Computational Methods for Multiplicative Intensity Models Using Weighted Gamma Processes: Proportional Hazards, Marked Point Processes, and Panel Count Data

Hemant ISHWARAN and Lancelot F. JAMES

We develop computational procedures for a class of Bayesian nonparametric and semiparametric multiplicative intensity models incorporating kernel mixtures of spatial weighted gamma measures. A key feature of our approach is that explicit expressions for posterior distributions of these models share many common structural features with the posterior distributions of Bayesian hierarchical models using the Dirichlet process. Using this fact, along with an approximation for the weighted gamma process, we show that with some care, one can adapt efficient algorithms used for the Dirichlet process to this setting. We discuss blocked Gibbs sampling procedures and Pólya urn Gibbs samplers. We illustrate our methods with applications to proportional hazard models, Poisson spatial regression models, recurrent events, and panel count data.

KEY WORDS: Blocked Gibbs sampler; Dirichlet process; Hazard function; Intensity; Kernel; Nonhomogeneous Poisson process; Pólya urn Gibbs sampler; Recurrent events; Spatially correlated counts.

1. INTRODUCTION

Aalen (1975, 1978) developed a unified theory for nonparametric inference in multiplicative intensity models from a frequentist perspective. This treatment included, for example, the life-testing model, the multiple decrement model, birth and death processes, and branching processes. A Bayesian treatment for the real line was given by Lo and Weng (1989), who modeled hazard rates in the multiplicative intensity model as mixtures of a known kernel k with a finite measure μ modeled as a weighted gamma measure on the real line. That is, a hazard r is modeled as

$$r(x|\mu) = \int_{\Re} k(x, \nu)\mu(d\nu). \tag{1}$$

Lo and Weng (1989) showed that using arbitrary kernels provides the user with a great deal of flexibility; for instance, the choice of the kernel $k(x, \nu) = I\{\nu \leq x\}$ gives monotone-increasing hazards as considered by Dykstra and Laud (1981), whereas kernels $k(x, \nu) = I\{|x - a| \geq \nu\}$ and $k(x, \nu) = I\{|x - a| \leq \nu\}$ give U-shaped hazards (with minimum and maximum at a) similar to those of Glaser (1980), and normal density kernels $k(x, \nu) = \exp(-.5(x - \nu)^2/\tau^2)/\sqrt{2\pi\tau^2}$ can be used to estimate hazards without shape restriction.

In this article we develop a general approach to Bayesian inference for hazard (intensity) rates in nonparametric and semiparametric multiplicative intensity models by incorporating kernel mixtures of spatial weighted gamma process priors. This approach extends the work of Lo and Weng (1989) and Dykstra and Laud (1981) from a nonparametric setting on the real line to the nonparametric and semiparametric settings over general spaces and applies to the nonparametric multiplicative intensity models considered by Aalen (1975, 1978) and their semiparametric extensions developed by Andersen, Borgan, Gill, and Keiding (1993, chap. III). Models that fall within this framework that have been considered using gamma and weighted gamma processes from a Bayesian perspective include Markov models used in survival analysis subject to certain types of censoring, filtering, and truncation (Arjas and Gasbarra 1994; Laud, Smith, and Damien 1996; Gasbarra and Karia 2000), as well as Poisson point process models used in reliability (Kuo and Ghosh 1997), forest ecology (Wolpert and Ickstadt 1998a), and health exposure analysis (Best, Ickstadt, and Wolpert 2000). Another related application was given by Ibrahim, Chen, and MacEachern (1999) who used a weighted gamma process to select variables in Cox proportional hazards models.

A major contribution of this article is to develop a unified computational treatment of these problems from a Bayesian perspective. As was shown by Lo and Weng (1989) (see also Lo, Brunner, and Chan 1996) the posterior for multiplicative intensity models under weighted gamma processes share common structural features with posterior distributions for models subject to the Dirichlet process (i.e., Dirichlet process mixture models). (See Lo 1984 for background and posterior descriptions of Dirichlet process mixture models.) Recently, James (2003) extended these results to an abstract semiparametric setting (see Sec. 3), thus providing explicit calculus for relating posteriors for spatial semiparametric intensity models to posteriors for Dirichlet process mixture models. This equivalence, in combination with our use of a weighted gamma process approximation (Sec. 3), allows us to use efficient Dirichlet process computational procedures and to approximate the laws for general functionals of interest. An important aspect is that we avoid ad hoc methods used to approximate likelihoods. For example, in survival analysis problems we do not discretize time, as is often done to simplify computations. Another nice benefit of using Dirichlet process methods is that they are well understood and have a rich literature that can be drawn on. For example, the number of iterations and burn-in iterations suggested for Gibbs sampling Dirichlet process mixture models, and other such practical experience,

Hemant Ishwaran is Associate Staff, Department of Biostatistics and Epidemiology Wb4, Cleveland Clinic Foundation, Cleveland, Ohio, OH 44195 (E-mail: *ishwaran@bio.ri.ccf.org*). Lancelot F. James is a Visiting Scholar, Department of Information Systems and Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (E-mail: *lancelot@ust.hk*). The authors thank the referees and the associate editor for their many helpful comments.

Journal of the American Statistical Association
March 2004, Vol. 99, No. 465, Theory and Methods
DOI 10.1198/016214504000000179

can all be applied here. Moreover, computational strategies that have been developed to handle problems such as nonconjugacy and slow Markov chain mixing all have natural analogs in our setting. Handling nonconjugacy is especially relevant, because it will allow us to deal with complicated terms that can appear in our posteriors. Slow mixing is also an important consideration. Here we deal with it by adapting acceleration steps used to improve mixing in Dirichlet process problems. Computational methods that we consider include Pólya urn Gibbs samplers (Escobar 1988, 1994; Escobar and West 1995) and blocked Gibbs sampling methods (Ishwaran and Zarepour 2000; Ishwaran and James 2001).

The article is organized as follows. Section 2 presents the multiplicative intensity likelihood and defines the weighted gamma process. Applications to proportional hazards, nonhomogeneous Poisson process and marked processes used in spatial regression models are discussed. Sections 3 and 4 present the details of our Pólya urn Gibbs sampler and blocked Gibbs sampler. The methods are illustrated by applications to a long-term follow-up study of heart rate recovery and mortality (Sec. 3) and to simulations from a nonhomogeneous Poisson spatial process (Sec. 4). Section 2 also discusses applications to recurrent event data arising from conditionally independent nonhomogeneous Poisson processes. Section 5 illustrates this extension to the panel count data problem (Kalbfleish and Lawless 1985; Sun and Kalbfleish 1995), a generalization of the interval-censoring problem. Using a Poisson process model discussed by Wellner and Zhang (2000), we show how one can obtain smoothed kernel estimates for the intensity, thus providing a novel Bayesian nonparametric approach complementary to the iterative convex minorant algorithm used by frequentists (Wellner and Zhang 2000).

2. SEMIPARAMETRIC MULTIPLICATIVE INTENSITY LIKELIHOODS

Throughout, we work with a multiplicative intensity likelihood of the form

$$\begin{aligned}\mathcal{L}(\mu,\boldsymbol{\theta}) &= \exp\left\{-\sum_{i=1}^{n+m}\int_{\mathcal{S}}\left[\int_{\mathcal{X}} Y_i(x)k_i(x,\nu,\boldsymbol{\theta})\eta(dx)\right]\mu(d\nu)\right\}\\ &\quad\times\prod_{i=1}^{n}\int_{\mathcal{S}} k_i(X_i,\nu_i,\boldsymbol{\theta})\mu(d\nu_i), \qquad (2)\end{aligned}$$

where μ is a finite measure over a measurable space $(\mathcal{S},\mathcal{A})$, the value $\boldsymbol{\theta}$ is a Euclidean parameter with parameter space Θ (in the applications considered here, $\Theta=\Re^d$), and $\mathcal{X}$ is the sample space for the data $X_1,\ldots,X_{n+m}$. (In right-censoring survival analysis problems, n and m denote the sample sizes for the failure and censored times; however, the values for n and m vary depending on the specific application.) The functions $k_i(x,\nu,\boldsymbol{\theta})$ in (2) are known nonnegative kernels that are jointly measurable in $(x,\nu,\boldsymbol{\theta})$ and integrable with respect to μ and η, where η is some fixed σ-finite measure. Their role will be to smooth the unknown hazard (intensity) function via a mixing approach similar to (1).

Censoring and more general types of filtration are captured by the functions $Y_i(x)$ appearing in (2), which in the counting process literature are usually called *predictable functions*. In many applications in event history analysis, $Y_i(x)$ records whether an individual is still at risk just before time x. A key point is that using predictable functions allows the likelihood (2) to retain the same structure under different types of filtration, thus presenting a unified approach for studying the multiplicative intensity model (see Andersen et al. 1993, chap. III). In right censoring (see Sec. 2.2), $X_1,\ldots,X_n$ denote observed failure times and $X_{n+1},\ldots,X_{n+m}$ denote censored times. In the case of Poisson spatial processes, considered in Section 2.3, all observations are observed, so that $m=0$. Setting $Y_i(x)=1/n$ yields a likelihood (2) for a nonhomogeneous Poisson spatial process based on a single realization. More complex marked processes are also possible, as discussed in Section 2.3. Our methods can also be extended to more general likelihoods that arise from conditionally independent nonhomogeneous Poisson processes. We motivate this idea briefly in Section 2.4 for recurrent event data, and in Section 5 we illustrate the approach in depth for panel count data.

2.1 Missing Spatial Data and Weighted Gamma Processes

Let $\mathbf{v}=(\nu_1,\ldots,\nu_n)$. Note that the integrals on the right side of (2) index ν by a subscript of i, even though this is notationally redundant. This is done to emphasize that $\mathbf{v}$ is thought of as missing spatial data. The idea is to augment the parameter space and the likelihood function (2) to include these missing data. Placing a prior on the parameters $\boldsymbol{\theta}$, $\mathbf{v}$, and μ induces a posterior, which we can then use to compute various quantities, including estimates for the smoothed intensity function. The key is the prior for $(\mathbf{v},\mu)$, which is assumed to have a joint product measure such that $\nu_1,\ldots,\nu_n$, given μ, represent n conditionally independent realizations from the random finite measure μ, where μ has a weighted gamma process law. Let $\pi(d\boldsymbol{\theta})$ denote our prior for $\boldsymbol{\theta}$. We assume a prior on $(\boldsymbol{\theta},\mathbf{v},\mu)$ with the joint product measure

$$\pi(d\boldsymbol{\theta})\prod_{i=1}^{n}\mu(d\nu_i)\mathcal{G}(d\mu|\alpha,\beta).$$

The expression $\mathcal{G}(\cdot|\alpha,\beta)$ denotes a weighted gamma process law with shape parameter α (a finite measure over $\mathcal{S}$) and scale parameter β (a positive integrable function over $\mathcal{S}$). That is, for each Borel-measurable set $A\in\mathcal{A}$, the random measure μ, defined by

$$\mu(A)=\int_A \beta(s)\gamma_\alpha(ds),$$

is said to have a $\mathcal{G}(\cdot|\alpha,\beta)$ law, where γ_α is a gamma process over $\mathcal{S}$ with shape measure α. We call γ_α a gamma process with shape α if $\gamma_\alpha(A)$ is a gamma$(\alpha(A))$ random variable with mean $\alpha(A)$ and variance $\alpha(A)$.

Remark 1. The gamma process was described in an early paper by Moran (1956). A more complete description was given by Ferguson and Klass (1972), who discussed gamma processes over $\Re$, and Kingman (1975), who provided a description over arbitrary measurable spaces. Lo (1982) described weighted gamma processes over Polish spaces, and Dykstra and Laud (1981) provided a description over $\Re$.

It is important to keep in mind that μ is an arbitrary finite measure and thus is not necessarily a probability measure. Because of this, what the resulting posterior for (2) will look like is not at all obvious. For example, it is not even clear whether the posterior values for $\nu_1, \dots, \nu_n$ can be associated with a proper probability distribution, which is something we would need if we wanted to use Monte Carlo methods to approximate posterior quantities related to $\mathbf{v}$. In familiar Dirichlet process mixture models, these kinds of connections follow automatically because we always work with proper probability measures. It might be guessed, though, that these same features apply here, due to the intimate connection between the gamma process and the Dirichlet process. It is well known that if $\mathcal{P}(\cdot|\alpha)$ is a Dirichlet process law with a finite measure parameter α (Ferguson 1973, 1974), then $\mathcal{P}(\cdot|\alpha)$ can be expressed as a normalized gamma process, that is, the random probability measure

$$P(\cdot) = \frac{\gamma_\alpha(\cdot)}{\gamma_\alpha(\mathcal{S})} \tag{3}$$

has a $\mathcal{P}(\cdot|\alpha)$ law. Thus it stands to reason that there must be an intimate connection between Dirichlet process mixture models and multiplicative intensity posteriors derived from weighted gamma process priors. This is, of course, only an informal argument. The exact details are more subtle, as we spell out specifically in Section 3. Before going into these details, however, we provide some motivating examples.

