j)} \left(\alpha_{j,\theta} + \sum_{l=1}^{i-1} \delta_{v_{jl}}\right)(dv_{ji}).$$

A characterization of the posterior distribution of $\mu, \theta|\mathbf{T}$, which identifies the marginal conditional distribution of the missing data $\mathbf{V}|\mathbf{T}, \theta$ is now given in Lemmas 3 and 4 and Theorem 2 below. Applying Theorem 1, the following lemma is immediate,

LEMMA 3. *For $j = 1, \ldots, q$, let $g(\mathbf{v}_j, \mu_j)$ be a non-negative or integrable function, then for each $f_j \in BM(\mathcal{V})$,*

$$\int g(\mathbf{v}_j, \mu_j) \exp\left\{-\int f_j(s|\theta)\mu_j(ds)\right\} \pi(d\mathbf{v}_j|\mu_j, \theta)\pi(d\mu_j|\theta)$$
$$= E_j(\theta) \int g(\mathbf{v}_j, \mu_j)\pi(d\mu_j|\mathbf{v}_j, \theta)\pi(d\mathbf{v}_j|\theta),$$

where the posterior distribution of $\mu_j|\mathbf{v}_j, \theta$ is

$$\pi(d\mu_j|\mathbf{v}_j, \theta) = \mathcal{G}(d\mu_j|\alpha_{j,\theta} + \sum_{i=1}^{n(j)} \delta_{v_{ji}}, \beta^*_{j,\theta}),$$

and the marginal measure for $\mathbf{v}_j|\theta$ is,

$$\begin{aligned} \pi(d\mathbf{v}_j|\theta) &= \int_{\mathcal{M}} \prod_{i=1}^{n(j)} \mu_j(dv_{ji}) \mathcal{G}_{\alpha_{j,\theta}, \beta^*_{j,\theta}}(d\mu_j) \\ &= \left[\prod_{i=1}^{n(j)} \beta^*_{j,\theta}(v_{ji})\right] m(d\mathbf{v}_j|\theta). \end{aligned}$$

The joint posterior distribution $\mathbf{V}, \mu|\mathbf{T}, \theta$ is given as follows,

LEMMA 4. *Suppose that $\mathbf{T}|\mathbf{V}, \mu, \theta$ is proportional to $L(\mathbf{V}, \mu, \theta)$ then a joint posterior distribution of $\mathbf{V}, \mu|\mathbf{T}, \theta$ is given as follows,*

$$\pi(d\mathbf{v}, d\mu|\theta, \mathbf{T}) = \prod_{j=1}^{q} \mathcal{G}(d\mu_j|\alpha_{j,\theta} + \sum_{i=1}^{n(j)} \delta_{v_{ji}}, \beta^*_{j,\theta})\pi(d\mathbf{v}_j|\mathbf{T}_j, \theta),$$

where for $j = 1, \ldots, q$,

$$\pi(d\mathbf{v}_j|\mathbf{T}_j, \theta) = \frac{\left[\prod_{i=1}^{n(j)} k^*_j(T_{ji}, v_{ji}, \theta)\right] m(d\mathbf{v}_j|\theta)}{D_j(\mathbf{T}|\theta)}$$

with,

$$D_j(\mathbf{T}|\theta) = \int_{\mathcal{V}^{n(j)}} \left[\prod_{i=1}^{n(j)} k_j^*(T_{ji}, v_{ji}, \theta)\right] m(d\mathbf{v}_j|\theta),$$

and

$$k_j^*(T_{ji}, v_{ji}, \theta) = k_j(T_{ji}, v_{ji}, \theta)\beta_{j,\theta}^*(v_{ji}).$$

PROOF. The result follows from Lemma 3 and the use of a disintegration argument as in Lo and Weng (1989). □

Define,

$$D(\mathbf{T}) = \int_{\mathcal{R}^d} \left[\prod_{j=1}^{q} E_j(\theta) D_j(\mathbf{T}|\theta)\right] \pi(d\theta).$$

A characterization of the joint posterior behaviour of $\mathbf{V}, \mu, \theta|\mathbf{T}$ is given below,

THEOREM 2. *Let $g(\mathbf{v}, \mu, \theta)$ be an arbitrary non-negative or integrable function then a joint posterior of $\mathbf{V}, \mu, \theta$ given $\mathbf{T}$ is characterized by,*

$$\int g(\mathbf{v}, \mu, \theta)\pi(d\mathbf{v}, d\mu, d\theta|\mathbf{T}) = \int \left[\int g(\mathbf{v}, \mu, \theta)\pi(d\mathbf{v}, d\mu|\theta, \mathbf{T})\right] \pi(d\theta|\mathbf{T}),$$

where the joint posterior distribution of $\mathbf{V}, \mu|\mathbf{T}, \theta$ is,

$$\pi(d\mathbf{v}, d\mu|\theta, \mathbf{T}) = \prod_{j=1}^{q} \mathcal{G}(d\mu_j|\alpha_{j,\theta} + \sum_{i=1}^{n(j)} \delta_{v_{ji}}, \beta_{j,\theta}^*)\pi(d\mathbf{v}_j|\mathbf{T}_j, \theta),$$

and

$$\pi(d\theta|\mathbf{T}) = \frac{\left[\prod_{j=1}^{q} E_j(\theta) D_j(\mathbf{T}|\theta)\right] \pi(d\theta)}{D(\mathbf{T})},$$

denotes the posterior density of $\theta|\mathbf{T}$.

PROOF. Lemma 4 provides the form of the posterior conditional on θ and $\mathbf{T}$. The only remaining difficulty involves obtaining the expression for $\pi(d\theta|\mathbf{T})$. This is identified via Bayes rule as,

$$\pi(d\theta|\mathbf{T}) = \frac{\pi(d\mathbf{T}|\theta)\pi(d\theta)}{\int_{\mathcal{R}^d} \pi(d\mathbf{T}|\theta)\pi(d\theta)}$$

where

$$\pi(d\mathbf{T}|\theta) = C \int_{\mathcal{M}^q} L(\mu, \theta)\pi(d\mu|\theta),$$

and C is a constant independent of θ. Now one simply applies Lemma 1 and Theorem 1. □

REMARK 5. Naturally, if each model j depends on a separate subcomponent of θ, say θ_j, and

$$\pi(d\theta) = \prod_{j=1}^{q} \pi(d\theta_j),$$

then the posterior of $\theta|\mathbf{T}$ is,

$$\pi(d\theta|T) = \prod_{j=1}^{q} \pi(d\theta_j|\mathbf{T}_j),$$

with obvious adjustments to the corresponding expressions above. In addition, notice that unlike the case for the distribution $\pi(d\mathbf{v}, d\mu|\theta, \mathbf{T})$ the quantities $\prod_{j=1}^{q} E_j(\theta)$ do not cancel out in the numerator and denominator since they depend on θ.

4.2 *Models depending on a common* μ. In the previous section, the hazard rates (7) are modelled based on q different mixing measures $(\mu_1, \ldots, \mu_q)$. However, there are some instances where this assumption may not be appropriate. Let us consider the model,

$$\begin{aligned} L(\mu, \theta) \quad &= \quad \prod_{j=1}^{q} \left[\prod_{i=1}^{n(j)} \int_{\mathcal{V}} k_j(T_{ji}, v_{i+n(j-1)}, \theta) \mu(dv_{i+n(j-1)}) \right] \\ &\quad \times \exp\left\{ -\int \left[\int_{\mathcal{S}} Y_j(s) k_j(s, v, \theta) \tau(ds) \right] \mu(dv) \right\} \end{aligned}$$

where $n(0) = 0$. Furthermore the missing data is denoted as $\mathbf{v} = \{v_1, \ldots, v_N\}$ where $N = \sum_{j=1}^{q} n(j)$.

This framework includes the Cox regression model discussed in section 1.1. This is seen by setting $q = n$, and $n(j) = 1$ for $j = 1, \ldots, n$. More details are given in section 6. See also Andersen, Borgan, Gill and Keiding (1993), in particular Chapter VII, for possible applications to general semi-parametric regression models. The model could also represent a deconvolution problem where recovery of μ is of primary interest but one only observes partial data through q different types of filters $k_j, j = 1, \ldots, q$. As such, this framework can be applied to positron emission tomography (PET) problems discussed in chapter 3 of Snyder and Miller (1991). Now analogous to the previous section define,

$$\beta^*_\theta(v) = \frac{\beta_\theta(v)}{[1+\beta_\theta(v)f(v|\theta)]}, \tag{8}$$

$$f(v|\theta) = \int_{\mathcal{S}} \left[\sum_{j=1}^{q} Y_j(s)k_j(s,v,\theta)\right] \tau(ds),$$

and

$$E(\theta) = \exp\left\{-\int_{\mathcal{V}} \ln\left[1+\beta_\theta(v)f(v|\theta)\right]\alpha_\theta(dv)\right\}. \tag{9}$$

In addition define,

$$m(d\mathbf{v}|\theta) = \prod_{i=1}^{N}\left(\alpha_\theta + \sum_{l=1}^{i-1}\delta_{v_l}\right)(dv_i),$$

$$D(\mathbf{T}|\theta) = \int_{\mathcal{V}^N}\left[\prod_{j=1}^{q}\prod_{i=1}^{n(j)} k^*_j(T_{ji}, v_{i+n(j-1)}, \theta)\right] m(d\mathbf{v}|\theta),$$

and for $j = 1,\dots,q$,

$$k^*_j(t,v,\theta) = k_j(t,v,\theta)\beta^*(v).$$

The posterior model is now summarized in the following theorem,

THEOREM 3. *Let $g(\mathbf{v},\mu,\theta)$ be a non-negative or integrable function, then the joint posterior distribution of $\mathbf{V},\mu,\theta|\mathbf{T}$ is characterized by*

$$\int g(\mathbf{v},\mu,\theta)\pi(d\mathbf{v},d\mu,d\theta|\mathbf{T}) = \int g(\mathbf{v},\mu,\theta)\mathcal{G}(d\mu|\alpha_\theta + \sum_{i=1}^{N}\delta_{v_i}, \beta^*_\theta)\pi(d\mathbf{v}|\mathbf{T},\theta)\pi(d\theta|\mathbf{T}),$$

where the posterior distribution of $\mathbf{V}|\mathbf{T},\theta$ is,

$$\pi(d\mathbf{v}|\mathbf{T},\theta) = \frac{\left[\prod_{j=1}^{q}\prod_{i=1}^{n(j)} k^*_j(T_{ji}, v_{i+n(j-1)},\theta)\right] m(d\mathbf{v}|\theta)}{D(\mathbf{T}|\theta)}$$

and the posterior density of $\theta|\mathbf{T}$ is,

$$\pi(d\theta|\mathbf{T}) = \frac{D(\mathbf{T}|\theta)E(\theta)\pi(d\theta)}{\int_{\mathcal{R}^d} D(\mathbf{T}|\theta)E(\theta)\pi(d\theta)}.$$