2.2 Proportional Hazards

The Cox regression model (Cox 1972) is an important example of a multiplicative intensity model of the form (2). Under independent right censoring, the Cox proportional hazards likelihood may be written as

$$\prod_{i=1}^{n+m} \big(r_0(T_i|\mu)\exp(\boldsymbol{\theta}^T\mathbf{Z}_i)\big)^{\Delta_i} \times \exp\left\{-\int_{\Re} Y_i(t) r_0(t|\mu)\exp(\boldsymbol{\theta}^T\mathbf{Z}_i)\eta(dt)\right\}, \tag{4}$$

where T_i are failure times, $\Delta_i = I\{T_i \le C_i\}$ is a 0–1 indicator function indicating whether censoring occurred at times C_i, the $\mathbf{Z}_i$'s are covariate vectors with parameter vector $\boldsymbol{\theta}$, and $Y_i(t) = I\{T_i \ge t\}$ is the predictable function. To produce a smoothed estimate for the hazard, we model the unknown baseline hazard function $r_0(t|\mu)$ as a mixture of the form

$$r_0(t|\mu) = \int_{\mathcal{S}} k_0(t,\nu)\mu(d\nu), \tag{5}$$

where $k_0(t,\nu)$ is some prespecified kernel. Typically, $\mathcal{S} \subseteq \Re^+$, with the exact choice depending on the selected kernel and the specific problem. Now to express (4) in the form of (2), let $X_i = T_i$, for $i = 1, \dots, n+m$, and set $\mathcal{X} = \Re$. If $T_1, \dots, T_n$ are the uncensored observed failure times, then (4) can be expressed in the form of (2) if

$$k_i(t,\nu,\boldsymbol{\theta}) = \exp(\boldsymbol{\theta}^T\mathbf{Z}_i)k_0(t,\nu). \tag{6}$$

Observe that the underlying hazard is modeled so that it is a kernel mixture,

$$r(t|\mathbf{Z},\boldsymbol{\theta},\mu) = \exp(\boldsymbol{\theta}^T\mathbf{Z})\int_{\mathcal{S}} k_0(t,\nu)\mu(d\nu).$$

Kalbfleisch (1978) presented one of the earliest Bayesian approaches for estimating the Cox model based on a gamma process. In this method the baseline cumulative hazard function is modeled as a gamma process. Time is then discretized by chopping the time axis into a collection of nonoverlapping intervals, which, due to the use of a gamma process, results in a cumulative hazard function that can then be modeled as a collection of independent gamma random variables (see also Burridge 1981). In more recent work, Ibrahim et al. (1999) took a different approach and modeled the baseline hazard function as a mixture (5), where μ has a weighted gamma process prior. This method applies to kernels of the form

$$k_0(t,\nu) = I\{\nu \le t\} \tag{7}$$

and leads to smoothed increasing hazard functions similar to those of Dykstra and Laud (1981). To sample the posterior, Ibrahim et al. (1999) constructed a refined partition of the time axis and worked with the resulting approximate likelihood of (4) (see also Laud et al. 1996).

However, in Section 3 we take a different approach, showing, by exploiting an equivalence to Dirichlet process mixture models, that the Cox posterior under a weighted gamma process can be sampled using a Pólya urn Gibbs sampler without requiring any approximation to the likelihood or the prior. This is especially advantageous with large datasets containing, say, thousands of observations (see the example of Sec. 3.3), where with a discretized approach it becomes tricky to select a suitably refined partition of the time axis while keeping computations manageable. Moreover, this approach also applies to more general kernels, and allows us to estimate smoothed hazard functions with or without imposed shape restrictions.

2.3 Poisson Process Spatial Regression Models

Lo and Weng (1989) discussed the use of a weighted gamma process for estimating the intensity of a nonhomogeneous Poisson process (see also Kuo and Ghosh 1997). This approach also falls within our framework. Suppose that $\mathbf{X}_1, \dots, \mathbf{X}_n$ are the observed points from one realization of a Poisson random measure, PRM(Λ), with mean intensity Λ. Assume that the derivative of Λ exists, that is,

$$\Lambda(d\mathbf{x}|\boldsymbol{\theta},\mu) = \eta(d\mathbf{x})\int_{\mathcal{S}} k_0(\mathbf{x},\boldsymbol{\nu})\mu(d\boldsymbol{\nu}),$$

where k_0 is some positive integrable kernel but μ is unknown. The likelihood function is

$$\mathcal{L}(\mu) = \exp\left\{-\int_{\mathcal{S}}\left[\int_{\mathcal{X}} k_0(\mathbf{x},\boldsymbol{\nu})\eta(d\mathbf{x})\right]\mu(d\boldsymbol{\nu})\right\} \times \prod_{i=1}^{n}\int_{\mathcal{S}} k_0(\mathbf{X}_i,\boldsymbol{\nu}_i)\mu(d\boldsymbol{\nu}_i).$$

This is equivalent to (2) after setting $k_i(\mathbf{x},\boldsymbol{\nu},\boldsymbol{\theta}) = k_0(\mathbf{x},\boldsymbol{\nu})$, $Y_i(\mathbf{x}) = 1/n$, and $m = 0$. (See Snyder and Miller 1991, chap. 2.5, for background on multidimensional Poisson processes and their likelihoods.)

In many statistical settings, in addition to the locations $\mathbf{X}_1, \dots, \mathbf{X}_n$ of the observed points from the point process, covariate information $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ is also observed. In such studies the goals are to discover the underlying mechanism producing the spatial counts and to explore its dependence on covariates. Such problems can be naturally subsumed within our multiplicative intensity framework, providing an important extension to the method of Lo and Weng (1989) and leading to what can broadly be considered a class of semiparametric Poisson spatial regression models.

A good situation illustrating the potential of such models is a spatial setting in which spatial count data has been aggregated at a macroregional level. Disease mapping or aggregate spatial epidemiology studies that analyze count data are good illustrative examples. In such studies the data comprise the observed counts N_i of a rare event, such as an incident of a rare disease or a death, within a particular fixed geographical region R_i for $i = 1, \ldots, m$. There may also be covariate information $\mathbf{Z}_i$ specific to the region R_i. Standard models take the form

$$N_i \sim \text{Poisson}\big(\lambda_i E_i \exp(\boldsymbol{\theta}^T \mathbf{Z}_i)\big), \qquad i = 1, \ldots, m, \quad (8)$$

where E_i is the expected number of events based on the region size and λ_i is a region-specific intensity rate. The counts N_i are typically the aggregated counts $N_{i,j}$ over different substrata $R_{i,j}$ of the region R_i. Often the covariates $\mathbf{Z}_i$ are also aggregated; for example, they may be averaged population values for R_i (Prentice and Sheppard 1995).

One serious limitation with a model like (8) is that it does not account for potential spatial correlation in counts. Another problem is that using covariate information aggregated at the regional level can often lead to serious bias. One way of minimizing the potential bias is to use covariate values averaged over random samples (Prentice and Sheppard 1995). Correlation in counts also can be addressed, for example, Green and Richardson (2002) discussed a Potts model formulation for λ_i to model dependence. However, if instead of aggregated data suppose data are available at the substratum level, then both problems can be addressed more naturally within a point process framework by treating the spatial locations of the events as points from a Poisson process. That is, if the spatial location $\mathbf{X}_{i,j}$ of each event and corresponding covariate information $\mathbf{Z}_{i,j}$ are recorded at the substratum level, then a more powerful analysis using a marked Poisson process approach can be used, which eliminates the dependence on fixed regions R_i. In such an analysis, covariates $\mathbf{Z}_{i,j}$ become what are called the *marks of the process.* (See Daley and Vere-Jones 2002, chap. 6.4, for background on marked processes.) The marks can record either spatial covariate information or information specific to the unit. At the micro level, the analog to (8) is the "proportional intensity model" (Svensson 1990)

$$\Lambda(d\mathbf{x}, d\mathbf{z}|\boldsymbol{\theta}, \mu) = w(d\mathbf{x}, d\mathbf{z}) \exp(\boldsymbol{\theta}^T \mathbf{z}) \int_{\mathcal{S}} k_0(\mathbf{x}, \boldsymbol{\nu}) \mu(d\boldsymbol{\nu}), \quad (9)$$

where k_0 is some prespecified kernel. This corresponds to the intensity for a marked point process $\text{PRM}(\Lambda)$, where $w(d\mathbf{x}, d\mathbf{z})$ is the dominating measure for the points and covariates. Typically,

$$w(d\mathbf{x}, d\mathbf{z}) = \eta(d\mathbf{x}) P_n(d\mathbf{z}),$$

where $P_n(\cdot) = \sum_{i,j} \delta_{\mathbf{Z}_{i,j}}(\cdot)/n$ is the empirical distribution for the covariates (we use δ_z to denote a discrete measure concentrated at z). This ensures that the likelihood can be written in the form of (2).

An important feature of (9) is that it provides a smoothed nonparametric estimate for the intensity through the mixed kernel $\int k_0(\cdot, \boldsymbol{\nu}) \mu(d\boldsymbol{\nu})$. Using μ is a way of accounting for heterogeneity, and hence it models potential spatial correlation. Thus the marked Poisson process approach handles the issue of spatial dependence but also adjusts for covariates at the micro level. The idea can be applied quite generally to other problems as well. Other examples of marked process applications are in forest ecology (Wolpert and Ickstadt 1998a) and spatial regression models for analysis of health exposure data (Best et al. 2000). These examples are based on more complex nonproportional intensity models. In Section 4 we discuss how to handle all of these models. For example, in Section 4.2, we look at kernels of the form

$$k_i(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\theta}) = k_0(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\theta}_1) \exp(\boldsymbol{\theta}_2^T \mathbf{Z}_i) \quad (10)$$

and discuss how these can be computed using a blocked Gibbs sampling method.

2.4 Poisson Process Regression Models for Recurrent Events

With only slight modification, the methods can be extended to problems involving n conditionally independent Poisson processes. For example, a general problem of interest are studies involving n independent individuals, each experiencing recurrent events, where the data comprise the number of events and times of events for each individual. Covariate information on individuals may also be included. The goal is to estimate the intensity of the underlying counting process. One method for handling such situations was discussed by Lawless (1987), who considered semiparametric proportional intensity models based on Poisson processes. Using our Bayesian approach, this idea can be expanded to handle nonproportionality and provide smoothed estimates for the intensity. More precisely, suppose that individual i is observed over the time interval $(0, T_i]$ and has repeated events occurring according to a nonhomogeneous Poisson process with a smooth intensity function

$$\Lambda_i(dt|\boldsymbol{\theta}) = dt \int_{\mathcal{S}} k_i(t, \nu, \boldsymbol{\theta}) \mu(d\nu).$$

The choice for the kernel k_i can be quite general, accommodating either proportionality or nonproportionality assumptions. The subscript i is used to indicate that k_i depends on covariate information. As in our previous examples, $\boldsymbol{\theta}$ denotes the covariate parameter. For each individual i, suppose that n_i events are observed to occur at times $T_{i,1} < \cdots < T_{i,n_i}$. The likelihood for the data is

$$\mathcal{L}(\mu, \theta) = \exp\left\{-\sum_{i=1}^{n} \int_{\mathcal{S}} \int_0^{\infty} Y_i(t) k_i(t, \nu, \boldsymbol{\theta})\, dt\, \mu(d\nu)\right\}$$
$$\times \prod_{i=1}^{n} \prod_{j=1}^{n_i} \int_{\mathcal{S}} k_i(T_{i,j}, \nu_{i,j}, \boldsymbol{\theta}) \mu(d\nu_{i,j}), \quad (11)$$

where $Y_i(t) = I\{0 < t \le T_i\}$. Although (11) is slightly different than the likelihood (2), by introducing hidden variables as before, we can show that the posterior for (11) under a weighted gamma process prior has similar features to a Dirichlet process mixture model. Note that the augmented variables here are $\nu_{i,j}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$.

In Section 5 we discuss these kinds of details for posteriors arising in connection with panel count data. This is a slightly more complex setup than (11), because the actual times at which events occur are not observed. However, with only slight modifications, those methods automatically address the class of problems of (11).

3. POSTERIOR EXPRESSIONS AND WEIGHTED GAMMA APPROXIMATIONS

Fundamental to our exposition is the following posterior representation given by James (2003, thm. 3) for semiparametric intensity models subject to weighted gamma process priors. (Also see Lo and Weng 1989 for related work.) Write $\mathbf{X} = (X_1, \ldots, X_{n+m})$ for the vector of data values, and let $g(\mathbf{v}, \mu, \boldsymbol{\theta})$ be an integrable function. It follows from James (2003) that the posterior for $(\mathbf{V}, \mu, \boldsymbol{\theta})$ from (2) under a $\mathcal{G}(\cdot|\alpha, \beta)$ process is characterized by

$$\int g(\mathbf{v}, \mu, \boldsymbol{\theta})\pi(d\mathbf{v}, d\mu, d\boldsymbol{\theta}|\mathbf{X})$$
$$= \int\int g(\mathbf{v}, \mu, \boldsymbol{\theta}) \times \mathcal{G}\left(d\mu|\alpha + \sum_{i=1}^{n} \delta_{\nu_i}, \beta^*\right)\pi(d\mathbf{v}, d\boldsymbol{\theta}|\mathbf{X}), \quad (12)$$

where we write β^* for the function $\beta^*(\nu, \boldsymbol{\theta}) = \beta(\nu)/[1 + \beta(\nu) f(\nu, \boldsymbol{\theta})]$, where

$$f(\nu, \boldsymbol{\theta}) = \sum_{i=1}^{n+m} \int_{\mathcal{X}} Y_i(x) k_i(x, \nu, \boldsymbol{\theta})\eta(dx).$$

The expression $\pi(d\mathbf{v}, d\boldsymbol{\theta}|\mathbf{X})$, a key quantity in (12), denotes the conditional density for $\mathbf{V}$ and $\boldsymbol{\theta}$. It is defined by

$$\pi(d\mathbf{v}, d\boldsymbol{\theta}|\mathbf{X}) \propto m(d\mathbf{v})\pi(d\boldsymbol{\theta}) D(\boldsymbol{\theta}) B^*(\mathbf{v}, \boldsymbol{\theta}) \prod_{i=1}^{n} k_i(X_i, \nu_i, \boldsymbol{\theta}), \quad (13)$$

where

$$B^*(\mathbf{v}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \beta^*(\nu_i, \boldsymbol{\theta}),$$

$$D(\boldsymbol{\theta}) = \exp\left\{-\int_{\mathcal{S}} \log\bigl(1 + \beta(\nu) f(\nu, \boldsymbol{\theta})\bigr)\alpha(d\nu)\right\}, \quad (14)$$

and

$$m(d\mathbf{v}) = \int \prod_{i=1}^{n} \mu(d\nu_i)\mathcal{G}(d\mu|\alpha, 1) = \prod_{i=1}^{n}\left(\alpha + \sum_{j=1}^{i-1} \delta_{\nu_j}\right)(d\nu_i) \quad (15)$$

is the joint marginal density from a gamma process with shape α.