PROOF. The proof follows from arguments similar to the proof of Theorem 2. □

4.3 *Semiparametric Dirichlet process models* It was noted earlier that Lemma 2 implies Lemma 1 of Lo (1984) for the Dirichlet process. In this section a connection between the respective posterior distributions of mixtures of gamma and Dirichlet processes in the semiparametric setting is given. Note that the joint distribution of $\mathbf{V}, \theta$ given $\mathbf{T}$ may also be written as

$$\pi(d\mathbf{v}, d\theta|\mathbf{T}) \propto E(\theta)\left[\prod_{i=1}^{N}\beta_\theta^*(v_i)\right]\nu(d\mathbf{v}|\mathbf{T},\theta)\pi(d\theta)$$

where,

$$\nu(d\mathbf{v}|\mathbf{T},\theta) \propto \left[\prod_{j=1}^{q}\prod_{i=1}^{n(j)} k_j(T_{ji}, v_{i+n(j-1)}, \theta)\right] m(d\mathbf{v}|\theta). \tag{10}$$

The significance of this representation is that the density ν, for fixed θ, is virtually indistinguishable from a density which may have arisen from a Dirichlet process mixture model. Note however that in the gamma process setting (10) allows for σ -finite α. The following corollary follows from an application of Lemma 2 and arguments similar to the derivation of Theorem 3.

COROLLARY 1. *Suppose that P is a probability distribution which given θ is modelled as Dirichlet process with finite shape parameter α_θ and consider the likelihood,*

$$\prod_{j=1}^{q}\left[\prod_{i=1}^{n(j)}\int_{\mathcal{V}} k_j(T_{ji}, v_{i+n(j-1)}, \theta) P(dv_{i+n(j-1)})\right]. \tag{11}$$

Then $\nu(d\mathbf{v}|\mathbf{T},\theta)$ is the posterior distribution of the missing data $\mathbf{V}$ under the above model (11).

In other words, the presence of $E(\theta)$ and β_θ^* constitute the major difference between the posterior distribution of the missing data $\mathbf{V}, \theta$ under a gamma process model as opposed to similar missing data models under a Dirichlet process model. It will be shown how this fact may be used in the context of the proportional hazards model.

5. Partition Representations

The results in Theorems 2 and 3 allow one to compute expressions for Bayes estimation for a variety of functionals of (μ, θ) including as special cases hazard rates and other linear functionals. The partition representations given in Theorem 2 of Lo (1984) and Theorem 4.2 of Lo and Weng (1989) can be used to simplify the complicated resulting multifold integrals. Their results carry over to this setting without any difficulty since these representations are independent of the dimension of $\mathcal{V}$. Here this result is discussed in relation to Theorem 3.

Let $\mathbf{p} = (C_1, \ldots, C_{N(\mathbf{p})})$ denote a partition of size $N(\mathbf{p})$ of the integers $\{1, \ldots, N\}$, let e_l denote the cardinality of each cell C_l for $l = 1, \ldots, N(\mathbf{p})$. A Chinese restaurant partition distribution on the integers $\{1, \ldots, N\}$ with positive scalar parameter ω has the density

$$q(\mathbf{p}|\omega) = \frac{\omega^{N(\mathbf{p})} \prod_{i=1}^{N(\mathbf{p})} (e_i - 1)!}{\prod_{i=1}^{N} (\omega + i - 1)}. \tag{12}$$

This distribution arises in a variety of applications and can be described via a sequential seating of customers, labelled $\{1, \ldots, N\}$, to tables or cells C_l, [see Aldous (1985) and Pitman (1996)]. It will be shown shortly the connection between (12) and the posterior distributions in Theorem 3. As mentioned in the introduction Brunner, Chan, James and Lo (2001) [see also Lo, Brunner and Chan (1996)] discuss a Weighted Chinese Restaurant computational scheme exploiting the partition structure $\mathbf{p}$. An analogue of their Theorem 2 for the models in the previous section is presented. First, because of the complexity of the model additional sub-cells are introduced which do not appear in Brunner, Chan, James and Lo (2001). That is, for $l = 1, \ldots, N(\mathbf{p})$ and $j = 1, \ldots, q$, let C_{jl} denote sub-cells such that

$$C_l = \cup_{j=1}^{q} C_{jl}.$$

where for each fixed j, $C_{jl} \subset \{(j, 1), \ldots, (j, n(j))\}$. The notation $i \in C_{jl}$ is used to mean that i corresponds to a pair (j, i) where $(j, i) \in \{(j, 1), \ldots, (j, n(j))\}$. In addition, define,

$$D(\mathbf{T}|\mathbf{p}, \theta) = \prod_{l=1}^{N(\mathbf{p})} \int_{\mathcal{V}} \left[\prod_{j=1}^{q} \prod_{i \in C_{jl}} k_j^*(T_{ji}, v, \theta) \right] \alpha_\theta(dv).$$