3.1 The General Procedure

A key point underlying our computational algorithms is that the joint marginal density $m(d\mathbf{v})$ is the nonnormalized Blackwell and MacQueen (1973) Pólya urn density. That is, $m(d\mathbf{v})$ is proportional to

$$m_0(d\mathbf{v}) = \int \prod_{i=1}^{n} P(d\nu_i)\mathcal{P}(dP|\alpha) = \frac{\prod_{i=1}^{n}(\alpha + \sum_{j=1}^{i-1} \delta_{\nu_j})(d\nu_i)}{\prod_{i=1}^{n}(\alpha(\mathcal{S}) + i - 1)},$$

the Pólya urn density for a Dirichlet process $\mathcal{P}(\cdot|\alpha)$. This is, of course, an immediate consequence of the relationship (3). Thus, because $m(d\mathbf{v})$ is proportional to $m_0(d\mathbf{v})$, the measure (13), except for the expression $D(\boldsymbol{\theta})B^*(\mathbf{v}, \boldsymbol{\theta})$, has the same structural features as a semiparametric Dirichlet process mixture model. In such cases k_i are kernel densities, $\boldsymbol{\theta}$ is typically a regression coefficient, and $\pi(d\boldsymbol{\theta})$ is the prior for $\boldsymbol{\theta}$. Thus it stands to reason that simulating values from (13) can, with some modification, be implemented similarly to Dirichlet process mixture models.

This general principle can be used to derive various Gibbs sampling procedures. The technique for obtaining a posterior draw for $\mathbf{V}$ and $\boldsymbol{\theta}$ is as follows:

1. Draw $\mathbf{V}$ from its joint conditional distribution,
$$\pi(d\mathbf{v}|\boldsymbol{\theta}, \mathbf{X}) \propto m(d\mathbf{v}) B^*(\mathbf{v}, \boldsymbol{\theta}) \prod_{i=1}^{n} k_i(X_i, \nu_i, \boldsymbol{\theta}).$$
2. Draw $\boldsymbol{\theta}$ from its conditional distribution,
$$\pi(d\boldsymbol{\theta}|\mathbf{v}, \mathbf{X}) \propto \pi(d\boldsymbol{\theta}) D(\boldsymbol{\theta}) B^*(\mathbf{v}, \boldsymbol{\theta}) \prod_{i=1}^{n} k_i(X_i, \nu_i, \boldsymbol{\theta}).$$

Drawing $\boldsymbol{\theta}$ is straightforward using standard parametric methods, whereas drawing $\mathbf{V}$ is based on various Dirichlet process Monte Carlo algorithms. The resulting values $\mathbf{V}$ and $\boldsymbol{\theta}$ usually will be more than adequate for computing various posterior quantities. For example, these values can be used to estimate the posterior mean of the intensity in a Poisson regression model or the posterior mean hazard in a survival analysis. In some settings, however, we may need to compute more complicated functionals, or we also may want to produce confidence intervals. In such cases, a third step is required involving a draw from the posterior random measure μ (see Sec. 3.3 for details):

3. Draw μ from the conditional law $\mathcal{G}(\cdot|\alpha + \sum_{i=1}^{n} \delta_{\nu_i}, \beta^*)$.

3.2 Cox Regression via Pólya Urns

To illustrate, we return to the proportional hazards model discussed in Section 2.2. Substituting the Cox kernels (6) into the posterior (12), and using the previous notation from Section 2.2 (e.g., setting $X_i = T_i$), we have

$$\pi(d\mathbf{v}, d\boldsymbol{\theta}|\mathbf{X}) \propto m_0(d\mathbf{v})\pi(d\boldsymbol{\theta}) D_0(\boldsymbol{\theta}) B^*(\mathbf{v}, \boldsymbol{\theta}) \prod_{i=1}^{n} k_0(T_i, \nu_i), \quad (16)$$

where

$$D_0(\boldsymbol{\theta}) = D(\boldsymbol{\theta}) \exp\left(\boldsymbol{\theta}^T \sum_{i=1}^{n} \mathbf{Z}_i\right)$$

and $D(\boldsymbol{\theta})$ is defined by (14) with

$$f(\nu, \boldsymbol{\theta}) = \sum_{i=1}^{n+m} \exp(\boldsymbol{\theta}^T \mathbf{Z}_i) \int_0^{T_i} k_0(t, \nu)\eta(dt).$$

To simulate values for $\mathbf{V}$ from (16), we use a modification of the Pólya urn Gibbs sampler of Escobar (1988, 1994). Absorb the effect of β^* from $B^*(\mathbf{v}, \boldsymbol{\theta})$ into the kernel by defining $k_0^*(t, \nu, \boldsymbol{\theta}) = \beta^*(\nu, \boldsymbol{\theta}) k_0(t, \nu)$. Let $\mathbf{v}_{-i}$ denote the subvector of $\mathbf{v}$ with the ith coordinate removed. Write $H(\cdot)$ for the distribution $\alpha(\cdot)/\alpha(\mathcal{S})$.

To simulate $(\mathbf{V}, \boldsymbol{\theta})$ from (16), run the following Pólya urn Gibbs sampler:

1. Draw $(\nu_i | \boldsymbol{\theta}, \mathbf{v}_{-i}, \mathbf{X})$ for $i = 1, \ldots, n$. The required conditional densities are

$$\pi(d\nu_i | \boldsymbol{\theta}, \mathbf{v}_{-i}, \mathbf{X}) = \ell_{0,i} \lambda_i(d\nu_i) + \sum_{j=1}^{n_{0,i}} \ell_{j,i} \delta_{\nu_j^*}(d\nu_i),$$

where $\lambda_i(d\nu) \propto k_0^*(T_i, \nu, \boldsymbol{\theta}) H(d\nu)$, and

$$\ell_{0,i} = \frac{\alpha(\mathcal{S})}{c_i} \int_{\mathcal{S}} k_0^*(T_i, \nu, \boldsymbol{\theta}) H(d\nu),$$

$$\ell_{j,i} = \frac{e_{j,i}}{c_i} k_0^*(T_i, \nu_j^*, \boldsymbol{\theta}).$$

The value c_i is a normalizing constant chosen so that $\sum_{j=0}^{n_{0,i}} \ell_{j,i} = 1$, and $\{\nu_1^*, \ldots, \nu_{n_{0,i}}^*\}$ represents the set of $n_{0,i}$ unique values of $\mathbf{v}_{-i}$, with each value ν_j^* occurring with frequency $e_{j,i}$.

2. Draw $(\boldsymbol{\theta} | \mathbf{v}, \mathbf{X})$ from the density proportional to $\pi(d\boldsymbol{\theta}) \times D_0(\boldsymbol{\theta}) B^*(\mathbf{v}, \boldsymbol{\theta})$.

The sampled values can be used to compute various posterior quantities. Thus if $(\mathbf{v}^{(b)}, \boldsymbol{\theta}^{(b)})$, $b = 1, \ldots, B$, are simulated values from the Gibbs sampler, then we can, for example, approximate the posterior mean of a function $g(\mathbf{V}, \boldsymbol{\theta})$ by

$$E\big(g(\mathbf{V}, \boldsymbol{\theta}) | \mathbf{X}\big) \approx \frac{1}{B} \sum_{b=1}^{B} g\big(\mathbf{v}^{(b)}, \boldsymbol{\theta}^{(b)}\big).$$

For example, given a suitable choice for g, this can be used to estimate the posterior mean of the hazard. That is, by (12), the baseline hazard $r_0(t|\mu) = \int_{\mathcal{S}} k_0(t, \nu)\mu(d\nu)$ has posterior expectation

$$\begin{aligned}
&E\big(r_0(t|\mu) | \mathbf{X}\big) \\
&\quad = \iiint k_0(t, \nu)\mu(d\nu) \mathcal{G}\left(d\mu | \alpha + \sum_{i=1}^{n} \delta_{\nu_i}, \beta^*\right) \\
&\qquad \times \pi(d\mathbf{v}, d\boldsymbol{\theta} | \mathbf{X}) \\
&\quad = \iint k_0^*(t, \nu, \boldsymbol{\theta}) \left(\alpha(d\nu) + \sum_{i=1}^{n} \delta_{\nu_i}(d\nu)\right) \pi(d\mathbf{v}, d\boldsymbol{\theta} | \mathbf{X}) \\
&\quad = E\big(g(\mathbf{V}, \boldsymbol{\theta}) | \mathbf{X}\big),
\end{aligned}$$

where

$$g(\mathbf{v}, \boldsymbol{\theta}) = \alpha(\mathcal{S}) \int_{\mathcal{S}} k_0^*(t, \nu, \boldsymbol{\theta}) H(d\nu) + \sum_{j=1}^{n} k_0^*(t, \nu_j, \boldsymbol{\theta}).$$

Remark 2. An acceleration step can be added to enhance mixing for the Markov chain using a technique discussed by West, Müller, and Escobar (1994) for Dirichlet process mixture models. This requires adding the following simple step after completing step 1:

1(a). Let $\mathbf{p}$ be the partition of $\{1, \ldots, n\}$ of sets $C_1, \ldots, C_{n_0}$, where C_j consists of all i where $\nu_i = \nu_j^*$. Draw independent values for $(\nu_j^* | \mathbf{p}, \boldsymbol{\theta}, \mathbf{X})$ from

$$\pi(d\nu_j^* | \mathbf{p}, \boldsymbol{\theta}, \mathbf{X}) \propto H(d\nu_j^*) \prod_{i \in C_j} k_0^*(T_i, \nu_j^*, \boldsymbol{\theta}),$$

$$\text{for } j = 1, \ldots, n_0. \qquad (17)$$

Use the newly sampled ν_j^* values and the partition $\mathbf{p}$ to determine a new updated vector $\mathbf{v}$.

3.3 Weighted Gamma Process Approximations

Let $U_1, \ldots, U_N$ be iid from the distribution $H(\cdot) = \alpha(\cdot)/\alpha(\mathcal{S})$. Let $\mathbf{U} = (U_1, \ldots, U_N)$, and let H^N denote its joint distribution. The following theorem provides a method for accurately approximating a weighted gamma process that can be used for drawing values for μ from the posterior (12). The proof is given in Appendix A.

Theorem 1. For the mixture of weighted gamma processes

$$\mathcal{G}_{N,\beta}(\cdot) = \int \mathcal{G}(\cdot | \alpha_N, \beta) H^N(d\mathbf{u}),$$

where $\alpha_N(\cdot) = \alpha(\mathcal{S}) \sum_{k=1}^{N} \delta_{U_k}(\cdot)/N$, it follows that

(a) $\mathcal{G}_{N,\beta}$ is the law for the random measure defined by

$$\mu(A) = \int_A \beta(\nu) \gamma_N(d\nu), \qquad A \in \mathcal{A},$$

where $\gamma_N(\cdot) = \sum_{k=1}^{N} G_{k,N} \delta_{U_k}(\cdot)$ and $G_{k,N}$ are iid gamma$(\alpha(\mathcal{S})/N)$ random variables independent of U_k.

(b) $\mathcal{G}_{N,\beta}(\cdot) \overset{\mathrm{d}}{\Rightarrow} \mathcal{G}(\cdot | \alpha, \beta)$, where "$\overset{\mathrm{d}}{\Rightarrow}$" indicates weak convergence.

Part (a) of Theorem 1 indicates a method for drawing μ approximately from a weighted gamma process, and part (b) justifies such an approximation. In particular, Theorem 1 suggests the following method for an approximate draw of μ from the posterior in the Cox model: Draw $\mathbf{v}$ and $\boldsymbol{\theta}$ from (16) and then draw a random μ from

$$\int \mathcal{G}\left(\cdot | \alpha_N + \sum_{i=1}^{n} \delta_{\nu_i}, \beta^*\right) H^N(d\mathbf{u}).$$

For a suitably large value of N, this should provide an accurate approximation to $\mathcal{G}(d\mu | \alpha + \sum_{i=1}^{n} \delta_{\nu_i}, \beta^*)$. Thus to get an approximate draw for μ, include the following additional step in the Pólya urn Gibbs algorithm:

3. Using the current values for $\mathbf{v}$ and $\boldsymbol{\theta}$, draw a value for μ from the weighted gamma approximation $\mathcal{G}(\cdot|\alpha_N + \sum_{i=1}^n \delta_{\nu_i}, \beta^*)$. That is, if $\{\nu_1^*, \ldots, \nu_{n_0}^*\}$ represents the set of n_0 unique values of $\mathbf{v}$, with each value ν_j^* occurring with frequency e_j, then draw μ according to

$$\mu(\cdot) = \sum_{k=1}^N G_k \beta^*(U_k, \boldsymbol{\theta}) \delta_{U_k}(\cdot) + \sum_{j=1}^{n_0} G_j^* \beta^*(\nu_j^*, \boldsymbol{\theta}) \delta_{\nu_j^*}(\cdot), \quad (18)$$

where G_k are iid gamma$(\alpha(\mathcal{S})/N)$ variables and U_k are iid H and G_j^* are independent gamma(e_j) variables. All variables are mutually independent.

Remark 3. As an example, to simulate posterior values for the baseline hazard $r_0(t|\mu)$, draw μ from (18) and compute

$$r_0(t|\mu) = \sum_{k=1}^N G_k k_0^*(t, U_k, \boldsymbol{\theta}) + \sum_{j=1}^{n_0} G_j^* k_0^*(t, \nu_j^*, \boldsymbol{\theta}).$$

Averaging this is another way to obtain an estimate for the posterior baseline hazard $E(r_0(t|\mu)|\mathbf{X})$, but typically the draw would be used to estimate quantities like the survival function or the cumulative hazard function.

Remark 4. It is important to note that the second sum in the approximation (18) is exact. Moreover, this value is the dominant term in the expression, making the accuracy of the first sum noncritical. In practice, a value of $N = 50$ works quite well.

Remark 5. Another method for approximating the draw for μ can be based on the inverse Lévy measure algorithm of Wolpert and Ickstadt (1998b), which applies to spatial weighted gamma processes. Another potential technique is the method of Laud et al. (1996), which applies to weighted gamma processes over $\Re$.

3.4 Heart Rate Recovery and Mortality

In this section we apply the Bayesian Cox regression model to a long-term follow-up study of heart rate recovery. The data that we consider were collected from patients who underwent exercise testing at the Cleveland Clinic Foundation between 1990 and 2001. In all, our database contained more than 20,000 patients, all of whom were referred to our institution for symptom-limited exercise testing. Various exercise test measurements were recorded for each patient, including heart rate during and after exercise. An assortment of clinical measurements was all recorded. All patients were followed-up for survival, with a mean follow-up time of 5.6 years (mean survival of 3.8 years for uncensored data) and a range of .01 to 10.1 years. A key variable of interest is the heart rate recovery value, defined as the difference between the heart rate at peak exercise (measured in beats per minute) and heart rate 1 minute after cessation of exercise. As noted recently (see Cole, Blackstone, Pashkow, Snader, and Lauer 1999), heart rate recovery is a powerful independent predictor of mortality, with patients exhibiting abnormal heart rate recovery values considered to be at high risk. An abnormal recovery value is defined as a reduction of 12 beats per minute or less. Here we analyze data for the subset of patients considered at high risk as defined by these criteria. This yields a total of 5,658 patients with 803 failure events (deaths) and 4,855 censored events (86% censoring rate).

Because we have no prior information regarding the shape of the hazard function, we use a uniform rectangular kernel to estimate the hazard without shape restriction. Thus we use

$$k_0(t, \nu) = I\{|t - \nu| \le \tau\}, \quad (19)$$

where the value for $\tau > 0$ is a bandwidth value. The use of a rectangular kernel here is analogous to its use in kernel density estimation. As in density estimation, the choice of kernel is essentially a matter of taste and convenience, and thus, for example, we could have used a normal kernel, $k_0(t, \nu) = \exp(-.5(t-\nu)^2/\tau^2)/\sqrt{2\pi\tau^2}$, to estimate the hazard. We prefer to use a rectangular kernel, because it will lead to some helpful simplifications.

We can assume that all survival times lie in an interval $[0, T]$ for some finite value T. If we rescale time by dividing by T so that time lies in the unit interval $[0, 1]$, then the resulting hazard function r^* will be related to the unscaled time hazard function r by $Tr(t) = r^*(t/T)$. Thus, because we can always recover r from r^* (transforming back after computing the posterior), we can assume that $T = 1$ without loss of generality. Thus we assume that $0 \le t$ and $\nu \le 1$. For the weighted gamma process prior, we take $\beta(\nu) = \beta_0$ and $\alpha(d\nu) = \alpha_0 H(d\nu)$, where $H(d\nu) = I\{0 \le d\nu \le 1\}$ is a uniform distribution. The value for $\alpha_0 > 0$ in this specification controls the amount of smoothing, and thus its role is similar to the bandwidth value τ used in our kernel. Because its role is redundant, we set $\alpha_0 = 1$ and vary the value for τ in controlling overall smoothness. The value for $\beta_0 > 0$ can be used to reflect prior strength. Large values induce a noninformative prior.

Usually, the tricky aspect in implementing the Pólya urn Gibbs sampler is the draw ν from the density $\lambda_i(d\nu)$. Here $\lambda_i(d\nu)$ is proportional to

$$k_0^*(T_i, \nu, \boldsymbol{\theta}) H(d\nu) = \beta^*(\nu, \boldsymbol{\theta}) k_0(T_i, \nu) H(d\nu) = \frac{k_0(T_i, \nu)\, d\nu}{\beta_0^{-1} + f(\nu, \boldsymbol{\theta})}. \quad (20)$$

It turns out that this draw can be implemented fairly easily thanks to the simplifications for $f(\nu, \boldsymbol{\theta})$ that occur by using a rectangular kernel. Appendix B gives the necessary details. The Gibbs algorithm also requires drawing $\boldsymbol{\theta}$ in step 2. For this draw, it is important to easily compute

$$D(\boldsymbol{\theta}) = \exp\left\{-\alpha_0 \int_0^1 \log\bigl(1 + \beta_0 f(\nu, \boldsymbol{\theta})\bigr)\, d\nu\right\}. \quad (21)$$

Again, using a rectangular kernel greatly simplifies this calculation; see Appendix B for the details.

Figure 1 and Table 1 present the results of our analysis. Estimates are based on 3,000 sampled values from our Pólya urn Gibbs sampler after an initial 2,000-iteration burn-in. The method used the acceleration step outlined in Remark 2 (also see App. B). For the draw for $\boldsymbol{\theta}$, we used random-walk Metropolis–Hastings with a multivariate normal transition kernel, where $\boldsymbol{\theta}$ was assumed to have a flat multivariate normal prior $\mathrm{N}(\mathbf{0}, 10^4\mathbf{I})$. For the weighted gamma process, we used

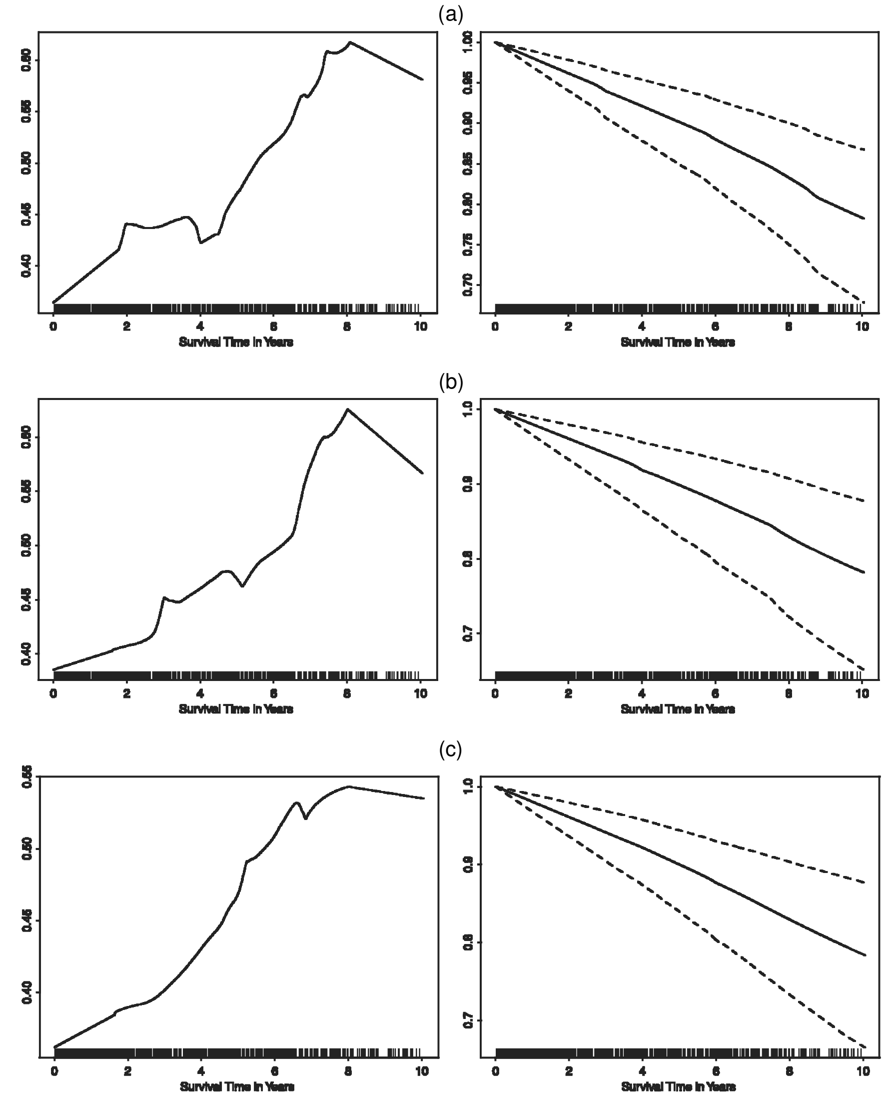

Figure 1. Posterior Baseline Hazard Function r_0 From Heart Rate Recovery Data (left) and Posterior Survival Function Evaluated at Mean Covariate Values (right) for Smoothing Parameters of (a) $\tau = 1.5$, (b) $\tau = 2$, (c) $\tau = 2.5$. The superimposed rug corresponds to observed failure times. The thick lines represent posterior mean values; the dashed lines given for survival functions represent a 95% probability band.

the priors discussed earlier with $\alpha_0 = 1$ and $\beta_0 = 10^5$. When drawing posterior values for μ, we used a truncation value of $N = 50$. (We also tried larger values for N, but found little difference in final estimates.)

Table 1. Parameter Estimates From Heart Recovery Data

Parameter	*Maximum likelihood estimation (MLE $\pm$ standard error)*	*Weighted gamma process, $\tau = 2.5$ (mean $\pm$ standard deviation)*
Asthma	$-.640 \pm .26$	$-.726 \pm .28$
Bronch	$.212 \pm .16$	$.213 \pm .16$
Betablock	$-.479 \pm .09$	$-.498 \pm .11$
Lung	$.861 \pm .12$	$.869 \pm .11$
Fitness	$.120 \pm .03$	$.119 \pm .03$
Hrrecov	$-.031 \pm .01$	$-.031 \pm .01$
Diabetes	$.514 \pm .12$	$.499 \pm .12$
Peakhr	$-.020 \pm .002$	$-.021 \pm .002$
Vascular	$.337 \pm .12$	$.340 \pm .12$
Smoking	$.037 \pm .08$	$.043 \pm .08$

NOTE: Parameters are asthma (no/yes), use of bronchodilators (no/yes), use of beta blockers (no/yes), history of chronic lung disease (no/yes), measure of fitness (1, 2, 3, or 4), heart rate recovery (beats per minute), diabetes (no/yes), peak heart rate (beats per minute), peripheral vascular disease (no/yes), smoking within past year (no/yes).

Figure 1 presents the posterior mean baseline hazard function r_0 (and corresponding survival function) estimated under various choices for the bandwidth value τ (we used $\tau = 1.5$, 2, and 2.5 years). Although it is possible to include τ in the Gibbs sampling scheme, we find that we can more clearly assess the shape of r_0 by manually adjusting the bandwidth value. Varying τ affects the posterior estimates for $\boldsymbol{\theta}$ very little, because the baseline hazard function is a nuisance parameter. (For concreteness, Table 1 is given for posterior estimates based on $\tau = 2.5$.) The plots for r_0 reveal that the hazard has a roughly linear shape, increasing slowly for the first 2 years and then flattening off until about 3 years, then increasing rapidly until approximately 8 years, and leveling off again. We find these plots for the hazard function quite useful for directly understanding the survival behavior. Survival functions shown on the right side of Figure 1 are also useful, but are more difficult to interpret. With respect to $\boldsymbol{\theta}$, Table 1 shows that the Bayes estimates agree closely with the maximum likelihood estimator

(MLE) based on the usual partial likelihood approach. (This is to be expected, because the partial likelihood approach gives valid inference for $\boldsymbol{\theta}$ without knowing the baseline hazard; the difference, of course, is our ability to nonparametrically estimate the baseline and the insight that we gain from this.) All variables except for the bronchodilator use and whether the patient smoked in the past year were significant at a 5% level. The heart recovery value is highly significant, as was anticipated, with a (posterior) relative risk of 1.03 with each decrease of 1 beat per minute recovery.

4. BLOCKED GIBBS SAMPLING

The rectangular kernel (19) used in the previous section provides a simple and flexible method for unrestricted hazard shape estimation in Cox regression settings. For shape-restricted hazards, convenient algorithms can be based on the 0–1 kernel (7). Thus the class of kernels based on indicator functions is sufficiently rich for Cox models. However, in extensions to the Cox model, as well as in extensions to more general multivariate counting processes, such as the Poisson spatial processes discussed shortly (see Sec. 4.2), we often need a wider collection of available kernels for modeling. With more complex kernels, however, posterior computations will become more tricky, because we will not always be able to rely on conjugacy and the kinds of simplifications seen in the previous example. Thus more general Monte Carlo method is needed to address these issues. Here we discuss a Gibbs sampling technique—the blocked Gibbs sampler—that can be used in general. (For a systematic comparison of Pólya urn Gibbs sampling to blocked Gibbs sampling, see Ishwaran and James 2001.)

The trick is to apply the weighted gamma process approximation $\mathcal{G}(\cdot|\alpha_N,\beta)$ (see Thm. 1) to the prior $\mathcal{G}(\cdot|\alpha,\beta)$ rather than the posterior, as was done earlier. This will replace integrals with sums if we use a form of data augmentation, thus greatly simplifying matters. Note that this method *does not involve approximating the likelihood* (2) and involves only the approximation to the prior.