After a bit of accounting a close examination of the arguments in Brunner, Chan, James and Lo (2001) [see also Lo, Brunner and Chan (1996)] combined with Theorem 3 gives the following result, THEOREM 4. *Suppose that* $\mathbf{V}|\mathbf{T},\theta$ *is distributed as* $\pi(d\mathbf{v}|\mathbf{T},\theta)$ *as in Theorem 3, then the conditional distribution of* $\mathbf{V}|\mathbf{T},\theta,\mathbf{p}$ *is such that the sequence* $V_1,\ldots,V_N$ *consists of* $N(\mathbf{p})$ *unique values* $\mathbf{V}^*=(V_1^*,\ldots,V_{N(\mathbf{p})}^*)$ *which are independent with respective distributions,*

$$\pi(dV_l^*|C_l)=\frac{\left[\prod_{j=1}^{q}\prod_{i\in C_{jl}}k_j^*(T_{ji},V_l^*,\theta)\right]\alpha_\theta(dV_l^*)}{\int_{\mathcal{V}}\left[\prod_{j=1}^{q}\prod_{i\in C_{jl}}k_j^*(T_{ji},v,\theta)\right]\alpha_\theta(dv)},$$

for $l=1,\ldots,N(\mathbf{p})$*. In adddition, the posterior distribution* $\mathbf{V},\mu|\mathbf{T},\theta$ *given in Theorem 3 is representable as,*

$$\pi(d\mathbf{v},d\mu|\mathbf{T},\theta)\propto\mathcal{G}(d\mu|\alpha_\theta+\sum_{i=1}^{N(\mathbf{p})}e_i\delta_{v_i^*},\beta^*)\prod_{i=1}^{N(\mathbf{p})}\pi(dV_i^*|C_i)\,W(\mathbf{p}|\theta),$$

where,

$$W(\mathbf{p}|\theta)=\frac{D(T|\mathbf{p},\theta)q(\mathbf{p}|1)}{\sum_{\mathbf{p}}D(T|\mathbf{p},\theta)q(\mathbf{p}|1)}$$

is a posterior distribution of $\mathbf{p}|\mathbf{T},\theta$.

Similar to Theorem 4.2 of Lo and Weng (1989), Theorem 4 implies for instance that the posterior expectation of $\lambda_j|\mathbf{T},\theta$ is

$$\begin{aligned}E[\lambda_j(t|\mu,\theta)|\mathbf{T},\theta] \;=\; &\textstyle\sum_{\mathbf{p}}\Big(\int_{\mathcal{V}}k_j^*(t,v,\theta)\alpha_\theta(dv)\\ &+\textstyle\sum_{i=1}^{N(\mathbf{p})}e_i\int_{\mathcal{V}}k_j^*(t,v,\theta)\,\pi(dv|C_i)\Big)W(\mathbf{p}|\theta).\end{aligned} \quad (13)$$

In addition,

$$\pi(d\theta|\mathbf{T},\mathbf{p})=\frac{D(\mathbf{T}|\mathbf{p},\theta)E(\theta)\pi(d\theta)}{\int_{\mathcal{R}^d}D(\mathbf{T}|\mathbf{p},\theta)E(\theta)\pi(d\theta)}.$$

Note that conditional on $\mathbf{p}$ and $\mathbf{T}$, the distributions of $\mathbf{V}^*$ and θ resemble parametric posterior distributions. This section is closed with an analogous result for $\mathbf{V}|\mathbf{T},\theta$ having distribution ν as in (10).

THEOREM 5. *Suppose that* $\mathbf{V}|\mathbf{T},\theta$ *is distributed as* $\nu(d\mathbf{v}|\mathbf{T},\theta)$*, then the conditional distribution of* $\mathbf{V}|\mathbf{T},\theta,\mathbf{p}$ *is such that the sequence* $V_1,\ldots,V_N$ *consists of* $N(\mathbf{p})$ *unique values* $\mathbf{V}^*=(V_1^*,\ldots,V_{N(\mathbf{p})}^*)$ *which are independent with respective distributions,*

$$\pi_\nu(dV_l^*|C_l)=\frac{\left[\prod_{j=1}^{q}\prod_{i\in C_{jl}}k_j(T_{ji},V_l^*,\theta)\right]\alpha_\theta(dV_l^*)}{\int_{\mathcal{V}}\left[\prod_{j=1}^{q}\prod_{i\in C_{jl}}k_j(T_{ji},v,\theta)\right]\alpha_\theta(dv)},$$

for $l = 1, \ldots, N(\mathbf{p})$*. Furthermore, the posterior distribution of* $\mathbf{p}|\mathbf{T}, \theta$ *is,*

$$W_\nu(\mathbf{p}|\theta) \propto q(\mathbf{p}|1) \prod_{l=1}^{N(\mathbf{p})} \int_{\mathcal{V}} \left[\prod_{j=1}^{q} \prod_{i \in C_{jl}} k_j(T_{ji}, v, \theta) \right] \alpha_\theta(dv).$$

REMARK 6. If one reparametrizes the shape measure α_θ as

$$\alpha_\theta(\cdot) = \omega G_\theta(\cdot)$$

for ω a positive scalar quantity and G_θ a sigma-finite measure, then one replaces $q(\mathbf{p}|1)$ and α by $q(\mathbf{p}|\omega)$ and G_θ respectively in the expressions above. This is similar to the form used in Brunner, Chan, James and Lo (2001).

6. Posterior Analysis for the Cox Regression Model

In this section it is shown how the results in sections 4.2 and 5 give an explicit posterior characterization of the Cox regression model (4). It is also shown how the iidWCR methods discussed in the introduction might be applied to approximate posterior distributional quantities.

The posterior characterization of the Cox regression model denoted as $\pi(d\mathbf{v}, d\mu, d\eta|\mathbf{T})$ can be expressed either via the partition representations given in Theorem 4 or alternatively using Theorem 3. Specifically this is seen by first settiing $q = n$, $N = n$, $\theta = \eta$, $\alpha_\theta(\cdot) = \alpha(\cdot)$ and $n(j) = 1$ for $j = 1, \ldots, n$. Without loss of generality, further set $\alpha(\cdot) = \omega G(\cdot)$ as in Remark 7. It then follows that,

$$f(v|\eta) = \sum_{l=1}^{n} \exp\{\eta^t Z_l\} \int_{[T_l, \infty)} k(u, v) du,$$

and

$$k_j^*(T_j, v, \eta) = \exp\{\eta^t Z_j\} k(T_j, v) \beta_\eta^*(v).$$

The necessary expressions for $\beta_\eta^*(v)$ and $E(\eta)$ can be deduced from (8) and (9).

The above quantities placed in Theorem 3 or 4 give an explicit characterization of the Bayesian Cox models presented in section 1.1. Now it is shown how special features of the Cox model, (4), allow one to simplify matters further. Notice that since

$$k_j(T_j|v, \eta) = k(T_j, v) \exp\{\eta^t Z_j\}, \tag{14}$$

it follows that

$$\pi_\nu(dV_l^*|C_l) \propto \left[\prod_{j\in C_l} k(T_j, V_l^*)\right] \alpha(dV_l^*). \tag{15}$$

In addition, $W_\nu(\mathbf{p})$ is defined as in Theorem 5 with $k_j^*(T_j, v, \theta)$ replaced by $k(T_j, v)$ as,

$$W_\nu(\mathbf{p}) \propto q(\mathbf{p}|\omega) D_\nu(\mathbf{T}|\mathbf{p})$$

where

$$D_\nu(\mathbf{T}|\mathbf{p}) = \prod_{j=1}^{n(\mathbf{p})} \int_{\mathcal{V}} \left[\prod_{i\in C_j} k(T_i, v)\right] G(dv).$$

Note importantly that π_ν and W_ν do not depend on η. Now define a distribution for η as,

$$\pi_\nu(d\eta|\mathbf{T}, \mathbf{p}) \propto \pi(d\eta) \prod_{l=1}^{n(\mathbf{p})} \prod_{j\in C_l} \exp\{\eta^t Z_j\}, \tag{16}$$

and let,

$$h(\mathbf{V}^*, \eta) = E(\eta) \prod_{l=1}^{n(\mathbf{p})} [\beta_\eta^*(V_l^*)]^{e_l}.$$

Notice that although the distribution in (16) is written to denote dependence on $\mathbf{p}$ it is actually invariant under all choices of $\mathbf{p}$. In addition, $h(\mathbf{V}^*, \eta)$ is equivalent to the Laplace functional of $\mathcal{G}(d\mu|\alpha + \sum_{i=1}^{n(\mathbf{p})} e_i \delta_{v_i^*}, \beta)$ evaluated at $f(\cdot|\eta)$, multiplied by

$$\prod_{l=1}^{n(\mathbf{p})} [\beta(V_l^*)]^{e_l}. \tag{17}$$

Now due to (14), an application of Theorem 3 and Theorem 5 yield the following result:

Corollary 2. *The posterior distribution $\pi(d\mathbf{v}, d\mu, d\eta|\mathbf{T})$, may be written as*

$$\pi(d\mathbf{v}, d\mu, d\eta|\mathbf{T}) \propto \pi_\nu(d\mathbf{v}, d\mu, d\eta|\mathbf{T}, \mathbf{p}) W_\nu(\mathbf{p}),$$

where,

$$\pi_\nu(d\mathbf{v}, d\mu, d\eta|\mathbf{T}, \mathbf{p}) \propto \mathcal{G}(d\mu|\alpha + \sum_{i=1}^{n(\mathbf{p})} e_i \delta_{v_i^*}, \beta^*) h(\mathbf{V}^*, \eta) \pi_\nu(d\mathbf{V}^*|\mathbf{p}) \pi_\nu(d\eta|\mathbf{T}, \mathbf{p})$$

and,

$$\pi_\nu(d\mathbf{V}^*|\mathbf{p}) = \prod_{l=1}^{n(\mathbf{p})} \pi_\nu(dV_l^*|C_l).$$

Let E_ν denote expectation with respect to $\pi_\nu(d\mathbf{v}, d\mu, d\eta|\mathbf{T}, \mathbf{p})$. Now let $E[g(\mathbf{V}, \mu, \eta)|T]$ denote the posterior expectations of integrable functionals $g(\mathbf{V}, \mu, \eta)$ taken with respect to $\pi(d\mathbf{v}, d\mu, d\eta|\mathbf{T})$. Corollary 2 implies the following equivalance;

$$E[g(\mathbf{V}, \mu, \eta)|\mathbf{T}] = \sum_{\mathbf{p}} s(\mathbf{p}) W_\nu(\mathbf{p}) \tag{18}$$