Hereafter, we take $\mathcal{G}(\cdot|\alpha_N,\beta)$ for the prior for μ. Similar to Section 3, it follows that for any integrable function $g(\mathbf{v},\mu,\boldsymbol{\theta})$, the posterior $\pi_N(d\mathbf{v},d\mu,d\boldsymbol{\theta}|\mathbf{X})$ under a $\mathcal{G}(\cdot|\alpha_N,\beta)$ prior is characterized by

$$\int g(\mathbf{v},\mu,\boldsymbol{\theta})\pi_N(d\mathbf{v},d\mu,d\boldsymbol{\theta}|\mathbf{X}) = \int\int g(\mathbf{v},\mu,\boldsymbol{\theta})\mathcal{G}\left(d\mu|\alpha_N+\sum_{i=1}^n\delta_{\nu_i},\beta^*\right) \times \pi_N(d\mathbf{v},d\mathbf{u},d\boldsymbol{\theta}|\mathbf{X}),$$

where

$$\pi_N(d\mathbf{v},d\mathbf{u},d\boldsymbol{\theta}|\mathbf{X}) \propto m_{0,N}(d\mathbf{v}|\mathbf{u})H^N(d\mathbf{u})\pi(d\boldsymbol{\theta}) \times D_N(\mathbf{u},\boldsymbol{\theta})\prod_{i=1}^n k_i^*(X_i,\nu_i,\boldsymbol{\theta}), \quad (22)$$

$$k_i^*(x,\nu,\boldsymbol{\theta}) = \beta^*(\nu,\boldsymbol{\theta})k_i(x,\nu,\boldsymbol{\theta}),$$

$$D_N(\mathbf{u},\boldsymbol{\theta}) = \exp\left\{-\frac{\alpha(\mathcal{S})}{N}\sum_{k=1}^N \log\bigl(1+\beta(u_k)f(u_k,\boldsymbol{\theta})\bigr)\right\},$$

and $m_{0,N}(d\mathbf{v}|\mathbf{u})$ is the Pólya urn density for $\mathcal{P}(\cdot|\alpha_N)$, a Dirichlet process with parameter α_N. The functions $\beta^*(\nu,\boldsymbol{\theta})$ and $f(\nu,\boldsymbol{\theta})$ are defined as in Section 3. Note how using the approximate prior has led to several simplifications by replacing potentially complex integrals with more manageable sums [compare the expression $D_N(\mathbf{u},\boldsymbol{\theta})$ to $D(\boldsymbol{\theta})$ defined by (14)].

To estimate posterior quantities, we need to draw $(\mathbf{V},\mathbf{U},\boldsymbol{\theta})$ from (22). As in the Cox regression, the appearance of a Pólya urn distribution is a signal to try a Dirichlet process approach. However, the presence of the augmented variables $u_1,\dots,u_n$ presents an additional wrinkle, making it no longer feasible to implement a Pólya urn Gibbs sampler. Instead, we use a blocked Gibbs sampling method, adapting a method discussed by Ishwaran and Zarepour (2000) and Ishwaran and James (2001) for Dirichlet process mixture models. This method works by augmenting the parameter space to include the underlying random measure.

Rewrite (22) by rexpressing $m_{0,N}(d\mathbf{v}|\mathbf{u})$ as the marginalized law obtained from integrating over $\mathcal{P}(\cdot|\alpha_N)$,

$$\pi_N(d\mathbf{v},d\mathbf{u},d\boldsymbol{\theta}|\mathbf{X}) \propto \left(\int\prod_{i=1}^n P(d\nu_i)\mathcal{P}(dP|\alpha_N)\right)H^N(d\mathbf{u})\pi(d\boldsymbol{\theta}) \times D_N(\mathbf{u},\boldsymbol{\theta})\prod_{i=1}^n k_i^*(X_i,\nu_i,\boldsymbol{\theta}).$$

Now, rather than sampling (22), we draw values from the augmented distribution

$$\prod_{i=1}^n P(d\nu_i)\mathcal{P}(dP|\alpha_N)H^N(d\mathbf{u})\pi(d\boldsymbol{\theta})D_N(\mathbf{u},\boldsymbol{\theta}) \times \prod_{i=1}^n k_i^*(X_i,\nu_i,\boldsymbol{\theta}). \quad (23)$$

Notice that P can be constructively defined as

$$P(\cdot) = \sum_{k=1}^N W_k\delta_{U_k}(\cdot), \quad (24)$$

where $\mathbf{W}=(W_1,\dots,W_N)$ has the Dirichlet distribution, Dirichlet$(\alpha(\mathcal{S})/N,\dots,\alpha(\mathcal{S})/N)$, independently of U_k. If $\nu_1,\dots,\nu_n$ is a sample obtained from P, then each ν_i can be expressed in terms of a classification variable that identifies its U_k value. In particular, let $\mathbf{K}=(K_1,\dots,K_n)$, where

$$\Pr\{K_i\in\cdot|\mathbf{W}\} = \sum_{k=1}^N W_k\delta_k(\cdot). \quad (25)$$

Then $\nu_i = U_{K_i}$. Now using $\mathbf{K}$, the identity $\nu_i = U_{K_i}$, and the construction for P, we can reexpress (23) in terms of $(\mathbf{K},\mathbf{W},\mathbf{U},\boldsymbol{\theta})$. Thus we end up with augmented variables $(\mathbf{K},\mathbf{W},\mathbf{U},\boldsymbol{\theta})$ with conditional density proportional to

$$\prod_{i=1}^n\left(\sum_{k=1}^N W_k\delta_k(dK_i)\right)\pi_{\mathbf{w}}(d\mathbf{W})\,H^N(d\mathbf{u})\pi(d\boldsymbol{\theta}) \times D_N(\mathbf{u},\boldsymbol{\theta})\prod_{i=1}^n k_i^*\bigl(X_i,u_{K_i},\boldsymbol{\theta}\bigr),$$

which is relatively simple to sample from. This gives us an efficient method for drawing $(\mathbf{V}, \mathbf{U}, \boldsymbol{\theta})$.

4.1 Blocked Gibbs Algorithm

To approximate the posterior law for a function $g(\mathbf{V}, \mu, \boldsymbol{\theta})$, cycle through the following steps:

1. Conditional draw for $\mathbf{K}$. Independently sample K_i according to
$$\Pr\{K_i \in \cdot|\mathbf{W}, \mathbf{U}, \boldsymbol{\theta}, \mathbf{X}\} = \sum_{k=1}^{N} W_{k,i}\delta_k(\cdot), \qquad \text{for } i = 1, \dots, n,$$
where $(W_{1,i}, \dots, W_{N,i}) \propto (W_1 k_i^*(X_i, U_1, \boldsymbol{\theta}), \dots, W_N \times k_i^*(X_i, U_N, \boldsymbol{\theta}))$.
2. Conditional draw for $\mathbf{W}$. Draw $\mathbf{W}$ from a Dirichlet$(\alpha(\mathcal{S})/N + e_1, \dots, \alpha(\mathcal{S})/N + e_N)$, where e_k is the number of K_i variables equal to k.
3. Conditional draw for $\mathbf{U}$. Let $\{K_1^*, \dots, K_{n_0}^*\}$ denote the unique set of K_i values. For each $k \notin \{K_1^*, \dots, K_{n_0}^*\}$, draw U_k from the density proportional to
$$H(du)\exp\left\{-\frac{\alpha(\mathcal{S})}{N}\log\bigl(1 + \beta(u)f(u, \boldsymbol{\theta})\bigr)\right\}.$$
Draw $U_{K_j^*}$, for $j = 1, \dots, n_0$, from the density proportional to
$$H(du)\exp\left\{-\frac{\alpha(\mathcal{S})}{N}\log\bigl(1 + \beta(u)f(u, \boldsymbol{\theta})\bigr)\right\} \times \prod_{\{i:\,K_i = K_j^*\}} k_i^*(X_i, u, \boldsymbol{\theta}).$$
4. Conditional draw for $\boldsymbol{\theta}$. Draw $\boldsymbol{\theta}$ from the density proportional to
$$\pi(d\boldsymbol{\theta}) D_N(\mathbf{u}, \boldsymbol{\theta}) \prod_{i=1}^{n} k_i^*\bigl(X_i, u_{K_i}, \boldsymbol{\theta}\bigr).$$
5. Conditional draw for μ. Setting $v_i = U_{K_i}$ gives the draw for $(\mathbf{V}, \mathbf{U}, \boldsymbol{\theta})$. Now to get a draw for $g(\mathbf{V}, \mu, \boldsymbol{\theta})$, draw μ according to
$$\mu(\cdot) = \sum_{k=1}^{N} G_k \beta^*(U_k, \boldsymbol{\theta})\delta_{U_k}(\cdot),$$
where G_k are independent gamma$(\alpha(\mathcal{S})/N + e_k)$ variables. Note that this draw is exact due to the use of the weighted gamma approximation for the prior.

4.2 Computations for Spatial Poisson Processes

To illustrate the blocked Gibbs sampler, we look at a simulation study with spatial data drawn from a Poisson process PRM(Λ) of the type discussed in Section 2.3. The method can be applied to proportional intensity models like (9) and also to more complex forms, such as the nonproportional models associated with kernels of the form (10). We consider a spatial model with an intensity of the form (10) defined by
$$\Lambda(d\mathbf{x}|\mu, \boldsymbol{\theta}) = \eta(d\mathbf{x}) \int_{\mathcal{S}} k_0(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\theta})\mu(d\boldsymbol{\nu}).$$
For brevity, we consider a setting without covariates. Adding a regression parameter, however, poses no difficulty and in fact proceeds in the same way as the draw for $\boldsymbol{\theta}$.

We simulated one realization from a bivariate Poisson process. We took $\mathcal{X} = \mathcal{S} = \Re^2$. For the kernel, we used a scaled bivariate normal density,
$$k_0(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \frac{1}{2\pi\theta_2}\exp\left(\theta_1 - \frac{1}{2\theta_2}(\mathbf{x} - \boldsymbol{\nu})^T(\mathbf{x} - \boldsymbol{\nu})\right),$$
where $\eta(d\mathbf{x})$, the dominating measure for $k_0(\cdot, \boldsymbol{\nu}, \boldsymbol{\theta})$, is Lebesgue measure on $\Re^2$. Here $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Re \times \Re^+$, with θ_1 representing a scaling term and θ_2 representing a positive bandwidth value, or dispersion parameter. The parameter θ_1 is in fact related to the number of values obtained in the Poisson process realization. A similar scaling effect could also be obtained by using a weighted gamma process prior with a shape parameter $\beta_{\boldsymbol{\theta}}$ depending on θ_1. It is a matter of convenience how one specifies this.

We took the true measure μ_0 to be a three-point bivariate normal mixture distribution, $\mu_0(d\boldsymbol{\nu}) = \sum_{j=1}^{3} p_j\,\phi(\boldsymbol{\nu}, \mathbf{M}_j, \tau_j)\,d\boldsymbol{\nu}$, where
$$\phi(\boldsymbol{\nu}, \mathbf{M}, \theta_2) = \frac{1}{2\pi\theta_2}\exp\left(-\frac{1}{2\theta_2}(\boldsymbol{\nu} - \mathbf{M})^T(\boldsymbol{\nu} - \mathbf{M})\right), \qquad \text{for } \boldsymbol{\nu}, \mathbf{M} \in \Re^2$$
(see Fig. 2 for details).

For priors, we used $\alpha(\cdot) = H(\cdot)$, where H is a bivariate N$(\mathbf{0}, \tau_\alpha\mathbf{I})$ distribution with $\tau_\alpha = 10^4$. For the scale parameter β, we took $\beta(\boldsymbol{\nu}) = C_\beta\phi(\boldsymbol{\nu}, \mathbf{0}, \tau_\beta)$, a bivariate N$(\mathbf{0}, \tau_\beta\mathbf{I})$ density scaled by its normalizing constant $C_\beta = 2\pi\tau_\beta$, where $\tau_\beta = 10^4$ is selected to be large to make β flat. For θ_1, we used a flat N(0, 10) prior, and for $1/\theta_2$, we used a noninformative gamma(.01, .01) prior. For these choices, we have
$$D_N(\mathbf{u}, \boldsymbol{\theta}) = \prod_{k=1}^{N}\bigl(1 + \exp(\theta_1)C_\beta\phi(\mathbf{u}_k, \mathbf{0}, \tau_\beta)\bigr)^{-1/N}$$
and
$$k_i^*(\mathbf{X}_i, \boldsymbol{\nu}_i, \boldsymbol{\theta}) = \beta^*(\boldsymbol{\nu}_i, \boldsymbol{\theta})k_0(\mathbf{X}_i, \boldsymbol{\nu}_i, \boldsymbol{\theta}) = \frac{\exp(\theta_1)C_\beta\phi(\boldsymbol{\nu}_i, \mathbf{0}, \tau_\beta)}{1 + \exp(\theta_1)C_\beta\phi(\boldsymbol{\nu}_i, \mathbf{0}, \tau_\beta)}\phi(\mathbf{X}_i, \boldsymbol{\nu}_i, \theta_2).$$
Each draw in the blocked Gibbs sampler can be done exactly in this setting except for the draws for $\mathbf{U}$ and $\boldsymbol{\theta}$, which were implemented using Metropolis–Hastings. For example, the conditional draw for $\boldsymbol{\theta}$ has density
$$\pi(d\theta_1)\exp(n\theta_1) \times \prod_{k=1}^{N}\bigl(1 + \exp(\theta_1)\,C_\beta\phi(\mathbf{u}_k, \mathbf{0}, \tau_\beta)\bigr)^{-(n_k + 1/N)} \times \pi(d\theta_2)\prod_{i=1}^{n}\phi\bigl(\mathbf{X}_i, \mathbf{u}_{K_i}, \theta_2\bigr).$$
Note that θ_2 has a gamma prior, so it can be drawn exactly,
$$\theta_2^{-1} \sim \text{gamma}\left(.01 + n, .01 + \frac{1}{2}\sum_{i=1}^{n}\bigl(\mathbf{X}_i - \mathbf{u}_{K_i}\bigr)^T\bigl(\mathbf{X}_i - \mathbf{u}_{K_i}\bigr)\right).$$
For the draw for θ_1, we used random-walk Metropolis–Hastings.