for the choice

$$s(\mathbf{p}) = E_\nu[g(\mathbf{V}, \mu, \eta)|\mathbf{T}, \mathbf{p})].$$

Corollary 2 is tailor-made to take advantage of the stucture (14) in the Cox model. It may be seen as a variation of the change of measure idea given in Brunner, Chan, James and Lo (2001). The significance of corollary 2, in terms of computational approximations, is that in many applications $\pi_\nu(d\mathbf{V}^*|\mathbf{p})$ and $\pi_\nu(d\eta|\mathbf{T}, \mathbf{p})$ will be easy to draw from. At the same time these models incorporate much of the information given in the true posterior. Oftentimes $\pi_\nu(d\mathbf{V}^*|\mathbf{p})$ will be based on conjugate models. In particular, the important models based on kernels, (1) and (2), are made conjugate by the choices of $\alpha(\cdot)$ (or $\beta(\cdot)\alpha(\cdot)$) proportional to a Pareto and gamma distribution respectively. The model $\pi_\nu(d\eta|\mathbf{T}, \mathbf{p})$ coincides with a parametric Bayesian model. The main task then reduces to obtaining draws of $\mathbf{p}$ from a distribution close to the $W_\nu(\mathbf{p})$. This may be accomplished by applying directly the (nonparametric) iidWCR mentioned in the introduction.

Now for clarity some additional notation is introduced. For $r > 1$, let $\mathbf{p}_r = \{C_{r,1}, \ldots, C_{r,n(\mathbf{p}_r)}\}$ denote a partition of $\{1, 2 \ldots, r\}$, where $C_{r,i}$ denotes the current configuration of table i after r customers have been seated and $e_{r,i}$ denotes the number of customers seated at $C_{r,i}$. The partition $\mathbf{p}_{r+1}$ then denotes the (updated) one step larger partition on $\{1, 2, \ldots, r+1\}$. Also, $\pi_\nu(du|C_{r,i})$ is defined as in (15) but is now based on $\{T_j : j \in C_{r,i}\}$ for $i = 1, \ldots, n(\mathbf{p}_r)$. Further, define

$$l(r) = \frac{\omega}{\omega + r} \int_{\mathcal{R}} k(T_{r+1}, v)\, G(dv) + \sum_{i=1}^{n(\mathbf{p}_r)} \frac{e_{r,i}}{\omega + r} \int_{\mathcal{R}} k(T_{r+1}, v)\, \pi_\nu(dv|C_{r,i}),$$

where in particular

$$l(0) = \frac{\omega}{\omega} \int_{\mathcal{R}} k(T_1, v) G(dv) = \int_{\mathcal{R}} k(T_1, v) G(dv).$$

One can now use the iidWCR procedure in Brunner, Chan, James and Lo (2001), to produce iid draws of $\mathbf{p}$, based on the following sequential "seating" algorithm:

6.1 *WCR seating rule.*

STEP 1: Seat the first customer to a table with probability

$$\frac{l(0)}{l(0)} = 1.$$

STEP $(r+1)$: Given $\mathbf{p}_r$, customer $r+1$ sits at table $C_{r,i}$ with probability

$$\Pr(\mathbf{p}_{r+1}|\mathbf{p}_r) = \frac{e_{r,i}}{\omega + r} \times \frac{\int_{\mathcal{R}} k(T_{r+1}, v)\pi_\nu(dv|C_{r,i})}{l(r)},$$

where $\mathbf{p}_{r+1} = \mathbf{p}_r \cup \{r+1 \in C_{r,i}\}$ for $i = 1, \ldots, n(\mathbf{p}_r)$. Otherwise, customer $r+1$ sits at a new table with probability

$$\Pr(\mathbf{p}_{r+1}|\mathbf{p}_r) = \frac{\omega}{\omega + r} \times \frac{\int_{\mathcal{R}} k(T_{r+1}, v)G(dv)}{l(r)}.$$

The completion of Step n produces a $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\} = \mathbf{p}_n$, where now $\mathbf{p}$ is drawn from the WCR density $q_\nu(\mathbf{p}|\omega, k)$ which satisfies,

$$L(\mathbf{p})q_\nu(\mathbf{p}|\omega, k) = q(\mathbf{p}|\omega)D_\nu(\mathbf{T}|\mathbf{p}),$$

where

$$L(\mathbf{p}) = \prod_{r=1}^{n} l(r-1).$$

That is,

$$W_\nu(\mathbf{p}) \propto L(\mathbf{p})q_\nu(\mathbf{p}|\omega, k). \tag{19}$$

This fact implies that for any integrable function $s(\mathbf{p})$,

$$\sum_{\mathbf{p}} s(\mathbf{p})W_\nu(\mathbf{p}) = \frac{\sum_{\mathbf{p}} s(\mathbf{p})L(\mathbf{p})q_\nu(\mathbf{p}|\omega, k)}{\sum_{\mathbf{p}} L(\mathbf{p})q_\nu(\mathbf{p}|\omega, k)}. \tag{20}$$

Using corollary 2, (18), (19) and the iidWCR, one can now use the following iid procedure to approximate posterior distributional properties of the Cox model:

6.2 *an iidWCR algorithm for Cox regression model.*

1. Draw B iid values $\mathbf{p}^{(b)}$ from $q_\nu(\mathbf{p}|\omega, k)$ and calculate $L(\mathbf{p}^{(b)})$.
2. Given $\mathbf{p}^{(b)}$, draw $\mathbf{V}^{*(b)}$ from $\pi_\nu(d\mathbf{V}^*|\mathbf{p}^{(b)})$.
3. Draw $\eta^{(b)}$ from $\pi_\nu(d\eta|\mathbf{T}, \mathbf{p}^{(b)})$

Now for instance to approximate $E[g(\eta)|\mathbf{T}]$ use,

$$\frac{\sum_{b=1}^{B} g(\eta^{(b)}) h(\mathbf{V}^{*(b)}, \eta^{(b)}) L(\mathbf{p}^{(b)})}{\sum_{b=1}^{B} h(\mathbf{V}^{*(b)}, \eta^{(b)}) L(\mathbf{p}^{(b)})}.$$

The choice $g(\eta) = I\{\eta \in C\}$ corresponds to the posterior probabilty that the regression coefficients are in a region C. From (13), the baseline hazard estimate $E[\lambda(t|\mu)|\mathbf{T}]$ can be approximated by

$$\frac{\sum_{b=1}^{B} g(\mathbf{V}^{*(b)}, \eta^{(b)}, \mathbf{p}^{(b)}) h(\mathbf{V}^{*(b)}, \eta^{(b)}) L(\mathbf{p}^{(b)})}{\sum_{b=1}^{B} h(\mathbf{V}^{*(b)}, \eta^{(b)}) L(\mathbf{p}^{(b)})},$$

where

$$g(\mathbf{V}^*, \eta, \mathbf{p}) = \omega \int_{\mathcal{R}} k(t, v) \beta_\eta^*(v) G(dv) + \sum_{i=1}^{n(\mathbf{p})} e_i k(t, V_i^*) \beta_\eta^*(V_i^*).$$

Approximations for products of general linear functions of μ may be carried out in a similar fashion. Ishwaran and James (2000), provide a general discusion on how to additionally approximate the posterior distribution of μ and related functionals. Wolpert and Ickstadt (1998a, b) also discuss a method to obtain draws from a weighted gamma distribution which can be applied here as well. The above algorithm can be easily modified in the special cases where direct incorporation of β and $E(\eta)$ pose no extra difficulties. For instance if $(\beta(\cdot)\alpha(\cdot))$ is conjugate then one multiplies $\pi_\nu(d\mathbf{V}^*|\mathbf{p})$ by (17), in step 2, and adjusts q_ν and h accordingly.

Remark 7. Note that because of (14), iidWCR approximations for the Cox regression model can be obtained by using the nonparametric methods in Brunner, Chan, James and Lo (2001). In general, the methods in Ishwaran and James (2000) are more directly applicable for semiparametric models. For instance, models based on kernels such as (5) and (6). Clearly other computational procedures which take advantage of the explicit posterior characterizations may be used here as well.

REMARK 8. In general it is assumed that the kernels generates a smooth hazard. For Poisson process models the likelihood $L(\mu, \theta)$, and hence the calculus, is valid for discrete choices of k.

ACKNOWLEDGEMENTS. I would like to thank the referees for their valuable suggestions which helped to improve the presentation of this work. Thanks also to Hemant Ishwaran for helpful comments.

References

AALEN, O.O. (1975). *Statistical inference for a family of counting processes.* Ph.D. Thesis, University of California, Berkeley.

AALEN, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 535-545.

ALDOUS, D.J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XII*, P.L. Hennequin, ed., Springer Lecture Notes in Mathematics, Vol. 1117, Springer-Verlag, Berlin.

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

ANTONIAK, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152-1174.

BADDELEY, A.J. and TURNER, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. N. Z. J. Statist.* **42**, 283-322.

BESAG, J. (1977). Some methods of of statistical analysis for spatial data. *Bull. Internat. Statist. Inst.* **47**, 77-92.

BEST, N., ICKSTADT, K., and WOLPERT, R.L. (2000). Spatial poisson regression for health and exposure data measured at disparate spatical scales. *J. Amer. Statist. Assoc.* **95**, 1076-1088.

BLACKWELL, D. and MACQUEEN, J.B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353-355.

BRUNNER, L.J., CHAN, A.T., JAMES, L.F. and LO, A.Y. (2001). Weighted Chinese restaurant processes and Bayesian mixture models. Preprint.

CLAYTON, D.G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, **47**, 467-485.

COX, D.R. (1972). Regression models and life tables (with discussion). *J.R. Statist. Soc. Ser. B* **34**, 187-220.

DALEY, D.J. and VERE-JONES, D. (1988). *An introduction to the theory of point processes.* Springer-Verlag, New York.

DAMIEN, P, LAUD, P.W., and SMITH, A.F.M. (1996). Monte Carlo methods for approximating a posterior hazard rate process. *Statist. Comput.* **6**, 77-83.

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

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

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

GASBARRA, D. and KARIA, S.R. (2000). Analysis of competing risks by using Bayesian smoothing. *Scand. J. Statist.* **27**, 605-617.

GELFAND, A. E. (1999). Approaches for semiparametric Bayesian regression. In *Asymptotics, Nonparametric, and Time Series: A tribute to Madan Lal Puri*, S. Ghosh, ed., Marcel Dekker, New York, 615-638.

HAYAKAWA, Y., PAUL, S. and VIGNAUX, G.A. (2000). Testing failure data for evidence of aging via a Bayesian nonparametric method. In *Proceedings of Second International Conference on Mathematical Methods in Reliability* Bordeaux, France, 509-512.

HAYAKAWA, Y., ZUKERMAN, J, PAUL, S. and VIGNAUX, G.A. (2001). Bayesian nonparametric testing of constant versus nondecreasing hazard rates . In *System and Bayesian Reliability: Essays in Honor of Professor Richard E. Barlow on his 70th birthday*, Y. Hayakawa, T. Irony and M. Xie, eds., World Scientific, River Edge, NJ, 391-406.