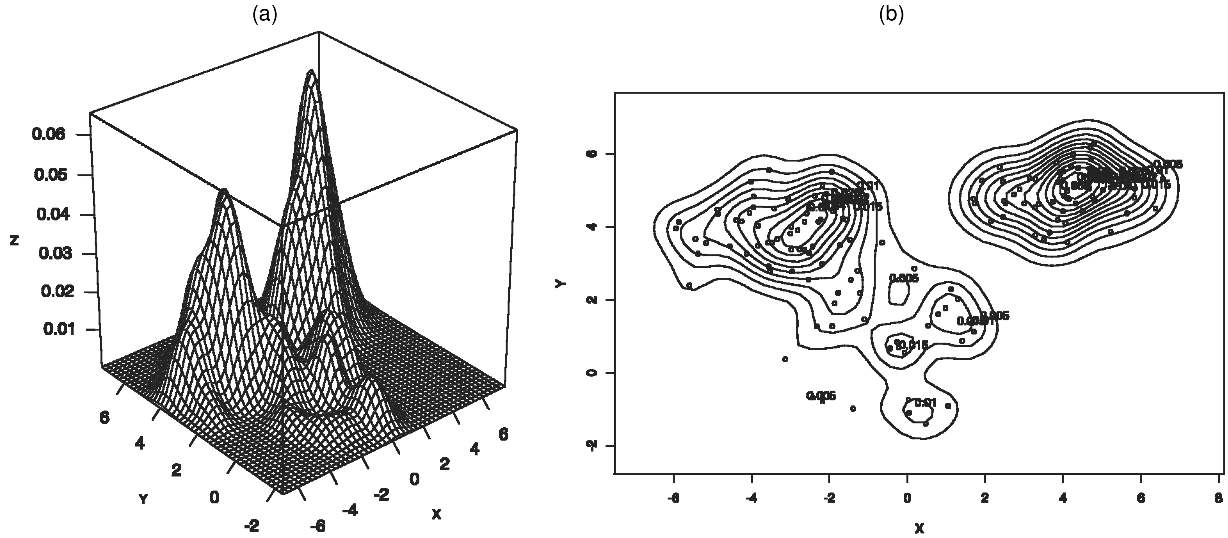

Figure 2. (a) Perspective Plot and (b) Contour Plot for Simulated Poisson Spatial Data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ ($n = 144$). For the measure μ_0, we used a three-point bivariate normal mixture with means $\mathbf{M}_1 = (0, 0)'$, $\mathbf{M}_2 = (4, 4)'$, and $\mathbf{M}_3 = (-3, 3)'$ and covariance matrices $\tau_j I$, where $\tau_1 = 1$, $\tau_2 = .5$, and $\tau_3 = .5$. The three components were weighted by $p_1 = .2$, $p_2 = .4$, and $p_3 = .4$. We took $\theta_1 = 5$ and $\theta_2 = 1$ for the scale and bandwidth parameters. Note that the dots superimposed on the contour plot are values for the data.

The results are given in Figures 3 and 4. Estimates are based on 5,000 sampled values after a 2,500-iteration burn-in. We used an approximation level of $N = 50$ in specifying the approximate weighted gamma prior. (We also tried larger values, with little difference.) Figure 3 shows the estimated posterior mean for μ_0, computed by averaging the measure

$$\mu(\cdot) = \sum_{k=1}^{N} G_k \beta^*(\mathbf{U}_k, \boldsymbol{\theta}) \boldsymbol{\delta}_{\mathbf{U}_k}(\cdot)$$

obtained in step 5 of the blocked Gibbs algorithm. The figure shows that the posterior is accurately recovering μ_0. The posterior values for $\boldsymbol{\theta}$ in Figure 4 also show that both θ_1 and θ_2 are well estimated.

5. PANEL COUNT DATA

As mentioned in Section 2.4, only some slight modifications are needed to handle problems involving conditionally independent nonhomogeneous Poisson processes. In Section 2.4 we discussed recurrent event data. In this section we consider a slightly more complex data setting in which exact times for events are unobserved. In this setup, we have n independent subjects, each observed several times during a study. The number of observations and observation times can vary for each individual. At an observation time, only the number of events up to that time is recorded for a subject, with the exact times for events unknown. This is slightly different than the setting discussed in Section 2.4, although the goal is similar in that we wish to estimate the underlying counting process. These kinds

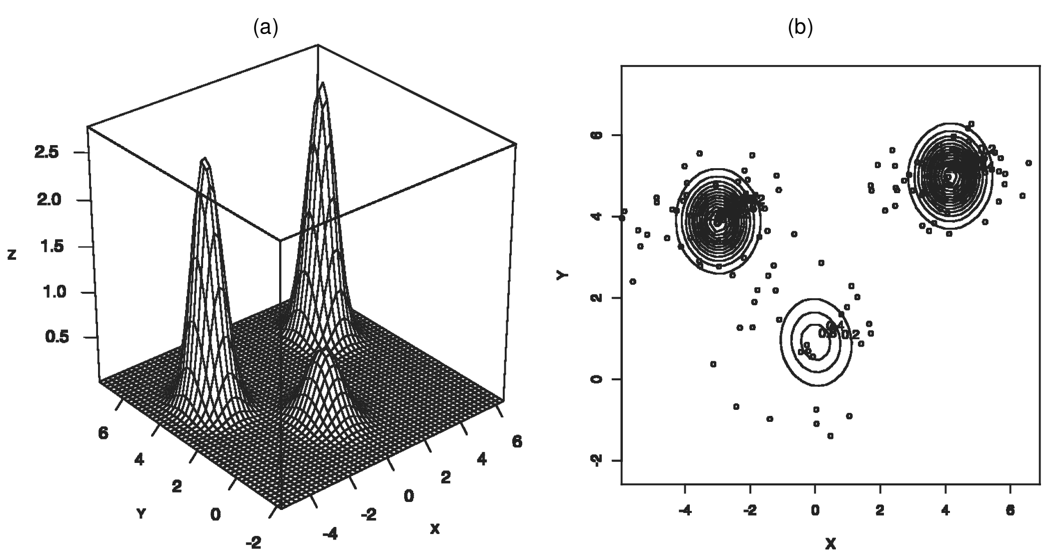

Figure 3. (a) Perspective Plot and (b) Contour Plot for the posterior Mean of μ. Estimate based on the blocked Gibbs sampler using 5,000 values after a 2,500-iteration burn-in. Dots in the contour plot are observed data values.

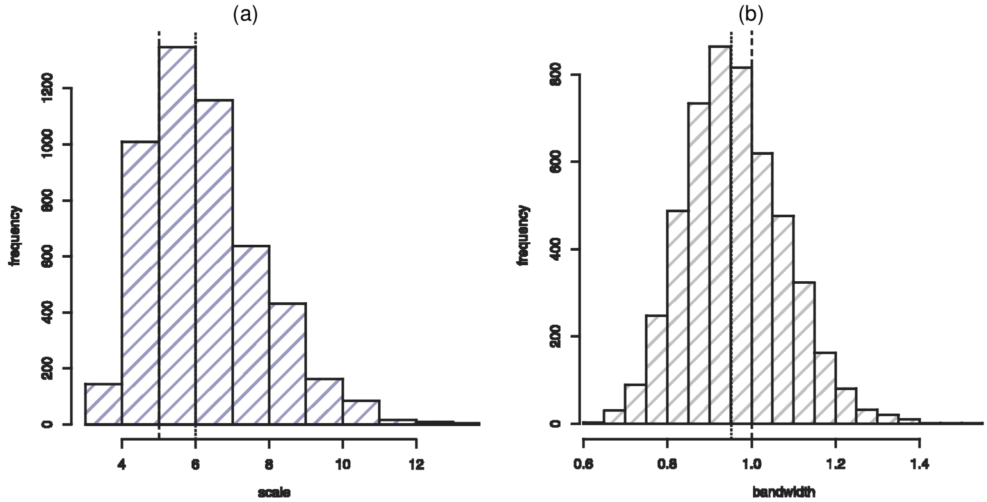

Figure 4. Posterior Values for (a) θ_1 and (b) θ_2 From the Blocked Gibbs Sampler. The thin-dashed lines indicate median values; the thick-dashed lines are true values used in simulation of the data.

of data are often called "panel count" data and are found in many scientific settings. Data where only one event is recorded per subject are commonly called "interval censored" data. (See Kalbfleish and Lawless 1985 and Sun and Kalbfleish 1995 for more background.)

Let $\mathbf{X}_{i,j}$ represent the data, where $i = 1, \ldots, n$ is the index for subjects and $j = 1, \ldots, J_i$ indexes the observation times $T_{i,j}$ for subject i. The data comprise the observation times $T_{i,j}$ as well as the number of events $N_{i,j}$ up to time $T_{i,j}$. Thus $\mathbf{X}_{i,j} = (T_{i,j}, N_{i,j})$. Wellner and Zhang (2000) discussed the use of a nonhomogeneous Poisson process for modeling such data. If $\Lambda(t)$ is the cumulative mean intensity of the process, then the complete likelihood that they considered is

$$\prod_{i=1}^{n}\prod_{j=1}^{J_i}\big(\Lambda(T_{i,j}) - \Lambda(T_{i,j-1})\big)^{n_{i,j}} \times \exp\big\{-\big(\Lambda(T_{i,j}) - \Lambda(T_{i,j-1})\big)\big\},$$

where $n_{i,j} = N_{i,j} - N_{i,j-1}$. (Note that $T_{i,0} = N_{i,0} = 0$.)

We use a kernel-smoothed version of this model, which will allow us to recover Λ as well as the intensity (a unique feature not addressed by Wellner and Zhang). In our smoothed version we model the cumulative intensity as the mixture

$$\Lambda(t|\mu) = \int_{\mathcal{S}}\int_0^t k_0(s, \nu)\,ds\,\mu(d\nu),$$

for some prespecified kernel k_0. Define F by $F(A|\nu) = \int_A k_0(s, \nu)\,ds$ for each Borel-measurable set A. Let $A_{i,j} = (T_{i,j-1}, T_{i,j}]$ and $A_i = (0, T_{i,J_i}]$. The likelihood is

$$\mathcal{L}(\mu) = \exp\left\{-\sum_{i=1}^{n}\int_{\mathcal{S}}\int_0^\infty Y_i(t)F(dt|\nu)\,\mu(d\nu)\right\} \times \prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{l=1}^{n_{i,j}}\int_{\mathcal{S}} F(A_{i,j}|\nu_{i,j,l})\,\mu(d\nu_{i,j,l}), \quad (26)$$

where $Y_i(t) = I\{t \in A_i\}$.

5.1 Posterior Characterization for Panel Count Data

Write $\mathbf{v}$ for the vector of missing values $\{\nu_{i,j,l}\}$. The posterior for the likelihood (26) under a $\mathcal{G}(\cdot|\alpha, \beta)$ prior can be addressed using theorem 3 of James (2003). For any integrable function $g(\mathbf{v}, \mu)$, the posterior for (26) is characterized by

$$\int g(\mathbf{v}, \mu)\pi(d\mathbf{v}, d\mu|\mathbf{X}) = \int\!\!\int g(\mathbf{v}, \mu)\mathcal{G}\Big(d\mu|\alpha + \sum_{i,j,l}\delta_{\nu_{i,j,l}}, \beta^*\Big)\pi(d\mathbf{v}|\mathbf{X}), \quad (27)$$

where

$$\pi(d\mathbf{v}|\mathbf{X}) \propto m_0(d\mathbf{v})\prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{l=1}^{n_{i,j}}\beta^*(\nu_{i,j,l})F(A_{i,j}|\nu_{i,j,l}) \quad (28)$$

and

$$\beta^*(\nu) = \beta(\nu)\Big/\left(1 + \beta(\nu)\sum_{i=1}^{n}F(A_i|\nu)\right).$$

5.2 Panel Count Data via Pólya Urns

The appearance of the Pólya urn density

$$m_0(d\mathbf{v}) = \int\prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{l=1}^{n_{i,j}}P(d\nu_{i,j,l})\mathcal{P}(dP|\alpha)$$

in (28) is again our signal that we can use a Dirichlet process approach to compute the posterior. Here we describe a Pólya urn Gibbs sampler, similar to that in Section 3. Let $\mathbf{v}_{-ijl}$ denote the subvector of $\mathbf{v}$ with the value for $\nu_{i,j,l}$ removed. Write $H(\cdot)$ for the distribution $\alpha(\cdot)/\alpha(\mathcal{S})$.

1. To draw $\mathbf{V}$ from (28), cycle through draws for $\nu_{i,j,l}$, where $\nu_{i,j,l}$ has conditional density

$$\pi(d\nu_{i,j,l}|\mathbf{v}_{-ijl}, \mathbf{X}) = \ell_0\lambda_{i,j,l}(d\nu_{i,j,l}) + \sum_{k=1}^{m}\ell_k\delta_{\nu_k^*}(d\nu_{i,j,l}),$$

where $\lambda_{i,j,l}(d\nu) \propto \beta^*(\nu)F(A_{i,j}|\nu)H(d\nu)$,

$$\ell_0 = \frac{\alpha(\mathcal{S})}{C}\int_{\mathcal{S}} \beta^*(\nu)F(A_{i,j}|\nu)H(d\nu),$$

and

$$\ell_k = \frac{e_k}{C}\beta^*(\nu_k^*)F(A_{i,j}|\nu_k^*).$$

Here C is a normalizing constant chosen to ensure that $\sum_{k=0}^m \ell_k = 1$ and $\{\nu_1^*,\dots,\nu_m^*\}$ represents the set of m unique values of $\mathbf{v}_{-ijl}$, with each value ν_k^* occurring with frequency e_k.