HO, M.-W. and LO A.Y. (2001). Bayesian nonparametric estimation of a monotone hazard rate . In *System and Bayesian Reliability: Essays in Honor of Professor Richard E. Barlow on his 70th birthday*, Y. Hayakawa, T. Irony and M. Xie, eds., World Scientific, River Edge, NJ, 301-314.

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

HUANG, J. and WELLNER, J.A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Statist.* **22**, 3-33.

IBRAHIM, J.G. and SINHA, D. (1998). Prior elicitation for semiparametric Bayesian survival analysis. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer Lecture Notes in Statistics, D. Dey, P. Mueller and D. Sinha, eds., Springer-Verlag, New York, 273-292.

IBRAHIM, J.G., CHEN, M-H. and MACEACHERN, S.N. (1999). Bayesian variable selection for proportional hazards model. *Canad. J. Statist.* **27**, 701-717.

ISHWARAN, H. (1998). Exponential posterior consistency via generalized Pólya urn schemes in finite semiparametric mixture models. *Ann. Statist.* **26**, 2157-2178.

ISHWARAN, H. and JAMES, L.F. (2000). Bayesian nonparametric methods for hazard mixture models using weighted gamma process approximations. Preprint.

JEWELL, N.I. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10**, 479-484.

KALBFLEISCH, J.D. (1978). Nonparametric Bayesian analysis of survival time data. *J.R. Statist. Soc. Ser. B* **40**, 214-221.

KINGMAN, J.F.C. (1975). Random discrete distributions. *J.R. Statist. Soc. Ser. B* **37**, 1-22.

KUO, L. (1986). Computations of mixtures of Dirichlet processes. *SIAM J. Sci. Statist. Comput.* **7**, 60-71.

LAUD, P.W., DAMIEN, P. and SMITH, A.F.M. (1998). Bayesian nonparametric and covariate analysis of failure time data. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer Lecture Notes in Statistics, D. Dey, P. Mueller and D. Sinha, eds., Springer-Verlag, New York, 213-226.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

LINDSAY, B.G. (1986). Exponential family mixture models (with least squares estimators). *Ann. Statist.* **14**, 124-137.

LINDSAY, B.G. (1995). *Mixture Models: Theory, Geometry and Applications. NSF-CBMS Regional Conference Series in Probability and Statistics Vol. 5.* Institute of Mathematical Statistics, Hayward, California.

LO, A.Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* , **59**, 55-66.

LO, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12**, 351-357.

LO, A.Y. (1992). Bayesian inference for Poisson process models with censored data. *J. Nonparametr. Statist.* **2**, 71-80.

LO, A.Y. (1993). A Bayesian method for weighted sampling. *Ann. Statist.* **21**, 2138-2148.

LO, A.Y. and WENG, C.S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rates estimates. *Ann. Inst. Statist. Math.* **41**, 227-245.

LO, A.Y., BRUNNER, L.J. and CHAN, A.T. (1996). Weighted Chinese restaurant processes and Bayesian mixture model. Research Report, Hong Kong University of Science and Technology.

MACEACHERN, S.N., CLYDE, M. and LIU, J.S. (1999). Sequential importance sampling for nonparametric Bayes models: the next generation. *Canad. J. Statist.* **27**, 251-267.

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, eds., IMS Lecture Notes-Monograph series, Vol 30., Institute of Mathematical Statistics, Hayword, CA, 245-267.

SAUNDERS, S.C. and MYHRE, J.M. (1983). Maximum likelihood estimation for two-parameter decreasing hazard rate distributions using censored data. *J. Amer. Statist. Assoc.* **78**, 674-683.

SINHA, D. and DEY, D. (1997). Semiparametric Bayesian analysis of survival data .*J. Amer. Statist. Assoc.* **92**, 1195-1212.

SINHA, D. and DEY, D. (1998). Survival analysis using semiparametric Bayesian methods. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer Lecture Notes, D. Dey, P. Mueller and D. Sinha, eds., Springer-Verlag, New York, 195-212.

SNYDER D.L. and MILLER, M.I. (1991). *Random Point Processes in Time and Space* (Second Edition). Springer-Verlag, New York.

TSILEVICH, N.V., VERSHIK, A.M. and YOR, M. (2001). An infinite-dimensional analogue of the Lebesque measure and distinguished properties of the gamma process. *J. Funct. Anal.* **185**, 274-296.

TSILEVICH, N.V., VERSHIK, A.M. and YOR, M. (2000). Distingusihed properties of the gamma process and related topics, Prépublication du Laboratoire de Probabilités et Modèles Aléatoires no. 575, Mars 2000.

WALKER, S.G., DAMIEN, P., LAUD, P.W. and SMITH, A.F.M. (1999). Bayesian nonparametric inference for random distributions and related functions. With discussion and reply by the authors. *J.R. Statist. Soc. Ser. B* **61**, 485-527.

Wolpert, R.L. and Ickstadt, K.(1988a). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251-267.

Wolpert, R.L. and Ickstadt K. (1998b). Simulation of Levy random fields. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer Lecture Notes in Statistics, D. Dey, P. Mueller and D. Sinha, eds., Springer, New York, 227-241.

Lancelot F. James,
Department of Information Systems and Management,
The Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon
Hong Kong
E-mail: lancelot@ust.hk