Remark 6. The draws for $\mathbf{V}$ from the Pólya urn Gibbs sampler can be used to estimate the intensity. For example, the posterior mean $E(\Lambda(t|\mu)|\mathbf{X})$ equals $E(g(\mathbf{V})|\mathbf{X})$, where

$$g(\mathbf{v}) = \alpha(\mathcal{S})\int_{\mathcal{S}} \beta^*(\nu)F([0,t]|\nu)H(d\nu) + \sum_{k=1}^m e_k\beta^*(\nu_k^*)F([0,t]|\nu_k^*),$$

and $\{\nu_1^*,\dots,\nu_m^*\}$ represents the set of m unique values of $\mathbf{v}$. Thus, averaging $g(\mathbf{V})$ over different draws for $\mathbf{V}$ provides an estimate for $E(\Lambda(t|\mu)|\mathbf{X})$.

5.3 Blocked Gibbs Sampling

The blocked Gibbs sampler of Section 4 can also be applied to this problem with suitable modification. In problems like this involving large numbers of missing variables, we tend to prefer its use over the Pólya urn sampler. This is because the counting needed to keep track of the different unique values and frequencies of all of the missing variables becomes quite challenging when using the Pólya urn approach. The blocked Gibbs sampler avoids these problems.

As before, let $\alpha_N(\cdot) = \alpha(\mathcal{S})\sum_{k=1}^N \delta_{U_k}(\cdot)/N$, where U_k are iid from $H(\cdot) = \alpha(\cdot)/\alpha(\mathcal{S})$. Approximate (27) by

$$\int g(\mathbf{v},\mu)\pi(d\mathbf{v},d\mu|\mathbf{X}) \approx \iint g(\mathbf{v},\mu)\mathcal{G}\Big(d\mu|\alpha_N + \sum_{i,j,l}\delta_{\nu_{i,j,l}},\beta^*\Big)\pi_N(d\mathbf{v},d\mathbf{u}|\mathbf{X}),$$

$$\pi_N(d\mathbf{v},d\mathbf{u}|\mathbf{X}) \propto m_{0,N}(d\mathbf{v}|\mathbf{u})H^N(d\mathbf{u}) \times \prod_{i=1}^n\prod_{j=1}^{J_i}\prod_{l=1}^{n_{i,j}} \beta^*(\nu_{i,j,l})F(A_{i,j}|\nu_{i,j,l}), \quad (29)$$

where $m_{0,N}(d\mathbf{v}|\mathbf{u})$ is the Pólya urn density for $\mathcal{P}(\cdot|\alpha_N)$.

Now augment the parameter space by using the construction (24) for P and using classification variables $K_{i,j,l}$ similar to (25) such that $\nu_{i,j,l} = U_{K_{i,j,l}}$. Thus, instead of drawing from (29), we draw $(\mathbf{K},\mathbf{W},\mathbf{U})$ from the density proportional to

$$\pi_{\mathbf{W}}(d\mathbf{W})H^N(d\mathbf{U})\prod_{i=1}^n\prod_{j=1}^{J_i}\prod_{l=1}^{n_{i,j}}\left(\sum_{k=1}^N W_k\delta_k(dK_{i,j,l})\right) \times \beta^*\big(U_{K_{i,j,l}}\big)F\big(A_{i,j}|U_{K_{i,j,l}}\big). \quad (30)$$

Sampling from this follows the same strategy discussed in Section 4.1.

Remark 7. Another way to approximate (27) is by using a stick-breaking approximation to the Dirichlet process. It is well known that there exists an infinite sequence of stick-breaking random weights $\{W_k\}$, independent of $\{U_k\}$, such that $P(\cdot) = \sum_{k=1}^\infty W_k\delta_{U_k}(\cdot)$ has a Dirichlet process law $\mathcal{P}(\cdot|\alpha)$. (See Ishwaran and James 2001 for more background on stick-breaking constructions.) Now it follows that if one takes T to be a gamma$(\alpha(\mathcal{S}))$ random variable, such that T is independent of $\{U_k, W_k\}$, then $T \times P(\cdot)$ is a gamma process, $\gamma_\alpha(\cdot)$. This follows from work of McCloskey (1965) as stated in theorem 1.1 of Perman, Pitman, and Yor (1992). As was shown by Ishwaran and James (2001, sec. 3.2), the Dirichlet process can be accurately approximated by a truncation approximation $P_N(\cdot) = \sum_{k=1}^N W_k\delta_{U_k}(\cdot)$. [The total variation distance between the two processes is order $\exp(-(N-1)/\alpha)$, an exponentially decreasing value in N.] Thus it follows that $T \times P_N(\cdot)$ is an approximation to a gamma process $\gamma_\alpha(\cdot)$. This is a new type of approximation for the gamma process that has not been discussed in the literature.

We apply this approximation to (27) in two ways. In the first approximation, we replace $\mathcal{G}(d\mu|\alpha + \sum_{i,j,l}\delta_{\nu_{i,j,l}},\beta^*)$ by its stick-breaking approximation. A draw from $\mathcal{G}(d\mu|\alpha + \sum_{i,j,l}\delta_{\nu_{i,j,l}},\beta^*)$ is of the form

$$\mu(\cdot) + \sum_{k=1}^m G_k^*\beta^*(\nu_k^*)\delta_{\nu_k^*}(\cdot),$$

where G_k^* are independent gamma(e_k) variables and μ is a weighted gamma process characterized by

$$\mu(A) = \int_A \beta(\nu)\gamma_\alpha(d\nu).$$

The approximation simply replaces γ_α, which is $T \times P$, with $T \times P_N$. As mentioned earlier, replacing P with P_N is exponentially accurate. Consequently, to draw an approximate value from $\mathcal{G}(d\mu|\alpha + \sum_{i,j,l}\delta_{\nu_{i,j,l}},\beta^*)$, we simulate T and $\{U_k, W_k : k = 1,\dots,N\}$ and $\{G_k^* : k = 1,\dots,m\}$ independently and draw a value from

$$T\sum_{k=1}^N W_k\beta^*(U_k)\delta_{U_k}(\cdot) + \sum_{k=1}^m G_k^*\beta^*(\nu_k^*)\delta_{\nu_k^*}(\cdot).$$

The second approximation to (27) involves approximating $\pi(d\mathbf{v}|\mathbf{X})$ by

$$\pi_N(d\mathbf{v}|\mathbf{X}) \propto m_N(d\mathbf{v})\prod_{i=1}^n\prod_{j=1}^{J_i}\prod_{l=1}^{n_{i,j}}\beta^*(\nu_{i,j,l})F(A_{i,j}|\nu_{i,j,l}),$$

where

$$m_N(d\mathbf{v}) = \int\prod_{i,j,l} P_N(d\nu_{i,j,l})\mathcal{P}(dP_N).$$

Justification for this again follows because $\mathcal{P}_N$ is an exponentially accurate approximation to $\mathcal{P}(\cdot|\alpha)$. Augmenting the parameter space to include the stick-breaking construction for $\mathcal{P}_N$ will lead to a blocked Gibbs sampling method similar to (30), but now based on stick-breaking weights.

5.4 Multiple Tumor Recurrence

To illustrate our Bayesian approach, we reanalyze data from a bladder tumor study considered by Wellner and Zhang (2000). This dataset was originally used by Andrews and Herzberg (1985, table 45.1) and was collected from a randomized clinical trial comprising patients with superficial bladder tumors. Patients were randomized to three treatments: placebo, treatment by pyridoxine, and treatment by a chemotherapeutic agent, thiotepa. Patients in the study were followed up. At each follow-up visit, any tumors noticed were counted and removed, the follow-up time was recorded, and treatment was then continued. Follow-up times and the number of follow-up visits varied among the patients.

As was done by Wellner and Zhang (2000), we focused on the analysis of tumor counts, which we modeled using the discussed nonhomogeneous Poisson process method. The posterior was sampled using the blocked Gibbs sampler outlined in the previous section. For a kernel, we used $k_0(t,\nu) = I\{\nu \le t\}$, although other kernels can be used with little modification. We set $\mathcal{S} = [0, T]$, where $T = \max\{T_{1,J_1}, \dots, T_{n,J_n}\}$. We used a flat prior by setting $\beta(\upsilon) = \beta_0$ with $\beta_0 = 10^5$. We also took $\alpha(\cdot) = \alpha_0 H(\cdot)$, for H a uniform distribution on $\mathcal{S}$ and $\alpha_0 > 0$.

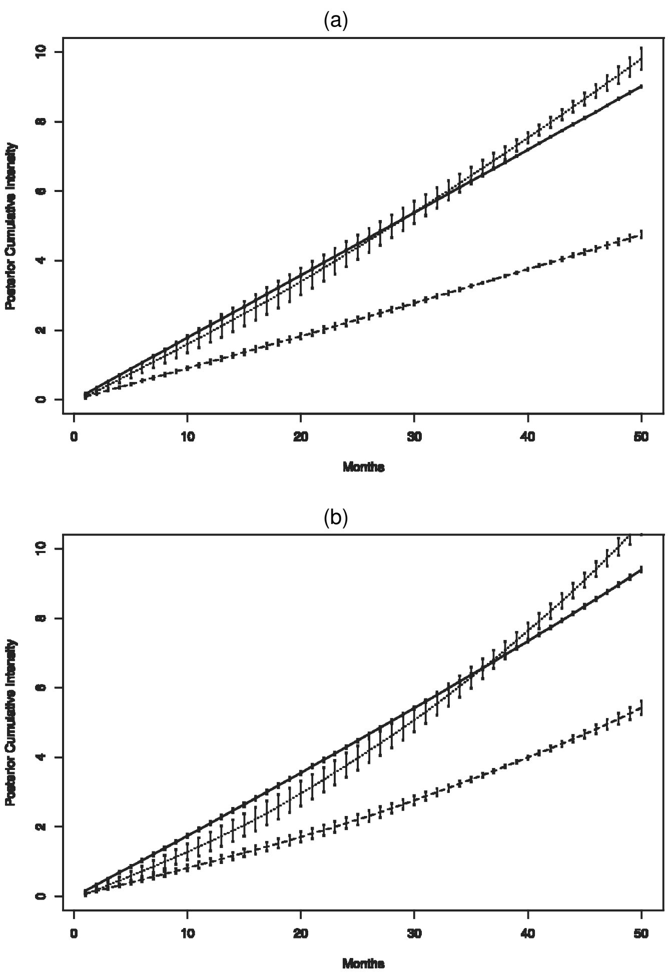

Figure 5. Posterior Means for (a) Cumulative Intensity $\Lambda(t|\mu)$ From Tumor Recurrence Data Using $\alpha_0 = 1$ and (b) $\Lambda(t|\mu)$ Using $\alpha_0 = 10$ (— placebo; ····· pyridoxine; - - - thiotepa). Error bars superimposed on plots are mean values plus or minus 1 standard deviation.

With these choices, the draws for **K**, **W**, and **U** are all straightforward.

We can estimate the posterior intensity $\Lambda(t|\mu)$ from the sampled values similar to Remark 6. For example, the posterior mean $E(\Lambda(t|\mu)|\mathbf{X})$ can be estimated by averaging

$$g(\mathbf{v}) = \frac{\alpha_0}{N}\sum_{k=1}^{N}\beta^*(U_k)F([0,t]|U_k) + \sum_{k=1}^{m} e_k\beta^*(\nu_k^*)F([0,t]|\nu_k^*)$$

over **V** and **U**, where $\{\nu_1^*, \dots, \nu_m^*\}$ represents the set of m unique values of **V** and U_k are iid H. Observe that $\beta^*(\nu)F([0,t]|\nu)$ has a fairly simple expression,

$$\beta^*(\nu)F([0,t]|\nu) = \frac{(t-\nu)I\{\nu \le t\}}{\beta_0^{-1} + \sum_{i=1}^{n}(T_{i,J_i} - \nu)I\{\nu \le T_{i,J_i}\}}.$$

Estimates for the posterior mean and standard deviation of $\Lambda(t|\mu)$ using this method are shown in Figure 5. These are based on a weighted gamma prior approximation with $N = 50$. The plots are calculated from 3,000 sampled values after a 2,500-iteration burn-in. Figure 5(a) was calculated based on a value of $\alpha_0 = 1$, whereas Figure 5(b) used $\alpha_0 = 10$. Both plots clearly show that the thiotepa treatment is the most effective in reducing tumor recurrence. The pyridoxine treatment seems to offer only a slight improvement over placebo, although there is substantially higher variability. Note that varying α_0 (which acts as a smoothing parameter) had little effect on estimates excepting for pyridoxine. The slight difference seen for pyridoxine might be due to its larger variability.

APPENDIX A: PROOF OF THEOREM 1

Proof of part (a) follows by noting that for a fixed $\mathbf{u} = (u_1, \dots, u_N)$, $\mathcal{G}(\cdot|\alpha_N, \beta)$ is a weighted gamma process with shape α_N and scale β. Thus $\mathcal{G}(\cdot|\alpha_N, \beta)$ is characterized by

$$\mu(A) = \int_A \beta(\nu)\xi_N(d\nu),$$

where ξ_N is a gamma process with shape parameter α_N. Now allow **u** to be random to obtain (a). To prove part (b), observe that the Laplace functional of $\mathcal{G}(\cdot|\alpha, \beta)$ is

$$\begin{aligned} L(g) &= \int \exp\left\{-\int g(\nu)\mu(d\nu)\right\}\mathcal{G}(d\mu|\alpha,\beta) \\ &= \exp\left\{-\int \log\big(1 + \beta(\nu)g(\nu)\big)\alpha(d\nu)\right\}, \end{aligned}$$

for $g > 0$ a bounded continuous function (see, e.g., Lo 1982). Meanwhile, for a fixed **u**, the Laplace functional for $\mathcal{G}_{N,\beta}$ is

$$L_N(g) = \exp\left\{-\frac{\alpha(\mathcal{S})}{N}\sum_{k=1}^{N}\log\big(1 + \beta(u_k)g(u_k)\big)\right\}.$$

By the law of large numbers, $L_N(g)$ converges to $L(g)$ for almost all **u** sequences. Thus, by the dominated convergence theorem, $E(L_N(g)) \to L(g)$, which implies $\mathcal{G}_{N,\beta}(\cdot) \overset{\mathrm{d}}{\Rightarrow} \mathcal{G}(\cdot|\alpha,\beta)$ as the Laplace functional uniquely characterizes the law.

APPENDIX B: COX PÓLYA URN DETAILS

Using a rectangular kernel leads to several simplifications when implementing the Cox Pólya urn Gibbs sampler. First, consider the draw from the density $\lambda_i(d\nu)$. Let $f_i(\nu) = \int_0^{T_i} k_0(t,\nu)\,dt$. Then

$$f_i(\nu) = \big(\nu + \tau - \max(0, \nu - \tau)\big) I\{\nu < T_i - \tau\} + \big(T_i - \max(0, \nu - \tau)\big) k_0(T_i, \nu), \qquad 0 \le \nu \le 1.$$

Although $f_i(\nu)$ may look complicated, it is in fact a simple piecewise function with a maximum of three inflection points occurring at $\{\tau, T_i - \tau, T_i + \tau\}$. Consequently,

$$f(\nu, \boldsymbol\theta) = \sum_{i=1}^{n+m} \exp(\boldsymbol\theta^T \mathbf{Z}_i) \int_0^{T_i} k_0(t,\nu)\,dt = \sum_{i=1}^{n+m} \exp(\boldsymbol\theta^T \mathbf{Z}_i) f_i(\nu)$$

is a piecewise linear function in ν, so we can write $f(\nu,\boldsymbol\theta) = \sum_{j=1}^K \psi_{j,\boldsymbol\theta}(\nu)$ as a sum of piecewise linear functions. Each function $\psi_{j,\boldsymbol\theta}$ has a slope $b_{j,\boldsymbol\theta}$ and is equal to 0 except over the interval $I_j = [t_{j,L}, t_{j,U}]$, where the $\{I_j\}$'s are constructed to form a partition of [0, 1]. An important point here is that the intervals $\{I_j\}$ depend only on the values $\{\tau, T_i - \tau, T_i + \tau\}$. Thus $\psi_{j,\boldsymbol\theta}$ and $f(\nu,\boldsymbol\theta)$ can be easily computed for different values of ν and $\boldsymbol\theta$.

Now to draw from $\lambda_i(d\nu)$ in (20), we need merely draw a value from the density proportional to

$$\frac{k_0(T_i,\nu)\,d\nu}{\beta_0^{-1} + f(\nu,\boldsymbol\theta)} = \sum_{j\in\mathcal{I}(i)} \frac{I\{|T_i - \nu| \le \tau\}\,d\nu}{\beta_0^{-1} + \psi_{j,\boldsymbol\theta}(\nu)},$$

where $\mathcal{I}(i)$ corresponds to the indices j for intervals I_j that intersect $A_i = \{\nu : |T_i - \nu| \le \tau\}$. It is straightforward to sample exactly from this. Suppose that $\mathcal{I}(i) = \{i_1, \ldots, i_m\}$. Draw an i_k from $\mathcal{I}(i)$ with probability $\gamma_{i_k} / \sum_{j=1}^m \gamma_{i_j}$, where

$$\gamma_{i_k} = \int_{I_{i_k}} \frac{k_0(T_i,\nu)\,d\nu}{\beta_0^{-1} + f(\nu,\boldsymbol\theta)} = \int \frac{I\{\nu \in I_{i_k} \cap A_i\}\,d\nu}{\beta_0^{-1} + \psi_{i_k,\boldsymbol\theta}(\nu)}.$$

Note that because $\psi_{i_k,\boldsymbol\theta}$ is linear, this can be computed in closed form. To complete the draw for ν, sample ν from $\gamma_{i_k}^{-1} I\{\nu \in I_{i_k} \cap A_i\} / (\beta_0^{-1} + \psi_{i_k,\boldsymbol\theta}(\nu))$.

Meanwhile, the value for $D(\boldsymbol\theta)$ needed to draw $\boldsymbol\theta$ can be computed explicitly. In particular, deduce that (21) equals

$$\exp\left\{-\alpha_0 \sum_{j=1}^K \int_{t_{j,L}}^{t_{j,U}} \log\big(1 + \beta_0 \psi_{j,\boldsymbol\theta}(\nu)\big)\,d\nu\right\}$$
$$= \exp\left\{\alpha_0\big(1 - \log(\beta_0)\big) - \alpha_0 \sum_{j=1}^K \left[\frac{(\beta_0^{-1} + \psi_{j,\boldsymbol\theta}(\nu))}{b_{j,\boldsymbol\theta}} \log\big(\beta_0^{-1} + \psi_{j,\boldsymbol\theta}(\nu)\big)\right]_{t_{j,L}}^{t_{j,U}}\right\}.$$

Remark B.1. A more efficient method for drawing ν from $\lambda_i(d\nu)$ uses the fact that $1/(\beta_0^{-1} + \psi_{j,\boldsymbol\theta}(\nu))$ is convex and has a very tight linear envelope function, say $g_{j,\boldsymbol\theta}$. Each $g_{j,\boldsymbol\theta}$ is a linear function over an interval I_j with slope $c_{j,\boldsymbol\theta}$ and intercept $a_{j,\boldsymbol\theta}$. From this, it is straightforward to develop an efficient rejection sampling scheme to simulate values from $\lambda_i(d\nu)$.

Remark B.2. Recall that to accelerate the Pólya Gibbs sampler, we need to resample the unique values for $\mathbf{v}$ according to (17). So, for example, to draw ν_j^*, we must draw from the density proportional to

$$\frac{1}{(\beta_0^{-1} + f(\nu_j^*, \boldsymbol\theta))^{e_j}} \prod_{i\in C_j} I\{T_i - \tau \le \nu_j^* \le T_i + \tau\}.$$

If $\mathcal{I}^*$ is the set of indices of the intervals for the indicator functions on the right side, then the density of interest is proportional to $\sum_{k\in\mathcal{I}^*} (\beta_0^{-1} + \psi_{k,\boldsymbol\theta}(\nu_j^*))^{-e_j}$. This can be sampled using the same methods just outlined.

[Received September 2002. Revised October 2003.]

REFERENCES

Aalen, O. O. (1975), "Statistical Inference for a Family of Counting Processes," unpublished Ph.D. thesis, University of California, Berkeley.

——— (1978), "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics*, 6, 701–726.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Andrews, D. F., and Herzberg, A. M. (1985), *Data: A Collection of Problems From Many Fields for the Student and Research Worker*, New York: Springer-Verlag.

Arjas, E., and Gasbarra, D. (1994), "Nonparametric Bayesian Inference From Right-Censored Survival Data, Using the Gibbs Sampler," *Statistica Sinica*, 4, 505–524.

Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000), "Spatial Poisson Regression for Health and Exposure Data Measured at Disparate Spatial Scales," *Journal of the American Statistical Association*, 95, 1076–1088.

Blackwell, D., and MacQueen, J. B. (1973), "Ferguson Distributions via Pólya Urn Schemes," *The Annals of Statistics*, 1, 353–355.

Burridge, J. (1981), "Empirical Bayes Analysis of Survival Time Data," *Journal of the Statistical Society*, Ser. B, 43, 65–75.

Cole, R. C., Blackstone, E. H., Pashkow, F. J., Snader, C. E., and Lauer, M. S. (1999), "Heart-Rate Recover Immediately After Exercise as a Predictor of Mortality," *New England Journal of Medicine*, 341, 1351–1357.

Cox, D. R. (1972), "Regression Models and Life-Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

Daley, D. J., and Vere-Jones, D. (2002), *An Introduction to the Theory of Point Processes* (2nd ed.), New York: Springer-Verlag.

Dykstra, R. L., and Laud, P. W. (1981), "A Bayesian Nonparametric Approach to Reliability," *The Annals of Statistics*, 9, 356–367.

Escobar, M. D. (1988), "Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means," unpublished Ph.D. thesis, Yale University.

——— (1994), "Estimating Normal Means With a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268–277.

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

——— (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629.

Ferguson, T. S., and Klass, M. J. (1972), "A Representation of Independent Increment Processes Without Gaussian Components," *Annals of Mathematical Statistics*, 43, 1634–1643.

Gasbarra, D. and Karia, S. (2000), "Analysis of Competing Risks by Using Bayesian smoothing," *Scandinavian Journal of Statistics*, 27, 605–617.

Glasser, R. E. (1980), "Bathtub and Related Failure Rate Characterizations," *Journal of the American Statistical Association*, 75, 667–672.

Green, P. J., and Richardson, S. (2002), "Hidden Markov Models and Disease Mapping," *Journal of the American Statistical Association*, 97, 1055–1070.

Ibrahim, J. G., Chen, M.-H., and MacEachern, S. N. (1999), "Bayesian Variable Selection for Proportional Hazards Models," *Canadian Journal of Statistics*, 37, 701–717.

Ishwaran, H., and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96, 161–173.

Ishwaran, H., and Zarepour, M. (2000), "Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-Parameter Process Hierarchical Models," *Biometrika*, 87, 371–390.

James, L. F. (2003), "Bayesian Calculus for Gamma Processes With Applications to Semiparametric Intensity Models," *Sankhyā*, Ser. A, 65, 196–223.

Kalbfleisch, J. D. (1978), "Non-Parametric Bayesian Analysis of Survival Time Data," *Journal of the Royal Statistical Society*, Ser. B, 40, 214–221.

Kalbfleisch, J. D., and Lawless, J. F. (1985), "The Analysis of Panel Count Data Under a Markov Assumption," *Journal of the American Statistical Association*, 80, 863–871.

Kingman, J. F. C. (1975), "Random Discrete Distributions," *Journal of the Royal Statistical Society*, Ser. B, 37, 1–22.

Kuo, L., and Ghosh, S. K. (1997), "Bayesian Nonparametric Inference for Nonhomogeneous Poisson Processes," Technical Report 9718, University of Connecticut, Department of Statistics.

Laud, P. W., Smith, A. F. M., and Damien, P. (1996), "Monte Carlo Methods for Approximating a Posterior Hazard Rate Process," *Statistics and Computing*, 6, 77–83.

Lawless, J. F. (1987), "Regression Methods for Poisson Process Data," *Journal of the American Statistical Association*, 82, 808–815.

Lo, A. Y. (1982), "Bayesian Nonparametric Statistical Inference for Poisson Point Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 59, 55–66.

——— (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates," *The Annals of Statistics*, 12, 351–357.

Lo, A. Y., Brunner, L. J., and Chan, A. T. (1996), "Weighted Chinese Restaurant Processes and Bayesian Mixture Models," Research Report 1, Hong Kong University of Science and Technology.

Lo, A. Y., and Weng, C. S. (1989), "On a Class of Bayesian Nonparametric Estimates: II. Hazard Rates Estimates," *Annals of the Institute of Statistical Mathematics*, 41, 227–245.

McCloskey, J. W. (1965), "A Model for the Distribution of Individuals by Species in an Environment," unpublished Ph.D. thesis, Michigan State University.

Moran, P. A. P. (1956), "A Probability Theory of a Dam With a Continuous Release," *The Quarterly Journal of Mathematics*, Ser. 7, 130–137.

Perman, M., Pitman, J., and Yor, M. (1992), "Size-Biased Sampling of Poisson Point Processes and Excursions," *Probability Theory and Related Fields*, 92, 21–39.

Prentice, R. L., and Sheppard, L. (1995), "Aggregate Data Studies of Risk Factors," *Biometrika*, 82, 113–125.

Snyder, D. L., and Miller, M. I. (1991), *Random Point Processes in Time and Space*, New York: Springer-Verlag.

Sun, J., and Kalbfleisch, J. D. (1995), "Estimation of the Mean Function of Point Processes Based on Panel Count Data," *Statistica Sinica*, 5, 279–290.

Svensson, A. (1990), "Asymptotic Inference for Multiplicative Counting Processes Based on One Realization," *Journal of Multivariate Analysis*, 33, 125–142.

Wellner, J. A., and Zhang, Y. (2000), "Two Estimators of the Mean of a Counting Process With Panel Count Data," *The Annals of Statistics*, 28, 779–814.

West, M., Müller, P., and Escobar, M. D. (1994), "Hierarchical Priors and Mixture Models, With Applications in Regression and Density Estimation," in *A Tribute to D. V. Lindley*, eds. A. F. M. Smith and P. R. Freeman, New York: Wiley.

Wolpert, R. L., and Ickstadt, K. (1998a), "Poisson/Gamma Random Field Models for Spatial Statistics," *Biometrika*, 85, 251–267.

——— (1998b), "Simulation of Lévy Random Fields," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Mueller, and D. Sinha, New York: Springer-Verlag, pp. 227–242.